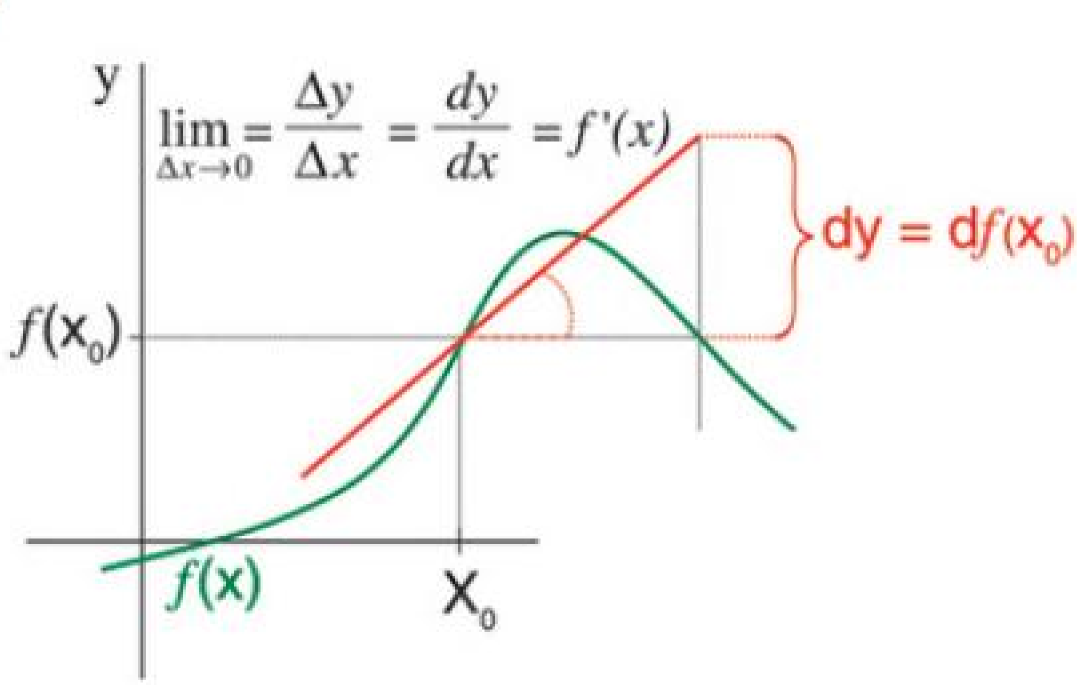
Ulrich L. Rohde G. C. Jain Ajay K. Poddar A. K. Ghosh

INTRODUCTION TO DIFFERENTIAL CALCULUS

Systematic Studies with Engineering Applications for Beginners



WILEY

INTRODUCTION TO DIFFERENTIAL CALCULUS

INTRODUCTION TO DIFFERENTIAL CALCULUS

Systematic Studies with Engineering Applications for Beginners

Ulrich L. Rohde
Prof. Dr.-Ing. Dr. h. c. mult.
BTU Cottbus, Germany
Synergy Microwave Corporation
Paterson, NJ, USA

G. C. Jain
(Retd. Scientist) Defense Research and Development Organization
Maharashtra, India

Ajay K. Poddar
Chief Scientist, Synergy Microwave Corporation,
Paterson, NJ, USA

A. K. Ghosh
Professor, Department of Aerospace Engineering
Indian Institute of Technology – Kanpur
Kanpur, India

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Introduction to differential calculus: systematic studies with engineering
applications for beginners / Ulrich L. Rohde... [et al.]. – 1st ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-118-11775-0 (hardback)
1. Differential calculus–Textbooks. I. Rohde, Ulrich L.
QA304.I59 2012
513'.33–dc23

2011018421

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTENTS

FOREWORD

"What is Calculus?" is a classic deep question. Calculus is the most powerful branch of mathematics, which revolves around calculations involving varying quantities. It provides a system of rules to calculate quantities, which cannot be calculated by applying any other branch of mathematics. Schools or colleges find it difficult to motivate students to learn this subject, while those who do take the course find it very mechanical. Many a times, it has been observed that students incorrectly solve real-life problems by applying Calculus. They may not be capable to understand or admit their shortcomings in terms of basic understanding of fundamental concepts! The study of Calculus is one of the most powerful intellectual achievements of the human brain. One important goal of this manuscript is to give beginner-level students an appreciation of the beauty of Calculus. Whether taught in a traditional lecture format or in the lab with individual or group learning, Calculus needs focusing on numerical and graphical experimentation. This means that the ideas and techniques have to be presented clearly and accurately in an articulated manner.

The ideas related with the development of Calculus appear throughout mathematical history, spanning over more than 2000 years. However, the credit of its invention goes to the mathematicians of the seventeenth century (in particular, to Newton and Leibniz) and continues up to the nineteenth century, when French mathematician Augustin-Louis Cauchy (1789–1857) gave the definition of the limit, a concept which removed doubts about the soundness of Calculus, and made it free from all confusion. The history of controversy about Calculus is most illuminating as to the growth of mathematics. The soundness of Calculus was doubted by the greatest mathematicians of the eighteenth century, yet, it was not only applied freely but great developments like differential equations, differential geometry, and so on were achieved. Calculus, which is the outcome of an intellectual struggle for such a long period of time, has proved to be the most beautiful intellectual achievement of the human mind.

There are certain problems in mathematics, mechanics, physics, and many other branches of science, *which cannot be solved by ordinary methods of geometry or algebra alone*. To solve these problems, we have to use a new branch of mathematics, known as *Calculus*. It uses not only the ideas and methods from arithmetic, geometry, algebra, coordinate geometry, trigonometry, and so on, but also *the notion of limit*, which is a *new idea* which, lies at the *foundation of Calculus*. Using this notion as a tool, *the derivative* of a function (which is a variable quantity) is defined as the limit of a particular kind.

In general, *Differential Calculus* provides a method for calculating "*the rate of change*" of the value of the variable quantity. On the other hand, *Integral Calculus* provides methods for calculating the total effect of such changes, under the given conditions. The phrase *rate of change* mentioned above stands for the actual rate of change of a variable, and *not its average rate of change*. The phrase "rate of change" might look like a foreign language to beginners, but concepts like *rate of change, stationary point*, and *root*, and so on, have precise mathematical meaning, agreed-upon all over the world. Understanding such words helps a lot in understanding the mathematics they convey. At this stage, it must also be made clear that whereas algebra, geometry, and trigonometry are the tools which are used in the study of Calculus, they should not be confused with the subject of Calculus.

This manuscript is the result of joint efforts by Prof. Ulrich L. Rohde, Mr. G. C. Jain, Dr. Ajay K. Poddar, and myself. All of us are aware of the practical difficulties of the students face while learning Calculus. I am of the opinion that with the availability of these notes, students should be able to learn the subject easily and enjoy its beauty and power. In fact, for want of such simple and systematic work, most students are learning the subject as a set of rules and formulas, which is really unfortunate. I wish to discourage this trend.

Professor Ulrich L. Rohde, Faculty of Mechanical, Electrical, and Industrial Engineering (RF and Microwave Circuit Design & Techniques) Brandenburg University of Technology, Cottbus, Germany has optimized this book by expanding it, adding useful applications, and adapting it for today's needs. Parts of the mathematical approach from the Rohde, Poddar, and Böeck textbook on wireless oscillators (*The Design of Modern Microwave Oscillators for Wireless Applications: Theory and Optimization*, John Wiley & Sons, ISBN 0-471-72342-8, 2005) were used as they combine differentiation and integration to calculate the damped and starting oscillation condition using simple differential equations. This is a good transition for more challenging tasks for scientific studies with engineering applications for beginners who find difficulties in understanding the problem-solving power of Calculus.

Mr. Jain is not an educator by profession, but his curiosity to go to the roots of the subject to prepare the so-called *concept-oriented notes for systematic studies in Calculus* is his contribution toward creating interest among students for learning mathematics in general, and Calculus in particular. This book started with these concept-oriented notes prepared for teaching students to face real-life engineering problems. Most of the material pertaining to this manuscript on calculus was prepared by Mr. G. C. Jain in the process of teaching his kids and helping other students who needed help in learning the subject. Later on, his friends (including me) realized the beauty of his compilation and we wanted to see his useful work published.

I am also aware that Mr. Jain got his notes examined from some professors at the Department of Mathematics, Pune University, India. I know Mr. Jain right from his scientific career at Armament Research and Development Establishment (ARDE) at Pashan, Pune, India, where I was a Senior Scientist (1982–1998) and headed the Aerodynamic Group ARDE, Pune in DRDO (Defense Research and Development Organization), India. Coincidently, Dr. Ajay K. Poddar, Chief Scientist at Synergy Microwave Corp., NJ 07504, USA was also a Senior Scientist (1990–2001) in a very responsible position in the Fuze Division of ARDE and was aware of the aptitude of Mr. Jain.

Dr. Ajay K. Poddar has been the main driving force towards the realization of the conceptualized notes prepared by Mr. Jain in manuscript form and his sincere efforts made timely publications possible. Dr. Poddar has made tireless effort by extending all possible help to ensure that Mr. Jain's notes are published for the benefit of the students. His contributions include (but are not limited to) valuable inputs and suggestions throughout the preparation of this manuscript for its improvement, as well as many relevant literature acquisitions. I am sure, as a leading scientist, Dr. Poddar will have realized how important it is for the younger generation to avoid shortcomings in terms of basic understanding of the fundamental concepts of Calculus.

I have had a long time association with Mr. Jain and Dr. Poddar at ARDE, Pune. My objective has been to proofread the manuscript and highlight its salient features. However, only a personal examination of the book will convey to the reader the broad scope of its coverage and its contribution in addressing the proper way of learning Calculus. I hope this book will prove to be very useful to the students of Junior Colleges and to those in higher classes (of science and engineering streams) who might need it to get rid of confusions, if any.

My special thanks goes to Dr. Poddar, who is not only a gifted scientist but has also been a mentor. It was his suggestion to publish the manuscript in two parts (Part I: *Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners* and Part II: *Introduction to Integral Calculus: Systematic Studies with Engineering Applications for Beginners*) so that beginners could digest the concepts of Differential and Integral Calculus without confusion and misunderstanding. It is the purpose of this book to provide a clear understanding of the concepts needed by beginners and engineers who are interested in the application of Calculus of their field of study. This book has been designed as a supplement to all current standard textbooks on Calculus and each chapter begins with a clear statement of pertinent definitions, principles, and theorems together with illustrative and other descriptive material. Considerably more material has been included here than can be covered in most high schools and undergraduate study courses. This has been done to make the book more flexible; to provide concept-oriented notes and stimulate interest in the relevant topics. I believe that students learn best when procedural techniques are laid out as clearly and simply as possible. Consistent with the reader's needs and for completeness, there are a large number of examples for self-practice.

The authors are to be commended for their efforts in this endeavor, and I am sure that both Part I and Part II will be an asset to the beginner's handbook on the bookshelf. I hope that after reading this book, the students will begin to share the enthusiasm of the authors in understanding and applying the principles of Calculus and its usefulness. With all these changes, the authors have not compromised our belief that the fundamental goal of Calculus is to help prepare beginners enter the world of mathematics, science, and engineering.

Finally, I would like to thank Susanne Steitz-Filler, Editor (Mathematics and Statistics) at John Wiley & Sons, Inc., Danielle Lacourciere, Senior Production Editor at John Wiley & Sons, Inc., and Sanchari Sil at Thomson Digital for her patience and splendid cooperation throughout the journey of this publication.

Ajoy Kanti Ghosh
Professor & Faculty Incharge (Flight Laboratory)
Department of Aerospace Engineering
IIT KANPUR, INDIA

PREFACE

In general, there is a perception that Calculus is an extremely difficult subject, probably because the required number of good teachers and good books are not available. We know that books cannot replace teachers, but we are of the opinion that good books can definitely reduce dependence on teachers, and students can gain more confidence by learning most of the concepts on their own. In the process of helping students to learn Calculus, we have gone through many books on the subject and realized that whereas a large number of good books are available at the graduate level, there is hardly any book available for introducing the subject to beginners. The reason for such a situation can be easily understood by anyone who knows the subject of Calculus and hence the practical difficulties associated with the process of learning the subject. In the market hundreds of books are available on Calculus. All these books contain a large number of important solved problems. Besides, the rules for solving the problems and the list of necessary formulae are given in the books, without discussing anything about the basic concepts involved. Of course, such books are useful for passing the examination(s), but Calculus is hardly learnt from these books. Initially, the coauthors had compiled *concept-oriented notes for systematic studies in differential and integral Calculus,* intended for beginners. These notes were used by students in school- and undergraduate-level courses. The response and the appreciation experienced from the students and their parents encouraged us to make these notes available to the beginners. It is due to the efforts of our friends and well-wishers that our dream has now materialized in the form of two independent books: Part I for Differential Calculus and Part II for Integral Calculus. Of course there are some world class authors who have written useful books on the subject at introductory level, presuming that the reader has the necessary knowledge of prerequisites. Some such books are: ***What is calculus about?*** (By Professor WW Sawyer), ***Teach yourself calculus*** (By P. Abbott, B.A), ***Calculus Made Easy*** (By S.P. Thomson) and ***Calculus Explained*** (By W.J. Reichmann). Any person with some knowledge of Calculus will definitely appreciate the contents and the approach of the authors. However, a reader will be easily convinced that most of the beginners may not be able to get (from these books) the desired benefit, for various reasons. From this point of view, both Parts (Part-I & Part-II) of our book would prove to be unique since this provide comprehensive material on Calculus for the beginners. The first six chapters of Part-I would help the beginner to come up to the level, so that one can easily learn ***the concept of limit*****,** which is in the foundation of calculus. The purpose of these works is to provide *the basic (but solid) foundation of Calculus to beginners*. The books aim to show them the *enjoyment in the beauty and power of Calculus and develop the ability to select proper material needed for their studies in any technical and scientific field, involving Calculus.*

One reason for such a high dropout rate is that at beginner levels, Calculus is so poorly taught. Classes tend to be so boring that students sometimes fall asleep. Calculus textbooks get fatter and fatter every year, with more multicolor overlays, computer graphics, and photographs of eminent mathematicians (starting with Newton and Leibniz), yet they never seem easier to comprehend. We look through them in vain for simple, clear exposition, and for problems that will hook a student's interest. Recent years have seen a great hue and cry in mathematical circle over ways to improve teaching Calculus to beginner and high-school students. Endless conferences have been held, many funded by the federal government, dozens of experimental programs are here and there. Some leaders of reform argue that a traditional textbook gets weightier but lacks the step-by-step approach to generate sufficient interest to learn Calculus in beginner, high school, and undergraduate students. Students see no reason why they should master tenuous ways of differentiating and integrating by hand when a calculator or computer will do the job. Leaders of Calculus reform are not suggesting that calculators and computers should no longer be used; what they observe is that without basic understanding about the subject, solving differentiation and integration problems will be a futile exercise. Although suggestions are plentiful for ways to improve Calculus understanding among students and professionals, a general consensus is yet to emerge.

The word "Calculus" is taken from Latin and it simply means a "stone" or "pebble," which was employed by the Romans to assist *the process of counting*. By extending the meaning of the word "*Calculus*," it is now applied to wider fields (of calculation) which involve processes other than mere counting. In the context of this book (with the discussion to follow), the word "*Calculus*" is an abbreviation for *Infinitesimal Calculus* or to one of its two separate but complimentary branches—*Differential Calculus* and *Integral Calculus*. It is natural that the above terminology may not convey anything useful to the beginner(s) until they are acquainted with the processes of *differentiation* and *integration*. What is the ***Calculus*?** What does it calculate? Is ***Calculus*** different from other branches of Mathematics? What type(s) of problems are handled by ***Calculus***?

The author's aim throughout has been to provide a tour of Calculus for a beginner as well as strong fundamental basics to undergraduate students on the basis of the following questions, which frequently came to our minds, and for which we wanted satisfactory and correct answers.

(i) *What is Calculus?*
(ii) *What does it calculate?*
(iii) *Why do teachers of physics and mathematics frequently advise us to learn Calculus seriously?*
(iv) *How is Calculus more important and more useful than algebra and trigonometry or any other branch of mathematics?*
(v) *Why is Calculus more difficult to absorb than algebra or trigonometry?*
(vi) *Are there any problems faced in our day-to-day life that can be solved more easily by Calculus than by arithmetic or algebra?*
(vii) *Are there any problems which cannot be solved without Calculus?*
(viii) *Why study Calculus at all?*
(ix) *Is Calculus different from other branches of mathematics?*
(x) *What type(s) of problems are handled by Calculus?*

At this stage, we can answer these questions only partly. However, as we proceed, the associated discussions will make the answers clear and complete. To answer one or all of the above questions, it was necessary to know: *How does the subject of Calculus begin?*; *How can we learn Calculus?*, and *What can Calculus do for us?* The answers to these questions are hinted at in the books: *What is Calculus about?* and *Mathematician's Delight*, both by W.W. Sawyer. However, it will depend on the curiosity and the interest of the reader to study, understand, and absorb the subject. The author use *very simple and nontechnical language to convey the ideas involved*. However, if the reader is interested to learn the operations of Calculus faster, then he may feel disappointed. This is so, because the nature of Calculus and the methods of learning it are very different from those applicable in arithmetic or algebra. Besides, one must have a real interest to learn the subject, patience to read many books, and obtain proper guidance from teachers or the right books.

Calculus is the higher branch of mathematics, which enters into the process of calculating changing quantities (and certain properties), in the field of mathematics and various branches of science, including social science. It is said to be the *Mathematics of Change*. We cannot begin to answer any question related with change unless we know: *What is that change and how it changes?* This statement takes us closer to the concept of function $y = f(x)$, wherein "y" is related to "x" through a rule "f." We say that "y" is a function of x, by which we mean that "y" depends on "x." (We say that "y" is a *dependent variable*, depending on the value of x, an *independent variable*.) From this statement it is clear that as the value of "x" changes, there results a corresponding change in the value of "y" depending on the *nature of the function* "f" or *the formula defining f.*

The *immense practical power of Calculus is due to its ability to describe and predict the behavior of the changing quantities* "y" *and* "x." In case of linear functions (which are of the form $y = mx + b$), an amount of change in the value of x causes a proportionate change in the value of y. However, in the cases of other functions (like $y = x^2 - 5$, $y = x^3$, $y = x^4 - x^3 + 3$, $y = \sin x$, $y = 3e^x + x$, etc.) which are not linear, *no such proportionality exists*. Our interest lies in studying the behavior of the dependent variable $y[= f(x)]$ with respect to the change in (the value of) the independent variable "x." In other words, we wish to find *the rate at which* "y" changes with respect to "x."

We know that *every rate* is the *ratio of change that* may occur in the quantities, which are related to one another through a rule. It is easy to compute *the average rate at which the value of y changes when x is changed from* x_1 to x_2. It can be easily checked that (*for the nonlinear functions*) these average rate(s) are different *between different values of* x. [Thus, if $|x_2 - x_1| = |x_3 - x_2| = |x_4 - x_3| = \ldots\ldots$, (for all $x_1, x_2, x_3, x_4, \ldots$) then we have $f(x_2) - f(x_1) \neq f(x_3) - f(x_2) \neq f(x_4) - f(x_3) \neq \ldots\ldots$]. Thus, we get that the rate of change of y is different *in between different values of* x.

Our interest lies in computing *the rate of change* of "y" at every value of "x." *It is* known as *the instantaneous rate of change of* "y" *with respect to* "x," and we call it the "*rate function*" *of* "y" *with respect to* "x." It is also called the *derived function* of "y" with respect to "x" and denoted by the symbol $y'[= f'(x)]$. The derived function $f'(x)$ is also called the derivative of $y[= f(x)]$ with respect to x. The equation $y' = f'(x)$ tells that the *derived function* $f'(x)$ is also a *function* of x, derived (or obtained) from the original function $y = f(x)$. There is another (useful) symbol for the *derived function*, denoted by dy/dx. This symbol *appears like a ratio, but it must be treated as a single unit*, as we will learn later. The equation $y' = f'(x)$ gives us the *instantaneous rate of change* of y with respect to x, for every value of "x," for which $f'(x)$ is defined.

To define the *derivative formally* and to *compute it symbolically* is the subject of *Differential Calculus*. In the process of defining the derivative, various subtleties and puzzles will inevitably arise. Nevertheless, *it will not be difficult to grasp the concept (of derivatives) with our systematic approach*. The relationship between $f(x)$ and $f'(x)$ is the *main theme*. We will study what it means for $f'(x)$ to be "*the rate function*" of $f(x)$, and what each function says about the other. It is important to understand clearly *the meaning of the instantaneous rate of change of* $f(x)$ *with respect to* x. These matters are systematically discussed in this book. Note that we have *answered the first two questions* and now proceed to answer the *third one*.

There are certain problems in mathematics and other branches of science, which cannot be solved by ordinary methods known to us in arithmetic, geometry, and algebra alone. In Calculus, we can study the properties of a function without drawing its graph. However, it is important to be aware of the underlying presence of the curve of the given function. Recall that this is due to the introduction of coordinate geometry by Decartes and Fermat. Now, consider the curve defined by the function $y = x^3 - x^2 - x$. We know that, the slope of this curve changes from point to point. If it is desired to find its slope at $x = 2$, then Calculus alone can help us give the answer, which is 7. No other branch of mathematics would be useful.

Calculus uses not only the ideas and methods from arithmetic, geometry, algebra, coordinate geometry, trigonometry, and so on, but also the *notion of limit*, which is a *new idea* that lies at the foundation of Calculus. Using the *notion of limit as a tool*, *the derivative of a function is defined as the limit of a particular kind.* (It will be seen later that the derivative of a function is *generally* a new function.) Thus, *Calculus provides a system of rules for calculating changing quantities which cannot be calculated otherwise*. Here it may be mentioned that the concept of limit is equally important and applicable in Integral Calculus, which will be clear when we study the concept of the definite integral in Chapter 5 of Part II. Calculus is the most beautiful and powerful achievement of the human brain. It has been developed over a period of more than 2000 years. *The idea of derivative of a function is among the most important concepts in all of mathematics and it alone distinguishes Calculus from the other branches of mathematics*.

The derivative and *an integral* have found many diverse uses. The list is very long and can be seen in any book on the subject. *Differential calculus* is a subject which can be applied to anything that *moves,* or *changes* or *has a shape*. It is useful for the study of machinery of all kinds - for electric lighting and wireless, optics, and thermodynamics. It also helps us to answer questions about the *greatest* and *smallest values* a function can take. Professor W.W. Sawyer, in his famous book *Mathematician's Delight*, writes: *Once* the *basic ideas of differential calculus have been grasped*, *a whole world of problems can be tackled without great difficulty*. *It is a subject well worth learning*.

On the other hand, *integral calculus* considers the problem of *determining a function from the information about its rate of change*. Given a formula for the velocity of a body, as a function of time, we can use integral calculus to produce a formula that tells us how far the body has traveled from its starting point, at any instant. It provides methods for the calculation of quantities such as areas and volumes of curvilinear shapes. It is also *useful for the measurement of dimensions of mathematical curves*.

The concepts basic to Calculus can be traced, in uncrystallized form, to the time of the ancient Greeks (around 287–212 BC). However, it was only in the sixteenth and the early seventeenth centuries *that mathematicians developed refined techniques for determining tangents to curves and areas of plane regions*. These mathematicians and their ingenious techniques set the stage for *Isaac Newton* (1642–1727) and *Gottfried Leibniz* (1646–1716), who are usually credited with the "*invention*" of Calculus.

Later on the concept of the definite integral was also developed. *Newton and Leibniz* recognized the importance of the fact that finding derivatives and finding integrals (i.e., antiderivatives) are *inverse processes, thus making possible the rule for evaluating definite integrals.* All these matters are systematically introduced in Part II of the book. (There were many difficulties in the foundation of the subject of Calculus. Some problems reflecting conflicts and doubts on the soundness of the subject are reflected in "Historical Notes" given at the end of Chapter 9 of Part I.) During the last 150 years, Calculus has matured bit by bit. In the middle of the nineteenth century, French Mathematician *Augustin-Louis Cauchy* (1789–1857) *gave the definition of limit, which removed all doubts about the soundness of Calculus and made it free from all confusion.* It was then that, Calculus had become, mathematically, much as we know it today.

Around the year 1930, the increasing use of Calculus in engineering and sciences, created a necessary requirement to encourage students of engineering and science to learn Calculus.

During those days, Calculus was considered an extremely difficult subject. Many authors came up with introductory books on Calculus, but most students could not enjoy the subject, because the basic concepts of the Calculus and its interrelations with the other subjects were probably not conveyed or understood properly. The result was that most of the students learnt *Calculus* only as a *set of rules* and *formulas*. Even today, many students (at the elementary level) learn Calculus in the same way. For them, it is easy to remember formulae and apply them without bothering to know: *How the formulae have come and why do they work*?

The best answer to the question "*Why study Calculus at all*?" is available in the book: *Calculus from Graphical, Numerical and Symbolic Points of View* by Arnold Ostebee and Paul Zorn. There are plenty of good practical and "educational" reasons, which emphasize that one must study Calculus.

- Because it is good for applications;
- Because higher mathematics requires it;
- Because its good mental training;
- Because other majors require it; and
- Because jobs require it.

Also, another reason to study Calculus (according to the authors) is that Calculus is among our deepest, richest, farthest-reaching, and most beautiful intellectual achievements. This manuscript differs in certain respects, from the conventional books on Calculus for the beginners.

In both the Parts of the book (Part-I & Part-II), efforts have been made to ensure that the beginners do not face such situations. The concepts related with calculus and the interrelations between other subjects contributing towards learning calculus have been discussed in a simple language in both part of book (Part-I & Part-II), maintaining the interest and the enthusiasm of the reader. One such example is that of co-ordinate geometry, which is the merging of geometry with algebra and helps in visualizing an equation as representing a curve and vice-versa (Remember, calculus cannot be imagined without co-ordinate geometry.)

It is a fact that people can achieve many things in life even without learning calculus. It is really a big loss to all those who had an opportunity to learn calculus but unfortunately missed it for mere comfort and carelessness. Also, they would never know what really they have missed. It is hoped that this book will motivate the readers who may like to revise their basic knowledge of calculus to achieve the delayed benefit now.

Organization

The work is divided into two independent books: Book I—*Differential Calculus* (*Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners*) and Book II—*Integral Calculus* (*Introduction to Integral Calculus: Systematic Studies with Engineering Applications for Beginners*).

Part I consists of 23 chapters in which certain chapters are divided into two sub-units such as 7a and 7b, 11a and 11b, 13a and 13b, 15a and 15b, 19a and 19b. Basically, these sub-units are different from each other in one way, but they are interrelated through concepts. Also, there are Appendices A, B, and C for Part-I.

Part II consists of 9 chapters in which certain chapters are divided into two sub-units such as 3a and 3b, 4a and 4b, 6a and 6b, 7a and 7b, 8a and 8b, and finally 9a and 9b. The division of chapters is based on the same principle as in the case of Part I. Each chapter (or unit) in both the parts begins with an introduction, clear statements of pertinent definitions, principles, and theorems. Meaning(s) of different theorems and their consequences are discussed at length, before they are proved. The solved examples serve to illustrate and amplify the theory, thus bringing into sharp focus many fine points, to make the reader comfortable.

Illustrative and other descriptive material (along with notes and remarks) is given in each chapter to help the beginner understand the ideas involved. The CONTENTS of each chapter are reflected with all necessary details. Hence, it is not felt necessary to repeat the same details again. However, the following two points are worth emphasizing.

The Part-I (*Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners*):

- The first six chapters of Part I are devoted for revising the prerequisites useful for both the parts. The selection of the material and its sequencing is very important. The reader will find it quite interesting and easy to absorb. Once the reader has gone through these chapters carefully, the reader will be fully prepared to study the concept of limit in Chapters 7a and 7b. The reader will not find any difficulty in absorbing and appreciating the $\varepsilon - \delta$ definition of limit. This definition is generally considered very difficult by the students and therefore it is mugged up without understanding its meaning.
- Chapter 8 deals with the concept of continuity that can be easily learnt, once the concept of limit is properly understood. (Chapters 7a, 7b, and 8 are considered as prerequisites for the purpose of understanding the concept of derivative.)
- Chapter 9 deals with the concept of derivative and its definition including the method of computing the derivative, ***by the first principle*** of a given function using the definition of derivative. (The concepts of limit, continuity, and derivative are discussed at length in the above chapters and must be studied carefully and with patience.) Once the reader has reached upto chapter-9, 50% ideas related with differential calculus is being understood. Subsequently, the ideas related with the integral calculus will be found very simple for understanding in Part-II of the book.
- Chapter 10 deals with the *algebra of derivatives* offering different methods for computing derivatives of functions depending on their properties and the algebra of limits. The concepts discussed in the remaining chapters do not pose problems to the reader since every concept is introduced in a proper sequence suggesting its necessity and applications.
- Chapter 11 is sub-divided into two part (11a and 11b). Chapter 11a deals with basic understanding of the trigonometric limits and its application for computing the derivatives of these functions.
- Chapter 11b deals with the methods of computing limits of trigonometric functions.
- Chapter 12 deals with exponential form (s) of a positive real number and its logarithm(s): Prerequisite for understanding exponential and logarithmic functions.
- Chapter 13 is sub-divided into two part (13a and 13b). Chapter 13a deals with the *properties of exponential and logarithmic functions* including their derivatives.
- Chapter 13b deals with methods for computing limits of exponential and logarithmic functions
- Chapter 14 deals with the *inverse trigonometric functions* and *their properties* including derivatives of many other functions using trigonometric identities.
- Chapter 15 is sub-divided into two part (15a and 15b).
- Chapter 15a deals with implicit functions and their differentiation.
- Chapter 15a deals with parametric functions and their differentiation.

- Chapter 16 deals with the concept of differentials dy and dx, and their applications in the process of integration and for understanding differential equations. It is also discussed how the symbol dy/dx for the derivative of a function can be looked upon as a ratio of differential dy to dx.
- Chapter 17 deals with the derivatives of higher order, their meaning and usefulness. Chapter 18 deals with applications of derivatives in studying motion in a straight line.
- Chapter 19 is sub-divided into two part (19a and 19b). Chapter 19a deals with the concepts of increasing and decreasing functions, studied using derivatives of first and second order.
- Chapter 19b deals with the methods of finding maximum and minimum values of a function using the concept of increasing and decreasing functions.
- Chapters 20, 21, and 22 *are extremely important* dealing with Mean Value Theorems and their applications like L'Hospital's Rule and introduction to the expansion of simple functions.
- Chapter 23 deals with the introduction of hyperbolic functions and their properties.

Important advice for using both the parts of this book:

The CONTENTS clearly indicate how important it is to go through the prerequisites. Certain concepts [like $(-1)\cdot(-1) = 1$, and why division by zero is not permitted in mathematics, etc] which are generally accepted as rules, are discussed logically. The ***concept of infinity*** and its algebra are very important for learning calculus. The ideas and definitions of functions introduced in Chapter-2, and extended in Chapter-6, are very useful.

The role of co-ordinate geometry in defining trigonometric functions and in the development of calculus should be carefully learnt.

The theorems, in both the Parts are proved in a very simple and convincing way. The solved examples will be found very useful by the students of plus-two standard and the first year college. Difficult problems have been purposely not included in solved examples and the exercise, to maintain the interest and enthusiasm of the beginners. The readers may pickup difficult problems from other books, once they have developed interest in the subject.

Concepts of ***limit, continuity*** and ***derivative*** are discussed at length in chapters 7(a) & 7 (b), 8 and 9, respectively. The one who goes through from chapters-1 to 9 has practically learnt more than 60 % of differential calculus. The readers will find that remaining chapters of differential calculus are easy to understand. Subsequently, readers should not find any difficulties in learning the concepts of integral calculus and the process of integration including the methods of computing definite integrals and their applications in fining areas and volumes, etc.

The differential equations right from their formation and the methods of solving certain differential equations of first order and first degree will be easily learnt.

Students of High Schools and Junior College level may ***treat this book as a text book for the purpose of solving the problems and may study desired concepts from the book treating it as a reference book***. Also the students of higher classes will find this book very useful for understanding the concepts and treating the book as a reference book for this purpose. ***Thus, the usefulness of this book is not limited to any particular standard***. ***The reference books are included in the bibliography.***

I hope, above discussion will be found very useful to all those who wish to learn the basics of calculus (or wish to revise them) for their higher studies in any technical field involving calculus.

Suggestions from the readers for typos/errors/improvements will be highly appreciated.

Finally, efforts have been made to ensure that interest of the beginner is maintained all through. It is fact that reading mathematics is very different from reading a novel. However, we hope that the readers will enjoy this book like a novel and learn Calculus. We are very sure that if beginners go through first six chapters of Part I (i.e., prerequisites), then they may not learn Calculus, but will start loving mathematics.

Dr. -Ing. Ajay Kumar Poddar
Chief Scientist,
Synergy Microwave Corporation.
NJ 07504, USA.
Former Senior Scientist (Defense Research &
Development Organization (DRDO), INDIA

Spring 2011

BIOGRAPHIES

Ulrich L. Rohde holds a Ph.D. in Electrical Engineering (1978) and a Sc.D. (Hon., 1979) in Radio Communications, a Dr.-Ing (2004), a Dr.-Ing Habil (2011), and several honorary doctorates. He is President of Communications Consulting Corporation; Chairman of Synergy Microwave Corp., Paterson, NJ; and a partner of Rohde & Schwarz, Munich, Germany. Previously, he was the President of Compact Software, Inc., and Business Area Director for Radio Systems of RCA, Government Systems Division, NJ. Dr. Rohde holds several dozen patents and has published more than 200 scientific papers in professional journals, has authored and coauthored 10 technical books. Dr. Rohde is a Fellow Member of the IEEE, Invited Panel Member for the FCC's Spectrum Policy Task Force on Issues Related to the Commission's Spectrum Policies, ETA KAPPA NU Honor Society, Executive Association of the Graduate School of Business-Columbia University, New York, the Armed Forces Communications & Electronics Association, fellow of the Radio Club of America, and former Chairman of the Electrical and Computer Engineering Advisory Board at New Jersey Institute of Technology. He is elected to the "First Microwave & RF Legends" (Global Voting from professionals and academician from universities and industries: Year 2006). Recently Prof. Rohde received the prestigious "Golden Badge of Honor" and university's highest Honorary Senator Award in Munich, Germany.

G.C. Jain graduated in science (Major—Advance Mathematics) from St. Aloysius College, Jabalpur in 1962. Mr. Jain has started his career as a Technical Supervisor (1963–1970), worked for more than 38 years as a Scientist in Defense Research & Development Organization (DRDO). He has been involved in many state-of-the-art scientific projects and also responsible for streamlining MMG group in ARDE, Pune. Apart from scientific activities, Mr. Jain spends most of his time as a volunteer educator to teach children from middle and high school.

Ajay K. Poddar graduated from IIT Delhi, Doctorate (Dr.-Ing.) from TU-Berlin (Technical University Berlin) Germany. Dr. Poddar is a *Chief Scientist*, responsible for design and development of state-of-the-art technology (oscillator, synthesizer, mixer, amplifier, filters, antenna, and MEMS based RF & MW components) at Synergy Microwave Corporation, NJ. Previously, he worked as a Senior Scientist and was involved in many state-of-the-art scientific projects in DRDO, India. Dr. Poddar holds more than dozen US, European, Japanese, Russian, Chinese patents, and has published more than 170 scientific papers in international conferences and professional journals, contributed as a coauthor of three technical books. He is a recipient of several scientific achievement awards, including RF & MW state-of-the-art product awards for the year 2004, 2006, 2008, 2009, and 2010. Dr. Poddar is a senior member of professional societies IEEE (USA), *AMIE (India)*, and *IE (India)* and involved in technical and academic review committee, including the Academic Advisory Board member Don Bosco Institute of Technology, Bombay, India (2009–to date). Apart from academic and scientific activities, Dr. Poddar is involved in several voluntary service organizations for the greater cause and broader perspective of the society.

A.K. Ghosh graduated and doctorate from IIT Kanpur. Currently, he is a Professor & Faculty Incharge (Flight Laboratory) Accountable Manager (DGCA), Aerospace Engineering, IIT Kanpur, India (one of the most prestigious institutes in the world). Dr. Ghosh has published more than 120 scientific papers in international conferences and professional journals; recipient of DRDO Technology Award, 1993, young scientist award, Best Paper Award—In-house Journal "Shastra Shakti" ARDE, Pune. Dr. Ghosh has supervised more than 30 Ph.D. students and actively involved in several professional societies and board member of scientific review committee in India and abroad. Previously, he worked as a Senior Scientist and Headed Aerodynamic Group ARDE, Pune in DRDO, India.

INTRODUCTION

In less than 15 min, let us realize that calculus is capable of computing many quantities accurately, which cannot be calculated using any other branch of mathematics.

To be able to appreciate this fact, we consider a “nonvertical line” that makes an angle “θ” with the positive direction of x-axis, and that $\theta \neq 0$. We say that the given line is “inclined” at an angle “θ” (or that the inclination of the given line is “θ”).

The important idea of our interest is the “slope of the given line,” which is expressed by the trigonometric ratio “tan θ.” Technically the slope of the line tells us that if we travel by “one unit,” in the positive direction along the x-axis, then the number of units by which the height of the line rises (or falls) is the measure of its slope.

Also, it is important to remember that the “slope of a line” is a constant for that line. On the other hand “the slope of any curve” changes from point to point and it is defined in terms of the slope of the “tangent line” existing there. To find the slope of a curve $y = f(x)$ at any value of x, the “differential calculus” is the only branch of Mathematics, which can be used even if we are unable to imagine the shape of the curve.

At this stage, it is very important to remember (in advance) and understand clearly that whereas, the subject of Calculus demands the knowledge of algebra, geometry, coordinate geometry and trigonometry, and so on (as a prerequisite), but they do not form the subject of Calculus. Hence, calculus should not be confused as a combination of these branches.

Calculus is a different subject. The backbone of Calculus is the “concept of limit,” which is introduced and discussed at length in Part I of the book. The first eight chapters in Part I simply offer the necessary material, under the head: What must you know to learn Calculus? We learn the concept of “derivative” in Chapter 9. In fact, it is the technical term for the “slope.”

The ideas developed in Part I are used to define an inverse operation of computing antiderivative. (In a sense, this operation is opposite to that of computing the derivative of a given function.)

Most of the developments in the field of various sciences and technologies are due to the ideas developed in computing derivatives and antiderivatives (also called integrals). The matters related with integrals are discussed in “Integral Calculus.”

The two branches are in fact complimentary, since the process of integral calculus is regarded as the inverse process of the differential calculus. As an application of integral calculus, the area under a curve $y = f(x)$ from $x = a$ to $x = b$, and the x-axis can be computed only by applying the integral calculus. No other branch of mathematics is helpful in computing such areas with curved boundaries.

Prof. Ulrich L. Rohde

ACKNOWLEDGMENTS

There have been numerous contributions by many people to this work, which took much longer than expected. As always, Wiley has been a joy to work with through the leadership, patience, and understanding of Susanne Steitz-Filler, Jacqueline Palmieri, and Danielle Lacourciere. We would like to express our sincere appreciation to all the staff at John Wiley & Sons involved in this publishing project for their cheerful professionalism and outstanding supports.

It is a pleasure to acknowledge our indebtedness to Prof. Hemant Bhate (Department of Mathematics), Prof. M. S. Prasad, and Dr. Sukratu Barve (Center for Modeling and Simulation), University of Pune, India, who read the manuscript and gave valuable suggestions for improvements.

We wish to express our heartfelt gratitude to the Shri K.N. Pandey, Dr. P. K. Roy, Shri Kapil Deo, Shri D.K. Joshi, Shri S.C. Rana, Shri J. Nagarajan, Shri A. V. Rao, Shri Jitendra C. Yadhav, and Dr. M. B. Talwar for their logistic support throughout the preparation of the manuscripts.

We are thankful to Mrs. Yogita Jain, Dr. (Mrs.) Shilpa Jain, Mrs. Shubhra Jain, Ms. Anisha Apte, Ms. Rucha Lakhe, and Mr. Parvez Daruwalla for their support towards sequencing the material, proof reading the manuscripts and rectifying the same, from time to time.

We also express our thanks to Mr. P. N. Murali, Mrs. Shipra Jain, Mr. Nishant Singhai, Mr. Nikhil Nanawaty, Mr. Vaibhav Jain, Mr. Atul Jain, Mr. A.G. Nagul, and Ms. Radha Borawake who have who have helped in typing and checking it for typographical errors from time to time.

We are indebted to Dr. (Ms.) Meta Rohde, Mrs. Sandhya Jain, Mrs. Kavita Poddar, and Mrs. Swapna Ghosh for their patience, encouragement, appreciation, support and understanding during the preparation of the manuscripts. We would also like to thank Tiya, Pratham, Harsh, Devika, Aditi and Amrita for their compassion and understanding.

We wish to thank our family and friends for their love, inspiration, support, and encouragement. Finally, we would like to thank our reviewers for reviewing the manuscripts and expressing their valuable feedback, comments, and suggestions.

1 From Arithmetic to Algebra

1.1 INTRODUCTION

Numbers are symbols used for counting and measuring. Hindu–Arabic numerals 0, 1, 2, 3, ..., 9 are grouped systematically in units, tens, hundreds, and so on, to solve problems containing numerical information. This is the subject of *Arithmetic*. It also involves an understanding of the structure of the number system and the facility to change numbers from one form to another; for example, the changing of *fractions to decimals* and vice versa. A detailed discussion about the *Real Number System* is given in Chapter 3. However, it would be instructive to *recall some important subsets of real numbers*, known to us.

Numbers, which are used in counting, are called *natural numbers* or *positive integers*. The set of *natural numbers* is denoted by

$$N = \{1, 2, 3, 4, 5, \ldots\}$$

1.2 THE SET OF WHOLE NUMBERS

The set of *natural numbers* along with the number "0" makes the set of whole numbers, denoted by W. Thus,

$$W = \{0, 1, 2, 3, 4, \ldots\}$$

Note: "0" is a whole number but it is not a natural number.

1.3 THE SET OF INTEGERS

All *natural numbers*, their *negatives* and *zero* when considered together, form *the set of integers* denoted by Z. Thus,

$$Z = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

1.4 THE SET OF RATIONAL NUMBERS

The numbers of the form p/q where p and q are *integers*, and the denominator $q \neq 0$, form the set of *rational numbers*, denoted by Q.

Examples: $\frac{3}{5}, \frac{-7}{9}, \frac{8}{-15}, \frac{0}{15}, \frac{9}{1}, \frac{-121}{-12}, \frac{16}{2}$ and so on, are all rational numbers.

What must you know to learn calculus? 1-(The Language of Algebra)

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

Remarks:

(a) *Zero* is a *rational number*, but *division by zero* is not defined. Thus, $5/0$ and $0/0$ are meaningless expressions.
(b) *All integers are rational numbers*, but the converse is not true.
(c) *Positive rational numbers* are called *fractions*.

Let us discuss more about fractions.

Generally, "*fractions*" are used to represent the parts of a given quantity, under consideration. Thus, 3/7 tells us that a given quantity or an object is divided into seven equal parts and three parts are under consideration. A fraction is also used to express a ratio. Thus, 2:5 is also written as 2/5 and similarly 12:5 is written as 12/5. Since the ratio of two natural numbers can be greater than 1, *all positive rational numbers* are called *fractions*. This definition suggests that fractions could be classified more meaningfully as follows:

- When both numerator and denominator are positive integers, the fraction is known as a *simple, common*, or *vulgar fraction* (Examples: 1/2, 3/5, 9/7).
- A *complex fraction* is one in which either the numerator or the denominator or both are fractions (Examples: 3/(7/5), (5/9)/2, (7/3)/(11/4)).
- If the numerator is less than the denominator, the fraction is called a *proper fraction* (Examples: 4/7, 3/5, 1/4).
- If the numerator is greater than the denominator, the fraction is called an *improper fraction* (Examples: 7/4, 5/3, 9/2).
- A *unit fraction* is a special proper fraction, whose numerator is 1 (Examples: 1/7, 1/100).

Note (1): A fraction is said to be in *lowest terms*, if the only common factor of the numerator and denominator is 1. Thus, 3/4 is in lowest terms, but 6/8 is *not* in lowest terms since 6 and 8 have a common factor 2, other than 1. We say that a/b, $2a/2b$, $3a/3b$, ... all belong to the *same family of fractions*, described by a/b.

In fact, we use the fraction in lowest terms to describe the family of fractions. We define the set of all fractions by $F = \{a/b | a, b \in N\}$

1.5 THE SET OF IRRATIONAL NUMBERS

There are numbers that cannot be expressed in the form p/q, where p and q are integers. They are called *irrational numbers*, and the set is denoted by Q' or Q^c. (More details are given in Chapter 3.)

Examples:

$\sqrt{2}$, $\sqrt{5}$, $6\sqrt{3}$, $7\sqrt{11}$, e, π, 1.101001 ..., 5.71071007100071, ... and so on.

1.6 THE SET OF REAL NUMBERS

The set of *rational numbers* together with the set of *irrational numbers*, form the set of *real numbers*, denoted by R.[(1)]

[(1)] The square roots of negative numbers (i.e., $\sqrt{-1}$ or $\sqrt{-7}$, etc.) do not represent real numbers, hence we shall not discuss about such numbers at this stage.

1.6.1 Arithmetic and Algebra

In arithmetic, there are four fundamental operations, namely, addition, subtraction, multiplication, and division, which are performed on the set of natural numbers to make new numbers, namely, the number zero, negative integers, and rational numbers. For the formation of *irrational numbers,* we have to go beyond the four fundamental arithmetic operations given above.

The subject of algebra involves the study of equations and a number of other problems that developed out of the theory of equations. *It is in connection with the solution of algebraic equations that negative numbers, fractions, and rational numbers were developed. The number "0" could enter the family of numbers only after negative numbers were developed.*

In arithmetic, we deal with numbers that have one (single) definite value. On the other hand, in algebra we deal with symbols such as x, y, z, . . ., and so on, which represent variable quantities and those like a, b, c, . . ., and so on, which may have any value we chose to assign to them. These symbols represent *variable quantities and are hence called variables*. We may operate with all these symbols as numbers without assigning to them any particular numerical value. Note that, both *numbers and letters are symbols*, which were developed to solve various problems.

In fact, *traditional algebra* is a generalization of *arithmetic*. Hence, the symbols used in *arithmetic* have the *same meaning in algebra*. Thus, we use $+$ (plus for addition), $-$ (minus for subtraction), $\times$ and $\cdot$ (cross and dot for multiplication), / (slash for division), $=$ (equals for equality), $>$ (for greater than), $<$ (for less than) and so on, in algebra also.

Before we enter the true realm of algebra, it is useful to recall *some more subsets of real numbers*, which will be needed in various discussions.

1.7 EVEN AND ODD NUMBERS

Every integer that is *exactly divisible by 2*, is called an *even number*, otherwise it is *odd*. Thus, an even number is of the form $2n$, where n is an integer.

An odd number is of the form $(2n \pm 1)$. If number "a" is even, then $(a \pm 1)$ is odd and vice versa. *It follows that 0 is an even integer.*

1.8 FACTORS

Natural numbers that exactly divide a given integer are called the factors of that number. For example, the factors of 12 are 1, 2, 3, 4, 6, and 12. We also say that 12 is a *multiple* of 1, 2, 3, 4, 6, and 12. Similarly, the factors of 6 are 1, 2, 3, and 6, and the factors of zero are all the natural numbers.

Remark: The number "0" is not a factor of any number.[(2)]

1.9 PRIME AND COMPOSITE NUMBERS

A *natural* number that has *exactly two unique factors* (namely the number itself and 1) is called a *prime number*. A *natural* number that has three or more factors is called a *composite number*.

[(2)] Factors are considered from *natural numbers* only. Besides, note that *division by zero is not permitted in mathematics*. This is explained at the end of this chapter.

Some examples of prime numbers are 2, 3, 5, 7, 11, 13, 17, 19, ……, and so on.

- Each prime number, *except* 2, is *odd.*
- The number 1 is *neither* prime *nor* composite. Six is a composite number since it has *four factors*, namely 1, 2, 3, and 6.

A given *natural* number can be uniquely expressed as a product of primes.

1.10 COPRIME NUMBERS

Two natural numbers are said to be *coprime (or relatively prime) to each other* if they have no common factor except 1. For example, 8 and 25 are coprime to one another. Obviously, all prime numbers are coprime to each other.

Remark: Coprime numbers need not be prime numbers.[3]

1.11 HIGHEST COMMON FACTOR (H.C.F.)

The highest common factor (H.C.F.) of two or more *(natural)* numbers is the greatest number which divides each of them exactly. It is also known as the greatest common divisor (G.C.D.). [The H.C.F. of any two prime numbers (or coprime numbers) is always 1.]

1.12 LEAST COMMON MULTIPLE (L.C.M.)

The least common multiple (L.C.M.) of two or more *(natural)* numbers is the *smallest number* which is exactly divisible by each of them. To find the L.C.M. of two (or more) natural numbers, we find prime factors. If two (or more) numbers have a factor in common, we select it once. This is done for each such common factor and the remaining factors from each number are taken as they are. The product of all these factors taken together, gives the L.C.M. of the given numbers.

$$(\textit{Product of two numbers} = \textit{their } H.C.F. \times \textit{their } L.C.M.)$$

1.12.1 Continuous Variables and Arbitrary Constants

A changing quantity, usually denoted by a letter (i.e., x, y, z, etc.), which takes on any one of the possible values, in an interval, is called a *variable*. On the other hand, the set of letters a, b, c, d, and so on are used to denote *arbitrary constants*.

In the case of arbitrary constants, though there is no restriction to the numerical values a letter may represent, it is understood that in the same piece of work, it keeps the same value throughout. For example, in the expression, $f(x) = ax^2 + bx + c$, $(0 \leq x \leq 5)$, x is a continuous variable in the interval [0,5] and a, b, c are arbitrary constants. (The concept of an interval is discussed in Chapter 3.)

(3) There is one more term used in connection with prime numbers. A pair of prime numbers which differ by 2, are called *twin-primes* (Examples: 3 and 5, 5 and 7, 11 and 13, 17 and 19, and so on).
Remark: It is proved that the number of primes is infinite, but it is not yet proved whether the number of twin-primes is finite or infinite. This is because of the fact that, so far there is no formula that can generate all primes.

1.13 THE LANGUAGE OF ALGEBRA

Let us now recall the terminology used in algebra:

- an algebraic expression;
- factors, coefficients, index/exponent (or power) of a quantity;
- positive and negative terms;
- like and unlike terms;
- processes involving addition, subtraction, multiplication, and division among algebraic expressions;
- removal and insertion of brackets;
- simplification of an algebraic expression;
- polynomials and related concepts.

It is assumed that all these terms and processes are known to the reader. However, *it is proposed to extend the terminology and concepts related to polynomials,* since the same will be useful to us, in our discussions to follow.

1.13.1 Polynomials

A polynomial in x is an expression of the form

$$p(x) = a_n \cdot x^n + a_{n-1} \cdot x^{n-1} + \ldots + a_1 \cdot x + a_0$$

where a_0, a_1, a_2, ..., a_n are *real numbers* called the *coefficients of p(x)* and n in x^n is a *non-negative integer*.[4]

Usually, we write a polynomial in either *descending powers of x or ascending powers of x*. The form of a ploynomial written in this way is called the standard form. From the definition of a polynomial, it is clear that polynomials are *special types of algebraic expressions* involving only *finite number of terms* and one variable.[5]

1.13.2 Degree of a Polynomial

The exponent, in the highest degree term of a nonzero polynomial is called the *degree of the polynomial*. Thus, if $a_n \neq 0$, then n (in x^n) is *the degree of the polynomial*. In particular, the degree of $3x^5 + 2x^3 - x + 7$ is 5 and the degree of $(3/2)y^3 - \sqrt{2}y - 1$ is 3.[6]

A polynomial having only one term is called "*monomial*".

[4] *By definition, the power of x in each term* of a polynomial must be *a whole number*. If the power of any term is a negative integer or a fraction, then such an expression is not called a polynomial. Note that the power of x in $p(x)$ can be zero. Such a polynomial is called a *constant polynomial*. Another way for getting a constant polynomial could be to make all the coefficients (except a_0) equal to zero, so that we get $p(x) = a_0$, $a_0 \neq 0$. If each of the coefficients $a_0, a_1, a_2, \ldots, a_n$ in $p(x)$ is zero, then such a polynomial is called the *zero polynomial*.

Remark: The zero polynomial is included in the definition of a polynomial.

[5] A polynomial may have more than one variable but our interest lies in the polynomials involving only one variable.

[6] If $n = 1$, it is a linear expression [Example: $f(x) = 2x + 5$].
If $n = 2$, it is a quadratic expression [Example: $f(x) = x^2 + 3x + 1$].
If $n = 3$, it is a cubic expression [Example: $f(x) = x^3 + 3x^2 + 2x + 1$].
If $n = 4$, it is a quartic or biquadratic expression. If $n = 5$, it is a quintic expression.

1.13.3 The Zero Polynomial

We know that a polynomial having all coefficients as zero is called "*the zero polynomial*". Zero polynomial is *unique* and it is denoted by the symbol "0".[7]

The degree of "zero polynomial" is not defined. (Note that, $0 = 0 \cdot x = 0 \cdot x^5 \ldots = 0 \cdot x^{107}$, and so on. These are all zero polynomials and obviously, their degree cannot be defined.) In what follows, *a polynomial will mean a nonzero polynomial (in a single variable) with real coefficients.*

1.13.4 Polynomials Behave Like Integers

Many properties possessed by integers are also possessed by the polynomials. Therefore, *we extend the terminology, used in the algebra of numbers, to the algebra of polynomials*. Thus, if $p(x)$ and $q(x)$ are two polynomials, then the expression $p(x)/q(x)$, where $q(x)$ is a nonzero-polynomial, is called a *rational expression.*[8]

A rational expression must be expressed in its *lowest terms*, by canceling the common factors in the numerator and denominator. For this purpose, one has to *learn the process of factorization of a polynomial.*

1.13.4.1 Factors of a Polynomial A polynomial $g(x)$ is called a factor of polynomial $p(x)$, if $g(x)$ divides $p(x)$ exactly; that is, on dividing $p(x)$ by $g(x)$ we get zero as the remainder.

1.13.4.2 Division Algorithm (or Procedure) for Polynomials On dividing a polynomial $p(x)$ by a polynomial $g(x)$, let the quotient be $q(x)$ and the remainder be $r(x)$, then we have $p(x) = g(x) \cdot q(x) + r(x)$, where either $r(x) = 0$ or degree of $r(x) <$ degree of $g(x)$.

Remark: When a polynomial $p(x)$ is divided by a linear polynomial $(x - \alpha)$ then the remainder is a constant, which may be zero or nonzero. The value of the remainder can be obtained by applying the *remainder theorem.*

1.13.4.3 Remainder Theorem If a polynomial $p(x)$ is divided by a linear polynomial $(x - \alpha)$, then the remainder is $p(\alpha)$. (This theorem can be easily proved using the division algorithm.)

Remark: If $p(x)$ is divided by $(x + \alpha)$, then the remainder $= p(-\alpha)$. Similarly, when $p(x)$ is divided by $(ax + b)$ then the remainder $= p(-b/a)$.

It is sometimes possible to express a polynomial as a product of other polynomials, each of degree ≥ 1. For example, $x^3 - x^2 + 9x - 9 = (x - 1) \cdot (x^2 + 9)$ and $3x^2 - 6x - 9 = 3(x^2 - 2x - 3) = 3(x - 3)(x + 1)$.

1.13.5 Value of a Polynomial and Zeros of a Polynomial

We know that for every real value of x, a polynomial has a real value. For example, let $p(x) = 3x^4 - 2x^3 + x + 5$. Then, for $x = 1$, we have $p(1) = 7$ and for $x = 0$, $p(0) = 5$.

[7] The role of zero polynomial can be compared with that of number "0", in arithmetic. The symbol "0", in polynomial algebra represents the zero polynomial whereas in arithmetic it represents the real number "0".
[8] Every polynomial may be regarded as a rational expression but the converse is not true. Note that $(x+3)/(x - \sqrt{x})$ is *not* a rational expression. It is *an irrational algebraic expression.*

An important aspect of the study of a polynomial is *to determine those values of x for which p(x) = 0*. Such values of x are called *zeros of the polynomial p(x)*. Consider the quadratic polynomial $q(x) = x^2 - x - 6$. It may be seen that $q(3) = 0$ and $q(-2) = 0$. If $x = a$ is a zero of the polynomial $p(x)$ then *(x − a) is a factor of p(x)*. This is known as the *factor theorem* of algebra.

Thus, the factor theorem helps in finding *the linear factors* of a polynomial, provided such factors exist. There are no standard methods available for finding linear factors of polynomials of higher degrees, except in some very special cases.

Every *quadratic polynomial* can have *at most two zeros*, a *cubic polynomial at most three zeros*, and so on. *Some polynomials do not have any real zero. In other words, there may be no real number "x" for which the value of the polynomial becomes zero*. For example, there is no real number "x" for which $x^2 + 3$ will be zero.

Now the following question arises: How do we determine the zeros of a given polynomial $p(x)$?

This leads us to the question: *How to solve the equation p(x) = 0?*

1.13.6 Polynomial Equations and Their Solutions (or Roots)

If $p(x)$ is a *quadratic polynomial*, then the equation $p(x) = 0$ is called a quadratic equation. If $p(x)$ is a cubic polynomial, then the corresponding equation $p(x) = 0$ is called a cubic equation, and so on. If the numbers α and β are two zeros of the quadratic polynomial $p(x)$, we say that α *and* β *are the roots of the corresponding quadratic equation p(x) = 0.*[(9)]

Note: The fundamental theorem of algebra states that a nonzero nth degree polynomial equation has at most n roots, in which some roots may be *repeated roots*.

Thus, starting from the concept of an algebraic expression we have revised the concepts of *polynomials, zeros of a polynomial, and the solution of simple polynomial equations.*

1.14 ALGEBRA AS A LANGUAGE FOR THINKING

We know that algebra has a set of rules; but we should not feel satisfied to have learnt algebra merely as a set of rules. It is more important to have some understanding of: *What is algebra all about? How does it grow out of arithmetic? And how is it used to convey concepts of arithmetic?* For instance, the following statements belong to arithmetic:

3^2 is 1 bigger than 2×4

4^2 is 1 bigger than 3×5

5^2 is 1 bigger than 4×6

[(9)] It is easy to solve equations of degree one and two. Thus, we get from a$x + b = 0$, $(a \neq 0)$, $x = -b/a$ and from a$x^2 + bx + c = 0$, $x = \left(-b \pm \sqrt{(b^2 - 4ac)}\right)/2a$. Mathematicians also solved a number of particular equations of degree three but were finding it difficult to express x in terms of general coefficients a, b, c, and d. This problem was finally solved by the Italian mathematician Tartaglia (1499–1557). Later Lodovico Ferari (1522–1565) solved the general fourth degree equation. It seemed almost certain to the mathematicians that the general fifth degree equation and still higher degree equations could also be solved. For 300 years this problem was a classic one. The Frenchman Evariste Galois (1811–1832) showed that the general equation of degree higher than the fourth cannot be solved by algebraic operations including radicals such as square root, cube root, and so on. To establish this result Galois created the Theory of Groups, a subject that is now at the base of modern abstract algebra and that transformed algebra from a series of elementary techniques to a broad, abstract, and basic branch of mathematics. [*Mathematics and the Physical World* by Morris Kline (pp. 71–72).]

These results suggest that "the square of any natural number is 1 bigger than the result of multiplying two numbers of which one is less by one and the other is more by one, than the given number". Thus, we should guess that 87^2 would be 1 bigger than 86×88.

The general result is stated most conveniently in the language of algebra. Let n be any natural number. Then *"the number before n" will be written as (n – 1) and "the number after n" is (n + 1).* We shall now say, n^2 is 1 bigger than $(n-1)(n+1)$, or, completely in symbols,

$$n^2 = 1 + (n-1)(n+1) \tag{1}$$

Note that, *the above equation holds not only for natural numbers but also for all numbers. It expresses* what we guessed at by looking at particular results in arithmetic. The beauty of algebra lies in its utility. Here, it enables us to prove that our guess is correct. By the usual procedures of algebra, we can simplify the expression on the right-hand side of Equation (1) and see that it equals the left-hand side.

In algebra itself, we often pass from particular results to more general ones. For example, we get from Equation (1)

$$n^2 - 1 = (n-1)(n+1)$$

but we know that $n^2 - 1 = n^2 - 1^2 = (n-1)(n+1)$

In general, we have $a^2 - b^2 = (a-b)(a+b)$

$$\text{or } a^2 = (a-b)(a+b) + b^2 \tag{2}$$

This result is more general than the one expressed by Equation (1).

We can make use of Equation (2) in simple calculations. For example,

$$\begin{aligned} 27^2 &= (27-3)(27+3) + 3^2 \\ &= (24 \times 30) + 9 \\ &= 720 + 9 = 729 \end{aligned}$$

Similarly, $103 \times 97 = (100 + 3)(100 - 3)$

$$\begin{aligned} &= (100)^2 - 3^2 = 10000 - 9 \\ &= 9991 \end{aligned}$$

Now consider the following products:

$$\begin{aligned} (x+3)(x+4) &= x^2 + 7x + 12 \\ &= x^2 + (3+4)x + 3 \cdot 4 \\ (x+5)(x+3) &= x^2 + 8x + 15 \\ &= x^2 + (5+3)x + 5 \cdot 3 \end{aligned}$$

In algebraic symbols, we guess that:

$$(x+a)(x+b) = x^2 + (a+b)x + a \cdot b$$

We can easily prove that our guess is correct. *This type of thinking is very useful in the study of mathematics.*

1.14.1 Algebra is the Best Language for Thinking About Laws

Consider the following table

x:	0	1	2	3	4	5	...
y:	0	2	4	6	8	10	...

We can easily guess the law that lies behind this table. Each number in the bottom row is twice the number that lies above it. The law behind the table is $y = 2x$. In the same way, the law behind the following table is $y = x^2$.

x:	0	1	2	3	4	5	...
y:	0	1	4	9	16	25	...

Incidentally, as a rule, there is little point in putting a law into words. *It is far easier to see what the formula $y = 2x^2 - 5x + 7$ means (by preparing a table, as given above)* than to understand the same formula expressed in words.

1.15 INDUCTION

In mathematics, it is not always wise to proceed by analogy and draw conclusions. The process of reasoning from some particular results to general one is called "*induction*".

As we know, *induction begins by observation*. We observe particular result(s) and use our intuition to arrive at a tentative conclusion—*tentative, because it is an educated guess or a conjecture.* It may be true or false. If the *general result is proved by systematic deductive reasoning, then it is accepted as true*. On the other hand, the result will be considered false if we are able to show a *counter example* where the conjecture fails.

Remember that, a conjecture remains *a conjecture no matter how many examples we can find to support it.* The great French mathematician Pierre de Fermat (1601–1665) observed that:

$$(2^{2^1} + 1) = (2^2 + 1) = 5 \text{ is a prime number.}$$
$$(2^{2^2} + 1) = (2^4 + 1) = 17 \text{ is a prime number.}$$
$$(2^{2^3} + 1) = (2^8 + 1) = 257 \text{ is a prime number.}$$

Accordingly, *he conjectured that* $(2^{2^n} + 1)$ *is a prime number* for every natural number n and had challenged the mathematicians of his day to prove otherwise. It was several years later that the Swiss mathematician Leonhard Euler (1707–1783) showed that $(2^{2^5} + 1) = 4{,}294{,}967{,}297$ *is not a prime number since it is divisible by 641.* Another interesting example is the following:

We observe that the *absolute values of the coefficients of various terms* in each of the following factorization are equal to 1

$$x^1 - 1 = (x - 1); \qquad x^2 - 1 = (x - 1)(x + 1)$$
$$x^3 - 1 = (x - 1)(x^2 + x + 1); \qquad x^4 - 1 = (x - 1)(x + 1)(x^2 + 1)$$
$$x^5 - 1 = (x - 1)(x^4 + x^3 + x^2 + x + 1)$$

Therefore, it was conjectured that when $x^n - 1$ *(n, a natural number) is expressed into factors, with integer coefficients, none of the coefficients is greater than 1, in absolute value.*

All attempts to prove this general statement failed, until 1941, when a Russian mathematician, V. Ivanov came up with a *counter-example*. He found that one of the factors of $x^{105} - 1$ violates the conjecture. This factor is a polynomial of degree 48, as given below.

$$\left.\begin{aligned} &x^{48}+x^{47}+x^{46}-x^{43}-x^{42}-2x^{41}-x^{40}-x^{39}+x^{36}+x^{35}+x^{34}+x^{33}+x^{32}\\ &+x^{31}-x^{28}-x^{26}-x^{24}-x^{22}-x^{20}-x^{17}+x^{16}+x^{15}+x^{14}+x^{12}-x^{9}-x^{8}\\ &-2x^{7}-x^{6}-x^{5}+x^{2}+x+1. \end{aligned}\right\} \text{(10)}$$

In mathematics, we have several such conjectures, which have remained conjectures for lack of proof, even though literally thousands of examples have been found in support of them. *Having employed intuition and arrived at a conjecture, the very difficult task of proving the conjecture begins.* If the conjecture is in the form of a statement, say $P(n)$, involving natural numbers, a method of proof is provided by *the principle of mathematical induction*.(11) [For example, let $P(n)$ represent the statements: (i) $n(n+1)$ is even or (ii) $3^n > n$, or (iii) $n^3 + n$ is divisible by 3, or (iv) $2^{3n} - 1$ is divisible by 7, etc.]

1.16 AN IMPORTANT RESULT: THE NUMBER OF PRIMES IS INFINITE

There is no known formula that relates successive primes to successive integers. Therefore, it is not possible to use the principle of mathematical induction to prove this result. Yet, algebra provides a simple method to prove it. An indirect approach is needed.(12)

1.17 ALGEBRA AS THE SHORTHAND OF MATHEMATICS

Algebra can be compared to writing shorthand in ordinary life. *It can be used either to make statements or to give instructions in a concise form.* Mathematical statements in ordinary language can be translated into algebraic statements and similarly statements in algebra can be translated into ordinary language. For example, consider the following instructions translated into the language of algebra:

(10) *A Textbook of Mathematics for Classes XI–XII* (Book No. 1, p. 100) NCERT Publication, 1978.
(11) To prove that a statement $P(n)$ is true for all natural numbers, we have to go through two steps.

Step (1): We must verify that $P(1)$ is true.

Step (2): Assuming that $P(k)$ is true for some $k \in N$, we must prove that $P(k+1)$ is true. For this purpose, we obtain an algebraic expression for $P(k+1)$ and put it in desired form (if possible) to show that $P(k+1)$ is true. If this is achieved the result is proved to be true for all n.

Remark: If $P(1)$ is not true, the principle of induction does not apply. [See Example (iii) above.]
(12) We assume that every natural number greater than 1, which is not prime can be represented by a product P_1, P_2, P_3, P_4, ..., P_n of prime integers P_i. *This is known as the fundamental theorem of arithmetic.*
Proof: Assume that there is but a finite number of primes and hence a last (largest) prime, P.
Let N be *the product of all primes up to P*: i.e., $N = 2, 3, 5, 7, 11, \ldots, P$. Now consider $N + 1 = (2, 3, 5, 7, 11, \ldots, P) + 1$. Let r be one of the prime numbers 2, 3, 5, ..., P. If we divide $(N+1)$ by r then we will always get the remainder 1. Therefore, $N+1$ itself must be a prime, which is larger than P. This contradicts the assumption that P is the largest prime. [The largest known prime as of March 2011 is $(2^{43,112,609} - 1)$. It has about 700 digits and a modern computer was used to perform the necessary computation. *Mathematics can be Fun* by Yakov Perelman (p. 288), Mir Publishers, Moscow, 1985.]

Statements in Ordinary Language	Equivalent Statements in the Language of Algebra
(i) Think of a number, add 7 to it and double the result.	$2(x + 7)$
(ii) Choose a number, multiply it by 5, add 2, square this expression, and divide the result by 8.	$(5x + 2)^2/8$

Algebra puts mathematical statements in a small space. *The statement is shorter to write, easier to read, quicker to say, and simpler to understand, than the corresponding sentence in ordinary English.*

Next, though it is easy to say that $2n$ (where n is a natural number) represents an even number, *it is not obvious that the number* $(n^2 \pm n)$ *also represents an even number*. Yet, algebra tells us that $(n^2 \pm n) = n \cdot (n \pm 1)$ *must always be even (Why?).*

When we say that algebra is a language, we mean that it has its own words and symbols for expressing what might otherwise be expressed in ordinary language such as French or German. However, we do not look at algebra from this point of view. *For us, algebra is a special kind of language for the following two reasons*:

(a) Algebra is concerned primarily with statement(s) about numbers, items, symbols, or quantities.

(b) The language of algebra uses symbols in place of words.

For example, *to discuss about a class of numbers* (say the class of natural numbers) a mathematician may say: Let "α" be any natural number. Thereafter, in the entire discussion whenever he wishes to refer to *an arbitrary natural number, he will use the letter* α and thus save words and space. Of course, he will have to be careful because any statement(s) he makes about α applies to all natural numbers.

1.18 NOTATIONS IN ALGEBRA

One important difference between the notation of arithmetic and algebra is as follows.

In arithmetic, the product of 3 and 5 is written as 3×5, whereas *in algebra, the product of a and b may be written in any of the forms* $a \times b$, $a \cdot b$, *or* ab. The form ab is the most useful. *In arithmetic, this is not permitted since* 35 means $(3 \times 10) + 5$ and is read as "thirty-five". *Acceptance of such notations in algebra may be treated as a special feature of algebra.*

There are many notations in algebra with which the reader is familiar. For example,

- $a^n = a \cdot \mathrm{a} \cdot \mathrm{a} \cdot \mathrm{a} \cdot \mathrm{a} \ldots$ (n times)
  Example, $3^5 = 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 243$
  We know that, $a^7/a^4 = a^{7-4} = a^3$
  $\therefore \quad a^n/a^n = a^{n-n} = a^0 = 1$, (provided $a \neq 0$)
  ($a^0 = 1$), $a \neq 0$, since, 0^0 is not defined.
- Product of first n natural numbers is given by
  $n! = n \cdot (n-1) \cdot (n-2) \cdot (n = 3) \ldots 3 \cdot 2 \cdot 1$
  Example, $7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$

- Number of permutations (arrangements) of *n different things taken r at a time is given by*

$$^{n}p_r = \frac{n!}{(n-r)!} \quad \text{Example, } ^{5}p_3 = \frac{5!}{(5-3)!} = \frac{5!}{2!}$$

$$= \frac{\text{Product of first 5 natural numbers}}{\text{Product of first 2 natural numbers}}$$

$$= \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = 60$$

$$^{n}p_n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = \frac{\text{Product of first } n \text{ natural numbers}}{\text{Product of first "zero" natural numbers}}$$

$$= n!$$

It follows that $0! = 1$. (This is taken as the definition of 0!)

- Number of combinations of n different items taken r at a time; is given by $^{n}C_r = \frac{n!}{r!(n-r!)}$.

Example: $^{7}C_3 = \frac{7!}{3!(7-3)!} = \frac{7!}{(3!)(4!)} = \frac{7.6.5.4.3.2.1}{(3.2.1)(4.3.2.1)} = 35.$

$$^{n}C_r = {}^{n}C_{(n-r)},\ ^{n}C_0 = 1,\ ^{n}C_n = 1$$

Note that in all these notations, *n is a natural number and r is a whole number, with* $n \geq r$.

A beginner may complain about some difficulty in learning the language of algebra. However, *one who has mastered this language of mathematics* and has grasped the ideas and reasoning, *does appreciate the mathematical symbolism.* It is a relatively modern invention and *mathematicians should be complimented for designing "symbols" and "notations", out of necessity.*

It is important to realize that, while all the languages of the world are quite different from one another, *the language of algebra* is a common one (as is the language of mathematics) and *serves the purpose so well.*

1.19 EXPRESSIONS AND IDENTITIES IN ALGEBRA

The basic function of algebra is to convert expressions into more useful ones. For example, the sum.

$$\sum_{k=1}^{n} k = \sum n = 1 + 2 + 3 + 4 + \ldots\ldots + n$$

was converted by Gauss to the more useful form $(n(n+1)/2)$.

How do you prove this?

The method is not obvious and yet a simple idea does the trick, as follows:

$$\text{Let } S = 1 + 2 + 3 + 4 + \ldots\ldots + (n-1) + n \tag{3}$$

$$\text{Also, } S = n + (n-1) + (n-2) + \ldots\ldots + 2 + 1 \tag{4}$$

Adding corresponding terms in (3) and (4), we get

$$\begin{aligned} 2S &= (n+1)+(n+1)+(n+1)\ldots\ldots(n\,\text{times}) \\ &= n(n+1).\ \text{Hence,}\ S = (n(n+1)/2) \end{aligned}$$

The ideas in this proof must arouse some excitement in the reader's mind. *Here, it is important to realize that by simple means we have converted the cumbersome expression to a simpler and readily computable expression.*

Similarly, using algebra, many such useful expressions can be obtained easily. For example,

- $\sum n^2 = 1^2+2^2+3^2+4^2+\ldots+n^2 = \dfrac{n(n+1)(2n+1)}{6}$

- $\sum n^3 = 1^3+2^3+3^3+4^3+\ldots+n^3 = \dfrac{n^2(n+1)^2}{4}$

  $\left[\text{Note that } \sum n^3 = \left(\sum n\right)^2\right]$

- $a+ar+ar^2+ar^3+\ldots+ar^{n-1}$

  $= \dfrac{a(1-r^n)}{(1-r)},\ (r<1)$

  $= \dfrac{a(r^n-1)}{(r-1)},\ (r>1)$

It is sometimes possible that a question may have two answers which at first sight appear different, but which are actually the same. This can be checked by simplifying both the algebraic expressions. *An important part of algebra therefore consists in learning how to express any result in the simplest form.* Algebraic identities,[13] and methods available for factorizing polynomials, are helpful in simplifying algebraic expressions.

Some important identities are given below:[14]

- $(x+y)(x-y) = x^2-y^2.$

  Thus, $(a+b)(a-b) = a^2-b^2.$

- $(x+y)^2 = x^2+y^2+2xy.$

  Thus, $a^2+b^2 = (a+b)^2-2ab.$

[13] An algebraic statement expressed in two (or more) forms with a symbol of equality (=) between them is called an algebraic identity. *Obviously, an identity is true for all real value(s) of the variable(s) involved.*

[14] For some purpose, the expression a^2-b^2 is useful as it stands, but for others it may be better to write it in the equivalent form $(a+b)(a-b)$. This statement is also applicable for other expressions to follow.

- $(x-y)^2 = x^2+y^2-2xy$.

  Thus, $a^2+b^2 = (a-b)^2+2ab$,

  $(a+b)^2+(a-b)^2 = 2(a^2+b^2)$,

  and $(a+b)^2-(a-b)^2 = 4ab$.
- $(x+y)^3 = x^3+y^3+3xy(x+y)$.

  Thus, $a^3+b^3 = (a+b)^3-3ab(a+b)$,

  or $a^3+b^3 = (a+b)(a^2-ab+b^2)$.
- $(x-y)^3 = x^3-y^3-3xy(x-y)$.

  Thus, $a^3-b^3 = (a-b)^3+3ab(a-b)$,

  or $a^3-b^3 = (a-b)(a^2+ab+b^2)$.

From the expression(s) for $(a\pm b)^3$ and $(a\pm b)^2$ many useful identities can be obtained. For example,

$$\frac{a^3+b^3}{a^2+b^2-ab} = (a+b),\ \frac{a^3-b^3}{a^2+b^2+ab} = (a-b)$$

$$\frac{(a+b)^2+(a-b)^2}{(a^2+b^2)} = \frac{2(a^2+b^2)}{(a^2+b^2)} = 2,$$

$$\frac{(a+b)^2-(a-b)^2}{ab} = \frac{4ab}{ab} = 4.$$

Next, observe that,

- $$\left.\begin{aligned}\left(a+\frac{1}{a}\right)^2 &= a^2+\frac{1}{a^2}+2\\ \left(a-\frac{1}{a}\right)^2 &= a^2+\frac{1}{a^2}-2\end{aligned}\right\} \therefore \left(a+\frac{1}{a}\right)^2-\left(a-\frac{1}{a}\right)^2 = 4$$

- $(a+b+c)^2 = a^2+b^2+c^2+2(ab+bc+ca)$
- $a^3+b^3+c^3-3abc = (a+b+c)(a^2+b^2+c^2-ab-bc-ca)$
- If $a+b+c = 0$, then $a^3+b^3+c^3 = 3abc$.
- $\dfrac{1}{a\cdot b} = \dfrac{1}{b-a}\left[\dfrac{1}{a}-\dfrac{1}{b}\right]$
- If n is a *natural number*, then the expansion $(x+y)^n = {}^nC_0x^n+{}^nC_1x^{n-1}\cdot y+{}^nC_2x^{n-2}\cdot y^2+\ldots+{}^nC_ny^n$ is *called the binomial expansion*, where x and y can be any *real numbers.*
  - This expansion has $(n+1)$ terms.
  - The general term is of the form ${}^nC_rx^{n-r}y^r$ and *it is the* $(r+1)$th *term* in the expansion.
  - In each term, *the sum of the indices* of x and y, is n.
- If m is *a negative integer* or *a rational number*, then the *binomial expansion* is

$$\begin{aligned}(b+x)^m &= b^m+mb^{m-1}x+\frac{m(m-1)}{2!}b^{m-2}x^2+\ldots\\ &+\frac{m(m-1)(m-2)\ldots(m-r+1)}{r!}b^{m-r}x^r+\ldots \qquad (5)\end{aligned}$$

provided $|x|<b$

Remark (1): Note that the coefficients m, $(m(m-1)/2!)$, and so on, *look like combinatorial coefficients* (i.e., ${}^nC_0$, ${}^nC_1$, ${}^nC_2$, $\ldots$, ${}^nC_r$, and so on). However, recall that *${}^nC_r$ is defined for natural number n and whole number r* (with $n \geq r$), and as such has *no meaning* in other cases.

Remark (2): When m is *a negative integer or a rational number*, there are *infinite number of terms* in the expansion of $(b + x)^m$.

Remark (3): The following results are very useful and can be easily obtained by using the expansion in Equation (5).

- $\dfrac{1}{1+x} = (1+x)^{-1} = 1 - x + x^2 - x^3 + \ldots; \ |x| < 1$
- $\dfrac{1}{(1+x)^2} = (1+x)^{-2} = 1 - 2x + 3x^2 - 4x^3 + \ldots; \ |x| < 1$
- $\dfrac{1}{1-x} = (1-x)^{-1} = 1 + x + x^2 + x^3 + \ldots; \ |x| < 1$
- $\dfrac{1}{(1-x)^2} = (1-x)^{-2} = 1 + 2x + 3x^2 + 4x^3 + \ldots; \ |x| < 1$

1.20 OPERATIONS INVOLVING NEGATIVE NUMBERS

A good deal of the machinery of elementary algebra is concerned with *the solution of equations involving unknowns*. However, we should note that this simple machinery can lead directly to useful results in numerous other types of problems.

The most difficult item in algebra is that devoted to operations involving negative numbers. The difficulty is twofold:

(i) Why introduce negative numbers?
(ii) Why does multiplication of two negative numbers (or division of a negative number by another negative number) yield a positive number?

In fact, it is in connection with the solution of equations, that both questions can be answered. For example, note that if we do not accept negative numbers then even a simple equation, like $2x + 5 = 0$ cannot be solved. Next, consider the equation

$$7x - 5 = 10x - 11 \tag{6}$$

To solve this equation, *we can transpose the terms in two ways so that the unknowns are on one side and the knowns are on the other side*. (Of course, we will expect that in both the cases the solution should be same.)

Thus, we get

$$11 - 5 = 10x - 7x$$

$$\text{or} \quad 6 = 3x \quad \text{so} \quad x = 2$$

Also, we get

$$7x - 10x = -11 + 5$$
$$-3x = -6.$$

$$x = \frac{-6}{-3} = \frac{(-1)\cdot 6}{(-1)\cdot 3} = \frac{(-1)}{(-1)}\cdot\frac{6}{3} = \frac{(-1)}{(-1)}\cdot 2 \quad (7)$$

$$\text{Also,} \quad \frac{-6}{-3} = (-1)\cdot 6\cdot\frac{(-1)}{3}\left[\because \frac{1}{-3} = \frac{1}{(-1)3} = (-1)\frac{1}{3} = \frac{(-1)}{3}\right]$$
$$= (-1)\cdot(-1)\cdot\frac{6}{3} = (-1)\cdot(-1)\cdot 2 \quad (8)$$

Now in order that the solution of the Equation (6) should be same, it is necessary that $(-1)/(-1) = 1$ in (7) and $(-1)(-1) = 1$ in (8).

1.21 DIVISION BY ZERO

The question, *"Why is division by zero not permitted in mathematics?"* is answered through algebra.

In arithmetic (or more generally in algebra), the operation of division is defined in terms of the operation of multiplication. Thus according to the existing rule, the division of an arbitrary number "a" by another number "b" means to find a number x such that

$$a\cdot\frac{1}{b} = x \quad \text{where} \quad b \neq 0$$
$$b\cdot x = a$$

or

Let us see what happens if division by zero is permitted. If $b = 0$, then we must consider the following two cases.

(i) when $a \neq 0$, and
(ii) when $a = 0$

Case (i): We try to solve the equation

$$b\cdot x = a, \text{ (where } b = 0, \text{ but } a \neq 0)$$
We get $$0\cdot x = a$$

It follows that $a = 0$, which is against our assumption that $a \neq 0$. This situation arises because there is no number x, which could be multiplied by "0" to get a fixed (nonzero) number "a". It follows that if a nonzero number is divided by zero than we get a meaningless result.

Case (ii): We try to solve the equation

$$b\cdot x = a, \text{ (where } b = 0, \text{ and } a = 0)$$
We get $$0\cdot x = 0$$

Unfortunately, this is true. Here any number x satisfies this equation. Let us see the consequence of this situation.

If division by zero is permitted, then we get from the equation $0\cdot x = 0$, $x = 0/0$. Similarly from $0\cdot y = 0$, we get $y = 0/0$, where $x, y, \ldots$ are all different (nonzero) numbers. From the above, it follows that $0/0 = x = y = z \ldots$, which means that all different numbers are equal.

Thus, if $a = 0$, and $b = 0$, then we have $a/b = 0/0$ and it represents any number whatever we choose. But mathematicians require that the division of "a" by "b" should yield a unique number as a result. But this is again not achieved.

From the above, we observe that *division by zero leads either to no number or any arbitrary number*. (Note that this is the consequence of permitting division by the number zero.) Thus, *division by zero leads to meaningless results and hence it is not permitted in mathematics.*

2 The Concept of a Function

2.1 INTRODUCTION

The concept of a "function" is one of the most basic in all of mathematics. The meaning of *the word "function" has evolved and changed during the last three centuries*. Its modern meaning is much broader and deeper than its elementary meaning from earlier days. The statement: "*y is a function of x*" *means something very much like* "*y is related to x by some formula*". In fact, this statement gives some idea about a function, but it is incomplete. In traditional algebra, x and y stand for numbers. But *today, functions can be defined that have nothing to do with numbers*.

In our study of calculus, we shall be mostly concerned with functions, which are related to numbers. Like any other mathematical concept, *the concept of function is nicely expressed through the language of sets*. Therefore, it is useful to revise "*Elementary Set Theory*" (see Appendix "A").

Assuming the knowledge of Elementary Set Theory, we define two important terms: (i) *ordered pairs* and (ii) *Cartesian product of sets*. These terms are needed to define a "function" on the basis of set theory. Let us discuss:

(i) *Ordered Pairs*: When we wish to consider *a pair of things as a whole*, we may use the terms *couple* or just *pair*. If $A = \{1, 2, 3, 4\}$ then the subsets $\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 1\}, \{3, 1\}$ are some examples of pairs. Here we have listed *some pairs twice*; for example $\{1, 2\} = \{2, 1\}$ and $\{1, 3\} = \{3, 1\}$.
We know that *the order, in which the elements of a set are written, is immaterial*. If in a pair we wish to single out one element as being the first, then the other element becomes the second. *Once we define the procedure of fixing the position of first element (in a pair), we have example of an ordered pair*. To denote an ordered pair we use the following notation:

The ordered pair consisting of the element 1 and 2, *in which 1 is the first element will be written as (1, 2)*, whereas *the ordered pair consisting of the elements 1 and 2 in which 2 is the first element will be written as (2, 1)*. Obviously, then $(1, 2) \neq (2, 1)$.[1]

What must you know to learn calculus? 2-The concept of function (Relations and functions, one-to-one correspondence, equivalent sets, infinite sets, the notion of infinity, and its algebra)

[1] It is necessary to consider the sets of elements in which order is important. For example, in analytic geometry of the plane, the coordinates (x, y) of a point represent an ordered pair of numbers. The point (3, 4) is different from the point (4, 3). Similarly, in 3D geometry, an ordered triplet (a, b, c) gives the coordinates of a point in 3D space. There are some authors who use the notation $<\langle a, b\rangle$ for the ordered pair, and (a, b) for the open interval, but the ambiguity need not cause any alarm because it will always be made clear by context and we will know which role the symbol (a, b) is to play.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

(ii) *Definition: Cartesian Product of Two Sets A and B*: Let A and B be two sets. We define the Cartesian product of A and B, written $A \times B$, to be the set of all ordered pairs (x, y) where $x \in A$ and $y \in B$.
Thus, $A \times B = \{(x, y) | x \in A \text{ and } y \in B\}$

Example: Let $A = \{1, 2, 3\}$ and $B = \{5, 6\}$

Then $A \times B = \{(1, 5), (2, 5), (3, 5), (1, 6), (2, 6), (3, 6)\}$

Note: It is important to note that in the product $A \times B$ the first element in each ordered pair belongs to A and the second element to set B. Also note that if set A contains m elements and set B contains n elements, then the set $A \times B$ will have $m{\cdot}n$ ordered pairs.[2]

2.2 EQUALITY OF ORDERED PAIRS

Two ordered pairs (a, b) and (c, d) are equal if $a = c$ and $b = d$.

2.3 RELATIONS AND FUNCTIONS

We know that if set A contains m elements and set B contains n elements, then the set $A \times B$ will have $m{\cdot}n$ ordered pairs. *Any subset of these ordered pairs is called a relation from A to B.*

Consider the following example:

Example: Let $A = \{1, 2, 3, 4\}$ and $B = \{2, 4, 5\}$

Then, $A \times B = \{(1,2), (1,4), (1,5), (2,2), (2,4), (2,5), (3,2), (3,4), (3,5), (4,2), (4,4), (4,5)\}$
Now there are many *relations* from A to B as follows:

$R_1 = \{(1, 2), (1, 5), (2, 2), (3, 4), (3, 5), (4, 5)\}$
$R_2 = \{(1, 4), (4, 2), (4, 5)\}$
$R_3 = \{(3, 2), (3, 4), (3, 5), (1, 4)\}$
$R_4 = \{(1, 4), (2, 5), (3, 2), (4, 4)\}$
$R_5 = \{(1, 2), (2, 5), (3, 4), (4, 4)\}$

However, *if we select the ordered pairs in such a way that*:

(i) their first elements constitute the *entire* set A, and
(ii) no two *distinct* pairs have the same first element,

then such a collection of ordered pairs (from the set $A \times B$) constitute a *special relation from A to B*, which is called a *function* from A to B.

[2] Here is a tricky situation: Let $A = \{1, 2, 3\}$, $B = \phi$ then $A \times B = \phi$ (Why?) Note that $A \times B$ is not defined *as a set of ordered pairs* if either A or B is empty. However, $A \times \{\phi\} = \{(1, \phi), (2, \phi), (3, \phi)\}$. But this set of ordered pairs is of no use to us.

2.3.1 Domain of a Relation

In any relation (in the form of a set of ordered pairs), the set consisting of the "first" element of each pair constitutes the *domain* of the relation. Let us consider the above relations.

- The domain of $R_1 = \{1, 2, 3, 4\} = A$. But there are two ordered pairs (1, 2) and (1, 5) in which the first element is same. *Hence, R_1 does not represent a function.*
- The domain of $R_2 = \{1, 4\} \neq A$. Also, there are two ordered pairs (4, 2) and (4, 5) in which the first element is the same. Hence, R_2 is not a function, from A to B.[3] Obviously, R_3 is also not a function.
- The domain of $R_4 = \{1, 2, 3, 4\} = A$. And no two distinct pairs have the same first element. Hence, R_4 represents a function. Similarly, R_5 represents a function.

We can still define many functions from A to B, as follows:

$f_1 = \{(1, 2), (2, 2), (3, 2), (4, 2)\}$,
$f_2 = \{(1, 4), (2, 2), (3, 5), (4, 4)\}$,
$f_3 = \{(1, 5), (2, 4), (4, 2), (3, 2)\}$,
$f_4 = \{(1, 5), (2, 5), (4, 5), (3, 5)\}$, and so on.

We are now in a position to define a "*function*" *on the basis of set theory.*

2.4 DEFINITION

Let A and B be two nonempty sets.

A function f from A to B is a subset of $A \times B$ (involving the entire set A) with the property that *each* "a" belonging to A, belongs to precisely one ordered pair (a, b), in the subset of $A \times B$, under consideration. In other words, a function f from *set* A to *set* B consists of a set of ordered pairs $(a, b) \in A \times B$ such that *no two ordered pairs have the same first element.*

2.4.1 Alternative Definition of a "Function"

A function f from set A to set B (written as $f: A \to B$) *is a rule of correspondence that associates to each element of A, one and only one element of B.*

(A function is also called a *mapping* from A to B.)

We observe that

(i) Each element of B need not be in the association, *but every element of A must be involved in it.* Hence, a function is a *one way pairing process*. (Every element of A pairs off with some element of B but not conversely.)

(ii) One element of A cannot be associated to more than one element of B, but one element of B may correspond to two or more elements of A.

The correspondence from the elements of set A to set B, shown in Figures 2.1–2.4 represents function(s) whereas that shown in Figures 2.5 and 2.6 does not represent functions. (Why?)

To study functions in details, it is useful to fix certain terms, which will be needed frequently.

[3] Note that there are two reasons due to which R_2 is not a function. In fact, any one reason is sufficient for this conclusion.

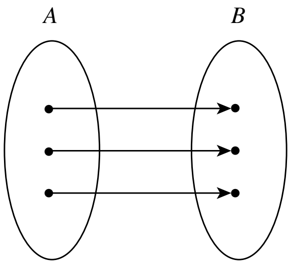


FIGURE 2.1

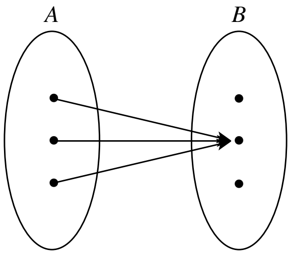


FIGURE 2.2

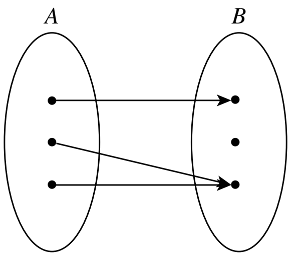


FIGURE 2.3

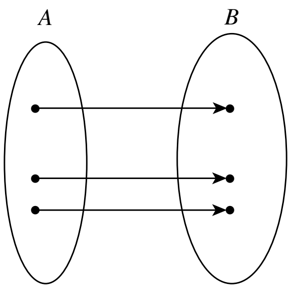


FIGURE 2.4

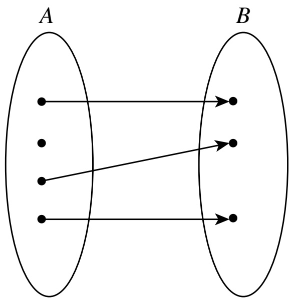


FIGURE 2.5

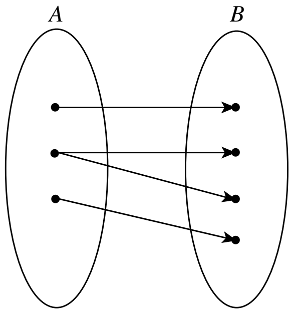


FIGURE 2.6

2.5 DOMAIN, CODOMAIN, IMAGE, AND RANGE OF A FUNCTION

Let f be a function from set A to set B (f: $A \to B$), then

- The (entire) set A is called *the domain* of f.
- The (entire) set B is called *the codomain* of f.
- An element of B that corresponds to some element x of A is denoted by $f(x)$, and it is called *the image* of x under f.
- The set of all images constitute *the range* of f. The *range of f is denoted by $f(A)$* and it is a *subset* of set B. In other words $f(A) \subseteq B$.

2.6 DISTINCTION BETWEEN "f" AND "$f(x)$"

Consider a function f: $A \to B$, and let (x, y) be an *arbitrary* ordered pair belonging to f. Then, instead of writing $(x, y) \in f$, we usually write $y = f(x)$ to mean that *y is related to x, through* "f", and we read it as y [or $f(x)$] is a function of x. In this notation, *x represents an arbitrary element of the domain* and *y represents the corresponding element of the range*. Remember that the element "$f(x)$" is selected by the rule of correspondence defined by the function "f". Thus, "f" represents the rule of correspondence.

It *is important to distinguish between the symbols f and $f(x)$*. It may be emphasized that a *single letter f* (or g, or h, or ϕ, etc.) is used to name a function. Remember that "f" *represents the* *rule of correspondence* (which may be a *statement*[4] or a *method* or a *formula*) relating the elements of set A to the elements of set B. *The symbol f(x) represents the result of applying "f" to "x". We call it the value of "f" at x.* If "f" is the function defined by the formula $x^2 - 2x + 3$, then we usually write

$$y = f(x) = x^2 - 2x + 3, \ (x \in A) \tag{I}$$

Equation (I) tells that the rule of correspondence is the formula "$x^2 - 2x + 3$", which converts every number $x \in A$, into a *new number* "$f(x)$" belonging to B, using the above formula. Thus, *the formula* "$x^2 - 2x + 3$" *must be identified with the function* "f". For computing the value "$f(x)$", we use the formula defining "f". Thus, if we choose $x = 0 \in A$, we get the corresponding element $y \in B$ as $f(0) = 3$. Similarly, for $x = 1$ we get $y = 2$, and for $x = 5$, $y = 18$.

To avoid the confusion, *possibly caused due to equation (I)* [wherein $f(x)$ is equated to $x^2 - 2x + 3$], we should write it as f: $x \to x^2 - 2x + 3$, which clearly states that "f" is a function that converts each x into $x^2 - 2x + 3$.

2.7 DEPENDENT AND INDEPENDENT VARIABLES

When the rule for a function is given by an equation of the form $y = f(x)$ (for example, $y = x^2 - 2x + 3$, or $y = \sin x$, or $y = e^x$, etc.), then x is called the *independent variable* and y [or $f(x)$] is called the *dependent variable. Note that for the dependent variable y* [or $f(x)$], *we look at the expression for f(x).* (For more details see Chapter 6.)

2.8 FUNCTIONS AT A GLANCE

(1) A function consists of three things:
   (i) A set known as *the domain* of the function.
   (ii) A set known as *the range* of the function.
   (iii) A *correspondence* (a *rule* or a *method* or a *procedure*), which associates with each member of the domain, precisely one member of the range.

(2) If we conceive of a function as being specified by a set of ordered pairs, then we must insist that *no two selected distinct pairs may have the same first element.*

2.9 MODES OF EXPRESSING A FUNCTION

A function is completely known if the objects and corresponding images are known. There are many ways in which this can be done. We give below four methods for describing a function.

(i) *Statement of the Rule of Association (By Formula or Otherwise)*: If the domain of the function is known and rule(s) of association between objects and images are known, the images can be found out and the correspondence is completely known.

[4] For example, nth digit in the decimal expansion of π, is a function, which has no formula.

As we go through the other modes of expressing a function, it will be realized that this mode of expressing a function is the most accurate and complete, when the domain is an infinite set.

(ii) *Description by Tables*: When the domain of a function consists of a small number of elements (say 10–20 numbers), it may be described by tabulating the objects and their corresponding images.
This is *specifically useful when the objects and images cannot be connected by a fixed rule due to irregular variations, and so on*. (This mode of expressing a function however, is *not useful when the domain has a large number of elements*.)

(iii) *Description in Terms of Ordered Pairs*: A function is expressible as a set of ordered pairs. *In fact, the set-theoretic definition of a function is the basis for this mode of expressing a function*. We have already discussed about this mode earlier. In general, this mode is *not useful* in handling problems in calculus.

(iv) *Description by Graphs*: We know that *a function is a set of ordered pairs, and each ordered pair is liable to be represented as a point in the plane*. Therefore, the function itself is represented by the set of these points. If $f: A \to B$ is a function then the set $\{(a, f(a)) | a \in A\}$ is called *the graph of f and is a subset of* $A \times B$. In particular if $A, B \subseteq R$, the graph of f can be represented by points in the plane. The graph of a function may be a set of distinct points or it may be a (continuous) curve.

Example: Consider the function: $f(x) = 2x + 1,\ x \in \{0, 1, 2, 3\}$

Here, $A = \{0, 1, 2, 3\}$.
The graph of this function consists of four *isolated points* with the coordinates (0, 1), (1, 3), (2, 5), (3, 7) and is *not a continuous curve*. However, the graph of the function $f(x) = 2x + 1$, $x \in R$ is a *continuous curve* (*line*) passing through the above four points.

2.10 TYPES OF FUNCTIONS

We know that a *relation* $f: A \to B$, which satisfies the following *two* conditions, *will be called a function*:

(I) Each element of the domain A is involved in the relation.

(II) Each element of A is associated to exactly one element of B, and not more than one element of B.

Remarks:

(a) Note that both the specifications are imposed on the elements of A, and that no restriction is imposed on the elements of the codomain B.

(b) If we make restrictions (I) and (II) on the codomain B, we get *two special kinds of functions namely* (i) *one–one function* and (ii) *onto function* as discussed below in (A) and (C), respectively.

(A) *One–One Function*: A function is one-one provided *distinct elements of the domain* are related to *distinct element of the codomain*. In other words, a function $f: A \to B$ is

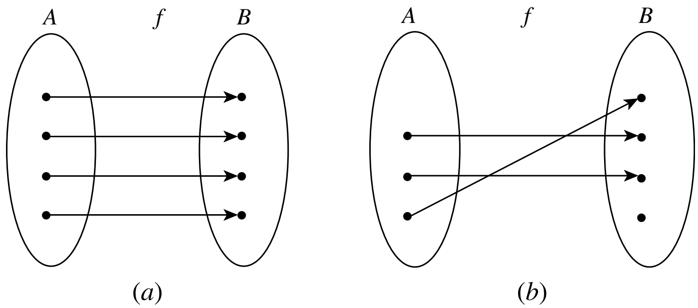


FIGURE 2.7

defined to be one-one if the images of distinct element of A under f are distinct, that is, for every $a_1, a_2 \in A$, $f(a_1) = f(a_2) \Rightarrow a_1 = a_2$. [It also means that, $f(a_1) \neq f(a_2) \Rightarrow a_1 \neq a_2$.] A *one–one function* is also called *injective function* (Figure 2.7a and b).

Note: If there is at least one pair of *distinct* elements, $a_1, a_2 \in A$, such that

$$f(a_1) = f(a_2) \text{ [though } a_1 \neq a_2\text{]}$$

then, such a function is called *many–one*. We define many-one function as follows:

(B) *Many–One Function*: If the codomain of the function has at least one element, which is the image for two or more elements of the domain, then the function is said to be *many–one function* (Figure 2.8a and b).
A constant function is a special case of many–one function (Figure 2.9).

(C) *Onto Function*: A function f: $A \to B$ is called an *onto function if each element of the codomain is involved in the relation.*
(Here, range of f = codomain B.)
In other words, a function f: $A \to B$ is said to be onto *if every element of B is the image of some element* of A, under f, that is, for every $b \in B$, there exist an element $a \in A$ such that $f(a) = b$ (Figure 2.10a and b). Onto function is also called *surjective function.*
The most important functions are those which are both one–one and onto. In a function that is one–one and onto, *each image corresponds to exactly one element of the domain*

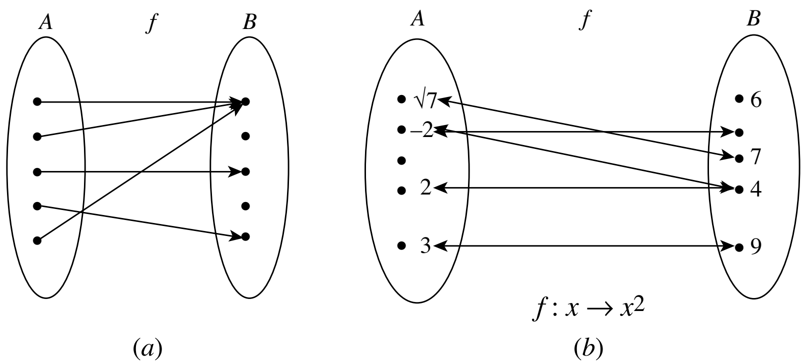
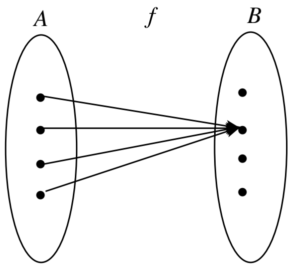


FIGURE 2.8



Constant function
(many–one function)

FIGURE 2.9

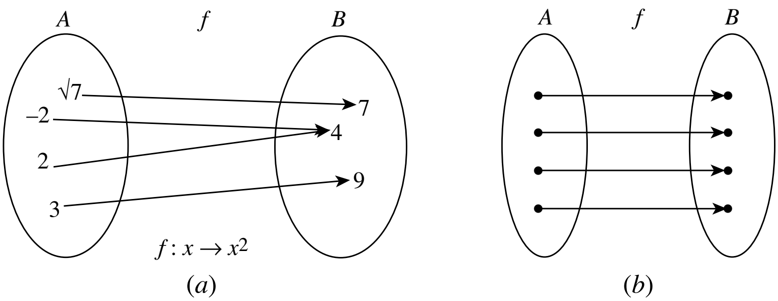
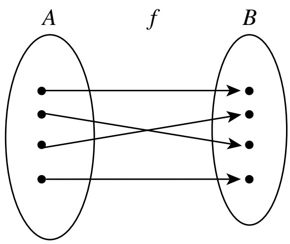


FIGURE 2.10



One–one and onto function

FIGURE 2.11

and *each element of codomain is involved in the relation* as shown in Figure 2.11. Such a *function* is also called *one-to-one correspondence* or a *bijective function.*

(D) *Bijective Function (or One-to-One Correspondence)*:

Definition (1): Consider a function f: $A \to B$ with "A" as the domain of definition (i.e., the admissible set of the values of x) and "B" as the range (i.e., the set of corresponding values of y).

We say that the function $y = f(x)$ specifies a mapping of the set "A" *onto* the set B: *if two different points of the set A, there correspond different points of the set B, and the entire set B is the range of* "f". Such a mapping is called one-to-one onto mapping.[(5)]

Definition (2): Let A and B be two *nonempty* sets. A rule f, which, associates with *each element* "a" *of the set A, exactly one element* "b" *of the set B,* and under which, *each element* "b" *of set B corresponds to exactly one element* "a" *of the set A*, is called a *one-to-one correspondence between the sets A and B*.[(6)]

Example (1): Consider the function $y = f(x) = x^3$. Here, for every value of $x \in R$, there corresponds a single value of y, and, conversely, to each $y \in R$, there corresponds a single value of x given by $\sqrt[3]{y}$. Therefore, f specifies a one-to-one mapping, from R onto R.

Example (2): Consider the function $y = g(x) = x^2$. Here, *for every value of* $x \in R$, *there corresponds a single value of* $y \in (0, \infty)$. However, to *every* $y > 0$, there correspond two values of x: $x = \pm\sqrt{y}$. Therefore, "g" *is not one-to-one correspondence.*

Example (3): Consider the exponential function $y = f(x) = e^x$. It can be shown that the function $f(x) = e^x$ is one-to-one mapping from $(-\infty, \infty)$ onto $(0, \infty)$. Note that for $x_1 \neq x_2$, we have $e^{x_1} \neq e^{x_2}$, where $x_1, x_2 \in R$, and $e^{x_1}, e^{x_2} \in R^+$. Consider $e^{x_1}/e^{x_2} \neq 1 \Rightarrow e^{x_1 - x_2} \neq 1$ or $e^{x_1 - x_2} \neq e^0$ (since $e^0 = 1) \Rightarrow x_1 - x_2 \neq 0 \Rightarrow x_1 \neq x_2$. In other words, $e^{x_1} \neq e^{x_2} \Rightarrow x_1 \neq x_2$. Thus, $x_1 \neq x_2 \Leftrightarrow e^{x_1} \neq e^{x_2}$.

Therefore, "f" defines a one-to-one correspondence from $(-\infty, \infty)$ onto $(0, \infty)$. (Here it is important to note that a one-to-one mapping has been defined from the entire real line on to the positive part of the real line.)[(7)]

(5) We distinguish between one–one mapping (which need not be onto mapping and one-to-one correspondence, which is a one-to-one and onto mapping). In the case of one–one mapping

$$x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2) \tag{III}$$

On the other hand, in the case of one-to-one onto mapping

$$x_1 \neq x_2 \Leftrightarrow f(x_1) \neq f(x_2) \tag{IV}$$

(6) If a function $y = f(x)$ performs a one-to-one mapping of a set A onto a set B, then the same correspondence considered in the reverse order assigns to every y belonging to the set B, a corresponding element x belonging to the set A. This reverse correspondence may be looked upon as defining a function $x = \phi(y)$, whose domain of definition is the set B and range the set A. Such a reverse correspondence has a special name—the inverse of f, to be discussed shortly.

(7) If we have two sets A and B, each with infinite number of elements, then it is possible to define one-to-one and onto mapping on them, irrespective of the observation that the one might appear smaller than the other. This will be clear shortly when we discuss the concept of infinity and define such functions on infinite sets.

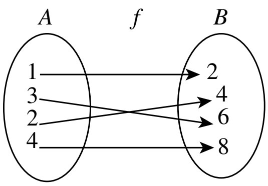
2.11 INVERSE FUNCTION f^{-1}

If a function "f" is one-to-one and onto, *then the correspondence associating the same pairs of elements in the reverse order is also a function.* This reverse function is denoted by f^{-1}, and we call it *the inverse of the function f.* Note that, f^{-1} *is also one-to one and onto.*

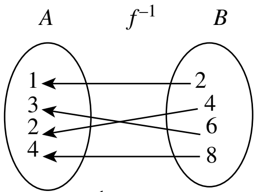
Remark: A function f has an inverse provided that there exists a function, f^{-1} such that

(i) the domain of f^{-1} is the range of f and
(ii) $f(x) = y$ *if and only if* $f^{-1}(y) = x$ *for all* x in the domain of "f" and *for all* y in the range of "f".

Note (1): *Not every function has an inverse.* If a function $f: A \rightarrow B$ has an inverse, then f^{-1}: $B \rightarrow A$ is defined, such that, the domain of f^{-1} is the range of f, and the range of f^{-1} is the domain of f, associating the same pairs of elements.



$f: A \rightarrow B$
One–one and onto



$f^{-1}: B \rightarrow A$
One–one and onto

It can be shown that if f has an inverse, then the inverse function is uniquely determined. *Sometimes, we can give a formula for* f^{-1}.

For example if $y = f(x) = 2x$, then $x = f^{-1}(y) = (1/2)y$. Similarly, if $y = f(x) = x^3 - 1$, then $x = f^{-1}(y) = \sqrt[3]{y+1}$, In each case, *we simply solve the equation that determines x in terms of y*. The formula in y expresses the (*new*) function f^{-1}.

We cannot always give the formula for f^{-1}. For example, consider the function $y = f(x) = x^5 + 2x + 1$. *It is beyond our capabilities to solve this equation for x.* (Why?)[(8)]

Note that, in such cases, we cannot decide whether a given function has an inverse or not. *Fortunately, there are criteria that tell whether a given function y = f(x) has an inverse, irrespective of whether we can solve it for x.*[(9)]

In the case of simple functions (*like linear functions*, etc.) there is a three-step process that gives a formula for the inverse.

Step (1): Solve the equation $y = f(x)$ for x, in terms of y.
Step (2): Use the symbol $f^{-1}(y)$ to name the resulting expression in y.
Step (3): Replace y by x to get the formula for $f^{-1}(x)$.

[(8)] A general polynomial equation of degree ≥ 5 cannot be solved in terms of the coefficients involved (see the relevant footnote in Chapter 1)

[(9)] A practical criterion is that "f" be strictly monotonic (i.e., either strictly increasing or strictly decreasing). For this purpose the simple and practical way is to check the sign of derivative of the function f. This will be clear when we have discussed the concept of derivative of a function (In Chapter 9 to follow) and its applications in Chapter 19a.

Example (1): Consider the function $y = f(x) = 3x - 2$, $x \in R$, and let us find its inverse function.

Solution:

Step (1): $y = f(x) = 3x - 2 \qquad \therefore x = \dfrac{y+2}{3}$

Step (2): $f^{-1}(y) = \dfrac{y+2}{3}$

Step (3): $f^{-1}(x) = \dfrac{x+2}{3}$

Example (2): Let us find the formula for $f^{-1}(x)$ if $y = f(x) = \dfrac{x}{1-x}$.

Step (1): $y = \dfrac{x}{1-x}$

$$\therefore (1-x)y = x \text{ or } y - yx = x \text{ or } y = x + yx = x(1+y)$$

$$\therefore x = \frac{y}{1+y}$$

Step (2): $f^{-1}(y) = \dfrac{y}{1+y}$ $(y \neq -1)$

Step (3): $f^{-1}(x) = \dfrac{x}{1+x}$ $(x \neq -1)$

$$\text{Whenever a function } y = f(x) \qquad \text{(i)}$$

has an inverse function, *that we can solve for x*, then we can write it as

$$x = f^{-1}(y) \qquad \text{(ii)}$$

We see that in this expression (of the inverse function) the roles of variables x and y are interchanged: Both the functions at (i) and (ii) *describe one and the same curve in the xy-plane*, and they are said to be mutually inverse functions.

For the function f, the axis of the independent variable is the x-axis, while for the function f^{-1} this role is played by the y-axis. If we want to construct the graphs of mutually inverse functions so that *the axis of arguments (i.e., the axis of independent variables) for both of them is the x-axis*, then we should denote the independent variable in formula (ii) by x and express the inverse function in the form: $y = f^{-1}(x)$.

In this notation, the letter x designates the independent variable and the letter y the dependent variable for both the mutually inverse functions. Thus the functions $y = x^3$ and $y = \sqrt[3]{x}$, represent a pair of *mutually inverse functions*. Also $y = 10^x$ and $y = \log_{10} x$ are *mutually inverse functions*.

There is a *simple relationship between the graphs of two mutually inverse* functions $y = f(x)$ and $y = f^{-1}(x)$: *They are symmetric with respect to the line* $y = x$.

A little thought convinces us that to interchange the roles of x and y on a graph, is to reflect the graph across the line $y = x$ (see Figure 2.12b and c).

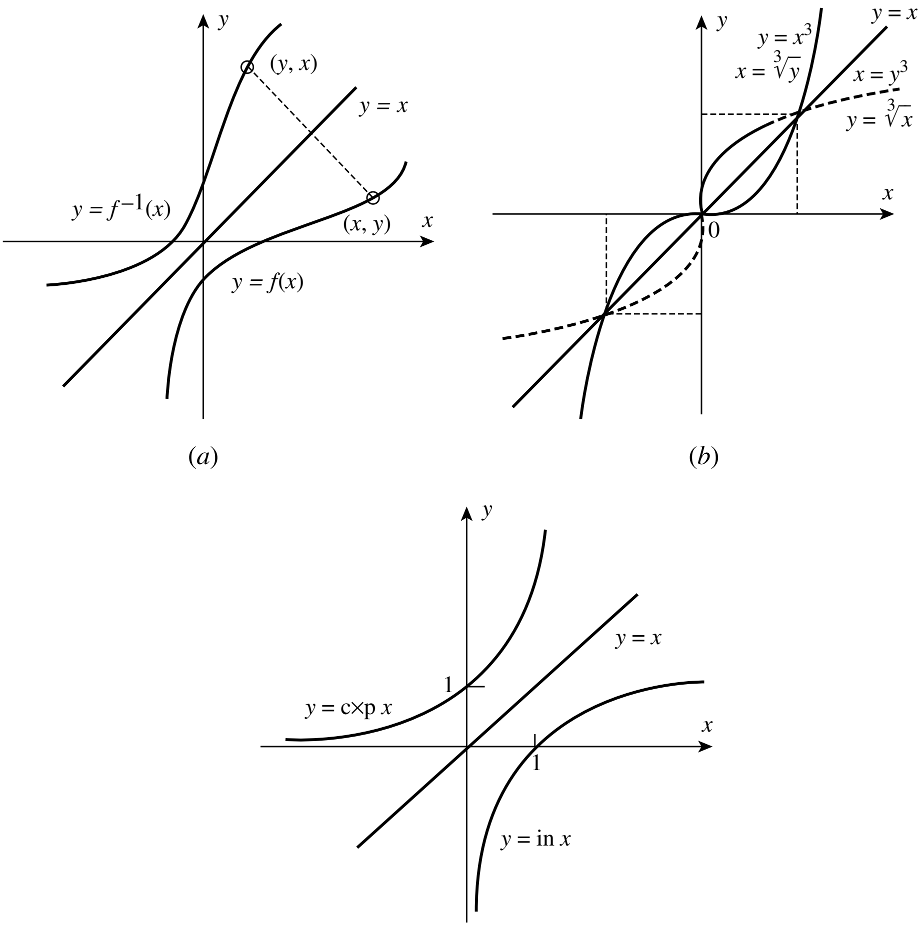


FIGURE 2.12

Caution: The symbol "f^{-1}" stands for inverse function.

Here we are using *the superscript* " -1" *in a new way*. The number $f^{-1}(x)$ is almost always different from $[f(x)]^{-1} = 1/(f(x))$. Thus, the symbol f^{-1} *does not stand for* $1/f$. This may be clear from the Examples (1) and (2) above. All mathematicians use the superscript " -1" to name the inverse function.

Each pair of inverse functions (i.e., f and f^{-1}) behave in such a way that one function undoes (or reverses) what the other does, that is, suppose that $f(x) = y$, then $f^{-1}(f(x)) = x$, and if $f^{-1}(y) = x$, then $f(f^{-1}(y)) = y$.

Remark: From the definition of one-to-one correspondence between the sets A and B, one might get the feeling that the number of elements in both the sets must be same. *Of course this is true if the number of elements in their domains is finite*. With the same line of thinking, when considering a pair of mutually inverse functions defined on intervals one might expect that the length of the *domain interval* and that of *range interval* should be the same, *but this need not be true*. For example, recall that the exponential function $y = e^x$, defines a one-to-one mapping from $(-\infty, \infty)$ onto $(0, \infty)$. *The inverse of exponential function $y = e^x$ is called logarithmic function* (expressed by $x = \log_e y$), which is defined from $(0, \infty)$ onto $(-\infty, \infty)$.

Remark: This is possible due to the fact that infinite sets can have proper subsets (which are also infinite) such that a one-to-one mapping can be defined from one set on to the other. Such sets are said to be *equivalent*, and they are said to have "in a sense" same number of elements. These matters are discussed below at length.

2.12 COMPARING SETS WITHOUT COUNTING THEIR ELEMENTS

The concept of one-to-one correspondence helps in comparing sets (for their sizes) without counting their elements. *It also helps in distinguishing between infinite sets and in answering a query whether all infinite sets share the "same degree of infinity"* or whether *some infinite sets are "larger" than others*. This discussion also helps in defining the notion of "*Infinity*."

Let A and B be two *nonempty* sets. We say that these sets are equal ($A = B$) if and only if they contain the *same elements* [For example, if $A = \{1, 2, 3, 4\}$ and $B = \{1, 2, 3, k\}$ then $A \neq B$.] Suppose A and B are not equal, then it is natural to ask *whether or not the number of elements in these sets is the same? The number of elements in a set is known as cardinality of the set.*[(10)]

In the case of finite sets, we can *count* the elements of each set and then observe whether or not the numbers obtained as a result of counting are same. *However, this question can also be answered without actually counting the elements of the sets. This is possible by using the concept of one-to-one correspondence* as explained below.

2.13 THE CARDINAL NUMBER OF A SET

Let A be a set of Latin letters, that is, $A = \{a, b, c, d, e\}$, and B be a set of Greek letters, that is, $B = \{\alpha, \beta, \gamma, \delta, \varepsilon\}$. It is clear that $A \neq B$ (why?). We can arrange these sets as shown below.

Sets	Elements				
A	a	b	c	d	e
B	α	β	γ	δ	ε

Now we can say without counting that both the sets A and B, have the same number of elements. *What is the characteristic of this method of comparing sets?* For each element of one set, there appears one and only one element corresponding to it in the other set, and conversely.

(10) One does not talk about the number of elements in a set. One only talks about the cardinal number of a set. The cardinal number of a *finite set is the number of elements* in the set. (The cardinal numbers for infinite sets are discussed later in this chapter.)

Observe that in this example both the sets A and B have only a finite number of elements. The strength of this method lies in that it can be applied even when the sets to be compared have an infinite number of elements.(11)

In the case of infinite set(s), though it is not possible to count their elements entirely, yet it is possible to compare them using the concept of one-to-one correspondence as in the above example of sets A and B that have finite number of elements.

Now, consider the set N of all natural numbers and M is the set of all numbers of the form $1/n$ where $n \in N$, then the second method of *comparison shows at once that the number of elements in both the sets is the same* (in some sense) *though the process of counting is endless and accordingly it is never completed.* More clearly, it is sufficient to arrange our sets as follows:

N:	1	2	3	4	5	6	.	.	.
	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$			
M:	1	1/2	1/3	1/4	1/5	1/6	.	.	.

and pair off the numbers n and $1/n$. We now turn to precise definition.

2.14 EQUIVALENT SETS (DEFINITION)

If it is possible to establish a one-to-one onto correspondence between two sets A and B, then these sets *are said to be equivalent* (or to have the *same cardinality*) and we write $A \sim B$.

Note: *If two finite sets are equivalent then it means that both the sets have the same number of elements*. Infinite sets can also be equivalent. For example, $N = \{1, 2, 3, 4, 5, \ldots\}$, $W = \{0, 1, 2, 3, 4, 5, \ldots\}$ are equivalent sets, since the function f defined by $f(1) = 0$, $f(2) = 1, f(n) = n - 1$ is *a one-to-one correspondence between N and W*.

Now, comparing with the case of finite sets, these two sets N and W can be thought to have the same number of elements. *In mathematics, one says that they have the same cardinal number*.

2.15 FINITE SET (DEFINITION)

A set S is called *finite* and is said to contain n elements, if

$$S \sim \{1, 2, 3, \ldots, n\}$$

Remarks:

- *The empty set is considered finite.*
- It is easily seen that two finite sets are equivalent if, and only if, they consist of the same number of elements. *So the concept of equivalence is a direct generalization of the concept of having same number of element, for finite sets.*

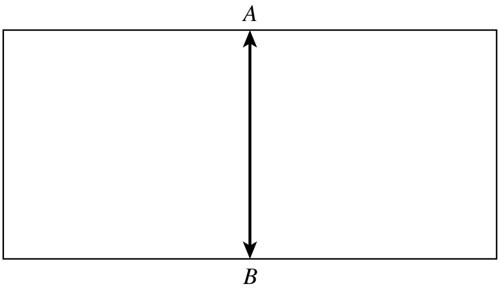
(11) We have not yet introduced the notion of "infinity". However, we define a finite set as the one for which the process of counting the elements ends for some number $n \in N$. Also, for the time being we agree to define an infinite set as the one, which is not finite. Though this is a negative definition of an infinite set, it coveys that in the case of infinite set the process of counting the elements must be endless. These observations will become clear shortly, when we try to count the elements of an infinite set using the concept of one-to-one correspondence.

2.16 INFINITE SET (DEFINITION)

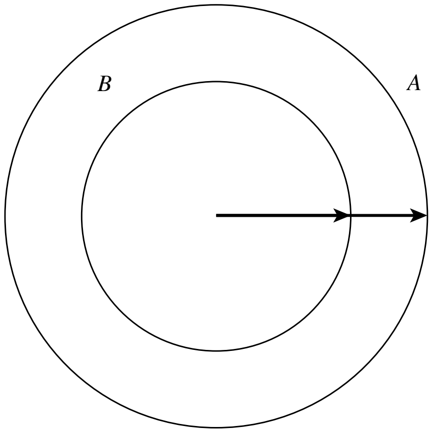
Sets that are *not* finite are called *infinite*.

We now give a few examples of *pairs of equivalent infinite sets.*

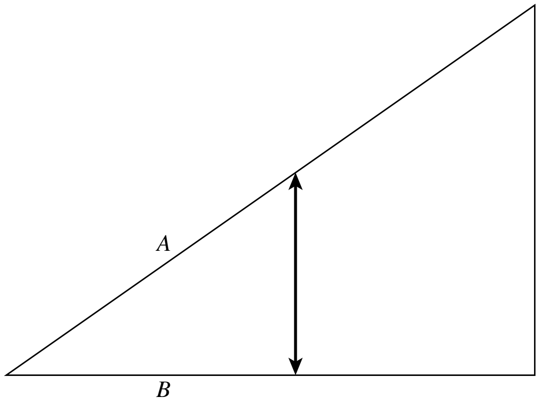
Example (1): Let A and B be the sets of points on two parallel sides of a rectangle. It is easy to see that $A \sim B$.



Example (2): Let A and B be the sets of points of two concentric circles. Here also it is clear that $A \sim B$. *This example is less trivial than the earlier*. If we cut and straighten out our circles, one of them is transformed into a shorter line segment than the other. *It would seem that there ought to be more points on the longer segment*. We see that this is not so.



Example (3): Here is an example that is more surprising. In a right angle triangle, let A be the set of points of the hypotenuse and B the set of points of a base. From the figure, it is clear that $A \sim B$, *despite the fact that the base is shorter than the hypotenuse. If we lay off the base on the hypotenuse,* the set B appears to be a *proper subset* of the set A, and hence different from A itself. *In this example, we encounter a set A containing a proper subset B, which is equivalent to A itself.*



We know that a finite set *cannot* contain a proper subset, which is equivalent to the given (finite) set. *It is thus the infiniteness of the set A that produces this curious phenomenon*. Let us consider one more example.

Example (4): Let N be the set of all natural numbers and let M be the set of all even natural numbers, then we can show that $N \sim M$.

$$\text{We have} \quad N = \{1, 2, 3, 4, 5, \ldots\}$$
$$\text{and} \quad M = \{2, 4, 6, 8, 10, \ldots\}$$

Let us define a *one-to-one correspondence* $f: N \to M$ given by $f(x) = 2x$, for each $x \in N$. This function "f" makes N and M equivalent. *Observe that M is a proper subset of N*. We may therefore say that there are as many positive even numbers as there are natural numbers. Similarly,

- $N \sim \{1, 4, 9, 16, \ldots\}$, where the *one-to-one correspondence* is defined by the function.

$$f(x) = x^2, \quad \forall x \in N$$

- $N \sim \{1, 8, 27, 64, \ldots\}$, where $f(x) = x^3$, $\forall\, x \in N$
- $W \sim N$, where $f(x) = x + 1$, $\forall\, x \in W$
- Let Z be the set of all integers and consider the correspondence shown below:

N:	1	2	3	4	5	6	7	...
	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	
Z:	0	1	−1	2	−2	3	−3	...

Here the function $f: Z \to N$ is defined by two formulas.

$$f(n) = 2n \text{ for all } n > 0, \text{ and } f(n) = -2n + 1 \text{ for } n \leq 0.$$

The function is a one-to-one correspondence from Z to N.

2.17 COUNTABLE AND UNCOUNTABLE SETS

Definition (1): If a set A is equivalent to *a subset of all positive integers*, then A is said to be *countable*. Thus, by definition *a finite set is countable.*

Remark: It is important to note that even *an infinite set will be called a countable set* provided its elements can be put in *one-to-one correspondence with the set of natural numbers*. Thus, the sets $\{2, 4, 6, 8, \ldots\}$, $\{1, 4, 9, 16, \ldots\}$, $\{1, 8, 27, 64, \ldots\}$, and so on, all are *countable infinite sets*. *A* set that is not countable is called *uncountable*. For instance, the set *of all points on a line segment* is an *uncountable set*. Another useful definition that distinguishes *countable* and *uncountable sets* is given below:

Definition (2): For any positive integer n, let $J_n = \{1, 2, 3, 4, 5, \ldots, n\}$. Let $J = \{1, 2, 3, 4, 5, \ldots\} = N$.

(Note that J_n is a *finite subset* of the set of natural numbers.)

Then, for any set A, we say

(a) A is *finite* if $A \sim J_n$, for some n. (*The empty set* "ϕ" is also considered *finite*.)

(b) A is *infinite* if it is *not finite*.

(c) A is countable, if $A \sim J$. (In fact, we may call the set A as *countably infinite* if $A \sim J$.)

Remark: *A finite set is always countable*. Those infinite sets, which are equivalent to set J, will be called *countably infinite sets*.

(d) A is *uncountable* if A is *neither* finite *nor* countable.

(e) A is *at most countable* if A is *finite* or *countably infinite*.

Note (1): It is clear that the definition (1) is equivalent to the statement (e) above, of the definition (2).

Note (2): *Countable sets* are sometimes called *enumerable* or *denumerable*.

Note (3): If $A \sim B$, then we say that A and B have the same cardinal number or the same cardinality.

Remark: *All countable infinite sets are equivalent among themselves and hence all of them have the same cardinality*. One might ask whether all infinite sets have the same cardinality that is whether they share the "*same degree of infinity*," or whether *some infinite sets are "larger" than others?*

2.18 CARDINALITY OF COUNTABLE AND UNCOUNTABLE SETS

The cardinality of any *countable infinite set* is denoted by the symbol $\aleph_o$, read as "aleph-null." The *symbol* "c" is used to denote the *cardinal number of the set R of all real numbers* (or of all points on the real line) *that is uncountable*. (*An uncountable set is necessarily infinite*.)

Note: "c" is the *cardinal number of R and of any set that is numerically equivalent to R*. It can be demonstrated that c is the cardinal number of any open interval, or any subset of R that contains an open interval.

Thus, our list of cardinal numbers has grown to 1, 2, 3, 4, ..., $\aleph_o$, c. Just as the positive integers, *we can order the cardinal numbers*, and they are related to each other by $1 < 2 < 3 < \cdots < \aleph_o < c$. At this stage, the following question arises.

Are there any infinite cardinal numbers greater than "c"? Yes, there are; for example, the cardinal number of the class of all subsets of R. This answer is the outcome of the axiom that if X is any *nonempty set*, then the *cardinal number of X* is less than the *cardinal number of the class of all* subsets of X. [If A is a finite set with n elements, then the set of all subsets of A, denoted by $P(A)$, has 2^n elements. P(A) is called the power set of A.]

We also have a cardinal arithmetic. One can add, multiply, exponentiate cardinal numbers. For example, the cardinal number of the power set $P(X)$ of a set X, with cardinal number $|X|$ is known to be $2^{|X|}$. Thus, the cardinal number of the power set of natural numbers $P(N)$ is $2^{\aleph_0}$ and it can be shown that $2^{\aleph_0} = c$.

If we follow up the hint contained in the fact that $2^{\aleph_0} = c$, and successively form 2^c, 2^{2^c}, ... *we get a chain of cardinal numbers*

$$1 < 2 < 3 < \cdots < \aleph_o < c < 2^c < 2^{2c} < \cdots$$

in which *there are infinitely many infinite cardinal numbers*. Clearly, *there is only one kind of countable infinity,* symbolized by $\aleph_o$, and beyond this *there is an infinite hierarchy of uncountable infinities* that are all *distinct from one another.*

At this point we bring our discussion to a close. However, with a view to introduce the "*Notion of Infinity*," which will be frequently needed in our study of *Calculus*, we state below one more definition of *an infinite set*.

2.19 SECOND DEFINITION OF AN INFINITE SET

A set A is infinite if, and only if, it is equivalent to one of its proper subsets.

2.20 THE NOTION OF INFINITY

In the history of mathematics the term "infinite" was obscure for a long period. The symbol for infinity is "∞". In modern mathematics, the symbol "∞" *is not a number, and not all algebraic operations are defined for this symbol.*[12]

Often we shall have to study the behavior of functions of x, as x becomes infinitely large, that is, when *x is permitted to attain larger and larger values exceeding any bound K, no matter how big K is chosen.* For example, take $f(n) = 1/n$. Then if n takes the values 1, 2, 3, ..., 100, we have an aggregate (i.e., the class, or set, consisting of the values of $f(n)$, for various values of n consisting of the fractions 1, 1/2, 1/3, ... 1/100.

We wish to discuss the behavior of this function for very large values of n. It is immediately obvious that $1/n$ *becomes very small when n is very large.*

Note: It is wrong to say that $1/n = 0$ when $n = \infty$. Remember that ∞ *is not a number*, so it cannot be equated to any number, howsoever large. *Further,* $1/n$ *can never be equated to zero, however big n is chosen, since* $1 \neq 0$. However, *it makes sense to say that the function* $f(n) = 1/n$ tends to zero for values of n that tend to infinity. Now, we define precisely what we mean by this statement.

[12] The symbol "∞" for "infinity" was proposed by the English mathematician and theologian John Wallis (1616–1703).

Suppose we take a positive real number ε, *however small*, then we can certainly choose a number N so that whenever $n > N$, the function $1/n$ is less than ε. For example, if we choose $\varepsilon = 0.001$, then $f(n) = 1/n$ can be made less than ε by choosing $N \geq 10^3$. (Note that if $n > N \geq 10^3$ then $1/n < 0.001$.) Similarly if we choose ε as 0.00000001, then $f(n) = 1/n$ can be made less than this by choosing $n > N \geq 10^8$.[13]

If we now consider the function $f(n) = n^2$, it is clear that this function can be made as large as we please by taking sufficiently large values of n. We may therefore, say that the function $f(n) = n^2$ tends to infinity when n tends to infinity.

Now, let us consider the function

$$f(n) = -n^2$$

In this case, we say that $f(n)$ tends to $-\infty$ when n tends to ∞. We would usually write these statements briefly as given below:

$$n^2 \to \infty \quad \text{as } n \to \infty$$

and

$$-n^2 \to -\infty \quad \text{as } n \to \infty$$

(These notations will be used when we introduce the concept of limit, in Chapter 7.)

2.21 AN IMPORTANT NOTE ABOUT THE SIZE OF INFINITY

We must not confuse the word "*infinite*" with the "*very large finite*". For example, think of *the number of inhabitants of the earth* at any particular instant, or *the number of leaves on all the trees of the earth* at any instant, or *the number of blades of grass on the earth* at any instant or *the number of all these things put together*. *These are all very large numbers*, yet they are *finite*. That is to say, given sufficient patience and manpower, we could set out to count the numbers of these large classes with the assurance that we could finish the job.

As an example of an infinite set we have the set of "*natural numbers*." If we set out to count the natural numbers, 1, 2, 3, 4, . . ., we cannot do so with the assurance that if we continue until we die and pass the job on from generation to generation, neither we nor any of our descendants will ever exhaust the supply. Also we know that *there is the infinite in algebra, the infinite in geometry, the infinitely small, the infinitely large, and so on. Again, there is not only one infinite, but a whole hierarchy of infinites.*

2.22 ALGEBRA OF INFINITY (∞)

We accept the following properties of ∞.

(i) *If $x \in R$, then we have*

$$x + (+\infty) = +\infty, \quad x + (-\infty) = -\infty$$
$$x - (+\infty) = -\infty, \quad x - (-\infty) = +\infty$$
$$x/(+\infty) = 0 = x/(-\infty)$$

[13] The Greek letter ε (epsilon) has become a standard notation for an arbitrarily small positive number.

(ii) *If $x > 0$, then we have*

$$x(+\infty) = +\infty, x(-\infty) = -\infty$$

(iii) *If $x < 0$, then we have*

$$x(+\infty) = -\infty, x(-\infty) = +\infty$$

(iv) $(\infty) + (\infty) = (+\infty)(+\infty) = (-\infty)(-\infty) = +\infty$

$$(-\infty) + (-\infty) = (+\infty)(-\infty) = -\infty$$

(v) If $x \in R$, then we write $-\infty < x < +\infty$

or $x \in (-\infty, \infty)$

Remark: Compare the above properties of ∞ with those of real numbers, and note the distinction. Besides, it may be mentioned that there are expressions involving ∞ [i.e., like$(\infty/\infty), (\infty)^0, (\infty - \infty)$, etc.] that are not defined and such expressions are called indeterminate forms. Later on, it will be seen that limiting values of such expressions can be found by L' Hospitals Rule (Chapter 21).

3 Discovery of Real Numbers: Through Traditional Algebra

3.1 INTRODUCTION

Calculus is based on the real number system and its properties. *But what are real numbers and what are their properties?* To answer this question, we start with the simplest number system consisting of *Natural Numbers* or *Counting Numbers*. In fact, the first numbers known to the man were counting numbers.

In arithmetic, the four fundamental operations, namely *addition, subtraction, multiplication, and division* are used to make new numbers out of old numbers—that is, *to combine two numbers to create a third.* Accordingly, these operations are called *binary operations.*

Ordinary algebra (or so-called traditional algebra) is a branch of mathematics, in which symbols are used to represent numbers (or quantities), in all the arithmetical operations. In fact, *ordinary algebra* is a generalization of *arithmetic*.(1)

The subject of algebra involves the *study of equations* and a number of other problems that developed out of the theory of equations. It is in connection with the solution of algebraic equations that negative numbers, fractions, rational numbers, and irrational numbers were discovered. The set of *rational numbers together with that of irrational numbers make the set of real numbers.*

It is interesting to study the history of development of the real number system. For example, it took a long time for the *zero* to enter the family of numbers. The reason for this delay was due to the fact that *zero* has a physical meaning of "*nothing*", and therefore, in order to consider it as a number, we had to wait for *negative numbers* to appear and be accepted.

It was nearly 150 years ago that mathematicians adopted the correct viewpoint toward these various types of numbers. They recognized that the concept of numbers could be extended to include *negative* and *irrational* numbers.

When the *negative numbers* were accepted as respectable members of the number community, the remaining numbers were named as *positive numbers*. Thus, the concept of positive numbers was developed from that of negative numbers and then "*0*" *(zero) was*

What must you know to learn calculus? (3-Real numbers and their properties)

(1) Other forms of higher algebra also exist. These are connected with mathematical entities other than numbers. For example, algebra of matrices, algebra of vectors, algebra of sets, and so on. Definitions of different entities, their properties, and the rules for combining them are of course different in *different algebras*, but consistent with the physically observed facts.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

accepted as a number that must neither be positive nor negative. Thus, the development of algebra has contributed (in a way) to the development of the Number System.[2]

In Chapter 1, we have introduced certain important subsets of real numbers (namely, *natural numbers, whole numbers, integers, rational numbers, fractions, and irrational numbers*). Also, we have introduced *even and odd numbers* (as subsets of integers), *prime and composite numbers* (as subsets of natural numbers), and certain important concepts useful in selecting pairs of coprime numbers or finding factors and computing H.C.F. (from a set of natural numbers).

Now, it is proposed to throw some more light on the following subsets of real numbers.

3.2 PRIME AND COMPOSITE NUMBERS

A *natural number* which has exactly two (different) factors, namely the number itself and 1, is called a *prime number*. Some examples of prime numbers are 2, 3, 5, 7, 11, 13, 17, 19, ... *Each prime number except 2 is odd.*

The number 1 is neither prime nor composite, since it has only one factor.

Natural Nos.	Factors	Set of Factors	No. of Factors	Remark
1	$1 = 1 \times 1$	{1}	1	Neither prime nor composite
5	$5 = 1 \times 5$	{1,5}	2	Prime number
12	$12 = 1 \times 12$ $= 2 \times 6$ $= 3 \times 4$	{1,2,3,4,6,12}	6	Composite number

A number that has three or more factors is called a *composite number*.

There is no formula that generates prime numbers. However, the number of primes is infinite and this can be easily proved using algebra. The method of proof is indirect but it is beautiful and surprisingly simple, as we have already seen in a footnote in Chapter 1.

Remark(s):

(i) By definition, *no negative integer is prime.*[3]

(ii) *Every composite natural number can be expressed as a unique product of its prime factors.*

We recall the following subsets of real numbers, which will be needed frequently in our discussion.

[2] In the process of solving certain algebraic equations (such as $x^2 + 1 = 0$), new type of numbers were discovered. These numbers have the property that their squares are negative numbers (here $x^2 = -1$). The solutions of this equation were denoted by $x = \pm\sqrt{-1}$. Thus $\sqrt{-1}$was born, as a strange mathematical entity. Then the contemporary mathematicians thought that such a number was "useless", "imaginary", and "impossible." Euler thought that expressions like $\sqrt{-1}$ were "neither nothing", "nor greater than nothing", "nor less than nothing", which necessarily makes them "imaginary" or "impossible". Euler used imaginary numbers in some of his works and he was apparently the first to use the symbol "i" for $\sqrt{-1}$. It was not until Gauss adopted them, that imaginary numbers finally acquired legitimate status. Subsequently, by combining real numbers with so-called imaginary numbers using the operation of addition, complex numbers came into existence. [*The Spell of Mathematics* by W.J. Reichman (p. 156), Pelican Book.]

[3] "0" is a composite number. Since zero is divisible by all natural numbers, so that all natural numbers are the factors of zero. Recall that factors are only natural numbers, thus zero is not a factor of an integer. Also, since division by zero is not permitted in mathematics, the expressions like 1/0 and 0/0 are meaningless expressions.

3.3 THE SET OF RATIONAL NUMBERS

$$Q = \left\{\frac{p}{q}\,|p,\ q \text{ belong to } Z,\ q \neq 0\right\}$$

Examples: $\frac{7}{1}, \frac{0}{3}, \frac{-3}{8}, \frac{5}{-9}, \frac{-2}{-3}, \frac{4}{5}$, and so on, are all rational numbers.

Zero is a rational number but division by zero is not permitted.

Also, 0/0 (i.e., zero divided by zero) is a meaningless expression, which is nothing more than a mathematical drawing. Decimal representation of a *rational number* either terminates (as in $3/8 = 0.375$) or else repeats in regular *cycle, forever* (as in $13/11 = 1.18181818\ldots$ or in $3/7 = 0.428571428571\ldots$). A little experimenting with the long division process will show why this happens. (Note that, there can be only a finite number of different remainders.)

Note: In the decimal form, a number of the type 3.2613261326132613 ... or 6.32537537 ... or 7.000 ..., and so on with nonterminating but repeating string of digits, in the decimal part from anywhere onwards represents a *rational number* and it can be expressed in the form p/q, where p and q are integers, and the *denominator* $q \neq 0$. (*Positive rational numbers form the set "F" of fractions.* $F = \{a/b | a,\ b \in N\}$.)

If a number has a decimal representation which ends in zeros, for example, $\frac{1}{4} = 0.2500000\ldots$ then it can also be written in another decimal expansion that ends in nines. For this purpose, we must decrease the last nonzero digit by one and write the subsequent digits as 99999 .... Thus, we have $\frac{1}{4} = 0.2499999\ldots$. Similarly $7.000\ldots = 6.9999\ldots$. Except for such substitution, decimal expansions are unique.

3.4 THE SET OF IRRATIONAL NUMBERS

Those real numbers, which are not rational, are called irrational numbers. (They cannot be expressed in the form p/q, where *p, q are integers* with $q \neq 0$.) We denote the set of irrational numbers by Q^c or Q'. *Irrational numbers too can be expressed as decimals.*

Note *:* Irrational numbers: In the decimal form, a number of the type 5.7101001000100001 ..., or 7.3030030003 ..., or $\pi = 3.141592653589793\ldots$ with nonterminating and nonrepeating string of digits, in the decimal part, represents an irrational number. Obviously, it cannot be expressed in the form p/q. Check this.

The numbers $\sqrt{2},\ \sqrt{3},\ \ldots,\ \sqrt{n}$ where n is a natural number which is not a perfect square, have got decimal forms like that of π wherein no pattern is noticed $\left(\sqrt{2} = 1.4142135623\ldots,\ \sqrt{3} = 1.7320508075\ldots,\right)$. These are all irrational numbers. Indirect methods of algebra are available to prove that $\sqrt{2},\ \sqrt{3}, \ldots$, and so on are irrational numbers.

π and e ($e = 2.7182818284\ldots$) are special types of irrational numbers (called *transcendental numbers*), which arise naturally in geometry and calculus, respectively.

3.5 THE SET OF REAL NUMBERS

The set of rational numbers together with the set of irrational numbers forms the set of real numbers denoted by R. Thus, $R = Q \cup Q^c$.

Now, we define a real number.

3.6 DEFINITION OF A REAL NUMBER

A real number is one that can be written as an unending decimal, positive or negative or zero.[(4)]

3.7 GEOMETRICAL PICTURE OF REAL NUMBERS

We use the term *real line* very frequently without any explanation, and of course what we mean by *it is an ordinary geometric straight line whose points have been identified with the set R of real numbers.*

We use the letter R to denote the *real line* and *the set of all real numbers*. We say that, to every real number there corresponds a unique point on the number line, and conversely, to every point on the number line there corresponds a real number. It is due to this one-to-one correspondence that we often speak of real numbers as if they were points on the number line and we speak of the points on the real line as if they were real numbers. Yet, the fact remains that, *a real number is an arithmetical object, whereas, a point is a geometric object.*

For the purpose of representing rational numbers by points on a straight line, we label any point on the line with “0” (zero) and any other point to the right of “0” with “1”. *This fixes the scale*. With this scale as unit length, we can easily plot on the number line all those points, which represent rational numbers. For this purpose we use the *four fundamental operations of arithmetic* (namely addition, subtraction, multiplication, and division). *With regards to the irrational numbers, we have to go beyond these operations.*

Shortly, we will learn that *between any two rational numbers, there is always another rational number.* A similar statement is true in the case of irrational numbers. (This is known as the *property of denseness*, which is studied later in this chapter.) Thus, if we plot only rational numbers on the number line, then there will be infinite number of holes throughout the line. These unoccupied positions must represent irrational numbers. Geometric constructions can be used to find points corresponding to certain irrational numbers, such as $\sqrt{2}$, $\sqrt{3}$, $\sqrt{5}$, $\sqrt{7}$, and so on. Points corresponding to other irrational numbers can be found by using decimal approximations.

Every irrational number can be associated with a unique point on the x-axis, and every point that does not correspond to a rational number can be associated with an irrational number. This fact is guaranteed by the axiom of completeness and is discussed later at the end of this chapter.

Note: We are familiar with the simpler properties of real numbers. It is now proposed to discuss *some other properties of real numbers, which are not obvious.* (This study will be found useful for building up necessary terminology, required for defining the “*concept of limit*” in Chapters 7a and 7b.)

The beauty and power of mathematics can be appreciated only if the properties of real numbers are properly understood. We give below the necessary material to make the study systematic and interesting. This material should be sufficient to meet the study requirements of this book and also serve as a good background for studying these concepts at higher levels.

3.8 ALGEBRAIC PROPERTIES OF REAL NUMBERS

These properties are associated with two basic operations on real numbers, namely *addition and multiplication*. [The operations of subtraction and division (by nonzero numbers) can be

(4) *Calculus with Analytic Geometry* by John B. Fraleigh.

defined respectively in terms of addition and multiplication.] *We assume that the following statements in real numbers are both well defined and true.*

If a, b, and c are real numbers, then, we have

1.	Commutative Property	
	$a+b=b+a$	$a \cdot b = b \cdot a$
2.	Associative Property	
	$(a+b)+c = a+(b+c)$	$(a \cdot b) \cdot c = a \cdot (b \cdot c)$
3.	$a+0=0+a=a$	$a \cdot 1 = 1 \cdot a = a$
	Additive identity is "0"	Multiplicative identity is "1"
4.	$a+(-a)=0$	$\text{a} \cdot (1/a) = 1$ if $a \neq 0$
	Additive inverse of a is $(-a)$ and vice versa	Multiplicative inverse of a is $(1/a)$ and vice versa
Remark: Multiplicative inverse of the real number "0" does not exist.[5]		
5.	Multiplication distributes over addition	
	$a \cdot (b+c) = a \cdot b + a \cdot c$	

The other algebraic properties of real numbers can be proved from these five properties. For example, $a \cdot b = 0$, iff $a = 0$ or $b = 0$, $(-a) \cdot (-b) = a \cdot b$, and so on.

At this stage, however, we are not proving these properties.

3.9 INEQUALITIES (ORDER PROPERTIES IN REAL NUMBERS)

Between any two unequal real numbers a and b, there is a relation (called *the order relation*) which states whether a is less than b ($a < b$) *or* b is less than a ($b < a$).

Hence, *for any two real numbers a and b,* we have exactly one of the following statements true.

1. Either $a = b$ *or* $a < b$ *or* $b < a$. This property is called the *Law of Trichotomy*. A relation of the form $a < b$ (read "a is less than b") or $b > a$ (read "b is greater than a") is called *an inequality*.
   Other properties of inequalities are as follows:
2. If $a < b$ and $b < c$, then $a < c$. This is known as the Transitive Property.
3. If $a < b$ and c *is any real number*, then $a + c < b + c$.
4. If $a < b$ and $c > 0$, then $a \cdot c < b \cdot c$.
   If $a < b$ and $c = 0$, then $a \cdot c = b \cdot c$.
   If $a < b$ and $c < 0$, then $a \cdot c > b \cdot c$.

Note (1): If $c < 0$, and we multiply both sides of an inequality by c, then *the direction of the inequality changes.*

Note (2): If $a > 0$, $b > 0$, with $a < b$, then $1/a > 1/b$.

5. If $a < b$ and $c < d$, then $a + c < b + d$.
6. If a, b, c, d are all positive and $a < b$, $c < d$, then $a \cdot c < b \cdot d$.

[5] Division by "0" leads to contradictions as was seen in Chapter 1.

The properties (5) and (6) tell us that,

- *two (similar) inequalities can always be added, and*
- *two (similar) inequalities (involving positive numbers) can be multiplied.*

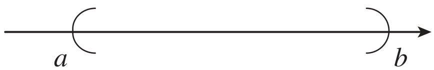
Note(s):

- If either $a < b$ or $a = b$, then we write $a \leq b$.
- If either $a > b$ or $a = b$, then we write $a \geq b$.
- If $a \geq 0$, then we say that a is *non-negative*.
- If $a < x$ and $x < b$, then we write $a < x < b$ and in this case we say that x is between a and b.
- If x is a real number between a and b (where $a < b$) and x may be equal to a or b, we write $a \leq x \leq b$.

3.10 INTERVALS

If we wish to *consider all the real numbers between a and b* (with or without including one or both the end points a and b) then such sets are called *intervals* as discussed below:

- Open Interval (a, b): If a and b are real numbers with $a < b$, we denote by (a, b) *the set of all real x such that* $a < x < b$. We call (a, b) an *open interval*. Thus $(a, b) = \{x | x \in \text{R}, a < x < b\}$. *It consists of all real numbers between a and b*. Obviously a and b are *not included* in the set.



- Closed Interval $[a, b]$: If $a \leq b$, then $[a, b]$ denotes the set of all real numbers x such that $a \leq x \leq b$. Thus $[a, b] = \{x | x \in \text{R}, a \leq x \leq b\} \cdot [a, b]$ is called a *closed interval*. It consists *of all real numbers between a and b, including the end points a and b.*

We occasionally need to use "*half-open*" intervals. For example, $(a, b]$ denotes the set of reals denoted by the interval $a < x \leq b$, which is open at a and closed at b. Similarly $[a, b)$ denotes the interval $a \leq x < b$.

By (a, ∞)*, we mean the set of all real x such that* $x > a$.
Thus, $(a, \infty) = \{x | x \in \text{R}, x > a\}$.

By $(-\infty, b)$*, we mean the set of all real x such that* $x < b$.

Thus, $(-\infty, b) = \{x | x \in \mathrm{R}, x < b\}$.

The set of all real numbers is sometimes denoted by $(-\infty, \infty)$.

Note: An open interval can be thought of as one that contains none of its end points, and a closed interval can be regarded as one that contains all of its end points. Consequently, the interval $[a, +\infty)$ is considered to be a closed interval, because it contains its only end point a. Similarly, $(-\infty, b]$ is *a* closed interval, whereas $(a, +\infty)$ and $(-\infty, b)$ are open. The intervals $(a, b]$ and $[a, b)$ are neither open nor closed. The interval $(-\infty,\infty)$ has no endpoints and it is considered both open and closed.

3.10.1 Bounded and Unbounded Intervals

The intervals in which the symbol "∞" (infinity) *does not appear*, are called *bounded intervals*, (They occupy a limited length of the real line.) The sets like (a, ∞), $(-\infty, b)$, (a, ∞), $(-\infty, b)$, and $(-\infty, \infty)$ are called *unbounded intervals*. (The usefulness of the concept of boundedness of sets is discussed later, in different contexts, in various chapters.)

3.10.2 Usefulness of Intervals

The *usefulness of intervals* can be seen from the following examples.

(i) The values of x for which the expression $\sqrt{(12 - x)}$ is a *real number*, are given by $x \leq 12$, i.e., $(-\infty, 12)$ and those for which $\sqrt{(16 - x^2)}$ is a *real number*, are given by the interval $-4 \leq x \leq 4$.

(ii) The function $f(x) = 1/(x - 2)$ is not defined for $x = 2$, but for all other real values of x, it is well defined. Thus, we say that $f(x)$ is defined for $x \in (-\infty, 2) \cup (2, \infty)$.

In calculus, we study the behavior of functions on intervals, which are defined using the absolute value of a real number. Hence, we introduce the concept of the absolute value of a real number.

3.10.3 Definition of Absolute Value of Real Number(s)

If a is any real number, *the absolute value* of a, denoted by $|a|$ is a, if a is *non-negative*, and $-a$ if a is *negative*. Thus, with symbols we write,

$$|a| = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{if } a < 0 \end{cases}$$

Examples: $|7| = 7$, $|0| = 0$, $|-3| = -(-3) = 3$

It is clear that $|a|$ is *never negative*; that is $|a| \geq 0$.[6]

[6] Since $|a|$ is never negative [i.e., $|a| \geq 0$], it follows that $x \leq |x|$ for any x, positive, zero, or negative. This observation will be useful in proving the triangle inequality, i.e., $|x + y| \leq |x| + |y|$, discussed later [as Theorem 3], in this chapter).

The absolute value of a real number can be considered as its distance (without regard to direction, left or right) from the origin.

3.10.4 The Geometric Interpretation of |a − b|

From the definition of the absolute value, we have,

$$|a-b| = \begin{cases} a-b & \text{if } (a-b) \geq 0 \\ -(a-b) & \text{if } (a-b) < 0 \end{cases} \text{ or equivalently, } |a-b| = \begin{cases} a-b & \text{if } a \geq b \\ b-a & \text{if } a < b \end{cases}$$

On the real line, $|a-b|$ units can be interpreted as the distance between a and b without regard to direction. In other words, the distance between a and b is said to be either $(a-b)$ or $(b-a)$, whichever is non-negative.

Examples:

1. $|7-3| = |4| = 4$
2. $|5-12| = |-7| = -(-7) = 7$
3. $|8-(-3)| = |8+3| = |11| = 11$
4. $|-2-(-7)| = |-2+7| = |5| = 5$
5. $|-9-(-6)| = |-9+6| = |-3| = -(-3) = 3$

Let us consider equations involving absolute values.

Example (6): Solve the equation $|x| = 5$

If $x \geq 0$, then $|x| = x = 5$
If $x < 0$, then $|x| = -x = 5 \therefore x = -5$.
Hence the solution set is $\{5, -5\}$.

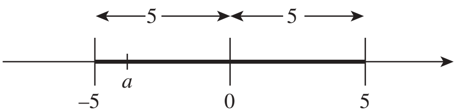
Example (7): Solve the equation $|x-8| = 7$

If $(x-8) \geq 0$, then $|x-8| = x-8 = 7$
So $x = 15$
If $(x-8) < 0$, then $|x-8| = -(x-8) = 7$
So $-x+8 = 7$ or $x = 1$.
Thus, the solution set is $\{1, 15\}$.

It turns out that the simplest way to describe an *interval* at the origin (or any other point) is by using *absolute value inequalities*.

3.10.5 Intervals Defined by Absolute Value Inequalities

The connection between absolute values and distance permits us to describe intervals using absolute value inequalities. An inequality like $|a| < 5$, says that the distance from "a" to the origin is less than 5 units. This is equivalent to saying that a lies between -5 and 5.



Thus, the set of numbers "a" with $|a| < 5$ *is the same as the open interval -5 to 5*. Accordingly, the inequality $|x| < a$, where $a > 0$, states that *on the real number line the distance from the origin to the point x is less than "a" units*; (Figure 3.1) that is,

$$-a < x < a \tag{I}$$

The inequality $|x| > a$, where $a > 0$, states that on the real number line the distance from the origin to the point x is greater than a units (Figure 3.2). that is, either

$$x > a \text{ or } x < -a \tag{II}$$

We state the above results (I) and (II), formally. The double arrow $\Leftrightarrow$ is used here and throughout the text to indicate that the statements on both sides of $\Leftrightarrow$ are equivalent.

$$|x| < a \Leftrightarrow -a < x < a, \text{ where } a > 0 \tag{1}$$

$$|x| > a \Leftrightarrow x > a \text{ or } x < -a, \text{ where } a > 0 \tag{2}$$

Example (8): Find

(i) The end points of the interval determined by the inequality $|x - a| < c$.
(ii) What is the geometric meaning of the inequality $|x - a| > c$?

Solution: (i) To find the end points of the interval $|x - a| < c$, we change $|x - a| < c$ to $-c < x - a < c$ adding "a" throughout, $a - c < x < a + c$. The end points are $a - c$ and $a + c$ (Figure 3.3a).

(ii) The points that satisfy the inequality $|x - a| > c$ are the points on the x-axis whose distances from a are greater than c (Figure 3.3b).

These are *the points outside the closed interval,* $|x - a| \le c$, that is, the points that lie on the right of $a + c$ and to the left of $a - c$.

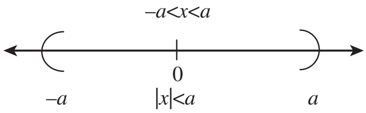


FIGURE 3.1

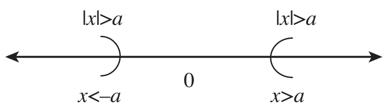


FIGURE 3.2

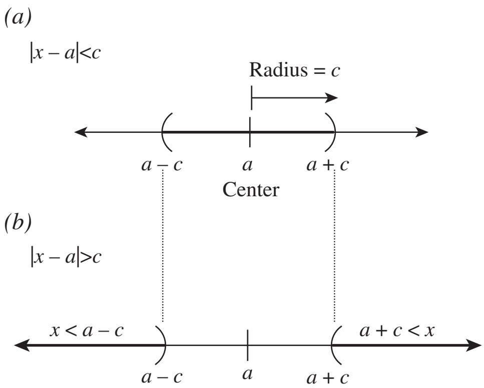


FIGURE 3.3 (a) and (b).

Geometrically, these points make up the *two infinite open intervals* $x < a - c$ and $x > a + c$.

Example (9): Find the values of x that satisfy the inequality

$$\left|\frac{3x+1}{2}\right| < 1 \tag{3}$$

Solution: Change $\left|\frac{3x+1}{2}\right| < 1$ to $-1 < \frac{3x+1}{2} < 1$ to $-2 < 3x+1 < 2$ to $-3 < 3x < 1$ to $-1 < x < \frac{1}{3}$. Thus the inequality (3) represents the open interval $(-1, \frac{1}{3})$.

3.10.6 Absolute Value Inequalities Used in Calculus

(a) The inequality $|x-a| < \delta$ means that the distance between "x" and "a" is less than the positive real number, δ. This in turn means that,

$-\delta < x - a < \delta$ or $a - \delta < x < a + \delta$

i.e., $x \in (a-\delta, a+\delta)$, which is an open interval containing x whose distance from a is less than δ.
Thus, the following four statements have one and the same meaning.

(i) $|x-a| < \delta$
(ii) $a - \delta < x < a + \delta$

(iii) $x \in (a-\delta,\ a+\delta)$

(iv) The distance between x and a is less than δ.

(b) The double inequality $0 < |x-a| < \delta$ may be broken up in two inequalities as $0 < |x-a|$ and $|x-a| < \delta$. We know that the value(s) of x satisfying $|x-a| < \delta$, lie in the interval $(a-\delta,\ a+\delta)$. Next, $0 < |x-a|$ means *that the distance between x and a is positive*, which in turn means that $x \neq a$.

These two observations tell us that $x \in (a-\delta,\ a+\delta)$ and $x \neq a$. In other words, *if we remove the midpoint from the interval $(a-\delta,\ a+\delta)$, then x belongs to the remaining set.*

Thus, $0 < |x-a| < \delta$ means $x \in (a-\delta,\ a) \cup (a,\ a+\delta)$. Therefore, the following four statements have one and the same meaning.

(i) $0 < |x-a| < \delta$

(ii) $a-\delta < x < a$ or $a < x < a+\delta$

(iii) $x \in (a-\delta,\ a) \cup (a,\ a+\delta)$

(iv) $x \neq a$ and the distance between x and a is less than δ.

[The inequality $|f(x)-l| < \varepsilon, (\varepsilon > 0)$ and the double inequality $0 < |x-a| < \delta, (\delta > 0)$ both will be used in *the definition of limit of a function* in Chapters 7a and 7b.]

3.11 PROPERTIES OF ABSOLUTE VALUES

Recall from algebra that the symbol $\sqrt{a}$, where $a \geq 0$, is defined as unique non-negative number x, such that $x^2 = a$. We read $\sqrt{a}$ as *the principal square root* of a (which is the *positive square root of a*).

For example,

$$\sqrt{4} = 2,\ \sqrt{0} = 0,\ \sqrt{\frac{9}{25}} = \frac{3}{5}$$

Note: Since $\sqrt{4}$ denotes only the positive square root of 4, therefore $\sqrt{4} \neq -2$, even though $(-2)^2 = 4$
The *negative square root* of 4 is designated by $-\sqrt{4}$.

Remark (1): Since we are concerned only with real numbers, $\sqrt{a}$ is not defined for $a < 0$.

Remark (2): From the definition of $\sqrt{a}$, it follows that $\sqrt{x^2} = |x|$, (since $|x|$ is always non-negative by definition).

Examples: $\sqrt{5} = |5|,\ \sqrt{(-3)^2} = |-3| = -(-3) = 3$

The properties of absolute value given in the following theorems are useful in *calculus*.

Theorem (1): Given $a, b \in R$

$$\left.\begin{aligned} |a \cdot b| &= |a| \cdot |b| \\ \text{and } \left|\frac{a}{b}\right| &= \frac{|a|}{|b|} \end{aligned}\right\}$$

(We shall prove the second result as Theorem 2.)

Proof: $|a{\cdot}b| = \sqrt{(ab)^2}\left[\quad \because \sqrt{x^2} = |x|\right]$

$= \sqrt{a^2b^2}$

$= \sqrt{a^2}\sqrt{b^2}$

$= |a| \cdot |b|$ (Proved)

Remark: The *absolute value of a product* is equal to the *product of the absolute values* of the factors:

$$|xyz| = |x|\,|y|\,|z|$$

In particular, $|b{\cdot}b| = |b| \cdot |b| = |b^2| = |b|^2$ and $|a \cdot a{\cdot}a| = |a|\,|a|\,|a| = |a|^3$.

Theorem (2): Given $a, b \in R \left|\frac{a}{b}\right| = \frac{|a|}{|b|}$

Proof:

$$\left|\frac{a}{b}\right| = \sqrt{\left(\frac{a}{b}\right)^2}$$

(Proved)

$$= \sqrt{\frac{a^2}{b^2}} = \frac{\sqrt{a^2}}{\sqrt{b^2}} = \frac{|a|}{|b|}$$

In other words, the absolute value of a quotient is equal to the quotient of the absolute values of the dividend and divisor.

Theorem (3): The Triangle Inequality

Given $x, y \in R$

$$|x+y| \leq |x| + |y|$$

To prove this result, *it is important to recall the definition of absolute value of a real number x* denoted by $|x|$.

We know that $|x|$ is a *non-negative real number x* that satisfies the conditions:

$$|x| = x, \text{ if } x \geq 0.$$
$$|x| = -x, \text{ if } x < 0.$$

From the definition, it follows that *the relationship $x \leq |x|$ holds for any x*. Now we consider the following two cases.

Case (i): Let $x+y \geq 0$, then

$|x+y| = x+y \leq |x|+|y|$ (since $x \leq |x|$ and $y \leq |y|$)

Case (ii): Let $x+y<0$, then
$|x+y| = -(x+y)$
$= (-x)+(-y) \leq |x|+|y|$ (since $(-x)<|x|$ and $(-y)<|y|$). Thus, $|x+y| \leq |x|+|y|$.

Remark: The absolute value of an algebraic sum of several real numbers is no greater than the sum of the absolute values of the terms.

Examples: Of $|a+b| \leq |a|+|b|$:

10. $|0+3| = 3 \leq |0|+|3| = 0+3 = 3$
11. $|-5+0| = 5 \leq |-5|+|0| = 5+0 = 5$
12. $|3+5| = 8 \leq |3|+|5| = 3+5 = 8$
13. $|-3-5| = 8 \leq |-3|+|-5| = 3+5 = 8$
    In all four cases, $|a+b|$ equals $|a|+|b|$. On the other hand,
14. $|-3+5| = |2| = 2 \leq |-3|+|5| = 3+5 = 8$ (Hence, $2<8$)
15. $|3-5| = |-2| = 2 \leq |3|+|-5| = 3+5 = 8$ (Hence also, $2<8$)

The general rule is that $|a+b|$ *is less than* $|a|+|b|$ *when a and b differ in sign*. In all other cases, $|a+b|$ *equals* $|a|+|b|$.

Note that *the absolute value bars* in expression like $|-3+5|$ *also work like parenthesis*. We do addition before taking the absolute value.[7]

Using theorem (3), we can easily prove the following theorem (4).

Theorem (4): Given $a, b \in R$

(i) $|a-b| \leq |a|+|b|$
(ii) $|a|-|b| \leq a-b$

Proof of (i):
$|a-b| = |a+(-b)| \leq |a|+|(-b)|$

$= |a|+|b|$

$\therefore |a-b| \leq |a|+|b|$ (Proved)

Proof of (ii):
Consider $|a| = |(a-b)+b| \leq |a-b|+|b|$

[7] The numbers $|a-b|$ and $|b-a|$ are always equal and give *the distance between a and b on the number line*. This is found to be *consistent with the square root formula for distance in the plane* between the two points whose coordinates are $(a, 0)$ and $(b, 0)$.

$$\sqrt{(a-b)^2+(0-0)^2} = \sqrt{(a-b)^2} = |a-b| \tag{II}$$

$$\sqrt{(0-0)^2+(b-a)^2} = \sqrt{(b-a)^2} = |b-a| \tag{III}$$

(We shall be studying the square root formula for distance in Chapter 4, on Coordinate Geometry.)

Thus from, subtracting $|b|$ from both sides of the inequality, we have
$|a| - |b| \leq |a - b|$ (Proved)

Note: In calculus, we often want to replace one inequality with an equivalent inequality, which is simpler, as in the following example.

Example (11): Show that the inequality $|(3x + 2) - 8| < 1$ is equivalent to $|x - 2| < 1/3$.

Solution: The following inequalities are equivalent:

$$|(3x + 2) - 8| < 1$$

$$|3x - 6| < 1$$

$$|3\,(x - 2)| < 1$$

$$|3|\,|x - 2| < 1$$

$$|x - 2| < \frac{1}{3}.$$

3.12 NEIGHBORHOOD OF A POINT

Many times we are interested in the values of a function near a point "a" (say) of the domain and not in the values throughout the domain. Hence, all *the points which are close to the point* "a" *on both sides of* "a" *are of interest to us and we shall call it a neighborhood of* "a".

3.12.1 Definition

Let (a, b) be any *open interval* and let "c" *be its midpoint* then, we say that (a, b) is a neighborhood of c. Specifically, if ε is any positive number, the open interval $(a - \varepsilon, a + \varepsilon)$ is called the *ε-neighborhood of* "a." Thus, ε-neighborhood of $a = \{x \mid a - \varepsilon < x < a + \varepsilon\}$.

When we say that x is in the ε-neighborhood of "a", we write $|x - a| < \varepsilon$ and it means that, the distance of the point x, from the point "a" (on the number line) is less than ε.

3.12.2 Right Neighborhood and Left Neighborhood of "*a*"

Definition: The open interval $(a, a + \varepsilon)$ is called a *right-hand ε—neighborhood* of "a", and the open interval $(a - \varepsilon, a)$ is called a *left-hand ε-neighborhood* of "a".

We know that any neighborhood of "a" is *an open interval containing "a" as the midpoint of the interval.* Note that, *in the case of one-sided neighborhood of "a" the point "a" is not included in the neighborhood. Thus, one-sided neighborhood of "a" is also an open interval.*

3.12.3 Deleted Neighborhood of "*a*"

Definition: If the point "a" is deleted from a neighborhood of "a", then the remaining part of the open interval is called *deleted neighborhood of* "a".

Thus, the deleted δ-neighborhood of "a"

$$= \{x \mid a - \delta < x < a + \delta,\ x \neq a\},\ (\delta > 0)$$

This neighborhood is defined by the inequality $0 < |x - a| < \delta$, as already discussed above.

3.12.4 A Useful Statement

If $k < x < K$, then there is a positive number M, such that $|x| < M$.

Of the numbers k and K, *consider the one, which is farther away from the origin*. Let its distance from origin be M. *Thus, M is the larger of the numbers $|k|$ and $|K|$*. Since x lies between k and K, *its distance, from the origin, denoted by $|x|$ must be less than M*. Therefore, $|x| < M$.

Example (12): If $-6 < x < 3$, then $-6 < x < 6$. Hence $|x| < 6$.

3.13 PROPERTY OF DENSENESS

This is a very important property of real numbers. It states that *between any two different real numbers, there is always a third real number*. It follows that between any two real numbers there are infinitely many real numbers. Observe that between any two real numbers a and b, the numbers, $(a + b)/2$, $(2a + b)/2$, $(3a + b)/2$, ..., all lie between a and b. It is important to consider the following problem: *What is the smallest real number greater than 3?* Note that this question cannot be answered since there is no such real number.

To see this, *suppose c is the smallest real number greater than 3*. Then we can always find *a* number c' between 3 and c. Thus, *c will not be the smallest number*. Similarly there is no greatest real number less than 3. Of course, there is nothing special about 3. We could replace it by any other real number.

This property is also found to hold for rational numbers. But in integers, we can find the smallest number greater than 3. It is 4. Thus, *the property of denseness does not exist in the set of integers*.

Note: The *property of denseness will be significant when we discuss* the *completeness property* (or *the least upper bound property*) of real numbers.

3.14 COMPLETENESS PROPERTY OF REAL NUMBERS

Roughly speaking, this property says that the real number system is complete in itself, in the sense that *it consists of rational and irrational numbers only and that no other type of number exists in R*.[8] (Of course, this property of real numbers is very important and useful, but a beginner may skip it at this stage. He may read it later after completing Chapter 7.)

To understand this property we introduce the concept of bounded and unbounded subsets of R.

[8] Technically speaking, the axiom of completeness states as follows.... If S is any nonempty set of real numbers, which has an upper bound in R, then S has the least upper bound in R.

3.14.1 Bounded and Unbounded Subsets of *R*

Upper Bound of a Set (Definition): Let $A \subset R$, we say that A is *bounded above* if there exists a real number u such that for every $x \in A$, $x \leq u$. Such a number u is called *an upper bound of A*.

Lower Bound of a Set (Definition): The subset A is said to be *bounded below* if there is a real number l such that for every $x \in A$, $l \leq x$. Such a number l is called *a lower bound of A*.

3.14.2 Bounded Set (Definition)

If A is both bounded above and bounded below then we say that A is *bounded.*

Remark:

(i) A is bounded, iff, $A \subset [l, u]$ for some interval $[l, u]$ of finite length.

(ii) A is bounded, iff, there is a positive integer K such that $1 \times 1 < K$ for all $x \in A$. Such a number K is called a *bound of the set A.*

(iii) Set A is said to be *unbounded* if A is *not* bounded.

(iv) *An upper* bound, *a lower* bound, and *a bound* of a set are *not* unique.

(v) A set may or may not have *an upper* bound (and/or *a lower* bound) and even if it has one (or both), *the bounds may not belong to the set.*

(vi) Any *real number* is an upper bound for *the empty set* and any real number is a lower bound *for the empty set. Therefore, the empty set is bounded.* (Here, the empty set ϕ must be looked upon with reference to the set of real numbers.)

Example (16):

(a) Consider the *finite* set $B = \{2, 12, 0, 5, -7, -2\}$. Here, 12 is an upper bound and –7 is a lower bound. Hence, B is bounded. From this example, it is clear that *every finite set is bounded.*

(b) The set N of natural numbers is bounded below but not bounded above.

(c) The interval [0,1] is bounded. (These examples show that the boundedness has nothing to do with countability.)

(d) Consider the set $A = \{1, 1/2, 1/3, \ldots\ldots\}$. This set consists of all numbers of the form $1/n$ where $n \in N$, the set of natural numbers. We observe that all the numbers in A are less than or equal to 1. In this case, 1 is an upper bound of A and thus A is bounded above. Also, we observe that no number of A is less than "0". Therefore, we shall say that "0" is a lower bound of A, and that A is bounded below. Thus, for any element $x \in A$, we have $0 < x \leq 1$. We therefore say that A is bounded.

We have mentioned above that a set "*A*" is *unbounded* if it is not bounded. To discuss about unbounded sets recall that a set is bounded iff it is bounded above and also bounded below. Therefore, we define unbounded set.

3.14.3 Unbounded Set

A set is unbounded if either it is not bounded above and/or it is not bounded below.

Now we ask the question: When *will you say that u is not an upper bound of A?* We know that a number *u is an upper bound of A if the relation $x \leq u$ holds for all $x \in A$*. Hence, *u* will *not* be an upper bound of *A* if there is some member of *A*, say $\alpha \in A$ such that $\alpha > u$.[9]

Example (17): Consider the sets

$C = \{4, 6, 8, 10, \ldots\}$, $D = \{0, -1, -2, -3, \ldots\}$

Observe that each element of *C* is greater than or equal to 4. Hence, 4 is a lower bound of *C* and thus *C* is bounded below. *Is C bounded above?*

From the nature of the elements of *C*, we note that *for any number u, however large,* there are always elements of *C* greater than *u*.

Therefore, *u cannot be an upper bound of C*. Hence, *no real number can be an upper bound of C.* Thus, *C* has no upper bound[10].

Similarly, it can be seen that the set *D* is not bounded below although it is bounded above. Hence, both the sets *C* and *D* are unbounded sets.

Remark: *If a set is bounded above, it has infinitely many upper bounds* [because if *u* is an upper bound so is $(u+1)$] and similarly *if it is bounded below, it has infinitely many lower bounds*.

Example (15): Consider the set $A = \{1, 1/2, 1/3, 1/4, \ldots\}$. It is bounded above and has 1 as an upper bound. But, 2, 3, 4, ... *are also its upper bounds*. In fact, any number greater than 1 is an upper bound. Similarly, *zero and any number less than zero is a lower bound of A.*

Now we may naturally ask whether there is *the smallest* (or the least) of all the upper bounds. This leads us to the concept of least *upper bound* of a set.

3.14.4 Definition: The Least Upper Bound (l.u.b.) of a Set

Let the subset *A* of *R* be bounded above. A number *M* is called the *least upper bound* (*l.u.b.*) for *A* if

(a) *M* is an upper bound for *A*, and
(b) No number smaller than *M* is an upper bound for *A*.

If such a number *M* exists, we write $M = l.u.b.\ A$. [The l.u.b. for a set is also called the supremum of *A* and is denoted by (sup) *A*.]

As in the case of upper bounds, we see that if a set is bounded below, it has *infinitely many* lower bounds. If there is the largest of these lower bounds, we call it the *greatest lower bound (g.l.b.)* of the set.

3.14.5 Definition: The Greatest Lower Bound (g.l.b.) of a Set

Let the subset *A* of *R* be bounded below.

[9] Note that the considerations of such negations are useful, as they throw a new light on concepts and help us to understand them better.
[10] Geometrically, it means that the "set of points of *C*" keeps on extending indefinitely on the right of 4 on the real line. Thus, no finite segment of the real line can contain all the points of an unbounded set.

A real number m is called the *greatest lower bound (g.l.b.)* of set A if

(a) m is a lower bound for A, and
(b) No number greater than m is a lower bound of A.

We write, $m = g.l.b. A$. [The $g.l.b.$ for a set is called infimum of the set and is denoted by (inf) A.]

Remark: If a set has *l.u.b.*, then it is *unique*. That is, a set can have only one *l.u.b.*

Proof: Suppose M and M' are two *l.u.b.s*, then we get
$M \leq M'$ (since M is a *l.u.b.*) and $M' \leq M$ (since M' is a *l.u.b.*)
It follows that $M = M'$.
This explains why in the definition we say *the l.u.b.* (and not "a *l.u.b.*") of a set.

Note: We know that the empty set ϕ is bounded. Further, since every real number is an upper bound for ϕ, so ϕ does not have a *l.u.b.* Similarly, ϕ does not have a *g.l.b.*

Example (18):

(a) Consider $A = \{1, 1/2, 1/3, \ldots\}$. Here, the *l.u.b.* is 1 and it belongs to A. But, the *g.l.b.* is 0, which is not in A.
(b) If $B = \{1/2, 3/4, 7/8, \ldots, (2 - 1)/2, \ldots\}$. It can be shown that the $g.l.b. = 1/2 \in B$ and the $l.u.b. = 1 \in B$.
(c) The set (3, 4) is an open interval. It does not contain its *g.l.b.* or its *l.u.b.,* which are 3 and 4, respectively.
(d) The *g.l.b.* and the *l.u.b.* for $\{5\}$ are both equal to 5.
(e) The empty set ϕ is bounded. However, ϕ has neither the least upper bound nor the greatest lower bound.
(f) The set $N = \{1, 2, 3, \ldots\}$ has the *g.l.b.*, $1 \in N$. There is no *l.u.b.*, since N is not bounded above.

Note: We have seen that,

(a) a set has the *l.u.b.* only if it is bounded above and
(b) the empty set ϕ, which is bounded (and hence bounded above), has no *l.u.b. So now the question is: If a nonempty set is bounded above, does it necessarily have the l.u.b.?*

3.14.6 Discovery of Real Numbers

On the basis of the algebraic properties of real numbers, it is difficult to answer this question. However, we can answer this question for the set of rational numbers. *We will show that this set does not possess the l.u.b. property*. For the moment, *let us assume that we know only rational numbers.*[11]

[11] We recall the terms *an upper bound and the l.u.b. for a subset of rational numbers, exactly as we did for the real numbers*. For example, a rational number M is called the l.u.b. of $A \subset Q$ if

(a) M is an upper bound of A, and
(b) No rational number less than M is an upper bound of A.

We will now show that *there is a set of rational numbers, which is bounded above, but it does not necessarily have the l.u.b. in rational numbers.* To show this, consider the set $A = \{x \in Q \mid x > 0 \text{ and } x^2 < 2\}$. Thus, A is the set of all those *positive rationals* whose square is less than 2.

Clearly, the set A is bounded above. *In fact, every positive rational number whose square is greater than 2* is an upper bound for this set. *But in this case, the l.u.b. is* $\sqrt{2}$. Some elements of set A (in the decimal form) are $1, 1.4, 1.414, \ldots$, which are all rational numbers, but the l.u.b. for A in R is $\sqrt{2}$, which is not in the set of rational numbers, as we know. Hence, among the rational numbers, this set has no l.u.b., but among the reals, it has the l.u.b. $\sqrt{2}$. Thus, if we had never heard of irrational numbers, then we would say that A has no l.u.b. *However, we have seen that a nonempty set A, of real numbers, bounded above necessarily has a l.u.b.—a fact that is not at all obvious.*

For real numbers, however, it is not possible to show that every nonempty set bounded above has the l.u.b. *Therefore, we take this property of the entire real number system as an axiom,* called *the completeness property or the axiom of l.u.b. or the axiom of least upper bound for real numbers.*

3.14.7 The Axiom of Least Upper Bound

If A is any nonempty subset of R, which has an upper bound in R, then A has the least upper bound in R.

The l.u.b. axiom does not hold if R is replaced by Q, the set of rational numbers. Thus, the l.u.b. property distinguishes real numbers from the rational numbers. *The axiom says roughly that R visualized as a set of points on a line has no gaps in it*. In other words, the real number system is complete in itself, in the sense that it does not have any other type of numbers different from rational and irrational numbers. For this reason, the *l.u.b. property* is also called *the completeness property* of real numbers.[(12)]

Remark: Note that, whereas the property of denseness is possessed by both the sets Q and R, *the property of completeness (or the l.u.b. property) is possessed only by the set R and not by the set Q of rational numbers.*

From the *l.u.b.* axiom, the following property of the *g.l.b.* can be proved.[(13)]

3.14.8 The Axiom of Greatest Lower Bound

If A is any nonempty subset of R, which has a lower bound in R, then A has the greatest lower bound in R.

Before considering the problems of the *l.u.b.*, let us examine the definition(s) of *l.u.b.* in detail and put them in a more convenient form. Let us recall the definition of the *l.u.b.*

Definition I (l.u.b.): (Old definition at 3.14.4)

Let $A \subset R$ be bounded above. A number M is called the l.u.b. of A if

(a) M is an upper bound for A; and
(b) No number smaller than M is an upper bound for A.

[(12)] It is not possible to show an example of a (nonempty) set of real numbers, which is bounded above but does not have a *l.u.b.* (In fact, any set of real numbers is nonempty.)

[(13)] One can deduce g.l.b. axiom from the *l.u.b.* axiom. *Remember that neither can be proved independently from the other properties of the real numbers.*

If such a number M exists, we write $M = l.u.b. A$. We see that a real number M is the $l.u.b.$ of A, *if it satisfies both the properties (a) and (b).*

First of all, M is an upper bound of A. Hence, every element of A must be less than or equal to M (i.e., no element of A is greater than M).

Second, *any number smaller than M is not an upper bound.* It means that if we choose *any number smaller than M* (for example, $M - \varepsilon$, where $\varepsilon > 0$) then there must be atleast one element of A greater than the number

$$(M - \varepsilon)^{(14)}.$$

Hence, *we may restate the definition of the l.u.b.* in Section 3.15.

3.15 (MODIFIED) DEFINITION II (l.u.b.)

M is the l.u.b. of A if

(a) $a \leq M$, for every $a \in A$, and
(b) *for any positive number* ε, there is *atleast one member* a_o *of* A such that $a_o > M - \varepsilon$.

Using arguments similar to those for the *l.u.b.*, we can restate the definition of *g.l.b.* in Section 3.16.

3.16 (MODIFIED) DEFINITION II (g.l.b.)

We say that "m" is the *g.l.b.* of A if

(a) $a \geq m$, for every $a \in A$ (i.e., *m is a lower bound of A*).
(b) *for any positive number* ε, there is atleast one member a_o of A which is less than $m + \varepsilon$ (i.e., $a_o < m + \varepsilon$).

Example (19): Let $A = (a, b) = \{x \mid a < x < b\}$

We observe that b is an upper bound of A. Also if we take any number c smaller than b, then there is atleast one member of A greater than c (for example, $b+c/2$). Hence, by the above definition (II), b is the *l.u.b.* of A. *Note that* $b \notin A$.

Example (20): Let $A = \{-6, -4, -2, 0, 2, 4\}$

Consider the upper bounds of this set. We observe that *every number* ≥ 4 *is an upper bound.* The smallest of these upper bounds is obviously 4. Hence, 4 is the *l.u.b.* Also it may be seen that if we take any number c less than 4, then it cannot be the *l.u.b.* of A, since there is one member of A greater than c (this member is 4 itself). Thus, 4 is the *l.u.b.* of A. It is also the greatest element of A, $4 \in A$.[15]

[14] Here ε represents an arbitrary positive real number, which can be used to get a number smaller than M.

[15] A set may or may not have the greatest (or the least) element of the set. For example, the set {1, 1/2, 1/3, ...} does not have the least element. In fact, the least element (i.e., the g.l.b.) of this set is "0," which is not defined by $1/n$ for any value of n. (Of course, the limit of the function $1/n$, as n approaches infinity equals zero, but this concept will be clear only when we discuss the concept of limit in Chapters 7a and 7b.)

Example (21): Let $A = \{1, 2, 3, 4, \ldots\}$

Here, A is not bounded above. It has no upper bound. *Hence, A has no l.u.b.*

Example (22): If $B = \left\{\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \ldots, \frac{2^n - 1}{2^n}, \ldots\right\}$ then $g.l.b. = \frac{1}{2}$ and $l.u.b. = 1$.

Proof: Observe that every number $\leq 1/2$ *is a lower bound for B*. The greatest of these lower bounds is obviously 1/2. *Hence, 1/2 is the g.l.b.*

Also it is clear that no real number y which is greater than 1/2 can be a lower bound for B, *since 1/2 which is an element of B is less than y*. Thus, 1/2 is *the greatest lower bound* (*g.l.b.*).

To prove that *l.u.b.* $B = 1$.

We observe that every element of B is less than 1. *Hence, 1 is an upper bound for B.* Then every real number bigger than 1 is also an upper bound of B.

Now, *we will show that any number less than 1 is not an upper bound of B*. For this purpose, suppose $\varepsilon > 0$ and let $1 - \varepsilon$ be an upper bound of B.

Then, we must have

$$\frac{2^n - 1}{2^n} < 1 - \varepsilon \; \forall \, n \in N \text{ or } 1 - \frac{1}{2^n} < 1 - \varepsilon$$

$$\text{or } \frac{1}{2^n} > \varepsilon \text{ or } 2^n < \frac{1}{\varepsilon}$$

But we *can always find an integer n such that $2^n$is greater than* $1/\varepsilon$ (for any $\varepsilon > 0$). *This is a contradiction*. Thus, $1 - \varepsilon$ *cannot be an upper bound* of B for any ε. Thus, *1 is the smallest* (or the least) *of all the upper bounds*. We, therefore, *write l.u.b. B = 1*,

Example (23): Let $A = \{0, 1-1/2, 1-1/3, 1-1/4, \ldots\}$
$= \{0, 1/2, 2/3, 3/4, 4/5, \ldots\}$

From the elements of set A, *we guess that 1 may be the l.u.b.* Let us check this. We observe that each member of A is less than 1. Hence, 1 is *an upper bound.*

Let $\varepsilon > 0$. *Consider the number 1 – ε, which is less than 1.* Let us suppose that $1 - \varepsilon$ is an upper bound. We note that *the nonzero elements of A are of the form (1–1/n), n ∈ N.*

If $1-\varepsilon$ is an upper bound of A, then we will have,

$$1 - \frac{1}{n} < 1 - \varepsilon \text{ or } \frac{-1}{n} < -\varepsilon \text{ or } \frac{1}{n} > \varepsilon \text{ or } n < \frac{1}{\varepsilon}$$

But we can always find a natural number n such that $n > \frac{1}{\varepsilon}$, for any $\varepsilon > 0$. *This is a contradiction.* Thus, $1 - \varepsilon$ cannot be an upper bound of A for any ε. Thus, 1 is the *l.u.b.* of A.

4 From Geometry to Coordinate Geometry

4.1 INTRODUCTION

Geometry appears to have originated from the need for measuring land. Today, *geometry* is a branch of mathematics in which we study the properties of various figures. It is believed that Egyptians and Babylonians (2000–1600 BC) were the first to use geometry, but mostly for practical purposes. They had discovered many geometrical properties of simple figures (i.e., triangles, rectangles, etc.) through actual measurements, but they never developed it as a systematic discipline. Geometry was also studied and taught by ancient Indians and references to this are contained in Vedic literature.

Later on, this knowledge was passed on to Greeks who developed the subject systematically. The pioneer in this science was the Greek mathematician Euclid who lived around 300 BC. He initiated a completely new approach in the study of geometry. He showed that by knowing certain measurements in geometrical figures, the remaining ones could be found out by calculation and thus one need not depend on actual measurements to know all the facts in geometry. He is said to be the father of geometry.

In Euclidean Geometry, the approach was to start with three undefined concepts (or terms) namely *point, line, and plane*. Suggested by physical experience, certain properties are attributed to these terms. These terms along with the properties attributed to them are called axioms or postulates. Euclid called them the *self-evident truths*.[1]

He showed that by accepting these axioms as true, other geometrical facts can be derived by logical reasoning. The new results, so derived are called theorems, which reveal to us the interesting and useful properties of various geometric figures. Now the question is: What is a proof? The process of establishing a conclusion by deductive logical reasoning on the basis of axioms and previously proved theorems is called proof.

What must you know to learn calculus? 4-Coordinate geometry (Cartesian coordinates, distance formula, inclination and slope of a line, loci and their equation(s), equations of lines and their slopes)

[1] It must be clear that axioms are simple and obvious *facts* that we observe. For example

(i) When two distinct lines intersect, their intersection is exactly *one* point.
(ii) Infinite number of lines can pass through a given point, and so on.
(iii) Two distinct points determine one and only one line, and so on.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

Before we proceed further, it is useful to understand clearly what we mean by the terms point, line, and plane in geometry.[(2)]

This was the only approach to geometry for some 2000 years till the French philosopher and mathematician Rene Descartes (1596–1650) published "La Geometrie" in 1637 wherein he introduced the analytic approach by systematically using algebra in his study of geometry.[(3)]

He combined algebra and geometry in a fashion that had not been accomplished previously and laid the foundations for *Calculus*. *This wedding of algebra and geometry is known as coordinate geometry or analytic geometry*. This was achieved by representing points in the plane by *ordered pairs of real numbers* (called Cartesian coordinates, named after Rene Descartes) and *representing lines and curves by algebraic equations*.

4.2 COORDINATE GEOMETRY (OR ANALYTIC GEOMETRY)

Coordinate Geometry differs in procedure from the geometry studied in high school, in that the former makes use of the coordinate system. *It includes the study of points, lines, curves, angles, and areas in a plane, with the help of algebra*.

We are familiar with the representation of real numbers on a line, which we call *the number line* denoted by R. Descartes and Fermat, introduced two perpendicular lines (called axes) and agreed to represent *any point in the plane by its directed distances (or signed distances) from the two axes*.

[(2)] From a practical point of view, we have got some ideas about the terms point, line, and plane. However, there are difficulties in defining these terms.

For example, we think of a point as a fine and tiny dot made by a sharp pencil on a paper. Also, the top of a needle or a very small hole made by a pointed pin on a sheet of paper can be considered as examples that are very close to the concept of a point. The most important idea involved in this concept is that a point is assumed to have no physical dimension (i.e., it has no length or width, etc.)

Similarly, the idea of a line comes in our mind by considering the edge of a paper in our note book, the intersection of two walls, a piece of thin wire, or a tight and stretched thread. The most important idea involved in this concept is that a line is assumed to have only one dimension, namely "length". It has no width and thickness. We use arrow heads at the ends of a line segment to say that a line has unlimited length.

To get an idea of a plane, one can think of surface of a smooth wall, a black board, top of a table, or a sheet of a paper. Again, one should keep in mind that a plane is assumed to have only two dimensions, namely "length" and "width". It has no thickness. Also, a plane is assumed to extend indefinitely in all sides.

Remark: If we try to define a "point " as a "mark" observed at the intersection of two distinct lines, then the term "line" enters into the definition, which is not defined. Similarly, if we try to define a line as a "mark" observed at the intersection of two planes, then the term plane enters into the definition and it must be defined. Thus, an attempt to define these terms makes them more difficult to understand, than what we already know about them.

It is therefore appropriate to accept these terms with the properties attributed to them, as "*self-evident truths*" so that they are not required to be proved. Moreover, this understanding has helped in obtaining many other facts (by logical reasoning) that are useful and interesting, as properties of many geometrical figures. This is how Euclid has contributed in the progress of geometry.

[(3)] Descartes complained that the geometry of Greeks was very much tied to figures, so he desired to have a simpler approach for understanding the subject by using algebra. Fortunately, for the world, a great deal of progress had been made in algebra during the latter half of sixteenth century and the early part of the seventeenth. Another French mathematician Pierre de Fermant (1601–1665) is also credited with the invention of coordinate geometry. His work was known after his death. Both Descartes and Fermant working independently of each other saw clearly the potential in algebra for the representation and study of curves.

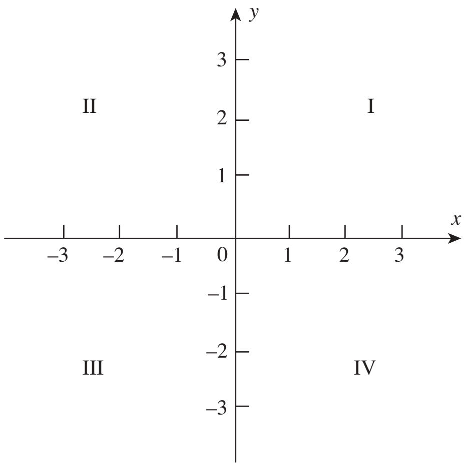


FIGURE 4.1 The number plane R^2.

4.2.1 The Plane and the Cartesian Coordinates

We take two copies of the number line (*with equal scales*) and place them perpendicular to each other in a plane, so that they intersect at the point "O" (say). We call this point of intersection, *the Origin*, from where all the distances on both the axes should be measured.

It is *customary* to have one of the lines *horizontal, with the positive numbers located to the right of* "O", and call it the *x-axis*. The other line is usually called the *y-axis*, *with the positive numbers lying above* "O". Then the points *to the left of* "O", on the x-axis and *those lying below* "O" on the y-axis represent *negative numbers*. These axes (called coordinate axes) divide the plane into *four regions, called quadrants*, labeled I, II, III, and IV numbered in the counter clockwise direction (Figure 4.1).

Now, let P be any point in the plane. Through P, we draw perpendicular(s) to respectively *x-axis* and *y-axis*. Let the foot of the perpendicular on the x-axis meet there at the point "a" and that on the y-axis at the point "b", then we associate P to the ordered pair (a, b) of numbers. If the point P is identified with the ordered pair (a, b) of real numbers, *we sometimes write P* (a, b) *for P*. Note that, in the ordered pair (a, b), the first number "a" is the *x-coordinate* (or *abscissa*); and the second number "b" is the *y-coordinate* (or *ordinate*). The origin is a point whose x and y coordinates are both "0". Hence, we identify the origin "O" by the ordered pair (0, 0) (Figure 4.2).

The set of all ordered pairs of real numbers is called the *number plane* denoted by R^2, and *each ordered pair (x, y) is called a point in the number plane*. Just as R, the set of real numbers, can be identified with points on an axis (a one-dimensional space), we can identify R^2 (i.e., the number plane) with points in a geometric plane (a two-dimensional space).

There is a one-to-one correspondence between the points in a geometric plane and the number plane R^2; that is, with each point in the geometric plane, there corresponds a unique ordered pair (x, y), and with each ordered pair (x, y) of real numbers there is associated only one point.[(4)]

[(4)] Because of this one-to-one correspondence, we identify the number plane R^2 with the geometric plane. Also, for this reason we call an ordered pair (x, y), a point.

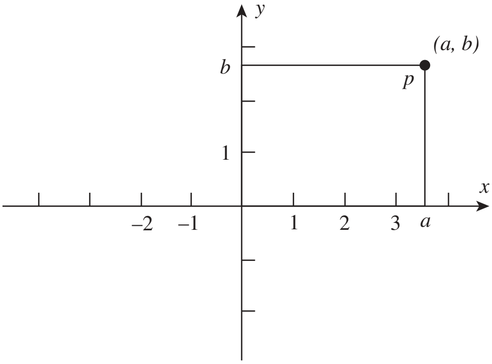


FIGURE 4.2

This system of coordinating ordered pair (x, y) with every point in the (geometric) plane is called *the rectangular Cartesian coordinate system*. Figure 4.3 illustrates a rectangular Cartesian coordinate system with some points.

The convention of the positive and negative signs, marked with the numbers in the ordered pairs, in different quadrants, follows from the very definitions of x-axis and y-axis in terms of signed lengths of line segments.

4.2.2 The Notion of Directed Distance (or Signed Length)

If A is the point (x_1, y_1) and B is the point (x_2, y_1) (i.e., A and B have *the same ordinate but different abscissas*), then the directed distance from A to B is denoted by $\overline{AB}$, and we define,

$$\overline{AB} = x_2 - x_1$$

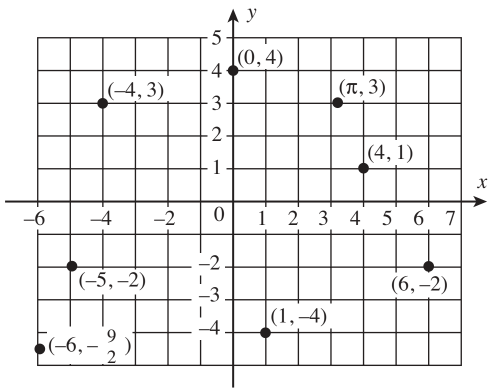
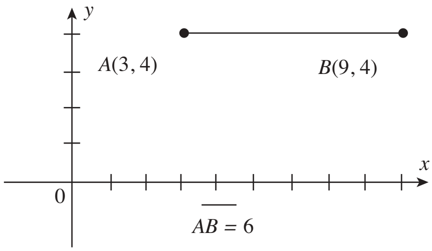


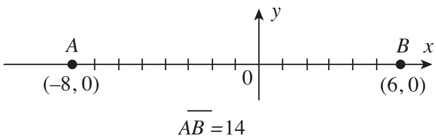
FIGURE 4.3

Illustration (1):

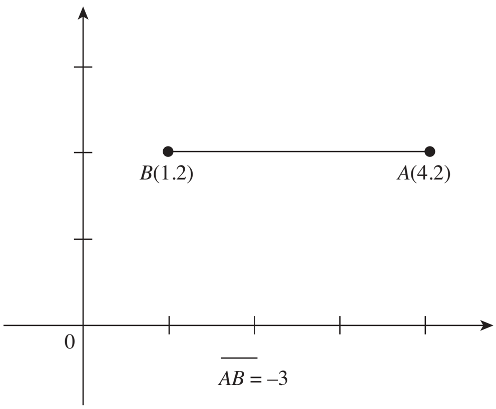
(a) If the two points are $A(3, 4)$ and $B(9, 4)$ then, $\overline{AB} = 9 - 3 = 6$.



(b) If the two points are $A(-8, 0)$ and $B(6, 0)$ then, $\overline{AB} = 6 - (-8) = 14$.



(c) If the two points are $A(4, 2)$ and $B(1, 2)$ then, $\overline{AB} = 1 - 4 = -3$.



We see that $\overline{AB}$ is positive if B is to the right of A, and $\overline{AB}$ is negative if B is to the left of A.

If C is the point $(x1, y1)$ and D is the point $(x1, y2)$, then the directed distance from C to D, denoted by $\overline{CD}$, is defined by

$$\overline{CD} = y2 - y1$$

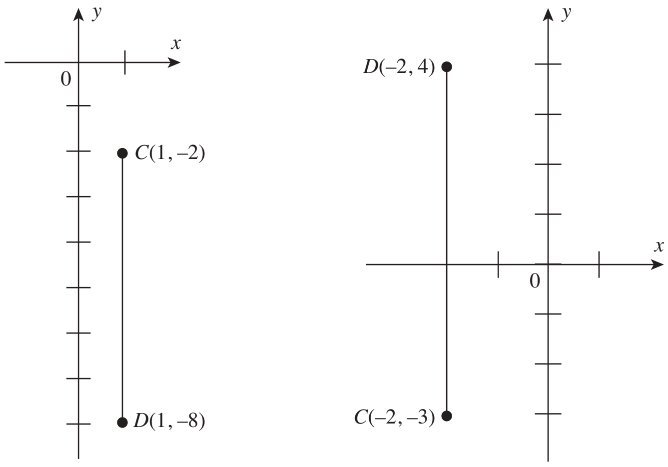
Illustration (2):

(a) The directed distance between the points $C(1, -2)$ and $D\ (1, -8)$, from C to D is given by

$$\overline{CD} = -8 - (-2) = -6$$

(b) The directed distance between $C(-2, -3)$ and $D\ (-2, 4)$ is given by

$$\overline{CD} = 4 - (-3) = 7$$



(*a*) (*b*)

The number $\overline{CD}$ is positive if D is above C, and $\overline{CD}$ is negative if D is below C.

Note (1): The terminology *directed distance* (or signed length) indicates both a *distance* and a *direction*, positive or negative. Note that we can talk about *positive or negative direction* only with reference to the horizontal and vertical line segments. *Thus, if a line segment parallel to the x-axis from x_1 to x_2 is denoted by the signed length* Δx *then the same line segment from x_2 to x_1 will be denoted by the signed length* $-\Delta x$. A similar statement is applicable to any line segment of signed length Δy, parallel to the y-axis.

Note (2): The introduction of Cartesian coordinates allows us to use numbers and their arithmetic as a tool in studying geometry. The term "analytic geometry" (or coordinate geometry) is used for *the study of geometry using coordinates*. This coordinate system also allows us to draw geometric pictures, which illustrates a great deal of numerical work.

Note (3): The notion of signed length is used for defining *the slope of a line*, to be studied shortly.

If we are concerned only with *the length of the line segment* between two points $P(x_1, y_1)$ and $Q\ (x_2, y_2)$, without regard to direction, then we use the word *distance* to mean an undirected distance (or unsigned length) between P and Q. The *horizontal distance* between the points

P and Q is denoted by

$$|x_2 - x_1| = |x_1 - x_2|$$

and the *vertical distance* is given by $|y_2 - y_1| = |y_1 - y_2|$.

If the line segment joining P and Q is neither horizontal nor vertical, then we can find the distance between the two points, as follows.

4.3 THE DISTANCE FORMULA

To find the distance between two points $P(x_1, y_1)$ and $Q(x_2, y_2)$ in the plane, we construct the right triangle ΔPRQ as shown in Figure 4.4.

The point $R(x_2, y_1)$ in the figure has the same x-coordinate as Q and the same y-coordinate as P.

$$\text{Therefore,}\quad |\Delta x| = |x_2 - x_1| = \text{length } PR \quad\text{and}\quad |\Delta y| = |y_2 - y_1| = \text{length } RQ$$

Now, *the distance* between $P(x_1, y_1)$ and $Q(x_2, y_2)$ is the length d of the hypotenuse of the right triangle QRP; so, by the *Pythagorean theorem*, we get

$$d^2 = |\Delta x|^2 + |\Delta y|^2 \tag{1}$$

Since the terms in (1) are squared, *the absolute-value symbols are not needed*, so that

$$d^2 = (\Delta x)^2 + (\Delta y)^2$$

and

$$d = \sqrt{[(\Delta x)^2 + (\Delta y)^2]} = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}$$

$$d = |PQ| = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}$$

This formula holds for all possible positions of P and Q in all four quadrants.

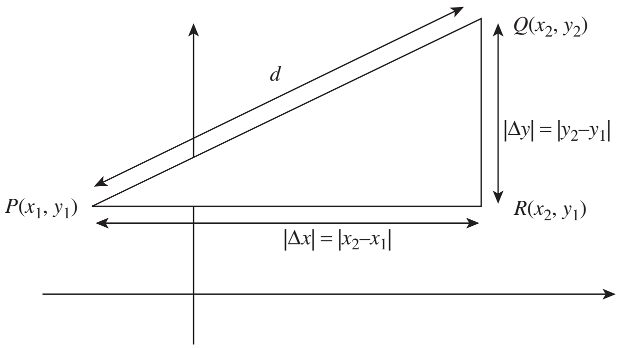


FIGURE 4.4

Example (1): Let $P(3, -4)$ and $Q(-2, 1)$. Find the distance between P and Q.

Solution: By the distance formula,

$$|PQ| = \sqrt{(-2-3)^2 + (1-(-4))^2}$$
$$= \sqrt{25+25} = \sqrt{50} = 5\sqrt{2}$$

(Note that, by knowing the coordinates of two points in a plane, we have been able to compute the distance between them.)

If P and Q are on the same horizontal line, then $y_2 = y_1$ and

$$d = \sqrt{(x_2 - x_1)^2 + 0^2}$$
$$= |x_2 - x_1| \left(\because \sqrt{a^2} = |a| \right)$$

Similarly, if P and Q are on the same vertical line then $x_2 = x_1$ and

$$d = \sqrt{0^2 + (y_2 - y_1)^2} = |y_2 - y_1|$$

4.4 SECTION FORMULA

We now obtain the formulas for finding the coordinates of the midpoint of a line segment. Let $M(x, y)$ be the midpoint of the line segment from $P_1(x_1, y_1)$ to $P_2(x_2, y_2)$.

Refer to Figure 4.5. Because ΔP_1RM and ΔMTP_2 are congruent,

$$|\overline{P_1R}| = |\overline{MT}| \quad \text{and} \quad |\overline{RM}| = |\overline{TP_2}|$$

Thus,

$$x - x_1 = x_2 - x \qquad y - y_1 = y_2 - y$$
$$\therefore 2x = x_1 + x_2 \qquad \therefore 2y = y_1 + y_2$$
$$\therefore x = \frac{x_1 + x_2}{2} \qquad \therefore y = \frac{y_1 + y_2}{2}$$

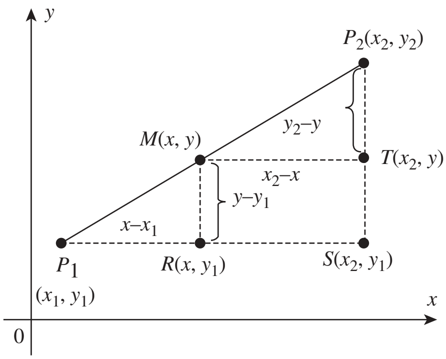


FIGURE 4.5

Similarly, it can be easily proved that the coordinates of the point $P(x, y)$ that divides the line joining $A(x_1, y_1)$ and $B(x_2, y_2)$ *internally* in the ratio $m{:}n$ are given by

$$x = \frac{mx_2 + nx_1}{m+n}, \qquad y = \frac{my_2 + ny_1}{m+n}$$

In the derivation of the above formulas, we assumed that $x_2 > x_1$ and $y_2 > y_1$. The same formulas are obtained by using any ordering of these numbers.

We now proceed to discuss about the *inclination* and *slope of a line*, which are two different concepts, but related to each other, as explained below.

4.5 THE ANGLE OF INCLINATION OF A LINE

In a coordinate plane, any line l will either intersect the x-axis or be parallel to that axis.

Definition: The angle of inclination (or simply inclination) of a line is the smallest positive angle θ (the part of the line above x-axis makes with the positive direction of x-axis)[(5)]

The angle of inclination may have any measure θ such that $0 \le \theta < 180°$. Note that the angle of measure 0° is included in the definition of inclination but the angle of 180° in not included (Figure 4.6a and b). (The angle of inclination of a line that does not cross the x-axis is taken to be 0°.)

Remark: From the above definition of inclination, it follows that *the inclination of x-axis* (or any line parallel to x-axis) is 0°, whereas the *inclination* of y-axis (or that of any line parallel to y-axis) is 90°.

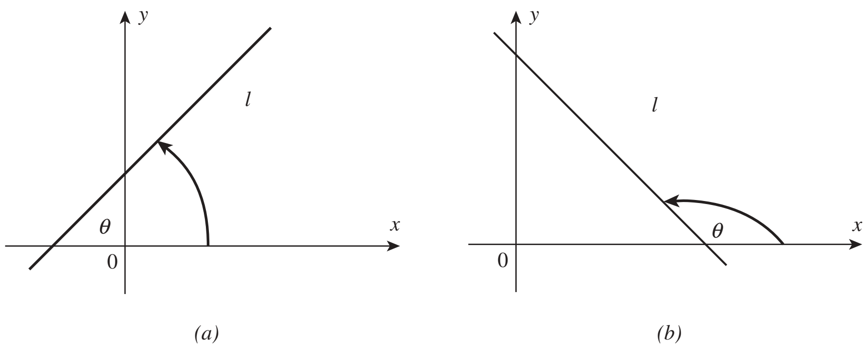


FIGURE 4.6 The angle of inclination θ $(0 \le \theta \le 180)$.

[(5)] The sense of an angle is derived from the direction of rotation of the initial side into the terminal side. If an angle is measured in the *anticlockwise direction*, its measure is said to be positive, whereas the one which is *measured in clockwise direction* is said to have negative measure.

4.5.1 Slope (or Gradient) of a Nonvertical Line

Definition (1): *The slope of a nonvertical line is defined* as the ratio of the *change in ordinates* to that of *change in abscissa*. For a given line, this ratio is a constant number denoted by *m*.

Consider a *nonvertical line segment* joining two points $P(x_1, y_1)$ and $Q(x_2, y_2)$. The number $y_2 - y_1$ gives the measure of the change in the ordinate from *P* to *Q*, and it may be positive, negative, or zero. (In the case of *horizontal line*, $y_1 = y_2$, and so $y_2 - y_1 = 0$.) The number $x_2 - x_1$ gives the measure of the change in abscissa from *P* to *Q*, and it may be *positive or negative* but *not zero* because the line is nonvertical so, $x_1 \neq x_2$ and therefore $x_2 - x_1 \neq 0$. Thus, the *slope of the line segment PQ* is expressed by

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{-(y_2 - y_1)}{-(x_2 - x_1)} = \frac{y_1 - y_2}{x_1 - x_2}, \quad (x_1 \neq x_2)$$

Remark: Parallel lines have equal angles of inclination and hence, if they are not vertical, they have the same slope. *The slope of a vertical line is not defined.* (Why?)

Remark: Since $\frac{y_2 - y_1}{x_2 - x_1} = \frac{y_1 - y_2}{x_1 - x_2}$, *it makes no difference* even if we label $P(x_2, y_2)$ and $Q(x_1, y_1)$. In other words, the slope of the line *PQ* or *QP* is same.

Definition (2): By the *slope of a line*, we mean the number of units the line *climbs up* or *falls down vertically*, for each unit of our *horizontal advance from left to right*.

From the definition of slope it follows that the slope "*m*" of a line will be *positive* if the line makes an *acute angle* with the *positive* direction of *x*-axis, and will be *negative* if it makes an *obtuse angle* with the positive direction of *x*-axis.[(6)]

Illustration (3):
If a line climbs upward three units *for each unit step we go to the right*, (as shown in Figure 4.7a), the line has the *slope 3*.
If the line falls two units downward per unit step to the right (as shown in Figure 4.7b), the line has *slope* $= -2$.

Remark:

- A horizontal line neither climbs nor falls, so it has slope 0.[(7)]
- *A vertical line climbs straight up over a single point, so it is impossible to measure how much it climbs per unit horizontal change.* Thus, *the slope of a vertical line* makes no sense. (Its calculation involves division by zero.) We say that *the slope of a vertical line is not defined.*[(8)]

(6) This becomes clear, if we use the lengths of directed line segments (i.e., the signed length).
(7) For a nonvertical line passing through $P(x_1, y_1)$ and $Q(x_2, y_2)$, the slope *m* is given by $m = (y_2 - y_1)/(x_2 - x_1)$. If the line is parallel to the *x*-axis $y_2 = y_1$; so the slope of the line is zero.
(8) If the line is vertical (i.e., parallel to y-axis), $x_2 = x_1$. Hence, the ratio defining m is meaningless, because division by zero is not permitted.

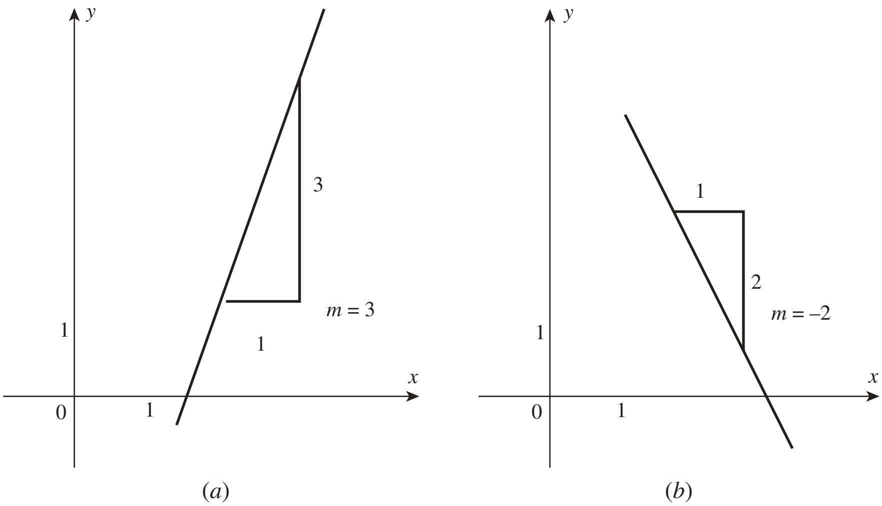


FIGURE 4.7 Slope of a non-vertical line.

Example (2): Let us find the slope of the line passing through the points $P(2, 4)$ and $Q(5, 16)$.

$$\frac{\Delta y}{\Delta x} = \frac{12}{3} = 4, \quad \text{which is the slope of the line PQ.}$$

Solution: As we go from $P(2, 4)$ to $Q(5, 16)$, we have $\Delta x = 5 - 2 = 3$ and $\Delta y = 16 - 4 = 12$. Thus, the line climbs $\Delta y = 12$ units, while we advance $\Delta x = 3$ units *to the right*. Therefore, the amount it climbs per unit horizontal advance to the right is 12/3 = 4 units.

Note: To find the slope of the line passing through the points $P(2, 4)$ and $Q(5, 16)$, in Example above, we may also proceed as follows.

As we go from $Q(5, 16)$ to $P(2, 4)$, we have *the signed lengths*

$$\Delta x = 2 - 5 = -3 \quad \text{and} \quad \Delta y = 4 - 16 = -12$$

$$\therefore \frac{\Delta y}{\Delta x} = \frac{-12}{-3} = \frac{12}{3} = 4$$

We know that, $m = \dfrac{\Delta y}{\Delta x} = \dfrac{y_2 - y_1}{x_2 - x_1} = \dfrac{-(y_1 - y_2)}{-(x_1 - x_2)} = \dfrac{y_1 - y_2}{x_1 - x_2}$

Hence, *the slope of a nonvertical line* is given by

$$m = \frac{\Delta y}{\Delta x} = \frac{\text{difference of } y\text{-coordinates}}{\text{difference of } x\text{-coordinates}}$$

provided both the differences are taken in the same order.

Example (3): The line through (7, 5) and (–2, 8) has slope

$$m = \frac{\Delta y}{\Delta x} = \frac{(8-5)}{(-2-7)} = \frac{3}{-9} = -\frac{1}{3}$$

The negative slope tells us that the line falls down vertically, as *we go horizontally from left to right*.

4.5.2 Relation Between the Inclination and the Slope of a Line

It can be shown *that if the angle of inclination of a line l is θ*, then $\tan\theta$ *gives the slope of the line*. Further two cases arise, which are as follows:

Case (i): θ is acute. Then it is clear from Figure 4.8a, that for the horizontal change Δx in the positive direction, the line climbs up by Δy *that is also positive*. Hence, *for the acute angle of inclination* θ, we have

$$\text{slope } m \text{ of the line } l = m = \frac{\Delta y}{\Delta x} = \tan\theta$$

This fact relates "*slopes of lines*" *with trigonometric functions*.

Case (ii): θ is obtuse. (see Figure 4.8b)

Here again, we consider the horizontal change Δx in the positive direction and observe that *the line l fall down by Δy, which is negative*. Hence, *for the obtuse angle of inclination*, we have slope m of the line $l = -(\Delta y/\Delta x)$, a negative number. There is another way to show that for an obtuse angle of inclination the slope of line l is given by $m = -(\Delta y/\Delta x)$.

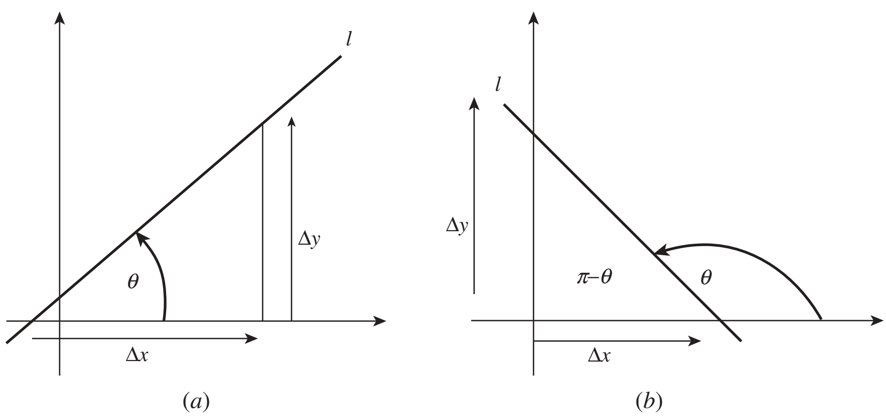


FIGURE 4.8 Another way of looking at the slope of a non-vertical line.

We write,

$$\begin{aligned}\frac{\Delta y}{\Delta x} &= \tan(\pi - \theta)\\ &= -\tan\theta \quad [\because \tan(\pi-\theta) = -\tan\theta]^{(9)}\\ \therefore \tan\theta &= -\frac{\Delta y}{\Delta x}\end{aligned}$$

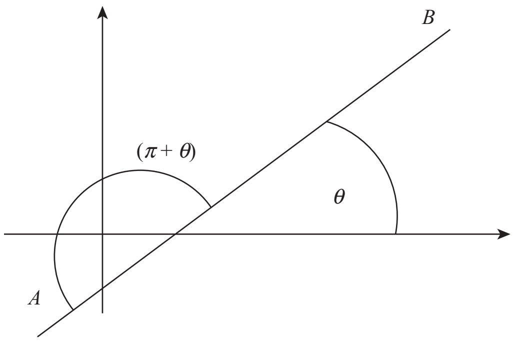
Note: For the purpose of analytic geometry, we associate a number with the inclination of the line in the following manner.

Definition: The slope of a *nonvertical* line having inclination θ, is defined to be tangent θ.

Points to Remember

Let a line with *angle of inclination* θ, have the slope m, then

1. Value of m is given by $\tan\theta$.
2. If θ is acute, m is positive.
3. If θ is obtuse, m is negative.
4. If $\theta = 0$; $m = 0$. The slope of x-axis (or any line parallel to x-axis) is 0.
5. If $\theta = \pi/2$, the line is vertical and *m is not defined.*
6. The slope m is *independent of the sense of the line segment*. [Note that if θ is the *angle of inclination* of line AB then $(\pi + \theta)$ is the angle of inclination of BA, which is the same line considered in opposite direction.]



   Now, we have slope of AB $= \tan\theta$, and slope of BA $= \tan(\pi + \theta) = \tan\theta$. Thus, *the direction of a line segment does not play any role in the measurement of its slope*.

7. A line that rises to the right has positive slope and the one that falls to the right has negative slope. [See (2) and (3) above.]
8. *The slope m is a measure of the steepness of a line* either up or down. The larger $|m|$ is, the steeper the line is (Figure 4.9).

(9) In fact, all trigonometric ratios are defined with reference to an acute angle in a right triangle. When these definitions are extended for angles of any magnitude and sign, the trigonometric ratios are still defined for acute angles. Of course, then the signed lengths of sides of the right triangle play their role. (This will be clear, when we study Chapter 5.)

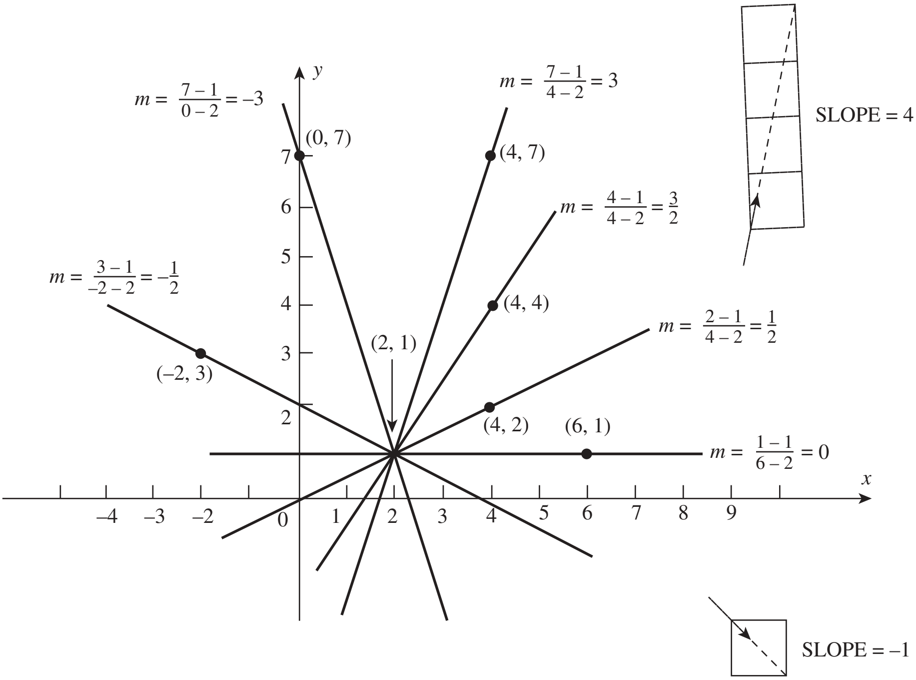


FIGURE 4.9 Lines of various slopes.

Remark (1): Note that, whereas the inclination of a *vertical line* is defined to be 90°, its slope is not defined.

Remark (2): Of all the curves, the line is the only curve having the property that, for any two distinct points $P(x_1, y_1)$ and $Q(x_2, y_2)$ on it, the value of the slope m is always constant and is given by the formula

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{y_1 - y_2}{x_1 - x_2}; \quad x_1 \neq x_2$$

For all other curves, the slope varies from point to point.

Note: "*Slope*" *is one of the central concepts of calculus*. In our study of calculus, an important concept to be learnt is *the slope of a curve at a point*. We shall return to this concept in Chapter 9, for the *derivative of a function*.

We have demonstrated how a rectangular Cartesian system can be used to obtain geometric facts (like length of a line segment and coordinates of the midpoint of a line segment, etc.) by algebra. We now show how such a coordinate system enables us to associate a *graph* (a geometric concept) *with an equation* (an algebraic concept).

4.6 SOLUTION(S) OF AN EQUATION AND ITS GRAPH

Consider *an algebraic equation in two variables x* and *y*. When *x* and *y* are replaced by specific numbers, say *a* and *b*, the resulting statement may be either *true* or *false*. If it is true, the ordered pair (a, b) is called *a solution of the equation*, and it represents a point in R^2. It can be easily seen that, in general, an equation in two variables has an unlimited number of solutions. All such ordered pairs can be graphed as points in a (geometric) plane and such a graph (which consists of an unlimited number of points) is said to represent the algebraic equation under consideration.[10]

4.6.1 Definition: Graph of an Equation

The *graph of an equation* in R^2 is the set of all those points in R^2 whose coordinates are solutions of the given equation.

The basic problem of coordinate geometry is to find algebraic equations for certain sets of points (geometrical objects), which satisfy the given geometric condition. Such a set of points is called *locus*.

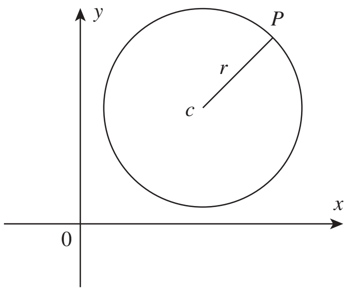
4.6.2 More About the Word "Locus"

As stated above, "locus" is a set of points satisfying a given geometric condition. Since each point of the set satisfies the given geometric condition, we may consider a representative point $P(x, y)$ of such a set. Sometimes, it is advantageous to think of a *locus as a path traced out by a moving point* satisfying at each position of its motion, the given geometric condition, characteristic of the locus. Consider the following examples:

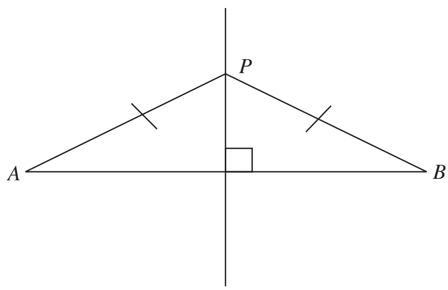
Example (4): Consider the set of points in a plane that satisfy the (geometric) condition that they are all at the same distance "r" from a fixed point "c". We can show that the set of all such points is the circle with center "c" and radius "r" (Figure 4.10a).

Example (5): Consider the *set of points P, in a plane* that satisfy the (geometric) condition that they all *are equidistant from two fixed points A and B* (Figure 4.10b).

We can show that *the set of all such points is the perpendicular bisector of the line segment AB*. We use the word *locus* for *such sets of points*.



(*a*)



(*b*)

FIGURE 4.10

[10] The equation $x^2 + y^2 = 0$, has only one solution, namely (0, 0). Hence, the graph of this equation consists of a single point. The equation $2x^2 + y^2 = -1$ has no solution, and hence it has no graph or its graph is a null set.

Note: With reference to Example (1) above, we may say that a *circle is the locus of a point P* [i.e., the path traced by $P(x, y)$] that moves in a plane, such that it always remains at a fixed distance "r" from a fixed point "c" in the plane.

In our discussion to follow, we will consider the loci (plural of locus) in planes only.

4.6.3 Locus and Its Equation

When the *locus* is looked upon as a path traced out by the (representative) moving point P, the geometric condition, which is satisfied at each position of the point $P(x, y)$, can be expressed in the form of a relation connecting the variables x and y—the coordinates of the point P. Such a relation is called the equation of the locus.

Since, a locus can be looked upon as a graph of all those points satisfying the given geometric condition, *the equation of a locus also stands for the graph representing the locus*. Thus, it is possible to study the properties of graphs with the help of their equations. *Our interest lies in finding the algebraic equation for a locus.*

4.6.3.1 Equation of a Locus Consider a circle with center $C(1, 3)$ and radius 4 units. Let $P(x, y)$ be any point on this circle.

$$\text{Then, } CP = 4 \tag{2}$$

By the distance formula, we get $CP = \sqrt{(x-1)^2 + (y-3)^2}$

Using (2), we get $\sqrt{(x-1)^2 + (y-3)^2} = 4$

On squaring both sides, we get

$$(x-1)^2 + (y-3)^2 = 16 \tag{3}$$

Equation (3) is called *the equation of locus* of $P(x, y)$. Also, Equation (3) is the equation of the circle with center $C(1, 3)$ and radius 4.

Note: *The equation of a locus is an algebraic relation between the variables x and y*, where $P(x, y)$ stands for *an arbitrary point of the locus.*

Two points should be noted:

(a) Coordinates of every point $P(x, y)$ of the locus must satisfy the equation of the locus.
(b) Any point $Q(x', y')$ satisfying the equation of the locus, must be on the locus.

The concept of *locus* forms *the basis of coordinate geometry*, and therefore a student has to first learn the following two things in coordinate geometry:

(i) *To find the locus, given an equation in x and y* (or only x, or only y) (i.e., the corresponding set of points).
(ii) *To find the corresponding equation*; in x and y (or only x, or only y), given a locus (i.e., a set of points defined by some geometric condition).[11]

[11] *Finding the algebraic equation* (in x and/or y) *for a set of points* defined by some (geometric) condition is the basic *problem of coordinate geometry.*

4.6.4 To Obtain the Equation of a Locus

To find the equation for a *set of points satisfying a given geometric condition*, we generally proceed as follows:

Step (1): Take any point $P(x', y')$ of the locus.

Step (2): Express the geometrical condition(s) of the locus by means of an algebraic relation between x' and y'.

Step (3): Replace x' by x and y' by y. The equation in x, y so obtained is *the required equation of the locus.*

Note (1): If there is no possibility of any confusion, we take *the coordinates of any point P of the locus* as (x, y) instead of (x', y').

Note (2): In the equation of a locus, *the variables x and y* [i.e., the coordinates of a (representative) moving point $P(x, y)$] are called *current coordinates.*

Note (3): *Locus represents a set of points*, satisfying a given geometric condition. *When locus is viewed as a path traced by a moving point, it represents a curve*. However, a locus, that is, a set of points satisfying a given geometric condition (defined algebraically), may also represent a region in a plane (Figure 4.11). (Thus, algebraic statement(s) defined geometric figures.)[12]

Note: Every locus (i.e., a set of points satisfying a given geometric condition) need not be represented by an equation. *It can also be an inequality*. For example, consider *the set of points lying inside a circle of unit radius with center at origin*. If $P(x, y)$ belongs to the *locus* (i.e., the set of points in question), then the condition to be satisfied by $P(x, y)$ is that

$$OP < 1[\text{i.e., } x^2 + y^2 < 1]$$

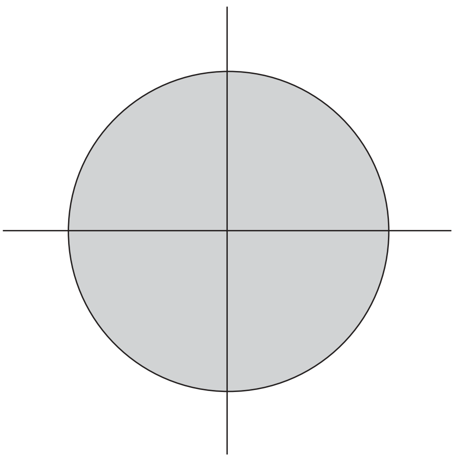


FIGURE 4.11

[12] It is for this reason that *coordinate geometry* is considered as a *wedding of geometry and algebra.*

Here, the set of points, representing locus of $P(x, y)$ *is not a curve*. Moreover, this locus is represented by an inequality. Thus, every locus is not expressed by an equality. Also, we can give examples to show that *every equation does not represent a* locus (i.e., a set of points).(13)

Illustrative Examples

Example (6): Let us find the equation of the locus of point $P(x, y)$ that satisfies the conditions as given below:

(i) Abscissa of P exceeds twice its ordinate by 7.

Solution: By the given condition x is greater than $2y$ by 7. $x = 2y + 7$, which is the required equation of the locus.

(ii) The sum of coordinates of P is 11.

Solution: Sum of coordinates of P is 11, $\therefore\ x + y = 11$, is the required equation of the locus.

(iii) P is always equidistant from $A\ (-2, 3)$ and $B\ (3, -5)$.

Solution: We have $P(x, y)$ as *a point of the locus.*

$$\therefore\ PA = PB \text{ (Given)} \tag{4}$$

$\therefore$ By the distance formula, we have

$$PA = \sqrt{[x-(-2)]^2 + (y-3)^2} = \sqrt{(x+2)^2 + (y-3)^2}$$

$$\text{and} \quad PB = \sqrt{(x-3)^2 + [y-(-5)]^2} = \sqrt{(x-3)^2 + (y+5)^2}$$

$$\therefore \text{(4) gives} \quad \sqrt{(x+2)^2 + (y-3)^2} = \sqrt{(x-3)^2 + (y+5)^2}$$

On squaring both sides, we get

$$(x+2)^2 + (y-3)^2 = (x-3)^2 + (y+5)^2$$

$$\text{i.e.,} \quad (y-3)^2 - (y+5)^2 = (x-3)^2 - (x+2)^2 \quad ^{(14)}$$

$$\text{i.e.,} \quad y^2 - 6y + 9 - (y^2 + 10y + 25) = x^2 - 6x + 9 - (x^2 + 4x + 4)$$

$$\text{i.e.,} \quad -6y + 9 - 10y - 25 = -6x + 9 - 4x - 4$$

$$\text{i.e.,} \quad -16y - 16 = -10x + 5$$

$$\text{i.e.,} \quad 10x - 16y = 21 \quad \text{or} \quad 16y = 10x - 21$$

which is the required equation of the locus (or the set of points).

(13) For example, the equation $x^2 + y^2 = -5$, does not represent any curve (nor a set of points).

(14) Here, we can use the identity $a^2 - b^2 = (a-b)\ (a+b)$ to simplify both sides. This will give us $(y-3-y-5)$ $(y-3+y+5) = (x-3-x-2)\ (x-3+x+2)$ or $-8(2y+2) = -5(2x-1)$ or $-16y-16 = -10x+5$ or $10x-16y = 21$ or $16y = 10x-21$

(iv) The sum of the squares of its distances from the axes is 9.

Solution: *The distance of* $P(x, y)$, from x-axis $= y$, and that from y-axis $= x$.

Now, according to the given condition, $x^2 + y^2 = 9$, which is the required equation of the locus.

(v) The sum of its distances from the coordinate axes equals to the square of its distance from the origin.

Solution: The distance of $P(x, y)$

(a) from x-axis $= y$,
(b) from y-axis $= x$, and
(c) from origin $0(0, 0) = \sqrt{(x-0)^2 + (y-0)^2} = \sqrt{x^2 + y^2}$

Now according to the given condition

$$x + y = \left[\sqrt{x^2 + y^2}\right]^2 \quad \text{or} \quad x + y = x^2 + y^2$$

which is the required equation of the locus.

Example (7): Derive the equation of the locus of a point $P(x, y)$, which moves so that the sum of the squares of its distances from points $A(0, 0)$, and $B(2, -4)$ is always 20.

Solution: $PA = \sqrt{(x-0)^2 + (y-0)^2} = \sqrt{x^2 + y^2}$

$$PB = \sqrt{(x-2)^2 + (y+4)^2}$$

It is given that $PA^2 + PB^2 = 20$

$$\therefore\ x^2 + y^2 + (x-2)^2 + (y+4)^2 = 20$$
$$\therefore\ x^2 + y^2 + x^2 - 4x + 4 + y^2 + 8y + 16 = 20$$
$$\therefore\ 2x^2 + 2y^2 - 4x + 8y + 20 = 20$$
$$\therefore\ x^2 + y^2 - 2x + 4y = 0 \qquad \text{(A)}^{(15)}$$

This is the equation of the locus.

Example (8): A point moves so that its distance from the y-axis is always equal to its distance from the given point $A(4, 0)$. Find the equation of the locus.

[15] This equation can also be written in the form

$$(x-1)^2 + (y+2)^2 = \left(\sqrt{5}\right)^2$$

Later, we will show that this equation represents a circle with center $C(1, -2)$ and radius $\sqrt{5}$. Also, we will show how the equation (A) can be put in this convenient form (see "shift of origin").

Solution: Let $P(x, y)$ be any point on the locus. Then the distance of $P(x, y)$ from the y-axis $= |x|$[16]

Let us denote it by PB, then $PB = |x|$

Again, the distance between $P(x, y)$, and the given point $A(4, 0)$

$$= \sqrt{(x-4)^2 + (y-0)^2} = \sqrt{(x-4)^2 + y^2}$$

Then, $P(x, y)$ must satisfy the geometric condition $PB = PA$.

$$\therefore \quad |x| = \sqrt{(x-4)^2 + y^2}$$

Squaring both sides, we get

$$x^2 = x^2 - 8x + 16 + y^2$$

$$\therefore y^2 - 8x + 16 = 0^{(17)}$$

This is the required equation of the locus.

4.6.5 Points on the Locus

If a point $P(x, y)$ belongs to a locus, then its coordinates satisfy the equation of the locus. *Conversely*, if the coordinates of a point $P(x, y)$ satisfy the equation of a locus, then the point $P(x, y)$ lies on the locus.

4.6.6 Points Not on the Locus

If a point $P(x, y)$ is *not* on the locus, then its coordinates *will not* satisfy the equation of that locus. *Conversely*, if the coordinates of a point $P(x, y)$ *do not* satisfy the equation of the locus, then the point $P(x, y)$ does not belong to the locus.

Example (9): Find the points on the x-axis, which lie on the curve whose equation is $x^2 + y^2 + 5x + 4 = 0$. Hence, find the length of the intercept (i.e., chord) made by the curve on the x-axis.

Solution: Any point on x-axis, has its y-coordinate zero. Let $(a, 0)$ be a point (on the x-axis), which lies on the curve. Therefore, coordinates $(a, 0)$ must satisfy the given equation.

$$a^2 + (0)^2 + 5(a) + 4 = 0$$

$$a^2 + 5a + 4 = 0 \quad \text{or} \quad (a+4)(a+1) = 0$$

$$\therefore a = -1 \quad \text{or} \quad a = -4$$

The required points are $(-1, 0)$ and $(-4, 0)$.

By the distance formula,

Length of the chord $= \sqrt{(-4+1)^2 + (0-0)^2} = 3$ Ans.

[16] The *distance of* $P(x, y)$ from y-axis is *the perpendicular distance of P from the y-axis, which is the absolute value of x-coordinate* of P.

[17] This is the equation of a parabola, as we will see later.

4.7 EQUATIONS OF A LINE

A line (i.e., a straight line) is a geometric object. When it is placed in a coordinate plane, the points (in the plane) through which the line passes, satisfy certain geometric conditions. For example, any two *distinct points* $P(x_1, y_1)$ and $Q(x_2, y_2)$ on the line, *determine it completely*. Also, if the line PQ is not vertical (i.e., if the line PQ is not parallel to y-axis), then its slope m is given by the number,

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{y_1 - y_2}{x_1 - x_2}, \ (x_1 \neq x_2)$$

which is a *constant*.

To obtain the equation of a line we use *the important fact that a point $P(x_1, y_1)$ and a slope m determines a unique line*. But, we know that the slope m of a horizontal line is zero and that of a vertical line is not defined. Hence, first we consider the equations of horizontal and vertical lines.

4.7.1 Equations of x-Axis, y-Axis, and the Lines Parallel to the Axes

- Observe that *the x-axis consists of all points of the form (x, 0)*. It means that *for any point on the x-axis, the y-coordinate is always zero*. Therefore, its equation is $y = 0$.
  Similarly, *the y-axis consists of all points of the form (0, y)*. Therefore, its equation is $x = 0$.
- Any line parallel to the x-axis consists of all points of the form (x, b). Therefore its equation is of the form $y = b$, for some number b. Similarly any vertical line is perpendicular to x-axis and consists of all points of the form (a, y), therefore, it has an equation of the form $x = a$, for some number a.

Remark: In our rectangular (x, y)-coordinate system, when we set coordinate variables equal to constants, we get two equations of lines: $x = a$ is vertical line and $y = b$ is a horizontal line (Figure 4.12).

Now we will consider only the equations of nonvertical lines in the following discussion.

4.7.2 Point–Slope Form of the Equation of a Line

[To find the equation of a line having the slope m, and passing through a given point $A(x_1, y_1)$]

Let a given line "l" have slope m and pass through the point $A(x_1, y_1)$ as shown in Figure 4.13. Let $P(x, y)$ be any point on the line l, other than A. Then, we have to find *an algebraic condition for the point P(x, y) to lie on the line* "l".

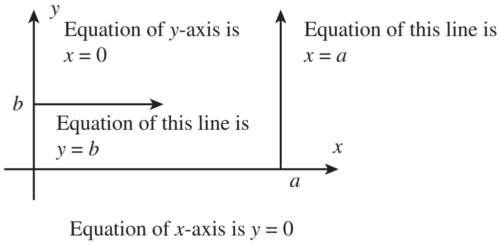


FIGURE 4.12

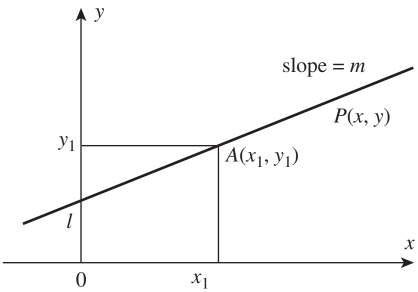


FIGURE 4.13

Since the slope of the line that joins (x_1, y_1) and (x, y) is also required to be m, therefore, the condition for $P(x, y)$ to lie on the given line is that

$$\frac{y - y_1}{x - x_1} = m \tag{5}$$

$$\text{or } y - y_1 = m(x - x_1) \tag{6}$$

Equation (6) is called *the point–slope form of the equation* of the line "l".

Example (10): Find a point–slope equation of the line passing through (2, 1) with the given slope, and sketch the line.

(a) Slope 0, (b) Slope $\frac{1}{2}$, (c) Slope -3

Solution: Point–slope form of the equation at (6) above is, $y - y_1 = m(x - x_1)$. Here, $x_1 = 2$ and $y_1 = 1$.

For (a), we have $m = 0$. Therefore equation (6) becomes $y - 1 = 0$. The line is horizontal. It is sketched in Figure 4.14.

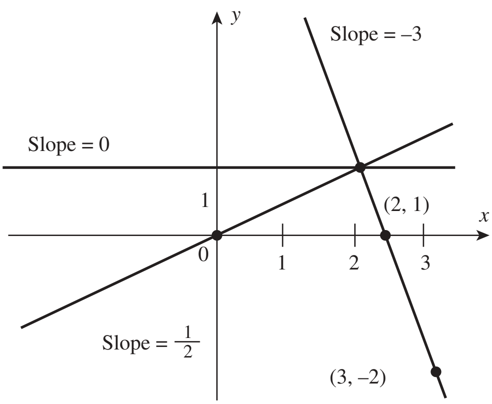


FIGURE 4.14

For (b), we have $m = 1/2$, so that equation (2) becomes $y - 1 = (1/2)(x - 2)$. To sketch this line, we need a second point on it. Now, if we put $x = 0$ in this equation, we obtain $y = 0$. Thus, the line passes through (2, 1) and (0, 0). It is also sketched in Figure 4.14.

For (c), $m = -3$, so that (2) becomes $y - 1 = -3(x - 2)$.

By choosing any (convenient) value for x, we can get a corresponding value for y, and thus obtain the coordinates of a second point on the line. However, we give below a very useful idea of finding the coordinates of a second point on the line.

Note that the slope of the line is -3. *It tells us that by moving 1 unit to the right causes a change of –3 units in the value of y*. Therefore, with reference to the point (2, 1) on the line, we easily get another point on the line, *with the x-coordinate as $2 + 1 = 3$ and y-coordinate as $1 + (-3) = -2$.*

[Note carefully the method of obtaining the coordinates of a new point on a line, when the slope of the line and a point $P(a, b)$ on the line are known.]

Now, it is easy to sketch the line passing through the points (2, 1) and (3, −2). This line also appears in Figure 4.14.

4.7.3 Slope–Intercept Form of the Equation of a Line

The point–slope form of the equation of a line, given at (6) above can also be rewritten in the form

$$y = mx + (y_1 - mx_1)$$

$$\text{or} \quad y = mx + b, \tag{7}$$

$$\text{where } b = y_1 - mx_1$$

[Note that $(y_1 - mx_1)$ is a *real number*.]

This form of the equation is very useful. The constant b in equation (7), has a *nice interpretation*. If we set $x = 0$, in Equation (7), we get $y = \text{b}$. So the point (0, b) lies on the line. Note that *the point (0, b) is on the y-axis* and therefore the line makes an intercept b on the y-axis. It is for this reason that equation (7) is called the *slope–intercept form of the equation of a line*.

Note: *The slope–intercept form* of equation of a line can also be obtained as follows.

Let a *nonvertical line* with *slope m* have the *y-intercept b*. Then, obviously, this line passes through the point (0, b). Now, using the available information, we may write its equation in point–slope form as

$$y - b = m(x - 0) \quad \text{or} \quad y = mx + b$$

which is in the slope–intercept form.

Note (1): For any *nonvertical line*, the equation of the line can always be put in the form *$y = mx + b$, in which the coefficient of x represents the slope of the line*. For the line $y = x$, the slope is 1, for $y = b$ (i.e., $y = 0 \cdot x + b$), the slope is 0, and for $3y = 7x - 5$, the slope is 7/3.

Note (2): A vertical line has undefined slope, so it does not have an equation of the form of equations (6) or (7).

Example (11): Let us find the equation of the line through (2, −3) with slope 7.

Solution: The required equation is given by

$$y - y_1 = m(x - x_1)$$

$$\text{or} \quad y - (-3) = 7(x - 2)$$

$$\text{or} \quad y + 3 = 7x - 14$$

$$\text{or} \quad y = 7x - 17$$

Here, the y-intercept of the line is -17. If it is desired to find the x-intercept, we set $y = 0$, in the equation and get $x = 17/7$, which is the x-intercept.

Remark: *Slope–intercept* form of the equation of a line involves only y-intercept of the line. There is another form of equation called "*intercept form*," which involves the x-intercept "a" and the y-intercept "b", both. We shall discuss about this form in Section 4.7.5.

4.7.4 Two-Point Equation of a Line

(Equation of a *nonvertical line passing through two given points.*)

Let l be any *nonvertical line* in the plane and $P(x_1, y_1)$, $Q(x_2, y_2)$ any two distinct fixed points on it. *Since l is nonvertical,* $x_1 \neq x_2$. The slope of the line is given by

$$m = \frac{y_2 - y_1}{x_2 - x_1}, \quad (x_1 \neq x_2)$$

Now, *using the available information*, we can easily write down the equation of the line using the *point–slope form*

$$(y - y_1) = m(x - x_1) \tag{8}$$

$$\text{or} \quad (y - y_2) = m(x - x_2) \tag{9}$$

where $m = \dfrac{y_2 - y_1}{x_2 - x_1}$

Here, it is important to note that equations (8) and (9) are equivalent.

We write the desired equation using either (8) or (9)

$$\text{Using (8), we get } (y - y_1) = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) \tag{10}$$

and it is called *two-point equation of the line.*

Similarly, using (9) we can write the two-point equation of the line as

$$(y - y_2) = \frac{y_2 - y_1}{x_2 - x_1}(x - x_2)$$

It can be verified that, on simplification, this equation and the one at (10) above, give the same equation of the line.

Note that the fraction$(y_2 - y_1)/(x_2 - x_1)$, is the slope of line l, *which is a constant independent of the points* (x_1, y_1) *and* (x_2, y_2).

Remark: For writing the equation of a nonvertical line it is a matter of convenience to use equations (6), (7), or (10), depending on the available information.

Example (12): Let us find the equation of the line through (−5, −3) and (6, 1).

Solution: We have

$$\text{Slope} \quad m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{1 - (-3)}{6 - (-5)} = \frac{4}{11}$$

Now, we may choose *any of the two given points for writing the equation of the line*. If we choose the point (−5, −3), then we get the required equation as

$$y - y_1 = m(x - x_1)$$

$$y + 3 = \frac{4}{11}(x + 5)$$

$$\text{or} \quad 11y + 33 = 4x + 20$$

$$\text{or} \quad 4x - 11y = 13 \tag{i}$$

$$\text{or} \quad y = \frac{4}{11}x - \frac{13}{11} \tag{ii}$$

If we choose the other point (6, 1), we get the equation of the line as

$$y - 1 = \frac{4}{11}(x - 6) \tag{iii}$$

Equation (iii) can be simplified to (i) or (ii).

If a line crosses the x-axis at $(a, 0)$, then "a" is called the x-intercept of the line. To find the x-intercept, we set $y = 0$, in the equation of the line and solve it for x. Similarly, if we set $x = 0$ in the equation of the line, and solve it for y, we get the y-intercept of the line. We shall now obtain the equation of a nonvertical line in the intercept form in which both the intercepts are reflected.

4.7.5 Equation of a Nonvertical Line in the Intercept Form (Showing Both the Intercepts)

Let l be any *nonvertical line*, which makes an intercept "a" on the x-axis and an intercept "b" on the y-axis ($a \neq 0$, $b \neq 0$).

Therefore, *by definition, the points (a, 0) and (0, b) are on the line l.* The slope of this line is given by

$$m = \frac{b - 0}{0 - a} = -\frac{b}{a}$$

Now, *using the point–slope form of the equation of a line*, we may write the equation of the above line *considering the point (a, 0) on the line as*

$$(y-0) = -\frac{b}{a}(x-a) \quad \text{or} \quad ay = -bx + ab$$

Dividing both sides of the equation by ab, we get

$$\text{or} \quad \frac{x}{a} + \frac{y}{b} = 1 \tag{11}$$

Equation (11) is called the "*intercept form*" of the equation of a line.

Similarly, equation (11) can *also be obtained by considering the point (0, b)* on the line, and the slope obtained above, we get $y - b = (-b/a)\,(x-0)$, which simplifies to equation (11).

4.7.6 General Linear Equation

(The equation of the line in the form $Ax + By + C = 0$.)[18]

It would be *nice* to have a form of the equation that covered *all lines, including vertical lines*. We have shown that the equation of a *nonvertical line* is of the form $y = mx + b$ [or $mx - y + b = 0$, where m is any real number including zero], and an equation of a vertical line is of the form $x = a$ [or $x + (0)\cdot y - a = 0$]. It can be shown that *each of these equations is a special case of an equation of the form*

$$Ax + By + C = 0 \tag{12}$$

where A, B, and C are constants and both A and B are *not zero simultaneously*. In other words, every line has an equation of the form (12).

Theorem (1): The equation $Ax + By + C = 0$, *always represents a straight line, provided A and B are not zero simultaneously*.

Proof: We consider the following three cases.

Case (I): If $B = 0$ (but $A \neq 0$), then the equation (6) becomes $Ax + C = 0$ or $x = -(C/A)$, which represents a *vertical line* (i.e., a *line parallel to y-axis*).

Case (II): If $A = 0$ (but $B \neq 0$), then equation (6) becomes $By + C = 0$ or $y = -(C/B)$, which represents a *horizontal line* (i.e., *a line parallel to x-axis*).

Case (III): If $A \neq 0$ and $B \neq 0$, we can solve the equation for y and obtain $y = -(A/B)x - (C/B)$, which represents the straight line with slope $-(A/B)$, and y-intercept $-(C/B)$.

The *converse* of the above theorem, given in the following theorem, is also true.

Theorem (2): Every straight line has an equation of the form $Ax + By + C = 0$, where A, B, and C are constants, *with the condition that both A and B are not zero simultaneously*.

[18] An equation of this type in which both x and y (or only x or only y) appear in degree one only, is called a linear equation, because its graph is a line.

Proof: Given a straight line, *either it cuts the y-axis or is parallel to it* (or coincident with it). We know that the equation of a line that has a *y-intercept* "b", can be put in the form

$$y = mx + b \tag{i}$$

Further, if the line is parallel to (or coincident with) the y-axis, its equation is of the form

$$x = x_1 \tag{ii}$$

(or $x = 0$ in case the line coincides with the y-axis).

Both equations (i) and (ii) are of the form given in the theorem; hence the proof.

4.7.7 Slope and Intercepts of the Line $Ax + By + C = 0$

The equation $Ax + By + C = 0$, can be written as

$$By = -Ax - C$$

$$\therefore\ y = -\frac{A}{B}x - \frac{C}{B} \quad (\text{if } B \neq 0)$$

Comparing this equation with the equation $y = mx + b$, we get

$$m = -\frac{A}{B}$$

$$\therefore \text{ Slope of the line} = -\frac{\text{coefficient of } x}{\text{coefficient of } y}$$

$$\left[\text{and } y \text{ intercept} = -\frac{C}{B} \quad (B \neq 0)\right]$$

[If the equation of a line is given in the form $Ax + By + C = 0$, then it is important to remember that its slope is given by the ratio $m = (A/B)$.]

Let the line $Ax + By + C = 0$, intersect the x-axis in $(a, 0)$ and y-axis in $(0, b)$, respectively. Then, $A(a) + B(0) + C = 0$ and $A(0) + B(b) + C = 0$.

$$\therefore\ a = -\frac{C}{A} \ (\text{if } A \neq 0), \qquad \therefore\ b = -\frac{C}{B} \quad (\text{if } B \neq 0)$$

$$\therefore\ x\text{-intercept} = -\frac{C}{A}, \quad \text{and} \quad y\text{-intercept} = -\frac{C}{B}$$

4.8 PARALLEL LINES

If two lines have the same slope, they are parallel. For example, $y = 2x + 1$ and $y = 2x - 3$ represent parallel lines, as both have the slope 2. The second line is 4 units below the first, for every value of x. Similarly, the lines with equations $-2x + 3y + 12 = 0$ and $4x - 6y = 5$ are parallel. (To see this, we must solve these equations for y.)

Example (13): Find the equation of the line through (6, 8) which is parallel to the line with equation $3x - 5y = 11$.

Solution: We solve $3x - 5y = 11$ for y, and we get $y = (3/5)x - (11/5)$. This equation shows that the slope of this line is 3/5. Since the desired line passes through the point (6, 8), its equation must be,

$$y - 8 = \frac{3}{5}(x - 6)$$

$$\text{or} \quad 5y - 40 = 3x - 18$$

$$\text{or} \quad 3x - 5y + 22 = 0 \qquad \text{Ans.}$$

Note: If the line $y = mx + b$ passes through origin, O (0, 0), then its equation will be $y = mx$. [A line drawn perpendicular to x-axis at the point (1, 0), will intersect the line $y = mx$ at the point (1, m) (since $y = m$, for $x = 1$). This *idea is found useful in the following derivation.*

4.9 RELATION BETWEEN THE SLOPES OF (NONVERTICAL) LINES THAT ARE PERPENDICULAR TO ONE ANOTHER

There is a simple slope condition between two nonvertical lines that are perpendicular *to one another.*

Method (1):
Consider two *nonvertical* lines l_1 and l_2 that are *perpendicular to one another and have the slope(s)* m_1*and* m_2*, respectively,* as shown in Figure 4.15. Without loss of generality, we may assume that these lines intersect at the origin, or we may translate them, so that they intersect at the origin, *without changing their slopes.*

We draw a line, *perpendicular to x-axis passing through the point* (1, 0). Then $P(1, m_1)$ and $Q(1, m_2)$ are points on the lines as shown in Figure 4.15.[19]

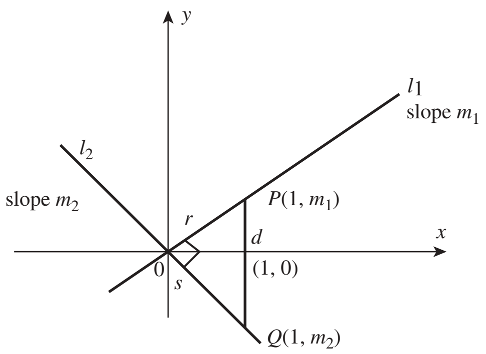


FIGURE 4.15 Pair of non-vertical lines (through origin) and perpendicular to one another.

[19] [Hint: We know that a line with slope m passing through the origin $O(0, 0)$ has the equation $y = mx$. Thus, the equation of l_1 is $y = m_1x$ and that of l_2 is $y = m_2x$. Therefore, the coordinates of P (on l_1) are (l, m_1) and those of Q (on l_2) are $(1, m_2)$.]

The lines are perpendicular, if and only if, the triangle with vertices (0, 0), $(1, m_1)$ and $(1, m_2)$ satisfies the Pythagorean relation $d^2 = r^2 + s^2$. Let us compute r^2, s^2, and d^2 using the *distance formula*. We obtain

$$r^2 = (1-0)^2 + (m_1 - 0)^2 = 1 + m_1^2$$

$$s^2 = (1-0)^2 + (m_2 - 0)^2 = 1 + m_2^2$$

$$d^2 = (1-1)^2 + (m_2 - m_1)^2 = (m_2 - m_1)^2$$

The Pythagorean condition becomes

$$(m_2 - m_1)^2 = (1 + m_1^2) + (1 + m_2^2)$$

$$m_2^2 - 2m_1m_2 + m_1^2 = 2 + m_1^2 + m_2^2$$

$$\text{Therefore,} \quad -2m_1m_2 = 2$$

$$m_1m_2 = -1$$

$$\text{or } m_2 = -\frac{1}{m_1}$$

Thus, if the slope of the line l is known, then we can write the slope of another line that is perpendicular to l. The above relation may also be obtained as follows.

Method (2):
Let l_1 and l_2 be two (nonvertical) lines perpendicular to one another, with slopes m_1 and m_2, respectively. Let θ_1 and θ_2 be their inclinations as shown in Figure 4.16.

$$\therefore \tan\theta_2 = \tan\left(\frac{\pi}{2} + \theta_1\right)$$

$$= -\cot\theta_1 = -\frac{1}{\tan\theta_1}$$

$$\therefore \tan\theta_1 \cdot \tan\theta_2 = -1$$

$$\therefore m_1 \cdot m_2 = -1 \quad [\text{From}(1)].$$

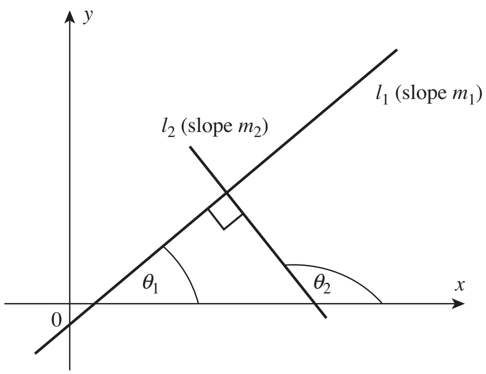


FIGURE 4.16 Two non-vertical lines with slopes m_1, m_2 and perpendicular to one another.

Example (14): Let us find the slope of a line perpendicular to the line passing through the points (6, −5) and (8, 3).

Solution: The slope of the given line is

$$\frac{\Delta y}{\Delta x} = \frac{-5 - (3)}{6 - (8)} = \frac{-8}{-2} = 4$$

Therefore, a line that is perpendicular to the given line has slope (−1/4).

Points to Remember

1. If a line with slope m passes through (0, 0), then $y = mx$ is the equation of the line.
2. Consider the line l, which passes through (0, 0), and is *equally inclined to both the axes*. Two cases arise. If the angle of inclination is $\pi/4$, then the slope $m = 1$, and if it is $3\pi/4$, then the slope $m = -1$. Hence $y = x$ or $y = -x$, will be the equations of the line, respectively.
3. The equation of a line may be written in any of the forms discussed above. The *choice is a matter of convenience and requirement*.

Note: *It is important to remember that*, for the line $y = mx + b$ (which has the slope m) the change of 1 unit in the value of x (i.e., from x_1 to $x_1 + 1$) causes a change of m units in the value of y (i.e., from y_1 to $y_1 + m$). In other words *if (x_1, y_1) is a point on the line, then* ($x_1 + 1$, $y_1 + m$), ($x_1 + 2$, $y_1 + 2m$), ($x_1 + 3$, $y_1 + 3m$), and so on, are other points on the line.

4.10 ANGLE BETWEEN TWO LINES

Suppose l_1 and l_2 are two intersecting lines. Then, we define *the angle from l_1 to l_2* to be the angle θ through which l_1 must be rotated counter clockwise *about the point of intersection* in order to coincide with l_2 (see θ in Figure 4.17). Thus, $0 \leq \theta < \pi$. By using trigonometric identities, we can express θ in terms of the slopes of l_1 and l_2.

Theorem: Let l_1 and l_2 be *two nonvertical lines* that are *not perpendicular*, with slopes m_1 and m_2, respectively. Then, the tangent of the angle θ from l_1 to l_2 is given by

$$\tan\theta = \frac{m_2 - m_1}{1 + m_1 m_2}$$

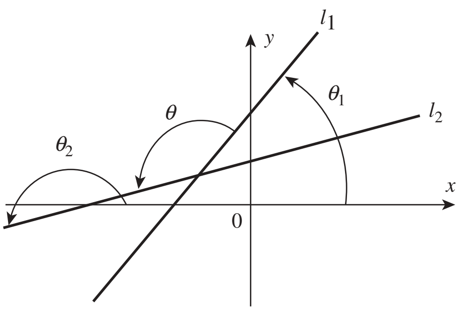


FIGURE 4.17

Proof: Let θ_1 and θ_2 be angles with initial sides along the positive x-axis and terminal sides along l_1 and l_2, respectively. For easy understanding, we choose θ_1 and θ_2 such that $0 \le \theta < \pi$ and $\theta_2 \ge \theta_1$. Then, $\theta = \theta_2 - \theta_1$. We have

$$m_1 = \tan\theta_1$$

and

$$m_2 = \tan\theta_2{}^{(20)}$$

Also, since $m_1 m_2 \neq -1$ (why?)

We have

$$\tan\theta = \tan(\theta_2 - \theta_1) = \frac{\tan\theta_2 - \tan\theta_1}{1 + \tan\theta_1 \tan\theta_2} = \frac{m_2 - m_1}{1 + m_1 m_2}. \text{ (Proved)}$$

Note: In numerical examples, the value of $\tan\theta$ will sometimes be found to be negative. This would merely mean that instead of acute angle of intersection, its supplement, which too is the angle of intersection of the lines, is being obtained.

Example (15): Let the equations of l_1 and l_2 be $y - 2x = 2$ and $2y + 5x = 17$.

Find the *tangent of the angle* θ from l_1 to l_2.

Solution: From the equation of l_1 and l_2, we find that $m_1 = 2$ and $m_2 = (-5/2)$

$$\tan\theta = \frac{m_2 - m_1}{1 + m_1 m_2} = \frac{(-5/2) - (2)}{1 + 2(-5/2)} = \left(\frac{-(9/2)}{-4}\right) = \frac{9}{8}$$

4.11 POLAR COORDINATE SYSTEM

So far, we have located a point in a plane by its rectangular Cartesian co ordinates. The position of a point in a plane may also be determined by means of a so-called polar coordinate system. This system is important because certain curves have simpler equations in the polar coordinate system.

Cartesian coordinates are *numbers*, the abscissa and ordinates, and these numbers are *directed distances from two fixed lines*. Polar coordinates consist of a *directed distance* and the measure of an angle related to a fixed point and a *fixed ray* (or half line).

The fixed point is called the *pole* (or origin), designated by the letter "O". The fixed ray is called the *polar axis* (or polar line), which we label OA. The ray OA is usually drawn *horizontally* and to the right and it extends indefinitely, (see Figure 4.18). Positive x-axis is generally taken as the *polar axis* and the origin (0, 0) as the *pole*.

Let P be any point in the plane *distinct* from "O". Let θ be the *radian measure* of a *directed angle* AOP, positive when measured counter clockwise, and negative when measured clockwise. Let the *initial side* of the angle θ be the ray OA and its *terminal side* the ray OP. Then the point P can be assigned the polar coordinates (r, θ), if r is taken as the *undirected distance* from O to P (i.e., $r = |\overline{OP}|$).

Actually, the coordinates $(r, \theta + 2k\pi)$, where k is any integer, give the same point as (r, θ). *Thus, a given point has an unlimited number of sets of polar coordinates unlike the rectangular*

(20) If $\theta_2 \ge \pi$ (as in Figure 4.17), then $\tan\theta_2 = \tan(\theta_2 - \pi) = m_2$

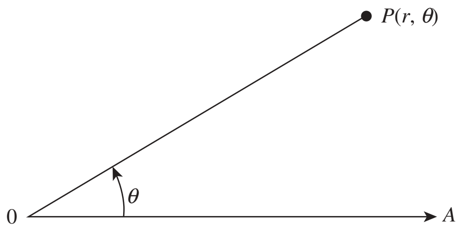


FIGURE 4.18

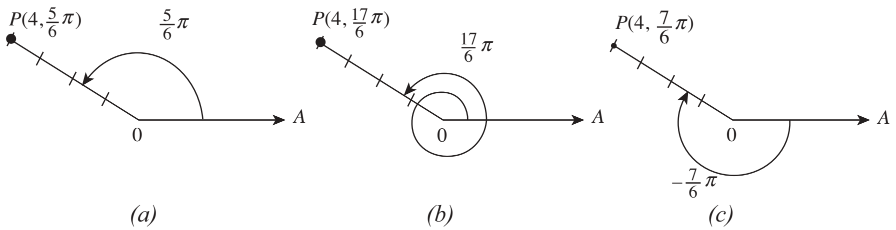


FIGURE 4.19

Cartesian coordinate system in which a one-to-one correspondence between the coordinates and the positions of points in the plane exists (see Figure 4.19a–c).

Polar coordinates of a point P are also defined by considering "r" as the *directed distance* from O to P. Thus, there can be a set of polar coordinates of P, denoted by (r, θ), where $r = -|\overline{OP}|$. Now, we consider polar coordinates for *which r is negative*. In this case, instead of being on the terminal side of the angle, the point is on the extension of the terminal side, which is the ray from the pole in the direction opposite to the terminal side (see Figure 4.20a and b).

Thus, the point $(-4, -(1/6)\pi)$ shown in Figure 4.20a is the same as $(4, (5/6)\pi)$, $(4, (17/6)\pi)$, and $(4, -(7/6)\pi)$ as shown in Figure 4.19a–c, and $(-4, (11/6)\pi)$, as shown in Figure 4.20b.

The angle is usually measured in *radians*. Thus, a set of polar coordinates of a point is an *ordered pair* of real numbers. For each ordered pair of real numbers, there is a unique point having this set of polar coordinates. However, we have seen that *a particular point can be given by an unlimited number of ordered pairs of real numbers*.

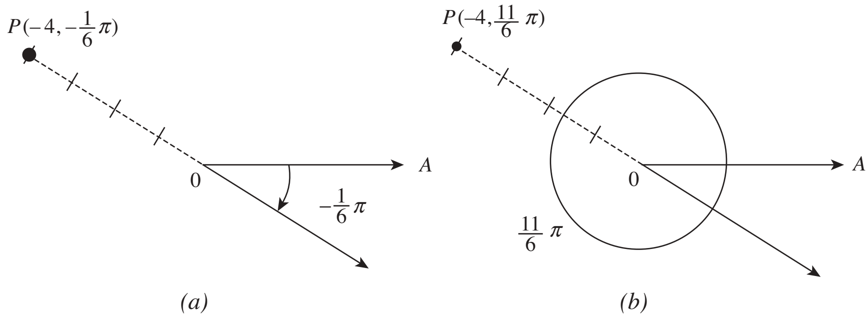


FIGURE 4.20

4.11.1 Relation Between the Rectangular Cartesian Coordinates and the Polar Coordinates of Point

To find the desired relation, we take the origin of the Cartesian coordinate system and the pole of the polar coordinate system coincident, the polar axis as the positive side of the x-axis and the ray for which $\theta = (1/2)\pi$ as the positive side of the y-axis.

Suppose P is a point, whose representation in the rectangular Cartesian coordinate system is (x, y) and (r, θ) is a polar coordinate representation of P. As a particular case, suppose P is in the second quadrant and $r > 0$, as indicated in Figure 4.21.

Then

$$\cos\theta = \frac{x}{\overline{|OP|}} = \frac{x}{r} \quad \text{and} \quad \sin\theta = \frac{y}{\overline{|OP|}} = \frac{y}{r}$$

Thus,

$$x = r\cos\theta \quad \text{and} \quad y = r\sin\theta \tag{13}$$

These equations hold *for P in any quadrant and r positive or negative.*

From equation (13), we can not only obtain the rectangular Cartesian coordinates of a point when its polar coordinates are known, but *we can also obtain a polar equation of a curve from its rectangular Cartesian equation.*

From equation (13) we get

$$x^2 + y^2 = r^2(\cos^2\theta + \sin^2\theta) = r^2$$

$$\therefore\ r = \pm\sqrt{x^2 + y^2} \tag{14}$$

Also, from equation (13), we get

$$\frac{r\sin\theta}{r\cos\theta} = \frac{y}{x}$$

$$\therefore\ \tan\theta = \frac{y}{x} \quad (\text{if } x \neq 0) \tag{15}$$

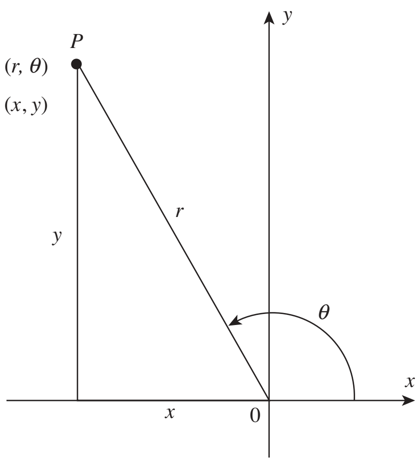


FIGURE 4.21

Note: *If a curve in a plane is expressed in polar coordinates, say* $r = f(\theta)$ *then* r *and* θ *both vary from point to point on the curve.*

Note: For the purpose of learning basic calculus, the material given in this chapter will prove to be sufficient. However, the subject of coordinate geometry is a very useful subject and can be easily learnt from standard books. This study will be found useful in realizing and appreciating the simpler methods, later offered by calculus, in studying many properties of curves, represented by functions. Some details about conic sections and their identification by Translation of Axes are given in Appendix B.

5 Trigonometry and Trigonometric Functions

5.1 INTRODUCTION

The word *trigonometry* is derived from two Greek words, together meaning *measuring the sides of a triangle*. The subject was originally developed to solve geometric problems involving triangles. One of its uses lies in *determining heights and distances*, which are not easy to measure otherwise. It has been very useful in *surveying*, *navigation*, and *astronomy*. Applications have now further widened.

At school level, in geometry, we have studied the definitions of *trigonometric ratios of acute angles* in terms of the ratios of sides of a right-angled triangle.

$$\left.\begin{array}{lll} \sin\theta = \dfrac{P}{H}, & \cos\theta = \dfrac{B}{H}, & \tan\theta = \dfrac{P}{B} \\ \operatorname{cosec}\theta = \dfrac{H}{P}, & \sec\theta = \dfrac{H}{B}, & \cot\theta = \dfrac{B}{P} \end{array}\right\}$$

Note that in the right-angled triangle *OAR*, if the lengths of the sides are respectively denoted by B (for base), P (for perpendicular), and H (for hypotenuse), as shown in Figure 5.1, then the angle θ (in degrees) is *an acute angle* (i.e., $0^\circ < \theta < 90^\circ$). It is for such angle(s) that we have defined trigonometric ratios in earlier classes.[(1)]

Now, in our study of trigonometry, it is required *to extend the notion of an angle in such a way that its measure can be of any magnitude and sign*. Once this is done, the trigonometric ratios are defined for angles of all magnitudes and sign. Finally, by identifying these magnitudes and signs of angles, with real numbers, we say that the *trigonometric ratios of directed angles represent trigonometric functions of real variables*. This is achieved by defining trigonometric ratios of any angle expressed in radians. To enjoy the subject of trigonometry, it is useful to start our study right from the concept of directed angles and the radian measure of an angle.

What must you know to learn calculus? 5-Trigonometry and trigonometric functions [Concept of angle, directed angle(s) of any magnitude and sign, extending the concept of trigonometric ratios (of acute angles) to trigonometric functions of real variable]

(1) For beginners, the trigonometric ratios are not considered for the angles of 0° and 90°, since the triangle does not exist for these values of θ. Also, since the above trigonometric ratios are found sufficient in solving the problems related to heights and distances and for studying trigonometric identities, the notion of *directed angles* is not introduced for beginners to avoid difficulties likely to be faced by them.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

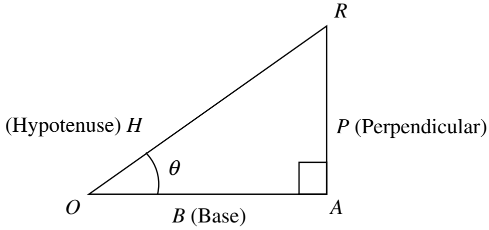


FIGURE 5.1 Right angled triangle defining trigonometric ratios.

5.2 (DIRECTED) ANGLES

Definition: In geometry, an angle is considered as the measure obtained by rotating a given ray about its end point.

- The original ray is called the *initial side* and the finial position of the ray (after rotation) is called the *terminal side* of the angle (Figure 5.2).
- The point of rotation is called the *vertex*.
- If the direction of rotation is *anticlockwise*, the angle is said to be *positive*; and if the direction of rotation is *clockwise*, the angle is *negative*.

5.2.1 An Angle in Standard Position

A directed angle is said to be in standard position if its *vertex* lies at the *origin* and the *initial side* lies on the *positive side* of the x-axis. Figure 5.3 shows an angle AOB in *standard position* with OA as the initial side.

We know that the angle AOB can be formed by rotating the side OA to the side OB and, *under such a rotation*, *the point A moves along the circumference of a circle* having its center at O and radius $|OA|$ to the point B.

5.2.2 Measure of an Angle

- The measure of an angle is the *amount of rotation* performed to get to the terminal side from the initial side.

Note (1): The definition of an angle suggests a unit, namely, *one complete revolution*, from the position of the initial side, as shown in Figure 5.4.

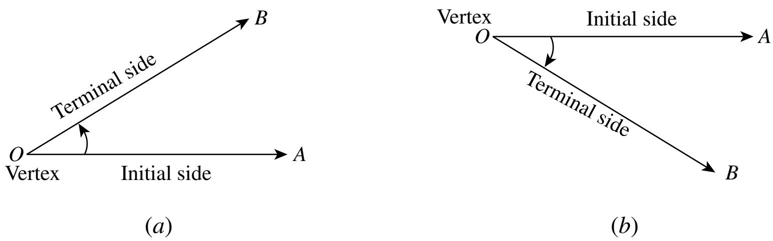


FIGURE 5.2 (a) Positive angle. (b) Negative angle.

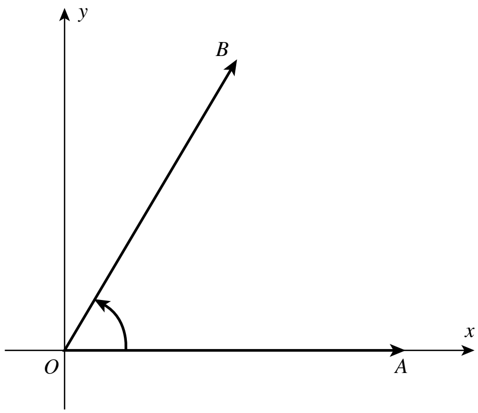


FIGURE 5.3 Angle in standard position.

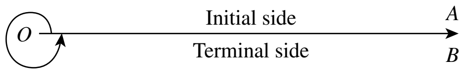


FIGURE 5.4

This unit that is based on a complete revolution is often convenient for *large angles*. For example, an engineer might speak of a spinning wheel making say 15 revolutions per minute.(2)

There are several units of measuring angles. We describe below two units of measuring of an angle that are most commonly used. One is the *degree* and the other is the *radian* measure of an angle.

5.2.3 Degree Measure of an Angle

If a rotation from the initial side to the terminal side is $(1/360)$th of a revolution, the angle is said to have a measure of 1°. For additional precision, we define two subunits of a degree by the following relations:

$$60 \text{ minutes (written as } 60') = 1^\circ \text{ (one degree)}$$
$$60 \text{ seconds (written as } 60'') = 1' \text{ (one minute)}$$

In dealing with problems involving angles of triangles, the measurement of an angle is usually given in degrees.(3)

(2) The idea of spinning wheel making large angle(s) suggests that one may generate angles of any magnitude and sign.
(3) There is no deep reason for choosing the number 360. Early astronomers, with their imperfect instruments, thought that the earth took 360 days to circle the sun, and hence divided a circle into 360 equal parts. One may wonder what they would have done had they known that this number was nearly $365\frac{1}{4}$, which itself is not accurate. So the choice is between 360 and $365\frac{1}{4}$. When an angle is measured in degrees, minutes, and seconds, the system of measurement is called the *sexagesimal system of measurement*, because it is based on *multiples of 60*.

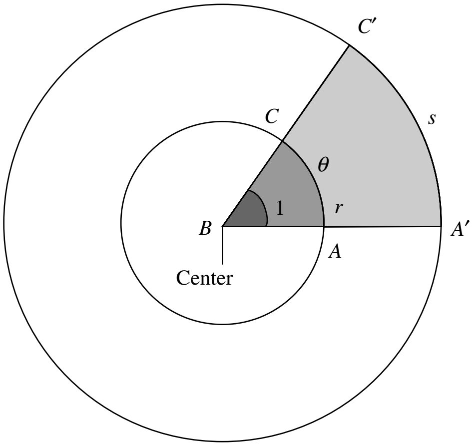


FIGURE 5.5 Radian measure of an angle.

However, in calculus, we are concerned with trigonometric functions of real numbers, and these functions are defined in terms of the radian measure of an angle.

5.2.4 Definition of Radian Measure of an Angle

Consider two concentric circles with B as their (common) center. Let the radius of the inner circle be one unit and that of the outer circle be r units, as shown in Figure 5.5.

The radian measure of angle ABC at the center B of the *unit circle* is defined to be the *length of the circular arc* AC.

If $A'C'$ $(= s)$ is the arc cut by the (same) angle from a second circle (which is the outer circle, in Figure 5.5), then the circular sectors $A'BC'$ and ABC are similar. In particular, their ratios of arc length to radius are equal. We denote this equality by the constant θ.

In the notation of Figure 5.5, this means that

$$\frac{\text{length of arc } A'C'}{r} = \frac{\text{length of arc } AC}{1} = \theta$$

or

$$\frac{s}{r} = \text{length of arc } AC = \theta \tag{1a}$$

This is true no matter how large or small the radius of the second circle may be. Thus, for any circle centered at B, the ratio s/r (of the length of the intercepted arc to the radius of the circle) *always gives the radian measure of the angle.*

Equation (1a) is sometimes written in the form

$$s = r\theta \tag{1b}$$

Equation (1b) can be used to find out any one of the related quantities (i.e., s, r, or θ) if the other two are known. Generally, this equation is used to compute the arc length s, when r and θ are known.

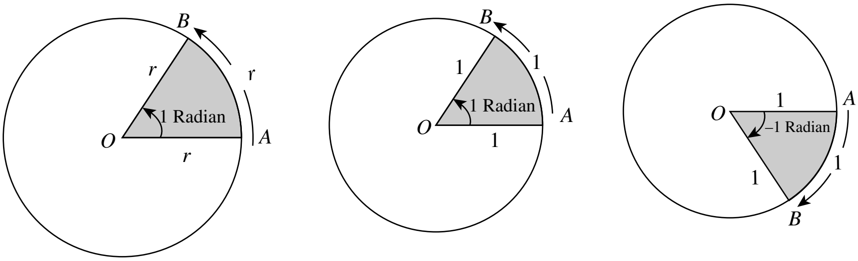


FIGURE 5.6 Angle of one radian.

Note (1): Equation (1b) is also useful in defining an angle of 1 rad, which is the unit angle in radian measure. If we put $r = 1$ in equation (1b), then the central angle θ, in radians, is just equal to the length of the circular arc AC, as defined above.

Definition: 1 rad is the measure of a central angle, subtended by a *circular arc whose length is equal to the radius of the circle* (Figure 5.6).

Note (2): Although angles can be expressed (or measured) in degrees or radians, we will here use only radian measure of angles, unless otherwise indicated. This will be convenient in our study of calculus.

The circumference of a circle is approximately 6.28 times its radius. In other words, the angle subtended by the circumference of a circle at the center is approximately 6.28 rad.[4]

Note (3): The length of circumference of a circle is given by $2\pi r$, so for a *unit circle*, the circumference equals the length 2π. Thus, π *can be interpreted in two ways*:

- When speaking in terms of the length of a circular arc of unit circle, π represents the length of half the circumference of unit circle, and so it stands for a real number $\approx 3.14159\ldots$ (note that this number is half of the number 6.28…, mentioned above).
- When speaking of angles, the unit circle subtends an angle of 2π rad at its center (or 360°). It follows that half the circle subtends the angle of π rad at the center. Thus, π stands for 180°.

Remark: It is important to clearly understand that π never represents the number 180. When expressing an angle in terms of π, the statement $180^\circ = \pi$ should be read as $180^\circ = \pi$ rad. Thus, the reader should mentally imply the word *radians* to avoid confusion between π and 180°.

(Note that, in the expression π rad, π is the coefficient of the unit *radian* and, therefore, it must be looked upon as a real number.)

5.2.4.1 Angle of Any Magnitude and Sign Suppose a ray starting from the initial position is rotated about the vertex, more than one rotation in positive (or negative direction). Then, we can generate angles of desired magnitude and sign, as indicated in Figure 5.7.

[4] Radian measure of an angle assumes that we know how to measure the length(s) of circular arc(s). Later on, when we discuss *Integration*, we will show how this can be done. For the present, we agree that the circumference of a circle is $2\pi r$, where r is the radius of the circle. In other words, we agree that π is the ratio of circumference of a circle to its diameter. It is true that π is related to the circle, but it also appears in many (definite) integrals and in sum (s) of certain infinite series. From this point of view, one should not carry an impression that π is related only to the circle. It arises in mathematics in the same way as the number e arises in calculus. (Both π and e are special types of irrational numbers, called transcendental numbers.)

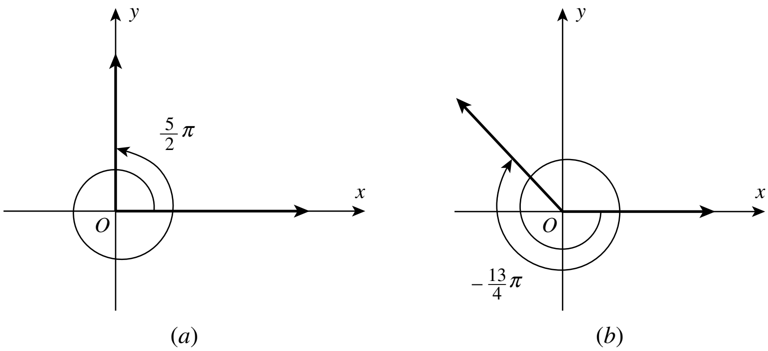


FIGURE 5.7 Angle of desired magnitude.

5.2.4.2 Zero Angle and Straight Angle

Zero Angle: Suppose the given ray OA is not rotated about the vertex O, then we say *that the measure of the angle at the point O is zero* (Figure 5.8).

Straight Angle: Suppose the given ray OA rotates half the circle (so that it occupies the final position OB opposite to the direction of OA), then the measure of the angle AOB will be 180° (or −180°) depending on the direction of rotation of the line OA about the vertex. It is called *straight angle* (Figure 5.9).

5.2.4.3 The Concept of Positive and Negative Arc Lengths Consider a circle centered at the origin, with *arbitrary radius* $r > 0$. We place an angle of θ rad (in the circle) *in standard position* (so that its vertex is at the origin and its initial side is on the positive side of the x-axis. If $\theta \geq 0$, then it opens *counterclockwise* (Figure 5.10a), and if $\theta < 0$, it opens *clockwise* (Figure 5.10b).

We allow θ to be greater than 2π (i.e., 360°). For example, an angle of 3π rad can be obtained by rotating a line through one full revolution (2π rad) and an extra half-revolution. Thus, an angle of 3π rad has the same initial and terminal side as an angle of π rad.

Since, the circle contains 2π rad and its circumference is $2\pi r$, an angle of 1 rad intercepts an arc of length r on the circle. If $\theta > 0$, an angle of θ rad intercepts an arc of length $r\theta$ on the circle. If we denote this arc length by s, we have

$$s = r\theta$$

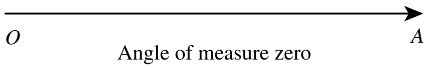


FIGURE 5.8 Angle of measure zero.

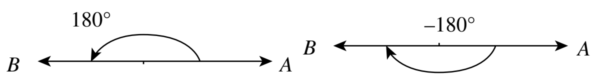


FIGURE 5.9 Straight angle.

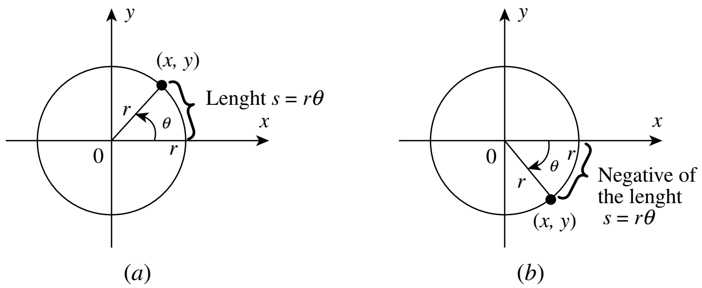


FIGURE 5.10 Positive and negative arc-lengths.

For $\theta < 0$, formula (1) holds *if we think of s as the negative of the length of the arc intercepted on the circle* (Figure 5.10b).

(Note that for the purpose of defining directed angles, we have agreed to accept the idea of positive and negative arc lengths, which is otherwise meaningless.)

5.2.5 Relation Between Degree and Radian Measures of an Angle

A circle subtends at the center an angle whose degree measure is 360° and radian measure is 2π rad (Figure 5.11). It follows that

$$360^\circ = 2\pi \text{ rad} \quad \text{(A)}$$

$$180^\circ = \pi \text{ rad} \approx 3.1415 \text{ rad}$$

$$\frac{180^\circ}{\pi} = 1 \text{ rad } \approx 57^\circ 17' 44.8''$$

and

$$1^\circ = \frac{2\pi}{360} = \frac{\pi}{180} \approx 0.01745 \text{ rad}$$

(Here the values of 1 rad and 1° are computed assuming $\pi \approx 22/7$.)

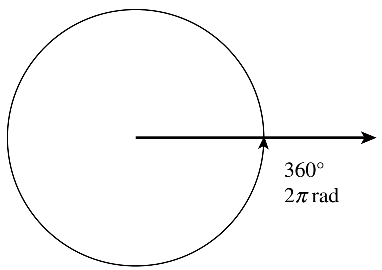


FIGURE 5.11 Angle of 360° or 2π radians.

TABLE 5.1 Measure of Some Useful Angles in Degrees with their Corresponding Measure in Radians

Degrees	0°	30°	45°	60°	90°	120°	135°	150°	180°	270°	360°
Radians	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{3\pi}{4}$	$\frac{5\pi}{6}$	π	$\frac{3\pi}{2}$	2π

Remark: In the equation (A), it will be more appropriate to use the symbol $\sim$ instead of using the symbol of equality (i.e., $=$), which tells that the given measurements are for the same or congruent angles.

Thus, we have[5]

$$\left.\begin{aligned} x \text{ rad} &= \left(\frac{180}{\pi}\right).(x)^\circ \\ &\text{and} \\ x^\circ &= \left(\frac{\pi}{180}\right).(x) \text{ rad} \end{aligned}\right\}$$

Table 5.1 on conversion is very often useful in trigonometry.

Now, note that

$$162^\circ = 162 \cdot \frac{\pi}{180} \text{ rad} = \frac{9}{10}\pi \text{ rad}$$

and

$$\frac{5}{12}\pi \text{ rad} = \frac{5}{12}\pi \cdot \frac{180^\circ}{\pi} = 75^\circ$$

It should be emphasized, however, that *the radian measure* of an angle is *dimensionless*. Note that r and s [in equation (1a)] represent lengths measured in identical units, so that the units get canceled.

5.2.6 Relation Between the Radian Measure and Real Numbers

Consider the *unit circle* with center at the origin O. Let A be any point on the circle so that OA is the radius of the circle and we consider it as the *initial side of an angle* (Figure 5.12a). We may imagine the circumference of the circle marked of with a scale from which we may read θ. The unit on this number scale is the same as the unit radius (Figure 5.12b).

Now, let a line PQ be tangent to the unit circle and let the point O be marked on it as 0 of the number scale based on the unit radius. We place the point O of the line at the point A of the circle so that the line PAQ is tangent to the circle at A.

We know that the length of an arc of the circle will give the radian measure of the angle, which the arc will subtend at the center of the circle. Thus, the point A represents the real number 0 on the tangent line, AP represents the positive side of tangent line, and AQ represents the negative side.

(5) For converting radians into degrees and vice versa, it is useful to remember that 1 rad is a bigger angle, nearly 57 times bigger compared to 1°. Hence, for converting radians into degrees, we must multiply the radian measure by a bigger factor $180/\pi$. On the other hand, the degree measure must be multiplied by a smaller factor $\pi/180$ to convert it to radians.

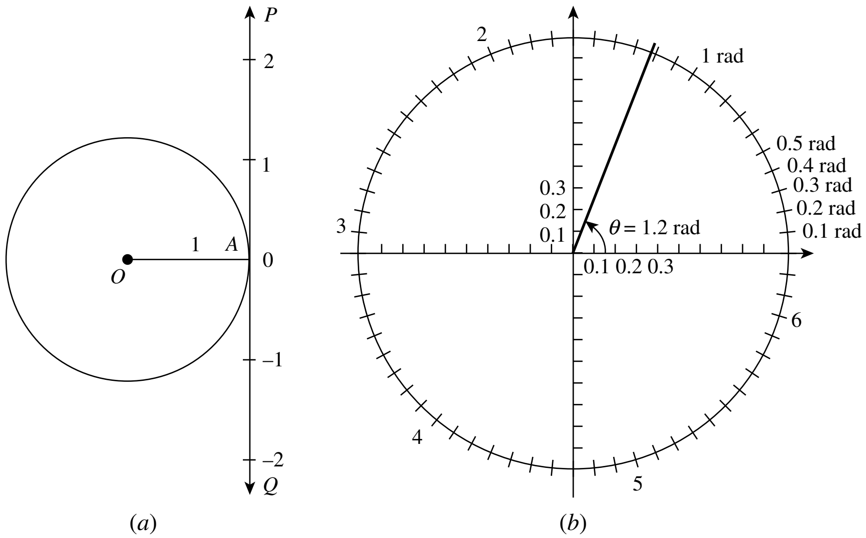


FIGURE 5.12 An angle in radians looked upon as a real number.

If we wrap the line AP around the circle in the counterclockwise direction and AQ in the clockwise direction, then every real number on the tangent line will correspond to a radian measure and conversely (every radian measure will correspond to a real number). Thus, radian measures of angles with reference to the unit circle can be considered to represent the real numbers having the same magnitude and sign.

5.2.6.1 Convention About the Notation If the angles are measured in degrees or radians, we adopt the convention that whenever we write angle θ°, we mean the angle whose degree measure is θ, and whenever we write angle θ (i.e., without superscript $^\circ$), we mean the angle whose *radian measure* is θ. Thus, in the expression $\sin 30^\circ$, measure of the angle is 30°; whereas in $\cos 75$, the number 75 represents the radian measure of the angle involved and also the number 75.

Now, we are in a position to extend the definitions of the trigonometric ratios for angles of any magnitude and sign. Such angles can be generated by rotating the initial side (about its vertex) in the desired direction to any desired extent.

5.2.7 Trigonometric Ratios for Angles of Any Magnitude and Sign: Definitions of Trigonometric Functions

Let an angle of θ rad be placed in *standard position* in a circle of radius r. The terminal side of the angle *intersects the circle at a unique point* (x, y) (see Figure 5.13a). We define the *sine and cosine* functions of θ by

$$\sin\theta = \frac{y}{r} \quad \text{and} \quad \cos\theta = \frac{x}{r} \tag{2}$$

In Figure 5.13a, x, y, and r represent the sides of a right-angled triangle and θ is the angle that the revolving line OP makes with the x-axis. In fact, the definitions of trigonometric functions at equation (2) are the same as the definitions of trigonometric ratios given for acute angle(s).

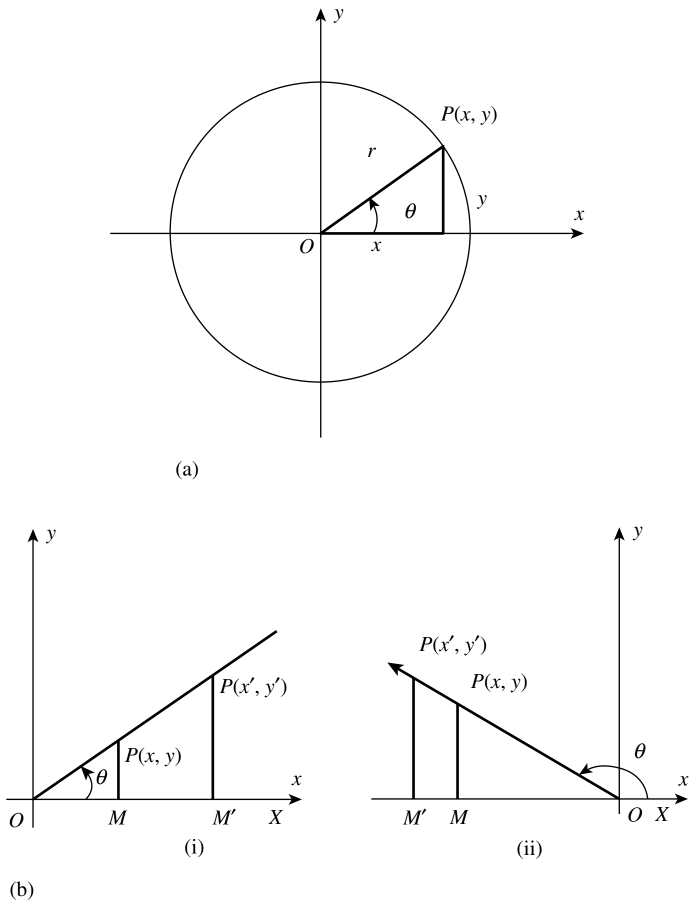


FIGURE 5.13 Angle θ in standard position.

Here, it is important to keep in mind that the angle θ can be of any magnitude and sign. Therefore, the terminal side *OP* can be in any quadrant. Thus, the angle θ that the revolving line makes with the x-axis need not be acute. However, we define the trigonometric function of the angle θ with reference to the right-angled triangle in which the revolving line (as hypotenuse) makes the angle θ with the x-axis. Obviously, θ may be acute or obtuse or negative.

The properties of similar triangles imply that $\sin\theta$ and $\cos\theta$ depend only on θ and not on the value of r (Figure 5.13b).

From equation (2) above, we get that if $r = 1$,

$$x = \cos\theta \text{ and } y = \sin\theta \tag{3}$$

Since, the angle θ (in radians) represents a real number, which can assume any real value in $(-\infty, \infty)$, the domains of both $\sin\theta$ and $\cos\theta$ are $(-\infty, \infty)$.

[Note that, in the expression $\sin\theta$, θ represents a number. Thus, we write sin 3 to mean sin (3 rad).]

5.2.7.1 Periodic Functions: sin θ and cos θ Since an angle of θ rad ($0 \le \theta < 2\pi$) and the one of $(\theta + 2\pi)$ radians have the same terminal side, we can write

$$\sin\theta = \sin(\theta + 2\pi) \text{ and } \cos\theta = \cos(\theta + 2\pi)$$

Thus, the values of sine and cosine functions repeat for an interval of 2π rad. We say that both sine and cosine functions are periodic; they both have a period of 2π. Consequently, for any integer n and *any number* θ,

$$\sin\theta = \sin(\theta + 2n\pi) \text{ and } \cos\theta = \cos(\theta + 2n\pi) \tag{4}$$

Coterminal Angles The angles that differ in their measure by an integral multiple of 360° $(= 2\pi^c)$ are called coterminal angles. They have the same initial arm and the same terminal arm.

5.2.8 Defining Other Trigonometric Functions using Sine and Cosine Functions

There are four other basic trigonometric functions that are defined in terms of $\sin\theta$ and $\cos\theta$. Remembering that $\sin\theta = y/r$, $\cos\theta = x/r$, we define

$$\left.\begin{array}{ll} \tan\theta = \dfrac{\sin\theta}{\cos\theta} = \dfrac{y}{x}, \quad x \neq 0 & \cot = \dfrac{\cos\theta}{\sin\theta} = \dfrac{x}{y}, \quad y \neq 0 \\ \sec\theta = \dfrac{1}{\cos\theta} = \dfrac{r}{x}, \quad x \neq 0 & \operatorname{cosec}\theta = \dfrac{1}{\sin\theta} = \dfrac{r}{y}, \quad y \neq 0 \end{array}\right\}$$

The values of these functions can be quickly computed from the corresponding values of $\sin\theta$ and $\cos\theta$.

5.2.9 A Simple Approach for Calculating the Values of sin θ and cos θ

Let (x, y) be a point on the standard unit circle. Then, using equation (3), we can express the coordinates (x, y) on the unit circle by $(x, y) = (\cos\theta, \sin\theta)$.

Thus, $x = \cos\theta$ and $y = \sin\theta$ (see Figure 5.14).

These values of $\cos\theta$ and $\sin\theta$ are called their *line values* and can be conveniently used for drawing their graphs. Also, graphical methods are available to find the line values of other trigonometric functions.

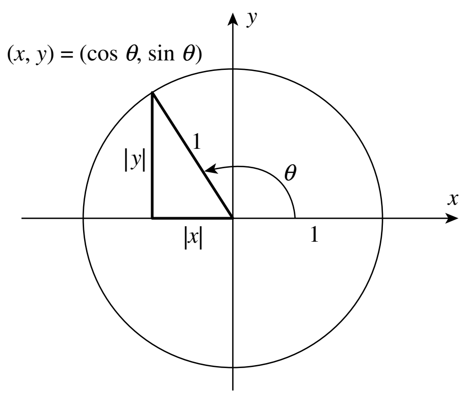


FIGURE 5.14 The acute reference triangle for an angle θ.

Observe that (in Figure 5.14), the revolving line that makes the angle θ with the x-axis lies in the *second quadrant*. If we drop a perpendicular from the point (x, y) on the x-axis, we get

$$|x| = \cos\theta$$

that is,

$$-x = \cos\theta \quad [\because\ x < 0]$$

and

$$|y| = \sin\theta$$

that is,

$$y = \sin\theta \quad [\because\ y > 0]$$

Thus, we get that the sign of cos θ is always the sign of x-coordinate and the sign of sin θ is the sign of y-coordinate. Now, recall that in the second quadrant, x-coordinate is *negative* and y-coordinate is *positive*. Thus, the values of cos θ *and* sin θ expressed above are consistent with their definitions at equation (3).

Important Note: Figure 5.14 suggests that the values of cos θ and sin θ can be calculated from an *acute reference triangle*, made by dropping a perpendicular to the x-axis, as shown in the figure. The ratios are read from *the triangle*, and the signs determined by the quadrant in which the angle lies.(6)

In fact, the method for calculating the values of sin θ and cos θ discussed above is applicable for any location of revolving line in the standard unit circle.

5.2.9.1 Values of sin θ and cos θ for Some Standard Angles. For certain values of θ, the values of sin θ and cos θ are easily obtained by placing the angle in a unit circle in standard position (see Figure 5.15).

We observe that

$$\begin{aligned}
\sin 0 &= 0 &\quad \text{and} \quad \cos 0 &= 1\\
\sin\frac{\pi}{4} &= \frac{1}{2}\sqrt{2} &\quad \text{and} \quad \cos\frac{\pi}{4} &= \frac{1}{2}\sqrt{2}\\
\sin\frac{\pi}{2} &= 1 &\quad \text{and} \quad \cos\frac{\pi}{2} &= 0\\
\sin\pi &= 0 &\quad \text{and} \quad \cos\pi &= -1\\
\sin\frac{3\pi}{2} &= -1 &\quad \text{and} \quad \cos\frac{3\pi}{2} &= 0
\end{aligned}$$

Table 5.2 gives these values and some others that are frequently used.

Note that in Table 5.2, a simple scheme is given for remembering the values of sin θ and cos θ.

(6) It must be clear that in the acute reference triangle, the trigonometric ratios are read with reference to the acute angle made by the revolving line with the x-axis. It is this acute angle that lies in the quadrant in which the revolving line lies.

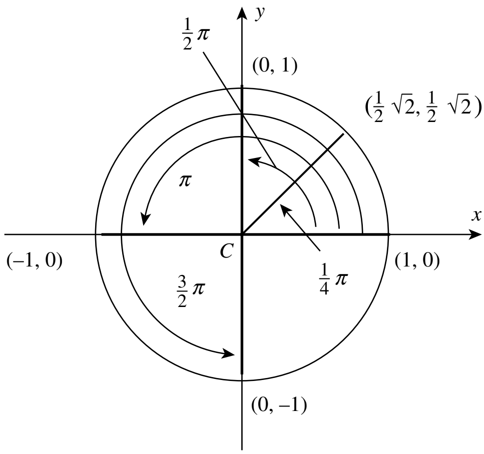


FIGURE 5.15

TABLE 5.2 A Simple Scheme Indicating the Values of Basic Trigonometric Functions, for Important Angles

0	0	$\pi/6$	$\pi/4$	$\pi/3$	$\pi/2$	$2\pi/3$	$3\pi/4$	$5\pi/6$	π
	(0°)	(30°)	(45°)	(60°)	(90°)	(120°)	(135°)	(150°)	(180°)
$\sin\theta$	$\sqrt{(\frac{0}{4})}$	$\sqrt{(\frac{1}{4})}$	$\sqrt{(\frac{2}{4})}$	$\sqrt{(\frac{3}{4})}$	$\sqrt{(\frac{4}{4})}$	$\sqrt{(\frac{3}{4})}$	$\sqrt{(\frac{2}{4})}$	$\sqrt{(\frac{1}{4})}$	$\sqrt{(\frac{0}{4})}$
	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0
$\cos\theta$	$\sqrt{(\frac{4}{4})}$	$\sqrt{(\frac{3}{4})}$	$\sqrt{(\frac{2}{4})}$	$\sqrt{(\frac{1}{4})}$	$\sqrt{(\frac{0}{4})}$	$-\sqrt{(\frac{1}{4})}$	$-\sqrt{(\frac{2}{4})}$	$-\sqrt{(\frac{3}{4})}$	$-\sqrt{(\frac{4}{4})}$
	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{3}}{2}$	-1

5.2.9.2 The Relations sin $(-\theta) = -$sin θ and cos $(-\theta) =$ cos θ Figure 5.16a and b shows two angles of opposite sign but of equal magnitude. The rays of the two angles t and $(-t)$ intersect the circle at the points (x, y) and $(x, -y)$, respectively. Each has equal x and y coordinates in magnitude, but the y-coordinates differ in sign.

From the above figure, we have

$$\sin(-\theta) = \frac{-y}{r} = -\frac{y}{r} = \sin\theta \text{ and } \cos(-\theta) = \frac{x}{r} = \cos\theta \qquad (6)$$

Remark: To define the radian measure of an angle, we use a circle. Hence, trigonometric functions of real variables are also called circular functions.

5.3 RANGES OF SIN *θ* AND COS *θ*

We know that the domains of sin θ and cos θ are $(-\infty, \infty)$ (see Section 2.7). In the reference right-angled triangle (Figure 5.13a), we have, $x^2 + y^2 = r^2$.

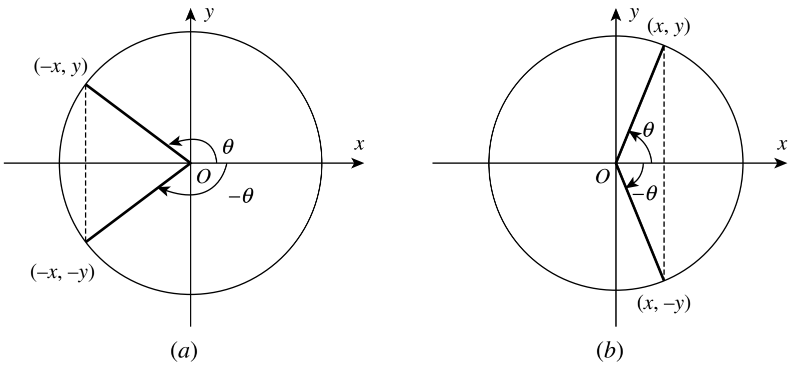


FIGURE 5.16

Since $r \neq 0$, we get

$$\frac{x^2}{r^2} + \frac{y^2}{r^2} = 1 \quad \text{or} \quad \left(\frac{x}{r}\right)^2 + \left(\frac{y}{r}\right)^2 = 1 \quad \text{or} \quad (\cos\theta)^2 + (\sin\theta)^2 = 1$$

We write $(\cos\,\theta)^2 = \cos^2\theta$, $(\sin\,\theta)^2 = \sin^2\,\theta$, and so on.

$$\therefore \cos^2\theta + \sin^2\theta = 1 \tag{7}$$

$$\therefore \cos^2\theta \leq 1 \quad \text{and} \quad \sin^2\theta \leq 1^{(7)}$$

Since $\cos^2\,\theta$ and $\sin^2\,\theta$ are nonnegative, the minimum value of $\cos^2\,\theta$ and $\sin^2\,\theta$ can be 0. Therefore, from equation (A) above, the maximum value of $\cos^2\,\theta$ and $\sin^2\,\theta$ is 1.

$$\therefore\ -1 \leq \cos\theta \leq 1 \text{ and } -1 \leq \sin\theta \leq 1$$

Thus, *range* of both these functions is the closed interval $[-1, 1]$.

5.3.1 Domains and Ranges of tan θ, sec θ, cot θ, and cosec θ

We have

$$\tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{y}{x},\ x \neq 0, \quad \text{and} \quad \sec\theta = \frac{1}{\cos\theta} = \frac{r}{x},\ x \neq 0$$

Thus, tan θ and sec θ are not defined for those values of θ for which $x = 0$.

(In radian measure, this means that $(\pi/2)$, $(3\pi/2)$, $\ldots$, $-(\pi/2)$, $-(3\pi/2), \ldots$ are excluded from the domains of the tangent and the secant functions.)

Similarly, cot θ and cosec θ are *not defined* for those values of θ, for which $y = 0$. Thus, $\theta = 0$, π, 2π, $\ldots$, $-\pi$, -2π, $\ldots$ are excluded from the domains of cot θ and cosec θ.

(7) This conclusion can also be drawn as follows. We know that the square root of a positive number is its principal square root, by which we mean its positive square root. Therefore, by taking the square root on both sides of the inequalities $\cos^2\theta \leq 1$ and $\sin^2\theta \leq 1$, we get $\sqrt{\cos^2\theta} = |\cos\theta| \leq 1$ and $\sqrt{\sin^2\theta} = |\sin\theta| \leq 1$ or $\cos\theta \leq |1|$ and $\sin\theta \leq |1|$. Then, by definition of absolute value we get $-1 \leq \cos\theta \leq 1$ and $-1 \leq \sin\,\theta \leq 1$.

TABLE 5.3 Domains and Ranges of Trigonometric Functions

	Function	Domain	Range
(1)	$\sin\theta = y/r$	All real numbers	$-1 \leq \sin\theta \leq 1$
(2)	$\cos\theta = x/r$	All real numbers	$-1 \leq \cos\theta \leq 1$
(3)	$\tan\theta = y/x$	All real numbers except $\pm\pi/2, \pm3\pi/2, \ldots$	All real numbers
(4)	$\cot\theta = x/y$	All real numbers except $0, \pm\pi, \pm2\pi, \ldots$	All real numbers
(5)	$\sec\theta = r/x$	All real numbers except $\pm\pi/2, \pm3\pi/2, \ldots$	$\sec\theta \leq -1$ and $\sec\theta \geq 1$
(6)	$\operatorname{cosec}\theta = r/y$	All real numbers except $0, \pm\pi, \pm2\pi, \ldots$	$\operatorname{cosec}\theta \leq -1$ and $\operatorname{cosec}\theta \geq 1$

Note: The values of θ for which these *functions are defined*, we have

$$\tan\theta = \frac{\sin\theta}{\cos\theta}, \quad \operatorname{cosec}\theta = \frac{1}{\sin\theta}, \quad \sec\theta = \frac{1}{\cos\theta}, \quad \cot\theta = \frac{1}{\tan\theta} \tag{8}$$

These relations (being the basic definitions of trigonometric functions) are very important. The domains and ranges of trigonometric functions are given in Table 5.3.

5.4 USEFUL CONCEPTS AND DEFINITIONS

(a) *Trigonometric Ratios of Coterminal Angles*: The trigonometric ratios are defined in terms of coordinates (x, y) of a point P and its (constant) distance r from the origin. Accordingly, the trigonometric ratios of coterminal angles are equal. Thus, the trigonometric ratios of 60° and any other angle of measure $60° + (n \times 360°)$ (where n is an integer) are same.

(b) *Trigonometric Ratios of an Angle of Large Measure*: To find the trigonometric ratios of an angle of large measure (which is greater than 360°), we find one coterminal angle whose measure θ is such that $0 \leq \theta < 360°$.
Then, the trigonometric ratios of the angle θ are the same as the trigonometric ratios of the given (large) angle. Furthermore, the concept of allied angles [discussed below in (e)] will be found useful in computing the trigonometric ratios of any angle in terms of the trigonometric ratios of any (small) angle θ where $0 < \theta < 90°$.

(c) *Quadrantal Angles: All angles that are integral multiples of π/2 are called quadrantal angles*: Some such angles are shown in the Figure 5.17.

(d) *Definition (Angle in a Quadrant)*: An angle is said to be in a quadrant in which the terminal side of the angle lies.[8]

(e) *Allied Angles*: Two angles are said to be *allied angles* if the *sum* or *difference* of their measures is either *zero* or an *integral multiple of* 90°. Thus, if θ is the measure of a given angle, then the angles whose measures are $-\theta$, $90° \pm \theta$, $180° \pm \theta$, $270° \pm \theta$, $360° \pm \theta$,

[8] Note that for an obtuse angle θ, the terminal side lies in the second quadrant, whereas the angle θ covers partly the second quadrant (see Figure 5.14). However, in view of this definition, the acute angle $(180° - \theta)$ in the second quadrant is considered for defining the trigonometric ratios of the angle θ. This definition is important because we can define $\sin\theta$ and $\cos\theta$ by considering the reference right-angled triangle in the second quadrant, in which the terminal arm (lying in the second quadrant) is taken as hypotenuse, and the acute angle made with the x-axis is taken as the reference angle. The same understanding is applicable for the location of the terminal arm in any quadrant.

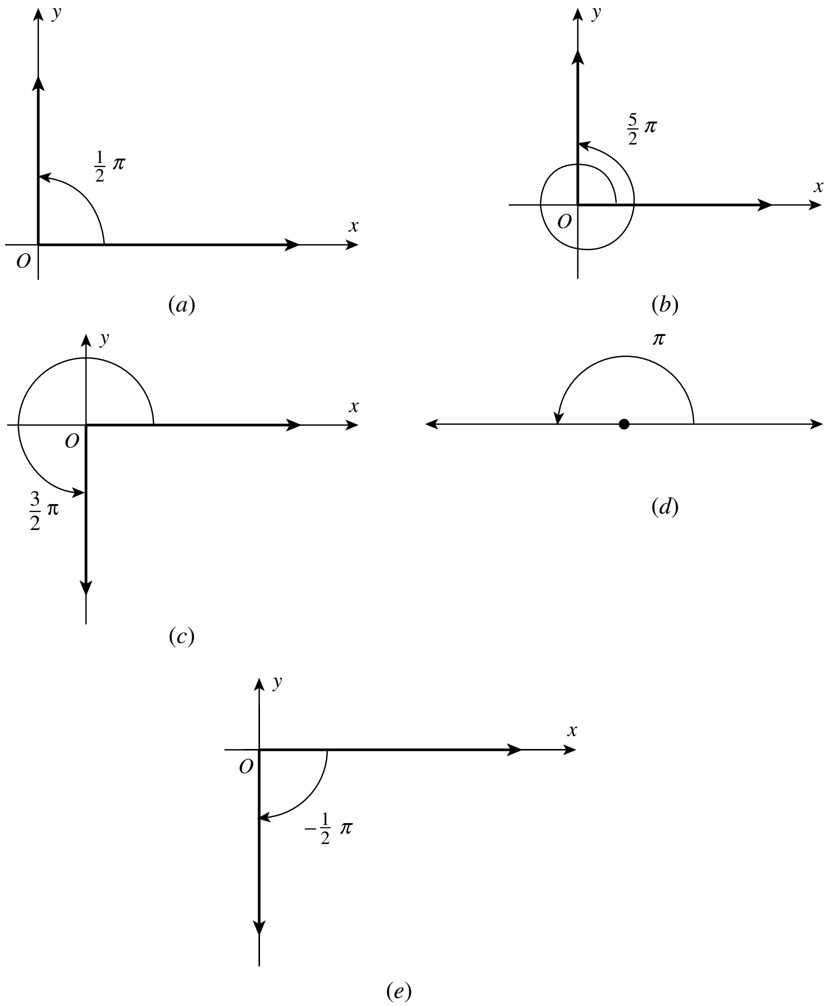


FIGURE 5.17 Coterminal angles.

and so on are its *allied angles*. Our interest lies in finding their trigonometric ratios in terms of those of θ.

If we are given an angle of any measure (large or small), then its trigonometric ratios can be found in terms of trigonometric ratios of a small angle θ (where θ lies between 0° and 90°). For this purpose, we must express the given angle in the form of an allied angle.

If the trigonometric ratios of the (small) angle θ are known, then the procedure that we are going to discuss will help us find the trigonometric ratios of the given angle in terms of those of θ. The procedure (i.e., the rules) under consideration suggests that for writing the trigonometric ratios of allied angles, we shall need the sine and cosine ratios of 90° and 180°.

In fact, we have already obtained the values of trigonometric ratios of 0°, 90°, and 180° using the unit circle in the standard position and have reflected them in Table 5.2 along with the values for some other angles frequently used. Besides, the tables for trigonometric ratios for angles of measures between 0° and 90° have been published. This can be used to find the trigonometric ratios of angles of large measures.

Note: Here, it may be mentioned that the trigonometric ratios of the angles of measure 30°, 45°, and 60° can be easily obtained by drawing the right-angled triangles and using geometry. It is convenient to take a hypotenuse of unit length.

We give three important points for expressing trigonometric ratios of *an allied angle* in terms of an angle θ *whose trigonometric ratios are known.*

(i) The signs of trigonometric ratios are governed by the location of terminal side in different quadrants. This is indicated in the following graph.

sin and cosec positive, all others negative (second quadrant) (–, +)	All ratios positive (first quadrant) (+, +)
(–, –) tan and cot positive, all others negative (third quadrant)	(+, –) cos and sec positive, all others negative (fourth quadrant)

Note: Observe that every trigonometric ratio (and its reciprocal) has a positive sign in two quadrants and a negative sign in the remaining two quadrants. Therefore, if a single trigonometric ratio is given, then it is not possible to determine exactly the quadrant in which the terminal side is located.

For example, both sin 30° and sin 150° have the same value 1/2. Similarly, both cos 45° and cos $-45°$ have the same value $\sqrt{2}/2$ and likewise tan 60° and tan 240° have the same value $\sqrt{3}/2$.

The location of the terminal side (and hence the measure of angle involved) can be uniquely determined iff the values of two independent trigonometric ratios are given. This will become more clear from the solved examples to follow subsequently.

(ii) If the revolving line bounds the angle θ with x-axis, then the trigonometric ratio remains unchanged, when expressed in terms of θ, while the sign of the ratio is governed by (i) above.

Example (1):

$$\sin 210^\circ = \sin\,(180^\circ + 30^\circ) = -\sin 30$$

$$\sec\,(150^\circ) = \sec\,(180^\circ - 30^\circ) = -\sec 30^\circ$$

(iii) If the revolving line bounds the angle θ with y-axis, then the trigonometric ratio changes to the corresponding coratio, when expressed in terms of θ, the sign of the ratio being governed by (i) above. Thus,

$$\sin 120^\circ = \sin(90^\circ + 30^\circ) = \cos 30^\circ$$

$$\tan 240^\circ = \tan(270^\circ - 30^\circ) = \cot 30^\circ$$

$$\cot 300^\circ = \cot(270^\circ + 30^\circ) = -\tan 30^\circ \text{ and so on}$$

5.5 TWO IMPORTANT PROPERTIES OF TRIGONOMETRIC FUNCTIONS

Now we introduce the following notions that will be needed to define *two important properties* of trigonometric functions.

5.5.1 Notion of Even and Odd Functions

5.5.1.1 Even Function A function is said to be even if $f(-x) = f(x)$ for all x.

Example (2):

(a) A polynomial function of the following form is an *even* function:

$$p(x) = a_0 + a_1x^2 + a_2x^4 + \ldots + a_nx^{2n}$$

Observe that the power of x in each term is an *even* integer.

(b) We have already seen that $\cos(-x) = \cos x$ for all x. Thus, the *cosine function is an even function.*

(c) A constant function is always *even* (how?).

5.5.1.2 Odd Function A function f is said to be odd if $f(-x) = -f(x)$ for all x.

Example (3):

(a) It can be easily verified that the functions $f(x) = x$ and $g(x) = x^3$ are odd functions. In fact, any polynomial function in which the power of each term is an odd integer is an odd function.

(b) We have also seen that for all x,

$$\sin(-x) = -\sin x$$

$$\tan(-x) = -\tan x$$

Thus, the sine and the tangent functions are odd functions.

Note: The property of functions whether even or odd is very useful. In particular, it helps in drawing graph of such functions.

5.5.2 The Notion of Periodic Function

Definition: A function $f: R \to R$ is said to be *periodic*, if there exists a real number p ($p \neq 0$) such that $f(x + p) = f(x)$ for all $x \in R$.

Period of a Periodic Function: If a function f is periodic, then the smallest $p > 0$, if it exists such that $f(x + p) = f(x)$ for all x, is called the period of the function. Obviously, the period of the sine and cosine functions is 2π. It can be shown that the period of the tangent function (and that of the cotangent function) is π.

Remark: *A periodic function may not have a period.* Note that a constant function f is periodic as $f(x + p) = f(x) =$ constant for all $p > 0$; however, there is no smallest $p > 0$ for which the relation holds. Hence, there is no period of this function, though it is periodic by definition. The periodicity of trigonometric functions helps us to compute their values for large angles greater than 2π.

5.6 GRAPHS OF TRIGONOMETRIC FUNCTIONS

The graph of a periodic function is completely known once we know it over an interval whose length is equal to the period of the function. We have already seen that values of sin x and cos x repeat after an interval of 2π. Hence, values of cosec x and sec x will also repeat after an interval of 2π. Also, we know that $\tan(\pi + x) = \tan x$. Hence, value of tan x is repeated after an interval of π. Using this knowledge and the behavior of trigonometric functions, we can sketch the graphs of these functions, as given in Figure 5.18a–f.

5.7 TRIGONOMETRIC IDENTITIES AND TRIGONOMETRIC EQUATIONS

Definition: An equation involving trigonometric functions, *which is true for all those angles for which the functions are defined*, is called a *trigonometric identity*. For example, the statements, $\sin^2\theta + \cos^2\theta = 1$ and $\sin 2\theta = 2\sin\theta\cdot\cos\theta$ are trigonometric identities. They are true for all values of θ. Similarly, the statement $\tan\theta = \sin\theta/\cos\theta$ is a trigonometric identity. It holds for all θ, except for those values for which $\cos\theta = 0$.

Note: In a trigonometric identity, two or more numbers (i.e., angles in radians) may be connected by a relation existing among their circular functions, as shown in the statement $\sin 2\theta = 2\sin\theta\cdot\cos\theta$.

Definition: An equation of the form

$$\sin\theta = \cos\theta$$

is a trigonometric equation but *not a trigonometric identity*, because it is not true for all θ. For example, if $\theta = \pi/2$, then $\sin(\pi/2) = 1$, whereas $\cos\pi/2 = 0$. Thus, $\sin\theta \neq \cos\theta$, for $\theta = \pi/2$.

Note: Trigonometric identities and solutions of trigonometric equations are very important and useful in various problems of engineering and science. Simple methods are available to obtain the solutions of trigonometric equations.[9]

[9] Solutions of trigonometric equations:

- The solutions of trigonometric equations for which $0 \leq \theta < 2\pi$ are called principal solutions.
- The solution involving integer n that gives all solutions of a trigonometric equation is called the general solution

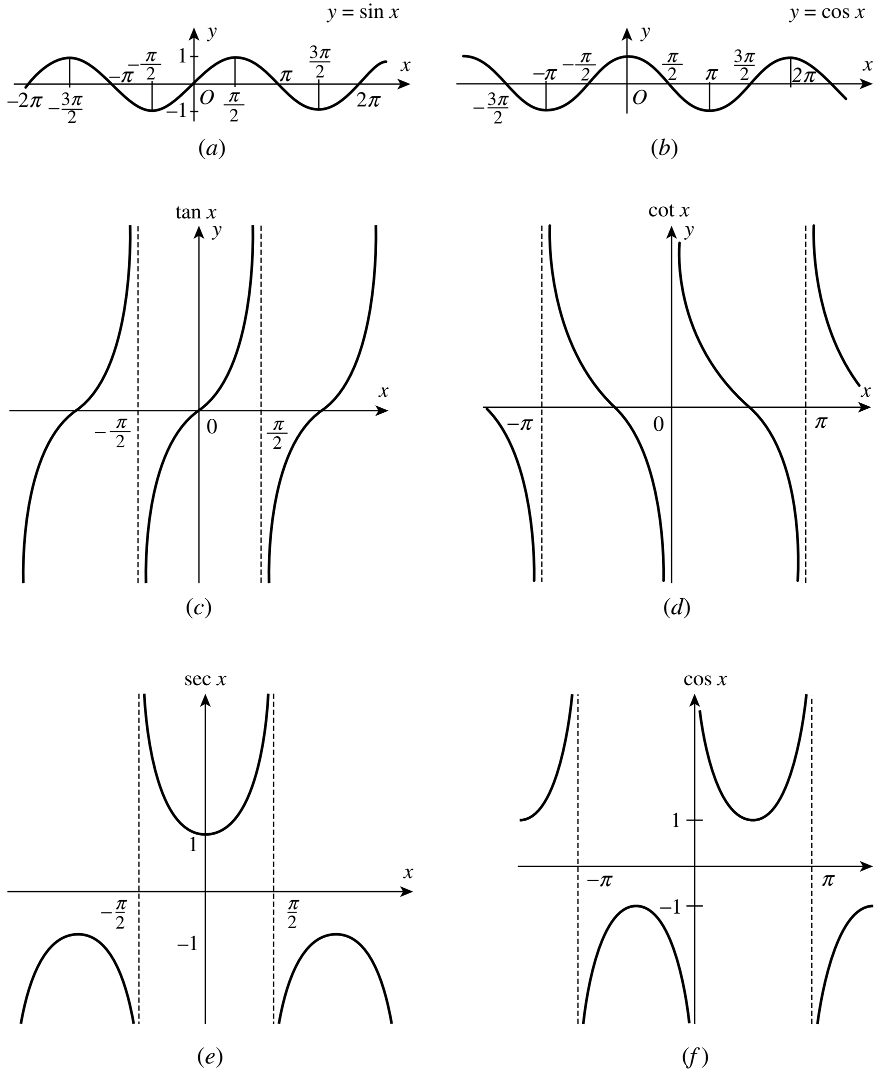


FIGURE 5.18 Graphs of trigonometric functions.

Here, we shall obtain some trigonometric identities and also list below some important identities used frequently.

Consider a circle of radius r, centered at the origin $O(0, 0)$. Then, distance r between the origin and any point $P(x, y)$ on the circle (see Figure 5.13a) is given by

$$x^2 + y^2 = r^2 \text{(using the distance formula)}$$

Substituting for x and y from the definitions $\cos\theta = x/r$ and $\sin\theta = y/r$, we get

$$r^2\cos^2\theta + r^2\sin^2\theta = r^2$$

This yields the famous Pythagorean identity

$$\cos^2\theta + \sin^2\theta = 1$$

Since we normally use x to represent points in the domain of a function, we will usually follow that convention for sine and cosine functions and replace θ *by* x. Thus, the above identity becomes

$$\sin^2 x + \cos^2 x = 1 \tag{9}$$

The next two identities are obtained by dividing both sides of equation (9) by $\cos^2 x$ and $\sin^2 x$, respectively. We have

$$\tan^2 x + 1 = \sec^2 x$$

$$1 + \cot^2 x = \operatorname{cosec}^2 x$$

These two identities are also called *Pythagorean identities*. Next, the following five important identities follow from the definitions of tan x, cot x, sec x, and cosec x:

$$\tan x = \frac{\sin x}{\cos x}, \cot x = \frac{\cos x}{\sin x}, \sin x \cdot \operatorname{cosec} x = 1, \cos x \cdot \sec x = 1, \text{and} \tan x \cdot \cot x = 1$$

These eight identities are called *fundamental trigonometric identities* (or basic trigonometric identities).

Also, we have discussed that $\sin(-x) = -\sin x$ and $\cos(-x) = \cos x$ $(*)$

Besides, it is easily proved (using geometry and the definitions of sine and cosine functions) that

$$\sin(x \pm y) = \sin x \cos y \pm \cos x \sin y$$

and

$$\cos(x \pm y) = \cos x \cos y \pm \cos x \cdot \sin y \; (**)$$

The list of trigonometric identities is very large. Memorizing every trigonometric identity is out of question. It is wiser to memorize only the *basic eight identities* and those indicated by $(^*)$ and $(^{**})$.

These are the most important ones. Other trigonometric identities can be derived using these identities. For convenience, we give below a list of several of the more useful trigonometric identities frequently used (including those mentioned above):

$$\sin^2 x + \cos^2 x = 1 \tag{10}$$

$$\sec^2 x = 1 + \tan^2 x \tag{11}$$

$$\operatorname{cosec}^2 x = 1 + \cot^2 x \tag{12}$$

$$\begin{cases} \sin(x+y) = \sin x \cos y + \cos x \sin y \\ \sin(x-y) = \sin x \cos y - \cos x \sin y \\ \cos(x+y) = \cos x \cos y - \sin x \sin y \\ \cos(x-y) = \cos x \cos y + \sin x \sin y \end{cases} \tag{13}$$

$$\begin{cases} \tan(x+y) = \dfrac{\tan x + \tan y}{1 - \tan x \tan y} \\ \tan(x-y) = \dfrac{\tan x - \tan y}{1 + \tan x \tan y} \end{cases} \tag{14}$$

Other important trigonometric identities, listed below [from (6) onward] can be derived from the above identities.

$$\begin{cases} \sin 2x = 2 \sin x \cos x, \quad \sin 2x = \dfrac{2 \tan x}{1 + \tan^2 x}, \quad \tan 2x = \dfrac{2 \tan x}{1 - \tan^2 x} \\ \left.\begin{aligned} \cos 2x &= \cos^2 x - \sin^2 x \\ &= 1 - 2\sin^2 x \\ &= 2\cos^2 x - 1 \\ \therefore\ \sin^2 x &= \frac{1 - \cos 2x}{2} \\ \text{and } \cos^2 x &= \frac{1 + \cos 2x}{2} \end{aligned}\right\} \cos 2x = \dfrac{1 - \tan^2 x}{1 + \tan^2 x} \end{cases} \tag{15}$$

$$\begin{aligned} \sin 3x &= 3 \sin x - 4 \sin^3 x \\ \tan 3x &= \frac{3 \tan x - \tan^3 x}{1 - 3 \tan^2 x} \\ \cos 3x &= 4 \cos^3 x - 3 \cos x \end{aligned} \tag{16}$$

$$\begin{cases} \sin(x + 2n\pi) = \sin x \\ \cos(x + 2n\pi) = \cos x \end{cases} \tag{17}$$

$$\begin{cases} \sin(-x) = -\sin x \\ \cos(-x) = \cos x \end{cases} \tag{18}$$

$$\begin{cases} \sin\left(\frac{\pi}{2}-x\right)=\cos x,\ \sin\left(\frac{\pi}{2}+x\right)=\cos x, \\ \cos\left(\frac{\pi}{2}-x\right)=\sin x,\ cos\left(\frac{\pi}{2}+x\right)=-\sin x, \\ \tan(x+y)=\dfrac{\tan x+\tan y}{1-\tan x\tan y}, \\ \tan(x-y)=\dfrac{\tan x-\tan y}{1+\tan x\tan y}, \end{cases} \qquad \sin\frac{\pi}{2}=1 \text{ and } \cos\frac{\pi}{2}=0 \tag{19}$$

$$\begin{cases} \tan 2x=\dfrac{2\tan x}{1-\tan^2 x}, \\ \tan\left(\frac{\pi}{4}+x\right)=\dfrac{1+\tan x}{1-\tan x},\left[\because \tan\frac{\pi}{4}=1\right], \qquad \tan\left[\frac{\pi}{2}-\left(\frac{\pi}{4}+x\right)\right]=\tan\left(\frac{\pi}{4}-x\right)=\cot\left(\frac{\pi}{4}+x\right) \\ \tan\left(\frac{\pi}{4}-x\right)=\dfrac{1-\tan x}{1+\tan x}, \end{cases} \tag{20}$$

$$\begin{cases} \sin A+\sin B = 2\sin\dfrac{A+B}{2}\cdot\cos\dfrac{A-B}{2} \\ \sin A-\sin B = 2\cos\dfrac{A+B}{2}\cdot\sin\dfrac{A-B}{2} \\ \cos A+\cos B = 2\cos\dfrac{A+B}{2}\cdot\cos\dfrac{A-B}{2} \\ \cos A-\cos B = -2\sin\dfrac{A+B}{2}\cdot\sin\dfrac{A-B}{2} \end{cases} \tag{21}$$

These identities are expressed in the following useful forms.

$$\begin{cases} \sin A\cos B=\dfrac{1}{2}[\sin(A+B)+\sin(A-B)] \\ \text{Usefulness in solving problems :} \\ \sin 5x\cos x=\dfrac{1}{2}[\sin 6x+\sin 4x] \\ \sin 5x\cos 7x=\dfrac{1}{2}[\sin 12x+\sin(-2x)]=\dfrac{1}{2}[\sin 12x-\sin 2x] \\ [\text{Note : } \cos A\cdot\sin B=\sin B\cdot\cos A] \end{cases} \tag{22a}$$

$$\begin{cases} \cos A \cos B = \frac{1}{2}[\cos(A+B) + \cos(A-B)] \\ \text{Usefulness in solving problems :} \\ \cos 5x \cos 3x = \frac{1}{2}[\cos 8x + \cos 2x] \\ \cos 2x \cos 5x = \frac{1}{2}[\cos 7x + \cos(-3x)] \\ = \frac{1}{2}[\cos 7x + \cos 3x] \quad [\because \cos(-x) = \cos x] \end{cases} \tag{22b}$$

$$\begin{cases} \sin A \cdot \sin B = \frac{1}{2}[\cos(A-B) - \cos(A+B)] \\ \text{Usefulness in solving problems:} \\ \sin 3x \cdot \sin 2x = \frac{1}{2}[\cos x - \cos 5x] \\ \sin 3x \cdot \sin 5x = \frac{1}{2}[\cos(-2x) - \cos 8x] \\ = \frac{1}{2}[\cos 2x - \cos 8x] \end{cases} \tag{22c}$$

5.8 REVISION OF CERTAIN IDEAS IN TRIGONOMETRY

It is useful to revise the following important points discussed in this chapter.

(a) In our school geometry, the definitions of trigonometric ratios are introduced for an acute angle in a right-angled triangle in terms of the ratios of its sides. Then, the concept of angle is extended to define directed angles that could have any magnitude (positive, zero, or negative).

Using the concept of radian measure of an angle, the directed angles are identified with real numbers, and vice versa. This helps in defining trigonometric ratios for angles having any magnitude (i.e., trigonometric functions of real numbers).

(b) The values of trigonometric functions (i.e., trigonometric ratios for angles of any magnitude) are still defined with reference to an acute angle in a right-angled triangle as follows:

We choose on the revolving line (generating an angle θ) a point P anywhere (other than the origin) and draw a perpendicular PM on the x-axis, as shown in Figure 5.19.

The values of trigonometric functions of an angle θ are then defined (as usual) for the acute angle $\angle POM$, made by the revolving line with the x-axis, the reference right-angled triangle being ΔOMP. Depending on the position of the revolving line, this triangle may be in any quadrant with $\overline{OM}$ as the *base* segment and $\overline{MP}$ as the *perpendicular* segment.

In the *standard unit circle*, if we change the (directed) angle θ, then the magnitude(s) of the (directed) line segment(s) $\overline{OM}$ and $\overline{MP}$ must change. The sign of any

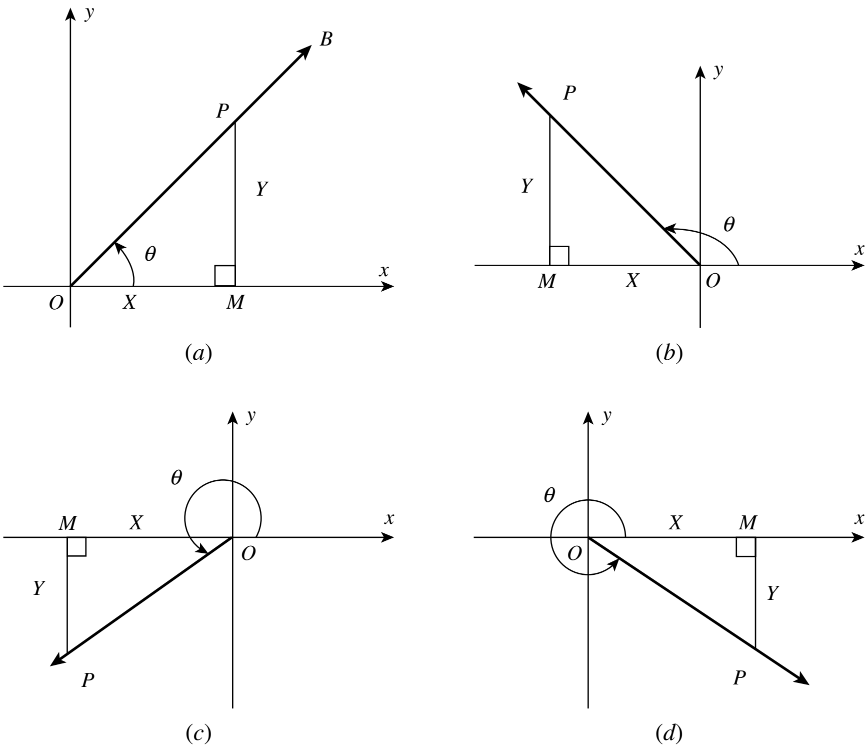


FIGURE 5.19

trigonometric ratio depends on the signs of these signed line segments. Accordingly, the values of trigonometric functions (for different angles) have different values, and their signs depend on the position of the terminal side. [The hypotenuse (being a line segment of the revolving line) is treated as *undirected segment,* and hence identified with *a positive number*.]

Recall that the functions sin θ and cos θ are the two basic trigonometric functions that are independent of each other. The remaining four trigonometric functions (i.e., tan θ, cot θ, sec θ, and cosec θ) are defined in terms of sin θ and cos θ.

(c) Coordinate geometry plays a very important role in defining trigonometric functions. We may choose the point *P* anywhere on the revolving line (except at the origin) and the trigonometric ratios are defined on the basis of coordinates of *P*.

Definitions of trigonometric functions are justified, based on the following two facts:

(i) The value of each trigonometric function is independent of the position of *P* on the revolving line [see Figure 5.13b: (i) and (ii)].

(ii) The values of trigonometric functions depend on the position of the *terminal side* in different quadrants.

(d) We know that *each trigonometric ratio* (and its reciprocal) is *positive* in two quadrants and *negative* in the remaining two quadrants. Hence, even if we are given the value of any one trigonometric function, it is *not possible* to find exactly the quadrant in which the terminal lies (depending on the quadrant in which the terminal side lies, we say that *the angle in question* lies in that quadrant).

Hence, to be able *to determine the position of the terminal side exactly*, it is necessary that the *values of two independent trigonometric functions* are given. Also, if the value of a single trigonometric function is given and *the quadrant* in which the angle x lies is also given, then we can find out *exactly* the values of all other trigonometric ratios for the angle x.

(e) Recall that if the Cartesian coordinates of a point P are (x, y) and its polar coordinates are (r, θ), then we have

$$x = r\cos\theta \text{ and } y = r\sin\theta \tag{23}$$

where

$$r = \sqrt{x^2 + y^2} \text{ and } \tan\theta = \frac{y}{x} \tag{24}$$

The relations (23) and (24) enable us to change the coordinates of a point from one system to the other. In view of the relations at (23), we have

$$P(x, y) = P(r\cos\theta,\ r\sin\theta)$$

We know (as already discussed in Chapter 4) that the polar coordinates differ from Cartesian coordinates as follows:

With each point P in the coordinate plane is associated a *unique pair* of Cartesian coordinates (x, y), and, conversely, with each ordered pair of real numbers is associated a unique point in the coordinate plane.

On the other hand, each ordered pair (r, θ) determines a point uniquely, but if θ is the amplitude of P (which means that if θ is the varying quantity) then all ordered pairs of the form $(r, \theta + 2n\pi)$, where n is an integer, correspond to the same point. Thus, an unlimited number of ordered pairs $(r, \theta + 2n\pi)$ represent the same point in the polar coordinate system.

Note: If θ is the radian measure of the angle involved and r is treated as the *undirected distance* from the origin O to P (i.e., $r = |\overline{OP}|$), then one set of polar coordinates of P is given by r and θ, denoted by $(r, \theta + 2n\pi)$.

On the other hand, if r is treated as the *directed distance* (with usual convention of signs), then we get another set of polar coordinates of P in which r is *negative* (details are already discussed toward the end of Chapter 4).

5.8.1 Illustrative Solved Examples: Revision of Useful Concept in Trigonometry

Example (4): Write the sign of sin 2.

Solution: sin 2 means sine of that angle whose measure is 2 rad. We know that π rad ≈ 3.14. Therefore, $(\pi/2) \approx 1.57$. Thus, $(\pi/2) < 2 < \pi$. It means that the angle of measure 2 rad lies in the second quadrant. Therefore, *sin* 2 *is positive*. Ans.

(Note that, cos 2 is negative and tan 2 must also be negative.)

Example (5): If the x-coordinate of a point on the unit circle is 8/17, find its y-coordinate.

Solution: Let the y-coordinate of the point on the unit circle be y.

$$\therefore x^2 + y^2 = 1$$

$$\left(\frac{8}{17}\right)^2 + y^2 = 1 \quad \text{or} \quad y^2 = 1 - \frac{64}{289} = \frac{225}{289} = \left(\frac{15}{17}\right)^2$$

$$\therefore y = \pm\frac{15}{17} \text{ Ans.}$$

Example (6): Given sin $x = -(3/5)$. State in which quadrants can the angle x lie.(10)

Solution: Since sin x is given to be a negative number, the angle x must lie in third or fourth quadrant. Ans.
(If tan x is a negative number, then x must lie in the second and fourth quadrants; and if cos x is negative, then x must lie in the second and third quadrants.)

Example (7): If sin $x = -(3/5)$ and cos $x = -(4/5)$, state the quadrant in which the angle x lies.

Solution: We know that sin x is *negative* in the *third* and *fourth* quadrants and cos x is *negative* in the *second* and *third* quadrants. Thus, *both* the given *conditions* are satisfied if x lies in the third quadrant. Ans.

Example (8): Find the values of trigonometric functions sin x, cos x, and tan x of *an angle x in standard position* whose *terminal arm* passes through the point $P(-3, 4)$.

Solution: The distance of the point $P(-3, 4)$ from the origin $O(0, 0)$ is given by

$$r = OP = \sqrt{(-3-0)^2 + (4-0)^2} = \sqrt{9+16} = 5$$

$$\therefore \sin x = \frac{y - \text{coordinate of } P}{r} = \frac{4}{5}$$

$$\cos x = \frac{x - \text{coordinate of } P}{r} = \frac{-3}{5} = \frac{3}{5}$$

(10) Recall that angle x is said to be in that quadrant in which the terminal side of the angle lies.

and

$$\tan x = \frac{\sin x}{\cos x} = \frac{4}{-3} = -\frac{4}{3} \text{ Ans.}$$

Example (9): If $\cos x = -(3/5)$, and x lies in the third quadrant, find the values of other five trigonometric functions.

Solution: We have $\cos x = -(3/5)$.

$$\therefore \sec x = -\frac{5}{3}$$

(Now we can use the identity $\sin^2 x + \cos^2 x = 1$ to compute the value(s) of $\sin x$.)
We have

$$\sin^2 x = 1 - \cos^2 x$$
$$= 1 - \left(-\frac{3}{5}\right)^2 = 1 - \frac{9}{25} = \frac{16}{5}$$
$$\therefore \sin x = \pm\sqrt{\frac{16}{25}} = \pm\frac{4}{5}$$

But, it is given that x lies in the third quadrant that means that $\sin x$ is negative. Therefore, we take the value of $\sin x = -(4/5)$. Accordingly, we have $\operatorname{cosec} x = -(5/4)$. Furthermore, we have $\tan x = (\sin x/\cos x) = 4/3$, and $\cot x = 3/4$ Ans.

Note: This example tells us about the usefulness of trigonometric identities.

Example (10): Find the values of the other five trigonometric functions if $\tan x = -(5/12)$ and x lies in the second quadrant.

Solution: Since $\tan x = -(5/12)$, we have

$$\cot x = -\frac{12}{5}$$

The identity $\sin^2 x + \cos^2 x = 1$ suggests that (by dividing both sides by $\cos^2 x$)

$$\tan^2 x + 1 = \sec^2 x \Rightarrow \left(-\frac{5}{12}\right)^2 + 1 = \sec^2 x$$

or

$$\sec^2 x = 1 + \frac{25}{144} = \frac{169}{144} = \left(\frac{13}{12}\right)^2$$

$$\therefore \sec x = \pm\frac{13}{12}$$

But x lies in second quadrant, which means that sec x *must be negative.*

$$\therefore \sec x = -\frac{13}{12}$$

$$\therefore \cos x = -\frac{12}{13}$$

Furthermore, we have

$$\tan x = \frac{\sin x}{\cos x} \Rightarrow \sin x = \tan x \cdot \cos x = \left(-\frac{5}{12}\right) \cdot \left(-\frac{12}{13}\right) = \frac{5}{13}^{(11)}$$

and

$$\text{cosec } x = \frac{1}{\sin x} = \frac{13}{5} \text{ Ans.}$$

Example (11): Given sin $x = 5/6$, find cos x.

Solution: We know that sin x is positive in the *first* and *second* quadrants. We have $\sin^2 x + \cos^2 x = 1$

$$\therefore \cos^2 x = 1 - \sin^2 x = 1 - \left(\frac{5}{6}\right)^2$$

$$= 1 - \frac{25}{36} = \frac{11}{36}$$

$$\therefore \cos x = \pm\frac{\sqrt{11}}{6} \text{ Ans.}$$

(Note that $\cos x = -\left(\sqrt{11}/6\right)$ when the angle x is in the second quadrant).

Example (12): Given sin $x = 5/6$, find cosec x and cot x.

Solution: Since sin $x = 5/6$, it follows that cosec $x = 6/5$. Furthermore, in Example (8), we have shown that for the given value of sin x, there are two values of $\cos x : \sqrt{11}/6$ and $-\left(\sqrt{11}/6\right)$.

Accordingly, cot x will have two values:

$$\frac{\sqrt{11}}{5} \text{ and } -\frac{\sqrt{11}}{5} \text{ Ans.}$$

(11) It is convenient to compute sin x using the identity sin = tan x·cos x, rather than using any other identity.

Example (13): Fill in the blanks:

(a) If cot $x = -1$ and sec $x = 1/2$, then x must lie in _ quadrant.
(b) If sin $x = -(3/5)$ and cos $x = 4/5$, then x must lie in _ quadrant.

Solution:

(a) cot x is negative in the *second* and *fourth* quadrants. On the other hand, sec x is positive in the *first* and *fourth* quadrants.
Therefore, x must lie in the fourth quadrant. Ans.
(b) sin x is negative in the *third* and *fourth* quadrants, while cos x is positive in the *first* and *fourth* quadrants.
Therefore, x must lie in the fourth quadrant. Ans.

Example (14): The coordinates of P are $(4 \cos \theta, 4 \sin \theta)$. Find $|OP|$ if O is the origin.

Solution: If (r, θ) are the polar coordinates of a point whose Cartesian coordinates are (x, y), then we have $x = r \cos \theta$ and $y = r \sin \theta$.

On comparing we get $r = 4$ in this problem.

$$\therefore |OP| = 4 \text{ Ans.}$$

Example (15): Convert $x = a$ into *polar form*.

Solution: By the relation between Cartesian coordinates and polar coordinates, we have $x = r \cos \theta$.

$$\therefore r \cos \theta = a. \text{ Ans.}$$

Exercises

Q. (1) Find the values of sin x, sec x, and tan x under the following conditions:

(a) cos $x = 12/13$, x lies in the first quadrant.

Ans. $\frac{5}{13}, \frac{13}{12}, \frac{5}{12}$

(b) tan $x = 1/3$, x lies in the third quadrant.

Ans. $-\frac{1}{\sqrt{10}}, -\frac{\sqrt{10}}{3}, \frac{1}{3}$

(c) sin $x = -(3/5)$ and tan x is positive.

Ans. $-\frac{2}{5}, -\frac{5}{\sqrt{21}}, \frac{2}{\sqrt{21}}$

Q. (2) Find the values of all trigonometric functions:

(i) If cot $x = 2/3$, x lies in the third quadrant.

$$\text{Ans. } \sin x = -\frac{3}{\sqrt{13}},\ \cos x = -\frac{2}{\sqrt{13}},\ \tan x = \frac{3}{2},\ \operatorname{cosec} x = -\frac{\sqrt{13}}{3},\ \sec x = -\frac{\sqrt{13}}{2}$$

(ii) If tan $x = -5$, x lies in the fourth quadrant.

$$\text{Ans. } \sin x = -\frac{5}{\sqrt{26}},\ \cos x = -\frac{1}{\sqrt{26}},\ \tan x = -5,\ \sec x = \sqrt{26},$$

$$\operatorname{cosec} x = -\frac{\sqrt{26}}{5}, \cot x = -\frac{1}{5}$$

6 More About Functions

6.1 INTRODUCTION

In Chapter 2, we defined a *function* as a special relation on the basis of set theory and discussed the related terminology. In this chapter, we think of a *function as a machine*. It will be found that this way of looking at a function is more useful than the earlier definitions.

The first important step toward learning the subject of calculus is to understand clearly the concept *of numerical function, by which we mean those functions* in which both the *domain* and the *range* consist of real numbers. The numerical functions of interest (in calculus) are those that are defined on intervals. They may be defined by one, or more formulas, given as follows:[(1)]

$$f(x) = 5x + 2, \; x \in R \tag{1}$$

$$g(x) = \begin{Bmatrix} 2x + 3, & x \leq 0 \\ x + 1, & x > 0 \end{Bmatrix} \tag{2}$$

$$h(x) = \frac{x^2 - 9}{x - 3}, \; x \neq 3 \tag{3}$$

Many functions arise as combinations of other functions. It is, therefore, necessary to discuss different methods of combining functions and find out the domain(s) of such combinations. Recall that *a single letter f* (or g or h or F, etc.) is used to name a function and that $f(x)$ denotes *the value* that the function "f" assigns to "x". We read $f(x)$ as "f of x" or "the value of f at x".

6.2 FUNCTION AS A MACHINE

We can think of a function as a *machine* (see Figure 6.1) that takes the members "x" of the domain and applies a rule (does something) to each "x", to produce the members "$f(x)$" of the range.

Consider the function at (1) above.

What must you know to learn calculus? 6-More about functions (Function as a machine, combinations, and compositions of functions and their domains)

(1) Though we are going to discuss only simple algebraic functions here, other functions of our interest (namely, trigonometric, logarithmic, exponential, and hyperbolic, etc.) and their properties are discussed later at appropriate places.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

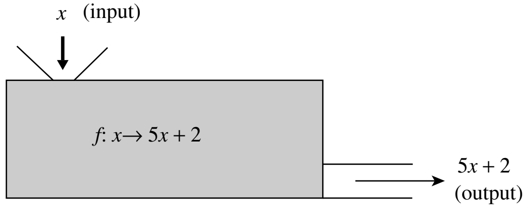


FIGURE 6.1 Function as a machine.

Here, the name of *the machine is* "f", and "rule of the machine" (or the operation of the machine) is given by

$$f: x \rightarrow 5x + 2$$

This operation converts each x (of the domain) into $5x + 2$. Thus, the number "0" fed into the machine is converted into the number 2. Similarly, 1/5 is converted into 3, –2/5 is converted into "0", $\sqrt{2}$ is converted into $5\sqrt{2} + 2$, and so on. In view of the above, we give the following definition:

Definition: *A function is an operation that assigns to each input number exactly one output number.*(2)

6.3 DOMAIN AND RANGE

The set of *all input numbers* that can be used in the operation is called the *domain* of the function. The set of *all output numbers* is called the *range.*

6.3.1 Natural Domain

When no domain is specified for a function, we always take *the domain as the largest set of real numbers for which the rule of the function makes sense* and gives *real number values*. This is called the *natural domain of the function*.

For example, *the natural domain* for $f(x) = 1/(x - 5)$ is $\{x \in R | x \neq 5\}$. We exclude 5 to avoid division by 0. Similarly, $g(x) = \sqrt{x}$ has the natural domain $[0, \infty)$ since this function is defined only for $x \geq 0$.

6.4 DEPENDENT AND INDEPENDENT VARIABLES

In calculus, we deal with functions, which are defined by *formulas* expressing dependence of one quantity on another. When two variables are related to one another, strictly speaking, either variable may be expressed in terms of the other. *In most situations, it is more natural to regard the variation of one as independent of other, in a way controlling the variation of the*

(2) We stress two key points in the definition:
(i) A function must make an assignment to *each* number in the domain.
(ii) A function *can assign only one number* to any given number in the domain.

other. (For example, it is more appropriate to say that income tax depends on the income, than it is to say it the other way round.) Similarly, we say that

- Area of a circle *depends* on its radius.
  $A(r) = \pi r^2$ [A is a function of "r"]
- Volume of a sphere *depends* on its radius.
  $V(r) = \frac{4}{3}\pi r^3$ [V is a function of "r"]
- Surface area of a cube *depends* on the length of its side
  $S(x) = 6x^2$ [S is a function of "x"]

When the *rule for a function is* given by an equation of the form $y = f(x)$ (e.g., $y = x^5 + 7x^2 - 2x + 3$), x is called the *independent variable* and y or $f(x)$, the *dependent variable*. This is a *numerical function* and its domain must be a set of real numbers. Any element of the domain must be chosen independently (as a value of the independent variable) and this choice completely determines the corresponding value of the dependent variable, y or $f(x)$. We say that the value $f(x)$ depends on the chosen value of x. In other words, the value $f(x)$ changes with x.

Note: We shall sometimes, *by abuse of notation, speak of the function $f(x)$*, but strictly speaking, "f" is the function and "$f(x)$" is the value of the function f at x. Whenever we speak of "the function $f(x)$", we shall generally mean "the value $f(x)$". However, if "$f(x)$" is used to stand for a function, we must read it as a function "f" of "x". The *meaning of the symbol $f(x)$* will be clear from the context.

Now, consider the function $f: x \to x^3 - 4$. Here, the function "f" converts each number "x" (of the domain) into $x^3 - 4$. We write $f(x) = x^3 - 4$.

Thus, $f(2) = 2^3 - 4 = 4$

$f(-1) = (-1)^3 - 4 = -5$,

$f(a) = a^3 - 4$

$f(a + h) = (a + h)^3 - 4 = a^3 + 3a^2h + 3ah^2 + h^3 - 4.$

Study the following examples carefully. They will play an important role later.

Example (1): For $f(x) = x^2 - 2x$, *find* and *simplify*

(a) $f(4)$,
(b) $f(4 + h)$,
(c) $f(4 + h) - f(4)$,
(d) $[f(4 + h) - f(4)]/h$, where $h \neq 0$.

Solution:

(a) $f(4) = 4^2 - 2(4) = 16 - 8 = 8$

(b) $f(4 + h) = (4 + h)^2 - 2(4 + h) = [4^2 + 2(4)h + h^2] - 2(4 + h)$

$= 16 + 8h + h^2 - 8 - 2h$

$= h^2 + 6h + 8$

(c) $f(4+h) - f(4) = (h^2 + 6h + 8) - 8 = h^2 + 6h$

(d) $\dfrac{f(4+h) - f(4)}{h} = \dfrac{h^2 + 6h}{h} = \dfrac{h(h+6)}{h} = h + 6 (\because h \neq 0)$

Example (2): Find the *natural domain* for $\phi(t) = \sqrt{9 - t^2}$.

Solution: Here, we must restrict "t" so that $9 - t^2 \geq 0$, in order to avoid nonreal values for $\sqrt{9 - t^2}$. This is achieved by requiring that $9 \geq t^2$ or $t^2 \leq 9$ or $|t| \leq 3$. Thus, the natural domain of ϕ is $\{t \in R: |t| \leq 3\}$. In interval notation, we can write the domain as $[-3,3]$.

6.5 TWO SPECIAL FUNCTIONS

We give below, two *very special functions* that will be used in many contexts.

(i) The Absolute Value Function | | (Figure 6.2) is defined by

$$|x| = \begin{cases} x & \text{if} \quad x \geq 0 \\ -x & \text{if} \quad x < 0 \end{cases}$$

Note that the graph of $|x|$ has a sharp corner at the origin.

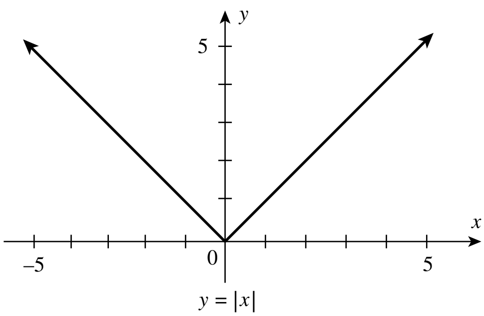


FIGURE 6.2 Absolute value function, $y = |x|$.

(ii) The Greatest Integer Function [ ] is defined by $[x]$ = the *greatest integer* less than or equal to x. Thus, $[2.1] = 2$, $[1.99] = 1$, $[-2.5] = -3$.

Its domain is the set of all real numbers and its range consists of all the integers. The graph of $[x]$ takes a jump at each integer (Figure 6.3).

6.6 COMBINING FUNCTIONS

Functions are *not* numbers. But, just as two numbers a and b can be added to produce a new number $(a + b)$, two functions f and g can be added to produce *a new function* $(f + g)$. This is just one of the several operations on functions. We shall consider the combinations and

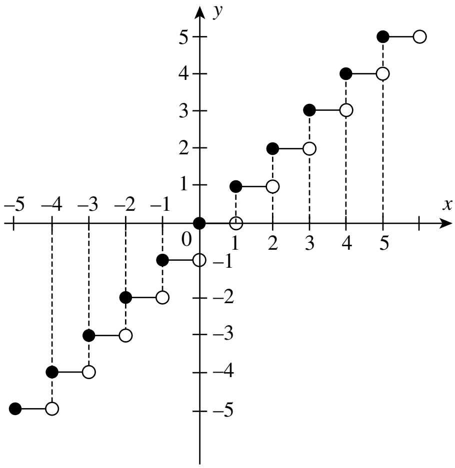


FIGURE 6.3 The greatest integer function, $y = [x] =$ the greatest integer less than or equal to x.

compositions of functions, together with some *special cases of combinations*, under the following heads:

- Sums, differences, products, and quotients of functions.
- Some simple functions and their combinations: constant function, identity function, polynomial function, linear function, and rational functions.
- Power functions.
- Root functions.
- Raising a function to a power.
- Composition of functions.

6.6.1 Sums, Differences, Products and Quotients of Functions

Let f and g be functions. We define the *sum* $f + g$, the *difference* $f - g$, and the product $f{\cdot}g$ to be the functions whose *domains consist of all those numbers that are common in the domains of both f and g* and whose *rules* are given by

$$(f+g)(x) = f(x) + g(x)$$
$$(f-g)(x) = f(x) - g(x)$$
$$(f \cdot g)(x) = f(x) \cdot g(x)$$

In each case, *the domain is the expected one*, consisting of *those values of x for which both f(x) and g(x) are defined*. Next, because division by 0 is excluded, we give the definition of quotient of two functions separately as follows:

The *quotient f/g* is the function whose domain consists of all numbers x in the domains of both f and g for which $g(x) \neq 0$, and whose rule is given by

$$\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}, \quad g(x) \neq 0$$

Example (3): Let $f(x) = 1/x$ and $g(x) = \sqrt{x}$. Let us find the *domain* and *rule* of $f + g$.

Solution: The domain of f is $\{x \in R | x \neq 0\}$ and the domain of g is $\{x \in R | x \geq 0\}$. *The only numbers in both domains* are the *positive numbers*, which constitute the *domain* of $f + g$.

For the rule, we have

$$(f+g)(x) = f(x) + g(x) = \frac{1}{x} + \sqrt{x} \ \ \textit{for} \ \ x > 0$$

Example (4): Let $f(x) = \sqrt{4 - x^2}$ and $g(x) = \sqrt{x - 1}$. Let us find the *domain* and *rule* of $f \cdot g$.

Solution: The domain of f is the interval $[-2, 2]$ and the domain of g is the interval $[1,\infty)$.

$\therefore$ The *domain* of $f \cdot g = [-2, 2] \cap [1, \infty] = [1,2]$. The rule of $f \cdot g$ is given by

$$\begin{aligned}(f \cdot g)(x) &= f(x) \cdot g(x) \\ &= \sqrt{4 - x^2}\ \sqrt{x-1} = \sqrt{(4 - x^2)(x - 1)} \ \text{ for } \ 1 \leq x \leq 2\end{aligned}$$

Caution: This example illustrates a *surprising fact* about the domain of combination of functions. We found that the domain of $f \cdot g$ is the interval [1,2]. Now observe that the expression $\sqrt{(4 - x^2)(x - 1)}$ *is also meaningful* for x in $(-\infty, -2]$. This is true because $(4 - x^2)\ (x - 1) \geq 0,\ x \leq -2$. *However,* $(-\infty, -2]$ *cannot be considered a part of the domain of* $f \cdot$g. By definition, the domain of the resulting function $f \cdot g$ consists of those values of x common to domains of f and g. It is not to be determined from the expression (or the rule) for $f \cdot g$. Similar comments hold for the domains of $f + g$ and $f - g$. For the domain of f/g, there is an additional requirement that the values of x, for which $g(x) = 0$, are excluded.

Example (5): Let $f(x) = x + 3$ and $g(x) = (x-3)\ (x + 2)$. Let us find the *domain* and *rule* of f/g.

Solution: Observe that the domains of f and g are all real numbers, but $g(x) = 0$, for $x = 3$ and -2. It follows that the *domain* of f/g consists of *all real numbers except* -2 and 3. The *rule* of f/g is given by

$$\left(\frac{f}{g}\right)x = \frac{f(x)}{g(x)} = \frac{x+3}{(x-3)(x+2)} \quad \text{for } x \neq -2 \quad \text{and} \quad x \neq 3$$

Note: We can *add* or *multiply* more than two functions. For example, if f, g, and h are functions, then for all x common to the domains of f, g, and h, we have $(f + g + h)\ (x) = f(x) + g(x) + h(x)$ and $(f \cdot g \cdot h)\ x = f(x) \cdot g(x) \cdot h(x)$.

6.6.2 Some Simple Algebraic Functions and Their Combinations

(a) *Constant Function*: A function of the form $f(x) = a$, where "a" is a *nonzero real number* (i.e., $a \neq 0$), is called a *constant function*.(3)

(The range of a constant function consists of only one nonzero number.)

(b) *Identity Function*: The function $f(x) = x$ is called the *identity function*.

From the functions at (a) and (b) above, we can build many important functions of calculus: polynomials, rational functions, power functions, root functions, and so on.

(c) *Polynomial Function*: Any function, that can be obtained from *the constant functions* and *the identity function* by using the operations of *addition, subtraction*, and *multiplication*, is called a *polynomial function*. This amounts to saying that "f" is a polynomial function, if it is of the form

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_3x^3 + a_2x^2 + a_1x + a_0$$

where $a_0, a_1, a_2, \ldots, a_n$ are real numbers $(a_n \neq 0)$ and *n is a nonnegative integer*.

If the coefficient $a_n \neq 0$, then "n" (in x^n), *the nonnegative integral exponent of x*, is called the *degree* of the polynomial. Obviously, the degree of *constant functions* is zero.(4)

- *Linear Function*: Polynomials of degree 1 are called *linear functions*. They are of the form $f(x) = a_1x + a_0$, with $a_1 \neq 0$. Note that, the identity function $[f(x) = x]$ is a particular *linear function*.
- $f(x) = a_2x^2 + a_1x + a_0$ is a *second degree polynomial*, called a *quadratic function*. If the degree of the polynomial is 3, the function is called a *cubic function*.
- *Rational Functions*: *Quotients of polynomials* are called *rational functions*. Examples are as follows:

$$f(x) = \frac{1}{x^2}, \quad f(x) = x^3 + \sqrt{5}x,$$

$$f(x) = \frac{x^3 - 2x + \pi}{x - \sqrt{2}}, \quad f(x) = \frac{x^2 + x - 2}{x^2 + 5x - 6}$$

Example (6): Let $f(x) = \frac{x^2+x-2}{x^2+5x-6}$. Let us find the *domain* of f.

Solution: We have $x^2 + 5x - 6 = (x - 1)(x + 6)$. Therefore, the denominator is 0 for $x = 1$ and $x = -6$. Thus, the *domain* of f consists of all numbers except 1 and -6.

Remark: Sometimes, it may happen that both the numerator and the denominator have a common factor. For example, we have $x^2 + x - 2 = (x - 1)(x + 2)$, and $x^2 + 5x - 6 = (x - 1)(x + 6)$

(3) Note that we do not call the function $f(x) = 0$, as a constant function. A special case of product occurs when one of the functions is a constant function: $g(x) = c$ for all x. For any function f, the domain of the product, $c \cdot f$, is the same as the domain of f.

(4) We distinguish between a zero-degree polynomial and a zero polynomial denoted by "0". Remember that while the degree of a constant polynomial is zero, the degree of zero polynomial is not defined. It can be easily seen why the degree of a zero polynomial cannot be defined. Accordingly, though some authors consider "0" as a special constant polynomial, but we shall not identify it as a constant polynomial.

$$\therefore\ f(x) = \frac{x^2 + x - 2}{x^2 + 5x - 6} = \frac{(x-1)(x+2)}{(x-1)(x+6)}$$

which may be simplified to read $(x+2)/(x+6)$, provided $x \neq 1$.

Note that, while the expression $(x + 2)/(x + 6)$ is meaningful for $x = 1$, *the number 1 is not in the domain of function f.*

(*This again suggests that the domain of a combination of functions must be determined from the original description of the function(s), and not from their simplified form.*)

6.6.3 Power Functions

These are functions, of the form $f(x) = x^n$, where n is an *integer*.

Examples are x^4, x^3, x^0, x^{-1}, x^{-4}, x^{-n}.(5)

We know that

$$x^3 = \frac{x^3}{1}, \quad x^0 = 1, \quad x^{-1} = \frac{1}{x}, \quad x^{-4} = \frac{1}{x^4}$$

The domain of x^n consists of all real numbers, if $n \geq 0$. If $n < 0$ (i.e., if n is a negative integer) then the domain consists of all real numbers *except* 0, since division by 0 is not defined.

Remark: Every power function is a rational function, but the converse is not true.

6.6.4 Root Functions

(a) *Square root function*: Consider the relation $y^2 = x$. We write it as $y = \sqrt{x}$ or $x^{1/2}$ and call it the square root function of x. *We know that there is no real number whose square is a negative number.* Hence, we define *square root* function $f(x) = \sqrt{x}$ that assigns *to each nonnegative* number x *the nonnegative number* $f(x)$.(6)

We *emphasize* that $\sqrt{x}$ is defined *only for* $x \geq 0$ and that $\sqrt{x} \geq 0$, for all $x \geq 0$. Accordingly, it is meaningful to write $\sqrt{8}$, $\sqrt{1/3}$, and $\sqrt{0}$, and so on, but $\sqrt{-5}$ has no meaning. Furthermore, while $\sqrt{4} = \pm 2$, we write $\sqrt{4} = 2$

(We never write $\sqrt{4} = -2$.)

(b) *Cube Root Function*: Consider the relation $y^3 = x$. We write it as $y = \sqrt[3]{x}$ or $x^{1/3}$, and call it the *cube root function*. It assigns to *any* number x, the *unique* number y such that $y^3 = x$. Of course, our interest lies only in real roots.

In contrast to the square root function, the cube root function has in its domain *all* real numbers, *including negative numbers*. For example, $\sqrt[3]{-1} = -1$, $\sqrt[3]{-8} = -2$, and $\sqrt[3]{-27/64} = -3/4$. Similarly $\sqrt[3]{8} = 2$, $\sqrt[3]{125} = 5$, and $\sqrt[3]{-125} = -5$. Thus cube root of any negative number is a negative number and that of any positive number is a positive number.

(c) *n*th *Root Function*: We note that cube root function "$\sqrt[3]{x}$" is defined for *all real numbers x*, whereas square root function "$\sqrt{x}$" is defined only for $x \geq 0$ *with the understanding that* $\sqrt{x} \geq 0$ (i.e., only nonnegative square roots are accepted). By

(5) Note that power functions are a special class of rational functions.

(6) There is a legitimate relation between square root and absolute value of a number, given by $|x| = \sqrt{x^2}$. This is obtained from the relation $|x|^2 = x^2$, which gives the definition of absolute value $|x|$.

extending these concepts to the roots of higher order, we get that *if* n *is odd*, then nth root function "$\sqrt[n]{x}$" is defined *for all real numbers*, and on the other hand, *if n is even*, then "$\sqrt[n]{x}$" is defined only for $x \geq 0$.[7]

Note (1): *In view of the* above, the expressions $\sqrt[3]{-1}$, $\sqrt[5]{-32}$, and $\sqrt[7]{-128}$ are *meaningful*, whereas the expressions $\sqrt[4]{-1}$, $\sqrt[6]{-64}$, and $\sqrt{-9/4}$ are *meaningless.*

Note (2): For *every positive integer n*, we also have

$$\sqrt[n]{1} = 1 \quad \text{and} \quad \sqrt[n]{0} = 0$$

Now, we can define the nth root function, by $f(x) = \sqrt[n]{x}, x \geq 0$, with the understanding that *whenever n is even*, we shall consider only positive nth root (i.e., for $x > 0$, $\sqrt[n]{x} > 0$).

6.7 RAISING A FUNCTION TO A POWER

We may also raise a function to a power. *By* f^n, *we mean the function that assigns to x the value* $[f(x)]^n$. Thus, if $f(x) = \frac{x-3}{2}$ and $g(x) = \sqrt{x}$, then

$$f^2(x) = [f(x)]^2 = \left[\frac{x-3}{2}\right]^2 = \frac{x^2 - 6x + 9}{4}$$

$$g^3(x) = [g(x)]^3 = \left(\sqrt{x}\right)^3 = x^{3/2}$$

and

$$f^{-2}(x) = 1/[f(x)]^2 = 1/[(x-3)/2]^2 = 4/(x^2 - 6x + 9)$$

Remark: There is one exception to the above agreement. *We never give the power* "-1" *to f.* We reserve the symbol f^{-1} for the *inverse function*, which we have already introduced in Chapter 2. *Thus,* f^{-1} *does not mean* $1/f$.

6.8 COMPOSITION OF FUNCTIONS

This is *another way of combining functions* that occur frequently in calculus. In fact, obtaining the composite function of two given functions is a new operation. This (new) operation consists of *carrying out two operations one after the other*, as illustrated by the following example.

[7] To understand this more clearly, consider $\sqrt[4]{16} = \sqrt[4]{(2)^4} = \left[(2)^4\right]^{\rightleftarrows 1/4} = 2$. Though it is also possible to write $\sqrt[4]{16} = \sqrt[4]{(-2)^4} = \left[(-2)^4\right]^{1/4} = -2$, we discard this negative fourth root of 16 (note that in $\sqrt[4]{16}$, n is 4, which is even).

Consider the function, $\phi(x) = f(x) = \sqrt{x+7}$

We may look at $\phi(x)$ as the result of carrying out the following two operations, one after the other:

(i) *Add 7 to x*. We express this operation by $f(x) = x + 7$.

(ii) *Take the square root of the above result*.
We express this operation by $g(x) = \sqrt{x}$.
(Here, it must be clearly understood that $\sqrt{x}$ stands for $\sqrt{f(x)}$.)

Thus, $\phi(x)$ is obtained by *first applying f to x* and then *applying g to the resulting value* $f(x)$.

To understand the method of *composition of functions*, think of two machines, put together, one after the other, thus making a more complicated machine. Let these machines represent functions f and g.

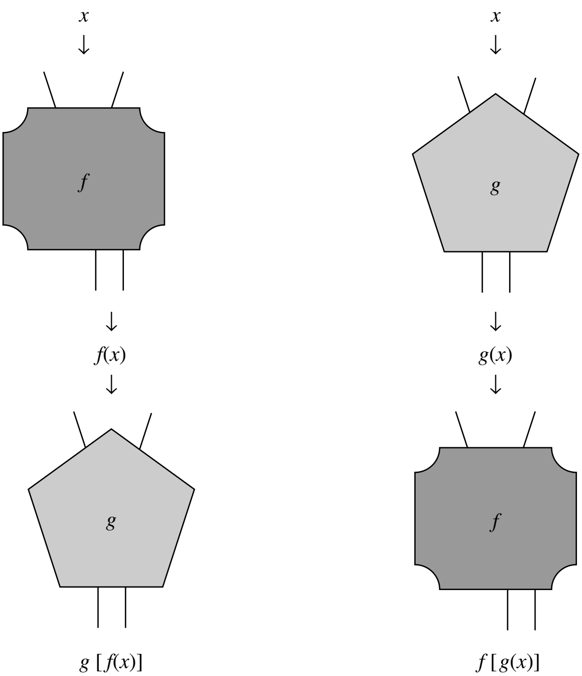
If f works on x to produce f(x) and then g works on f(x) to produce g(f(x)), we say that we have composed g with f (see Figure 6.4a).

The resulting function is called the *composite of g with f*, and we denote it by $g \circ f$. Thus,

$$(g \circ f)\,(x) = g(f(x))$$

If g works on x to produce g(x) and then f works on g(x) to produce f(g(x)), we say that we have composed f with g (see Figure 6.4b).

The resulting function is called the *composite of f with g*, and we denote it by $f \circ g$.



(*a*) (*b*)

FIGURE 6.4

6.8.1 Definition of a Composite Function

Given the two function f and g, the composite function denoted by $(g \circ f)$ is defined by

$$(g \circ f)\,(x) = g(f(x))$$

and the domain of $g(f(x))$ is the set of all numbers x in the domain of f such that $f(x)$ is in the domain of g.

The definition indicates that when computing $(f \circ g)(x)$, *we first apply* g *to x and then the function f to g(x).*

We write

$$(f \circ g)(x) = f(g(x))$$

Example (7): Let $f(x) = \frac{x-3}{2}$ and $g(x) = \sqrt{x}$. We may composite them as follows:

(i) $$(g \circ f)(x) = g(f(x)) = g\left(\frac{x-3}{2}\right), \left[\because\ f(t) = \frac{t-3}{2}\right]$$
$$= \sqrt{\frac{x-3}{2}}, \left[\because\ g(t) = \sqrt{t}\right]$$

(ii) $(f \circ g)(x) = f(g(x)) = f(\sqrt{x})\left[\because g(t) = \sqrt{t}\right]$

Now consider $f(t) = t - (3/2)$. From this definition of f, it follows that

$$f(\sqrt{x}) = \frac{\sqrt{x} - 3}{2}$$

$$\therefore\ (f \circ g)(x) = f(g(x)) = f(\sqrt{x}) = \frac{\sqrt{x} - 3}{2}$$

Remark: Note that $(g \circ f)\ (x) \neq (f \circ g)\ (x)$. Thus, composition of functions is *not commutative*; $g \circ f$ and $f \circ g$ are usually different.

6.8.2 Domain of a Composite Function

We must be more careful in describing the domain of a composite function. Let $f(x)$ and $g(x)$ be defined for certain values of x. Then, the domain of $g \circ f$ is that part of the domain of f (i.e., those values of x) for which g can accept $f(x)$ as input.[(8)]

In the above example, the domain of $g \circ f$ is $[3,\infty)$, since x must be greater than or equal to 3 in order to give a nonnegative number $x - 3/2$ for g to work on.

[(8)] Thus, the domain of $g \circ f$ is a subset of the domain of f. Similarly, the domain of $f \circ g$ is a subset of the domain of g.

In calculus, we shall often need to take a given function and decompose it (i.e., break it) into composite pieces.[9]

Usually this can be done in several ways.

Example (8): Consider the function $\phi(x) = \sqrt{x^3 + 7}$.

We can express ϕ as the composition of the two functions g and f, given by $f(x) = x^3 + 7$ and $g(x) = \sqrt{x}$.

Now, we have

$$\phi(\mathrm{x}) = (g \circ f)\,(x) = g(f(x)) = g(x^3 + 7) = \sqrt{x^3 + 7}.$$

Next, we can also express ϕ as the composition of another pair of functions g and f given by $f(x) = x^3$ and $g(x) = \sqrt{x + 7}$.

$$\text{Consider } \phi(x) = (g \circ f)(x) = g(f(x)) = g(x3) = \sqrt{x^3 + 7}.$$

Example (9): Given $\phi(x) = \frac{1}{\sqrt{x^2+3}}$

Express ϕ as the composition of two function f and g in two ways:

(i) The function f containing the radical.
(ii) The function g containing the radical.

Solution: To solve such problems, it is necessary to develop the ability of decomposing the given function into composite pieces.

(i) We choose $f(x) = 1/\sqrt{x + 3}$ and $g(x) = x^2$.
Now, $(f \circ g)(x) = f(g(x)) = f(x^2) = 1/\sqrt{x^2 + 3}$.

(Observe that to express $f(g(x))$, first we insert the expression for $g(x)$ and obtain $f(t)$, where t stands for $g(x)$. Next, we write the expression for $f(t)$ and replace t by $g(x)$.)

(ii) Now, we choose $f(x) = 1/x$ and $g(x) = \sqrt{x^2 + 3}$.
Then,

$$(f \circ g)(x) = f(g(x)) = f(\sqrt{x^2 + 3}) = \frac{1}{\sqrt{x^2 + 3}}$$

(Here again, to express f(g(x)), first we insert the expression for $g(x)$ and obtain $f(t)$, where $f(t)$ stands for $g(x)$. Now we look at the expression for $f(t)$, which suggests that we must take the reciprocal of t.)

Example (10): Let $f(x) = \sqrt{x - 1}$ and $g(x) = 1/x$. We shall determine the functions $g \circ f$ and $f \circ g$, and then find $g\ (f(5))$ and $f(g(1/4))$.

[9] An important theorem in calculus, called the chain rule (discussed in Chapter 10), involves composite functions. When applying the chain rule (for computing the derivative of a composite function), it is necessary to think of the given function as the composition of two other functions.

Solution: The function $g \circ f$ is given by

$$g(f(x)) = g\left(\sqrt{x-1}\right) = \frac{1}{\sqrt{x-1}}, \left[\because g(t) = \frac{1}{t}\right]$$

The domain of f is $[1, \infty)$. Therefore, *the domain of $g \circ f$ consists of those numbers x in* $[1, \infty)$ *for which g can accept f(x) as input*. This demands that

$g\left(\sqrt{x-1}\right) = \frac{1}{\sqrt{x-1}}$ must be defined, which requires that $x \neq 1$. Therefore, the domain of $g \circ f$ is $(1, \infty)$.

The rule for $f \circ g$ is given by

$$f(g(x)) = f\left(\frac{1}{x}\right), \left[\because \; g(t) = \frac{1}{t}\right]$$

$$= \sqrt{\frac{1}{x} - 1}, \left[\because \; f(t) = \sqrt{t-1}\right]$$

The domain of g is the set of nonzero numbers, that is, $(-\infty, 0) \cup (0, \infty)$. Therefore, *the domain of $f \circ g$ consists of those numbers x in the above domain for which f can accept g(x) as input*. This demands that $f(1/x) = \sqrt{(1/x) - 1}$ must be defined. It requires that

$$\frac{1}{x} - 1 \geq 0$$

$$\Rightarrow \frac{1}{x} \geq 1 (x \text{must be positive with } 1/x \geq 1)$$

$$\Rightarrow x \leq 1$$

The domain is (0, 1].

Finally, we have $g(f(x)) = 1/\sqrt{x-1}$ and $f(g(x)) = \sqrt{(1/x) - 1}$

$$\therefore \; g(f(5)) = \frac{1}{\sqrt{5-1}} = \frac{1}{\sqrt{4}} = \frac{1}{2} \text{ and}$$

$$f\left(g\left(\frac{1}{4}\right)\right) = \sqrt{\frac{1}{1/4} - 1}, \left[\because x = \frac{1}{4}\right]$$

$$= \sqrt{4-1} = \sqrt{3}$$

Note: We shall discuss about trigonometric, exponential, and logarithmic functions and their various properties (involving their combinations) at appropriate places in corresponding chapters.

6.9 EQUALITY OF FUNCTIONS

We say that two functions f and g are *equal* (or the *same*) if

(i) f and g have the *same domain* and
(ii) $f(x) = g(x)$, *for each x in the common domain.*

Thus, $f(x) = x^2$, $(1 \leq x \leq 3)$ and
$g(x) = x^2$, $(1 \leq x \leq 4)$, define two *distinct functions* because their domains are different. On the other hand, the equations

$$f(x) = x^2, x \geq 3,$$
$$g(x) = (x-1)^2 - 2x - 1, x \geq 3, \text{and } h(y) = y^2, y \geq 3$$

represent the same function because their domains are *identical* and their rules assign the same numerical number to each element (number) in the domain.

To summarize, if two functions have the same domain and assign the same value to each number in their domain, then they are equal.

6.10 IMPORTANT OBSERVATIONS

(i) *Two or more formulas may define a single function.* For example, consider

$$y = \begin{cases} \cos x, & x < 0 \\ 1 + x, & 0 \leq x \leq 2 \\ \log(x-1), & x > 2 \end{cases}$$

Note that, this is a *single function* defined on the real line, by three formulas.

(ii) *Not all functions can be written as formulas.* One such example is the Dirichlet function that is defined on the real line as follows:

$$y = \begin{cases} 1 & \text{if } x \text{ is a rational number} \\ 0 & \text{if } x \text{ is an irrational number} \end{cases}$$

This is certainly *an unusual function*, but still is a function. It maps the set of rational numbers to unity and the set of irrational numbers to zero. So far, no analytical expression is suggested for this function.

Similarly, the statement n^{th} *digit in the decimal representation of* π defines a function that *cannot* be expressed by any formula.

(iii) *Not every formula defines a function.* The rule of correspondence is the heart of a function. However, a function is not completely determined until its domain is given. We can write formulas whose domain is the empty set. Obviously, such formulas cannot represent any function.

For example, consider the formula

$$\phi(x) = \sqrt{x-2} + \sqrt{1-x} \qquad \text{(A)}$$

The domain of $y = \sqrt{x-2}$ is $[2, \infty)$ (i.e., $x \geq 2$), while that of $y = \sqrt{1-x}$ is $(-\infty, 1]$ (i.e., $x \leq 1$). *These intervals do not intersect*. Thus, the formula (A) does not define *any* function.

Remark: The above example also tells that if f and g define functions, then $f \pm g$ need not define a function. Following are some more examples of formulas, which do not define functions:

$$
\begin{aligned}
y = F(x) &= \frac{1}{\sqrt{x}} + \frac{1}{\sqrt{-x}} \\
y = G(x) &= \log x + \log(-x) \\
y = f(x) &= \sqrt{\sin x - 2} \\
y = h(x) &= \log(\sin x - 2), \text{and soon}^{(10)}
\end{aligned}
$$

In view of the above, *we must distinguish between a function and a formula*. Of course, *in calculus we shall generally be dealing with functions, which are expressed by formula(s)*. However, it must be remembered that there are certain functions, for which no formula exists. Furthermore, polynomials and rational functions are particular kinds of algebraic functions.[(11)]

In addition to algebraic functions that we have considered in this chapter, we shall also consider *transcendental functions* that are *trigonometric functions* discussed in Chapter 5, *inverse trigonometric functions* discussed in Chapter 14, *exponential and logarithmic functions* discussed in Chapter 13a, and *hyperbolic functions* discussed in Chapter 23.

6.11 EVEN AND ODD FUNCTIONS

We have introduced the notion of *even* and *odd* functions, in Chapter 5. We recall the formal definitions:

(i) A function is an *even function* if for every x in the domain of f

$$f(-x) = f(x)$$

(ii) A function is an *odd function* if for every x in the domain of f

$$f(-x) = -\mathrm{f}(x).$$

Remark: From both the definitions (i) and (ii) above, it is clear that $-x$ is in the domain of f whenever x is.

Note: It will be shown later that we can define *functions that are neither even nor odd.* For detailed discussion about even and odd functions, refer to Chapter 7b on integration (i.e., Part II

(10) *Calculus: Basic Concepts for High School* by L.V. Tarasov (pp. 52–54, English translation), Mir Publishers, Moscow, 1982.

(11) A complicated example of an algebraic function is the one defined by $f(x) = \left(x^3 - 3x^2 + x + 1)^3\right)/\sqrt{x^4 + 5}$.

of this book), wherein we have discussed some *special properties of definite integrals, restricted to even and odd functions.*

6.12 INCREASING AND DECREASING FUNCTIONS

Increase and *decrease* of a function are important characteristics of the behavior of a function.

Definition: A function $y = f(x)$ is said to be *increasing* on an interval *if to greater values of the argument* x belonging to that interval *there correspond greater values of the function*. Similarly, $f(x)$ is called *decreasing if to greater values* of the argument there correspond *smaller values* of the function.

If the graph of a function is traced from left to right (*this corresponds to the increase of the argument* x), then for an increasing function the *moving point* of the graph *goes upward* (*relative to the positive direction of OY)*, and for a decreasing function it moves downward.

6.12.1

The increase and the decrease of a function can be interpreted *in a broader sense*.

A function $f(x)$ is called *nondecreasing* on an interval $[a, b]$ if for any $x_1, x_2 \in [a, b]$, the condition $x_1 < x_2$ implies the *nonstrict inequality*

$$f(x_1) \leq \mathrm{f}(x_2)$$

Similarly, if $x_1 < x_2$ implies $f(x_1) \geq f(x_2)$, the function is said to be *nonincreasing*.

This type of *increase or decrease in the broad sense* is most often a characteristic of functions having *different analytic expressions on different intervals*. The definitions of *nondecreasing* (*nonincreasing*) functions cover bigger classes of functions than those of *increasing* (decreasing) functions (see Figures 6.5–6.10). Most often, they are defined by two or more different analytic expressions on different intervals. Note that such functions may stay constant on a subinterval, while on the remaining ones they must either be

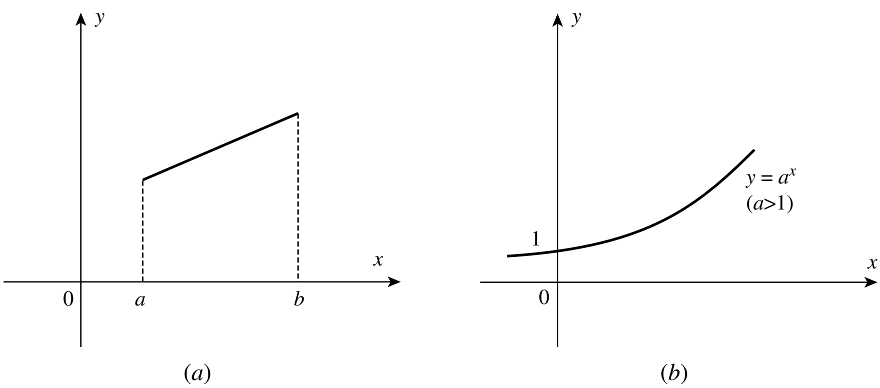


FIGURE 6.5 Graphs of increasing functions.

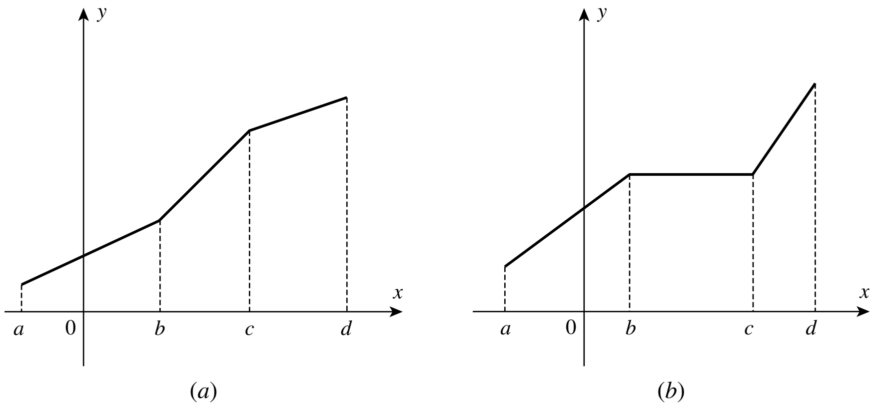


FIGURE 6.6 Graphs of nondecreasing functions.

increasing (see Figure 6.6b) or be decreasing (see Figure 6.8b). The following graphs of functions must clarify the distinction not only between the *increasing* and the *nondecreasing* functions but also between *decreasing* and *nonincreasing* functions.

Note that a nondecreasing function may be an increasing function (see Figure 6.6a), but the converse is not true (see Figure 6.6b). Similarly, a nonincreasing function may be a decreasing function (see Figure 6.8a), but the converse is not true (see Figure 6.8b). For obvious reasons, a nondecreasing function may be looked upon as an *increasing function* and similarly a nonincreasing function is considered a *decreasing function*.[12]

It is usually possible to break up the interval (on which a function is considered) into a number of subintervals on each of which the function is *either increasing or decreasing*. At times we use the terms *strictly increasing* (*or strictly decreasing*) function to mean increasing (or decreasing) function.

In Figure 6.9, we give the graphs of (strictly) increasing and (strictly) decreasing functions.

6.12.2 Monotonic Function

A function $f(x)$ is said to be monotonic on $[a, b]$ if $f(x)$ is *only nondecreasing*, in particular *increasing* on $[a, b]$, or *only nonincreasing*, in particular *decreasing* on $[a, b]$.

6.12.3 Strictly Monotonic Function

A function which is *either increasing or decreasing* on an interval will be called *strictly monotonic function* on that interval.

6.12.4 A Function Neither Increasing nor Decreasing

It is possible that a function is *neither increasing nor decreasing on a given interval*. For instance, see Figure 6.11 where the graph of a function defined on the interval $[a, b]$ is shown.

[12] Note that the question of using the terms nondecreasing and nonincreasing functions arises only if a function is defined on two or more subintervals, with different analytic expressions.

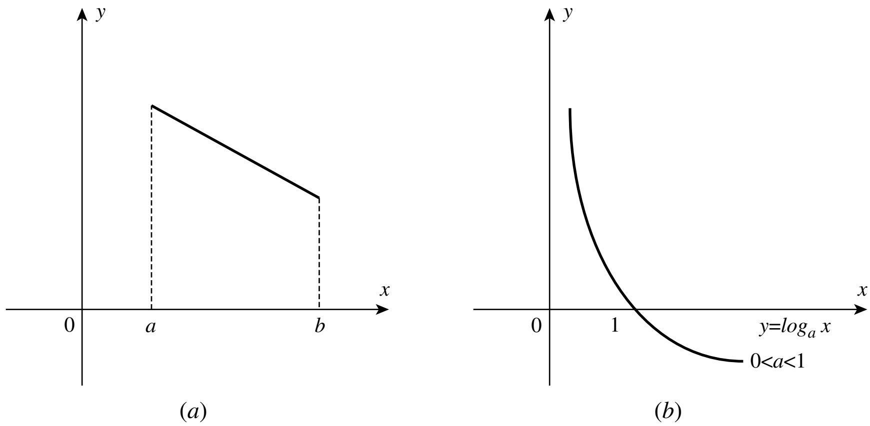


FIGURE 6.7 Graphs of decreasing functions.

This interval is split into the intervals $[a, x_1]$, $[x_1, x_2]$, $[x_2, x_3]$, and $[x_3, b]$, on which, respectively, the function decreases, increases, decreases, and increases. If a function *f* increases and decreases on different subintervals of its domain *I*, we say that the function is *neither increasing nor decreasing* on *I*.

Note: Later on, we will *find out the intervals on which a function is only increasing (or only decreasing)*. In Chapter 19a, *we shall use the properties of derivatives to find such intervals and study certain local properties of functions*. Furthermore, we will be able to investigate functions for maximum/minimum values of functions in Chapter 19b.

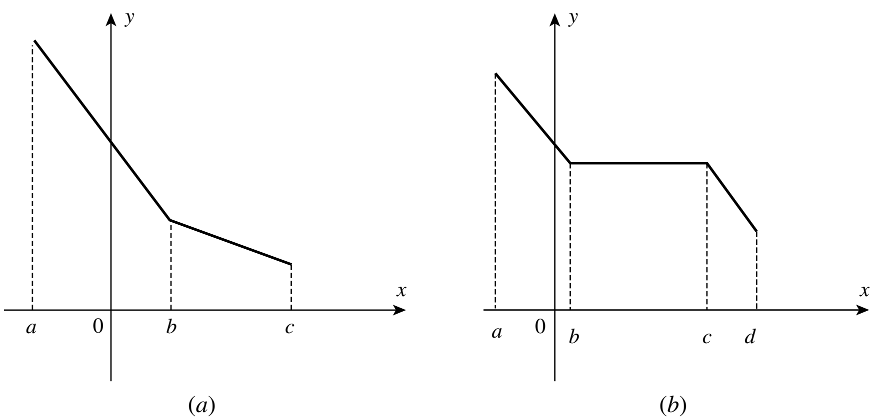


FIGURE 6.8 Graphs of nonincreasing functions.

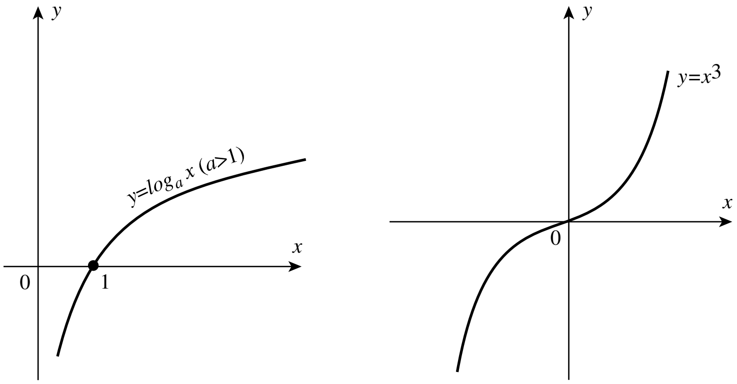


FIGURE 6.9 Graphs of (strictly) increasing functions.

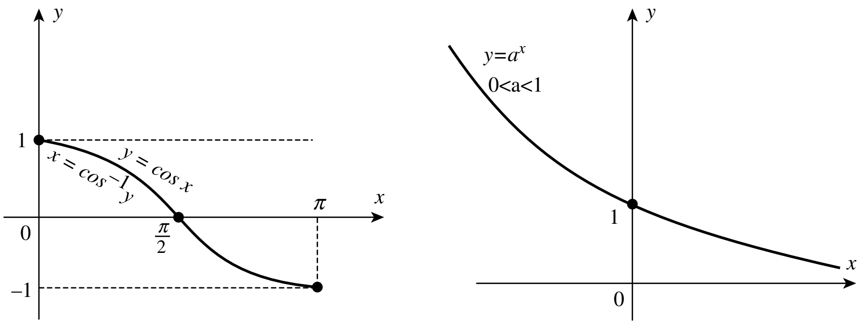


FIGURE 6.10 Graphs of (strictly) decreasing functions.

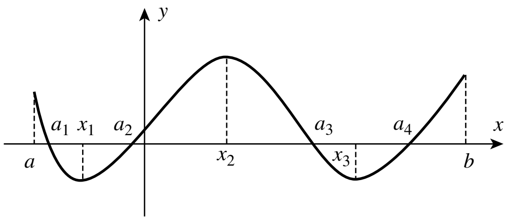


FIGURE 6.11 Graph of a function which is neither increasing nor decreasing.

6.13 ELEMENTARY AND NONELEMENTARY FUNCTIONS

First, we talk about the basic elementary functions by which we mean the following analytically represented functions:

(i) *Power Function*: $y = x^\alpha$, where α is a real number.

If α is irrational, this function is evaluated by taking logarithms and antilogarithms: $\log_{10} y = \propto \log_{10} x$. It is assumed that $x > 0$.

(For more details, refer to the definition of logarithm using *integral calculus* discussed in Chapter 6b of Part II)

(ii) *Exponential Function*: $y = a^x$, where $a > 0$ and $a \neq 1$.

(iii) *Logarithmic Function*: $y = \log_a x$, $a > 0$ and $a \neq 1$.
Throughout this book, the base of the logarithm will be either 10 or e, depending on the requirement of the problem. In case no base is indicated, the symbol log will stand for the logarithm to the base "e". (For a detailed discussion about the logarithmic function, refer to Chapters 12 and 13a.)

(iv) *Trigonometric Functions*: $y = \sin x$, $y = \cos x$, $y = \tan x$, $y = \cot x$, $y = \sec x$, and $y = \operatorname{cosec} x$.

(v) *Inverse Trigonometric Functions*: $y = \sin^{-1}x$, $y = \cos^{-1}x$, $y = \tan^{-1}x$, $y = \cot^{-1}x$, $y = \sec^{-1}x$, and $y = \operatorname{cosec}^{-1}x$.

Elementary Functions: Elementary functions are those that are represented analytically. In general, it is represented by a *single formula* of the type $y = f(x)$, where the expression on the right-hand side is made up *of basic elementary functions* and constants *by means of finite number of operations* of addition, subtraction, multiplication, division, and taking function of a function.

(An elementary function may also be represented by two formulas. The important point to be emphasized is that elementary functions are represented analytically.)

Examples of Elementary Functions:

$$y = |x| = \sqrt{x^2},\ y = \sqrt{1 + 4\sin^2 x},\ y = \frac{\log x + 4\sqrt[3]{x} + 2\tan x}{10^x - x + 10}, \text{ and so on.}$$

Examples of Nonelementary functions:

(a) The Dirichlet function defined on the whole real line, is *not an elementary function*. It is defined as follows:

$$y = 1 \text{ if } x \text{ is rational and } 0 \text{ if } x \text{ is irrational.}$$

(Note that this function is defined in terms of a property of real numbers and not in the form $y = f(x)$. Thus, it is not represented analytically.)

(b) The function $y = 1, 2, 3, \ldots, n$ $[y = f(n)]$ is not elementary because the number of operations that must be performed to obtain y increases with "n". Thus, *the number of operations is not finite*.

7a The Concept of Limit of a Function

7a.1 INTRODUCTION

Addition, subtraction, multiplication, division, raising to a power, extracting a root, taking a logarithm, or a modulus are operations of elementary mathematics. In order to pass from elementary mathematics to higher mathematics, we must add to this list one more mathematical operation, namely, "*finding the limit of a function*".

The notion of limit is an important *new idea* that lies at the foundation of Calculus. In fact, we might define *Calculus* as the study of limits. It is, therefore, important that we have a deep understanding of this concept. Although the topic of limit is rather theoretical in nature, we shall try to represent it in a very simple and concrete way.

7a.2 USEFUL NOTATIONS

Our work for understanding the concept of limit will be simplified if we use certain notations. Therefore, let us first get familiar with these notations:

- *Meaning of the notation* $x \to a$:
  Let x be a variable and "a" be a constant. If x assumes *values nearer and nearer to* "a" (without assuming the value "a" itself), then we say x *tends to* a (or x approaches a) and we write $x \to a$. In other words, the procedure of giving values to x *(from the domain of "f")* nearer and nearer to "a", *but not permitting x to assume the value* "a", is denoted by the symbol "$x \to a$".

  Thus, $x \to 1$ means, we assign values to x which are nearer and nearer to 1 (but not permitting x to assume the value 1), which means that x comes closer and closer to "1", reducing the distance between "x" and "1", in the process.

  Thus, by the statement "x" tends to "a", we mean that:

  (i) $x \neq a$,

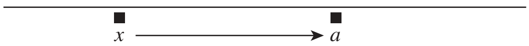
  (ii) x assumes values nearer and nearer to a, and

  (iii) *The way in which x should approach a is not specified.*
  (Different ways of approaching "a" are given below.)

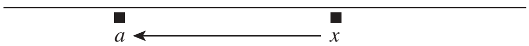
What must you know to learn calculus? 7a-The concept of limit of a function, development of epsilon (ε), delta (δ) definition of limit and its applications. Algebra of limits (limit theorems) and one-sided limits.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

- *Meaning of* $x \to a^-$
  If we consider x to be approaching closer and closer to "a" *from the left side* (i.e., through the values less than "a"), then we denote this procedure by writing $x \to a^-$ and read it as "x" *tends to "a minus"*.



- *Meaning of* $x \to a^+$
  If we consider x approaching closer and closer to "a" *through the values greater than* "a" (i.e., x approaching "a" *from the right side*), then this procedure is denoted by writing $x \to a^+$ and we read it as "x" tends to "a plus".



Example (1): Consider the function $F(x) = 3x + 5$, $x \in (2, 3) \cup (3, 5]$.
Note the following points:

(i) "4" is in the domain of F, and it can be approached from both the sides. Therefore, we can write $x \to 4$.

(ii) "5" is in the domain of F, but x can approach "5", only from the left of 5 (i.e., through values of $x < 5$).

(iii) "2" is *not* in the domain of F, but x can approach "2", *from the right of* "2" (i.e., through values of $x > 2$). Thus, in this case, *it is meaningful to write* $x \to 2^+$, but we cannot write $x \to 2^-$ or $x \to 2$.[(1)]

(iv) "3" is *not* in the domain of F, but x can approach "3" *from both the sides of* "3". Thus, we can write $x \to 3^+$ and $x \to 3^-$ or $x \to 3$.[(1)]

7a.2.1 What Happens When "x" Approaches "a"?

We know that the distance between "x" and "a" is denoted by $|x - a|$. Thus, as x tends to "a", $|x - a|$ becomes smaller and smaller for values of "x" nearer and nearer to "a". Mathematically, we say that *for an arbitrary small positive number* δ*, the absolute number* $|x - a|$ *can be made less than* δ*, if the number* x *is chosen nearer and nearer to* "a".

We write $x \to a \Rightarrow |x - a| < \delta$, for an *arbitrary small* $\delta > 0$. But, we also want that x *should never attain the value* "a" (i.e., $x \neq a$). This is expressed by the inequality $0 < |x - a|$. We can, therefore, combine these two inequalities and write $0 < |x - a| < \delta$, to mean $x \to a$.

In other words, $x \to a$ means

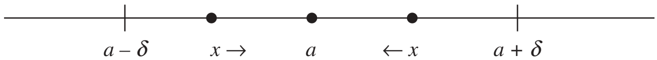
$$0 < |x - a| < \delta, \text{for an arbitrary small}\, \delta > 0 \qquad (1)^{(2)}$$

[(1)] Note the conditions under which "x" can approach "a", even when "a" does not belong to the domain of the function.
[(2)] This statement is true irrespective of whether "x" approaches "a" from one side or from both the sides.

Notes:

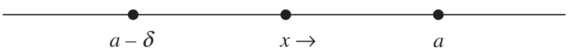
(1) The variable "x" may approach the fixed number "a" from either side (or both the sides, simultaneously). This approach may be along all the points of an interval (on either side) or by jumping on certain points, which are closer and closer to "a".[(3)]

(2) If x can approach "a" *from both sides*, then the statement (1) tells us that, for an arbitrary small $\delta > 0$, *x always belongs to the deleted δ-neighborhood of* "a", that is, $x \in (a-\delta, a+\delta)$, *with $x \neq a$.*

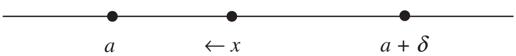


This is equivalent to assigning values to "x", closer and closer to "a" from both sides of "a". (*This procedure is useful for studying the values of a function in the neighborhood of the given point* "a".)

(3) If $x \to a^-$ (i.e., if x approaches "a" from the left) then, statement (1) means that for an arbitrary small $\delta > 0$, x *always* belongs to $(a-\delta, a)$.[(4)]



(4) If $x \to a^+$ (i.e., if x approaches "a" from the right) then, statement (1) means that for an arbitrary small $\delta > 0$, x *always* belongs to $(a, a+\delta)$.[(4)]



7a.3 THE CONCEPT OF LIMIT OF A FUNCTION: INFORMAL DISCUSSION

We know that the value of a function "f" for any given number "a" of its domain is denoted by $f(a)$. However, if "a" is not in the domain of "f", then we say that $f(a)$ does not exist or $f(a)$ is not defined. For example, consider the function

$$f(x) = 5x + 2, \ x \in [0, 2]$$

Note that, the numbers 0, 1, and 2 are in the domain of "f". Here, we have

$$f(0) = 2, f(1) = 7, \quad \text{and} \quad f(2) = 12.$$

Next, consider the function $\phi(x) = 5x + 2$, $x \in (0, 1) \cup (1, 2)$.[(5)]

Observe that 0, 1, and 2 are not in the domain of ϕ. Accordingly $\phi(0)$, $\phi(1)$, and $\phi(2)$ are *not defined.* We ask the following question:

If x is made to assume values closer and closer to 1 (from either side), *how will the value $\phi(x)$ change?* In other words, *to what number is $\phi(x)$ closest to when x is close to* 1*?*

[(3)] Whenever "x" approaches "a" through jumps, we are in effect considering the limit(s) of sequences, which are also functions of a particular type. Here, it may be mentioned that once we have learnt the concept of *limit of a function*, it is simpler to understand the concept of limit of a sequence, which is a function whose domain is the set of natural numbers. Here, we shall not discuss about the limit of a sequence $<a_n>$.

[(4)] Remember that in the one-sided neighborhood of "a", the point "a" itself is not included in the neighborhood (see Chapter 3).

[(5)] Recall that, whenever the domain of a function is changed, we get a new function. Thus, $f_1(x) = 5x + 2$, $x \in [0, 2]$ is different from $f_2(x) = 5x + 2$, $x \in [0, 2)$.

In this case, it is *easy to guess* that if x gets close to 1, then, $\phi(x)=5x+2$ gets close to 7. Similarly, if x gets close to 0, $\phi(x)$ gets close to 2 and if x gets close to 2, then $\phi(x)$ gets close to 12. *We say that the limit of the function* $\phi(x)=5x+2$, *when x approaches the number 1, is 7.*

We express this idea by the notation:

$$\lim_{x\to 1}\phi(x) = \lim_{x\to 1}(5x+2) = 7, \qquad \text{here } x \text{ can approach 1 from both the sides}$$

$$\lim_{x\to 0^+}\phi(x) = \lim_{x\to 0^+}(5x+2) = 2, \qquad \text{here } x \text{ can approach 0 only from the right}$$

$$\lim_{x\to 2^-}\phi(x) = \lim_{x\to 2^-}(5x+2) = 12, \qquad \text{here } x \text{ can approach 2 only from the left}$$

Note that, whereas the function $\phi(x)$ is not defined at the point 0, 1, and 2, yet the limit(s) as indicated above exist.

We agree that our discussion will be restricted to the real valued functions of real variables. *This restricts our choice of functions*. For example, the formula $g(x) = \sqrt{x}$ will be a function only for $x \geq 0$.

Now, it is easy to guess that as x approaches "9", $\sqrt{x}$ approaches 3 and $(\sqrt{x}+13)$ approaches 16. It follows that the reciprocal of $(\sqrt{x}+13)$ should approach 1/16 and $(\sqrt{x}+13)^{1/4}$ must approach 2. *Later on, we will be able to show that all our guesses are correct*.

Remark: In connection with limit of the function ϕ, we have considered only those points that are *not* in the domain of ϕ. However, *the concept of limit is equally applicable to the points (numbers), which are in the domain of* ϕ. For example, 1/5 is in the domain of ϕ (and it can be approached from either side), hence we can say that as x approaches the number 1/5 (from either side), the function ϕ approaches 3.

We write,

$$\lim_{x\to 1/5}\phi(x) = \lim_{x\to 1/5}(5x+2) = 3$$

Similarly,

$$\lim_{x\to \sqrt{2}}\phi(x) = \lim_{x\to \sqrt{2}}(5x+2) = 5\sqrt{2}+2$$

The point we are making here is that *the following two questions are different.*

(i) What is the value of $\phi(1)$?

(ii) What is the number which $\phi(x)$ is close to, when x is close to 1?

Note that, whereas $\phi(1)$ does not exist, the $\lim_{x\to 1}\phi(x)$ exists and it is the number 7.

The idea of limit indicated in (ii) above will be found useful when we compute the limit of the type $\lim_{x\to a}\left((x^2-a^2)/(x-a)\right)$. In fact, it is due to this type of function that we can understand the concept of the limit in a better way.

Remark: To be able to find the limit of a function at any point "a", (which may or may not be in its domain) it is necessary that there exists some neighborhood of "a" in which "f" is defined, except possibly at "a". This is necessary, since only then can x approach "a".

7a.4 INTUITIVE MEANING OF LIMIT OF A FUNCTION

Let $f(x)$ be a function. If x assumes values nearer and nearer to the number "a" except possibly the value "a" and $f(x)$ assumes the values nearer and nearer to l, which is a finite real number, then we say that $f(x)$ tends to the limit l as x tends to a, and we write $\lim_{x \to a} f(x) = l$.

Notice that we do not insist anything to be true at "a". The function f need not even be defined at "a". Since "a" may be approached from both the sides of a (i.e., left side and right side of a) *when we say that* $\lim_{x \to a} f(x) = l$, *we really mean to say that* $\lim_{x \to a^-} f(x) = l = \lim_{x \to a^+} f(x)$. *If these conditions are not satisfied simultaneously*, we say that $\lim_{x \to a} f(x)$ *does not exist.*[6]

The following examples will clarify the situation.

Example (2): Consider $f_1(x) = \dfrac{x^2 - 4}{x - 2}, \quad x \neq 2$

Observe that, here $f_1(x)$ is not defined for $x = 2$. Further, since $(x^2 - 4)$ and $(x - 2)$ both approach "0" as x approaches 2, it follows that limit of the quotient function is of the form 0/0, *which is not defined.* Therefore, *it is not possible to* compute $\lim_{x \to 2} \big((x^2 - 4)/(x - 2)\big)$. We, therefore, use an indirect method as explained below.

We have seen that $f_1(2)$ is *not defined.* However, since f_1 is defined for all other values of x, *there is no objection in computing the values of f_1 at all other points.* We, therefore, *study the values of f_1 when x is considered very close to the number 2.*

For this purpose, we prepare the following calculations, by choosing successive values of x from *a small neighborhood of 2* (say 0.1 neighborhood of 2) and compute corresponding values $f_1(x)$.

This involves the following calculations:

$$x \quad x^2 \quad x^2 - 4 \quad x - 2 \quad f_1(x) = \frac{x^2 - 4}{x - 2}$$

From the above calculations, we get the data of our interest, which is given in Table 7a.1.

Observe that *as x approaches 2, $f_1(x)$ takes up values closer and closer to* 4. We, therefore, say that the limit of $f_1(x)$, as x approaches 2, is 4. In symbols, we write $\lim_{x \to 2} f_1(x) = 4$.

Note that the preparation of Table 7a.1 is time consuming and tedious. On the other hand, a logical way of thinking (which is explained below) is found to be useful and simpler in evaluating the limits. We have

$$f_1(x) = \frac{x^2 - 4}{x - 2}, \quad x \neq 2 = \frac{(x - 2)(x + 2)}{(x - 2)}, \quad x \neq 2 \tag{2}$$

Note that, if $(x - 2) \neq 0$, (i.e., if $x \neq 2$) then we can cancel the factor $(x - 2)$ from the numerator and the denominator of the above expression on the right-hand side of Equation (2), and get,

$$f_1(x) = (x + 2), \quad x \neq 2 \tag{3}$$

[6] In other words, if $\lim_{x \to a^-} f(x)$ is different from $\lim_{x \to a^+} f(x)$ then we will say that $\lim_{x \to a} f(x)$ does not exist.

TABLE 7a.1

x	$f_1(x)$
1.91	3.91
1.92	3.92
1.96	3.96
1.99	3.99
1.997	3.997
1.9998	3.9998
1.999998	3.999998
1.99999999	3.99999999
2	Not defined
2.00000001	4.00000001
2.0000001	4.0000001
2.000001	4.000001
2.00001	4.00001
2.0001	4.0001
2.001	4.001
2.01	4.01
2.02	4.02

Thus, we have two Equations (2) and (3), *both representing* the same *function* $f_1(x)$, *when* $x \neq 2$. We may choose any of them for computing the limit of the function in question. Obviously, the Equation (3) is simpler to handle in view of the difficulty observed in connection with the expression $(x^2 - 4)/(x - 2)$, $x \neq 2$, in listing the values of $f_1(x)$ in the neighborhood of 2. Hence, we choose the expression $(x + 2)$ for computing the limit in question. We get

$$\begin{aligned} \lim_{x \to 2} f_1(x) &= \lim_{x \to 2} \frac{x^2 - 4}{x - 2}, \quad x \neq 2 \\ &= \lim_{x \to 2} (x + 2), \quad x \neq 2 \\ &= 2 + 2 = 4 \end{aligned}$$

Note that whereas $f_1(2)$ *does not exist* (since 2 is not in the domain of "f"), $\lim_{x \to 2} f_1(x)$ exists, and it is given by the number 4.

This shows that the existence or nonexistence of the limit of a function at a point does not depend on the existence or nonexistence of the value of the function at that point.

Example (3): Consider $F(x) = x^2 - 5x + 2$, $x \in (0, 3) \cup (3, 5)$

Here, F is *not* defined for $x = 0, 3$, and 5. Therefore, $F(0)$, $F(3)$, and $F(5)$ do not exist. However, the limit(s) of F at 0, 3, and 5 exist. (Of course, limits at 0 and 5 are one-sided limits, to be discussed later.)

Remark: Limit of a function at any point "a" may be considered if and only if it is possible to approach "a" from at least one side. Thus, if "a" is an isolated point in an interval (so that there exists an open interval which contains "a" alone) then limit of a function at "a" cannot be discussed.

In the following example, we observe that $\lim_{x \to a} G(x)$ should not exist even though "a" can be approached from both the sides.

Example (4): Consider $G(x) = \frac{x+2}{x-2}$, $x \neq 2$

Note that this function is defined for all real values of x, except $x = 2$, which is *not an isolated point*. However, the limit $\lim_{x \to 2} ((x+2)/(x-2))$, $x \neq 2$ does not exist.

This is because, as $x \to 2$, the numerator $(x+2)$ approaches the number 4 whereas the denominator approaches the number "0", so that $G(x)$ approaches arbitrary large values and hence not defined. *Whenever such a situation arises, we say that the limit of the function does not exist.*[7]

Further note that

$$\lim_{x \to 1} G(x) = \lim_{x \to 1} \frac{x+2}{x-2} = -3, \quad \text{and} \quad \lim_{x \to 3} G(x) = \lim_{x \to 3} \frac{x+2}{x-2} = 5$$

Remark: To evaluate the limit $\lim_{x \to a} (f(x)/g(x))$, where $f(a) = 0$ and $g(a) = 0$, we cannot put $x = a$, since it produces the expression 0/0, which is not defined. In such cases, we must search for a common factor in $f(x)$ and $g(x)$. If there is a common factor in both $f(x)$ and $g(x)$ whose limit is zero as $x \to a$, then we can reduce the quotient to a simpler form and finally evaluate the limit by using the direct method.[8]

Example (5): Let $F(x) = \begin{cases} \dfrac{x^2-4}{x-2}, & x \neq 2 \\ 6, & x = 2 \end{cases}$, and consider $\lim_{x \to 2} F(x)$.

We know that $\lim_{x \to 2} \left((x^2-4)/(x-2)\right) = 4$ [see Example (2)]. Here, $F(2)$ is defined to be 6. In fact, we may define $F(2)$ to be *any real number, artificially*. Thus, limit of the function $F(x)$ at $x = 2$ and the value $F(2)$ both exist, but they are not equal. This example shows that $\lim_{x \to a} F(x)$ *need not be equal to F(a), even when both exist.*

Next, consider the following example.

Example (6): Let $G(x) = \begin{cases} \dfrac{x^2-4}{x-2}, & x \neq 2 \\ 4, & x = 2 \end{cases}$, and consider $\lim_{x \to 2} G(x)$.

Here, we note that the limit of the function G(x) at x = 2 and the value of the function at x = 2, both exist and each is equal to 4.

(This property will be very useful in the next chapter, where we study the concept of continuity of a function.)

Example (7): Consider $f_2(x) = \begin{cases} x+5, & \text{for } x > 0 \\ x+2, & \text{for } x < 0 \end{cases}$

Observe that $f_2(0)$ is *not defined*. Let us study the values of $f_2(x)$ as $x \to 0$. We note that as $x \to 0^-, f_2(x) \to 2$. On the other hand, as $x \to 0^+, f_2(x) \to 5$. Thus, $\lim_{x \to 0^0} f_2(x) \neq \lim_{x \to 0^+} f_2(x)$. When this happens, *we say that the limit of the function does not exist.*

[7] Remember that "limit of a function" at any point must be a "finite" (real) number. Since $\lim_{x \to 2} \frac{x+2}{x-2}$ approaches infinity (∞), which does not represent a real number, we say that this limit does not exist. Later on in Chapter 7b, we shall introduce infinity as limit of a function.

[8] As regards other algebraic, trigonometric, exponential, and logarithmic functions or their combinations, different methods are available for evaluating their limit(s) in corresponding chapters.

TABLE 7a.2

$x<2$	$f_3(x)$	$x>2$	$f_3(x)$
1.9	2.8	2.1	3.4
1.99	2.98	2.01	3.04
1.999	2.998	2.001	3.004
1.9999	2.9998	2.0001	3.0004
1.9999	2.99998	2.00001	3.00004
As $x \to 2^-$	$f(x) \to 3$	As $x \to 2^+$	$f(x) \to 3$

Example (8):

$$f_3(x) = \begin{cases} 2x - 1, & \text{for } 1 \le x < 2 \\ 4x - 5, & \text{for } 2 < x \le 3 \end{cases}$$

Observe that $f_3(2)$ is not defined. Let us study the values of $f_3(x)$ as $x \to 2$. We prepare Table 7a.2. From Table 7a.2, we observe that $\lim_{x \to 2^2} f_3(x) = 3$, and $\lim_{x \to 2^+} f_3(x) = 3$. *Thus, the left-hand limit of $f_3(x)$ at $x = 2$ is equal to its right-hand limit at $x = 2$.* In this case, we say that the limit of $f_3(x)$ as $x \to 2$ exists, and we write

$$\lim_{x \to 2} f_3(x) = 3$$

The function $f_3(x)$ is really interesting. Moreover, *it is very simple to define any number of such functions.*[9]

Now we consider the following (*more complicated*) functions and the associated difficulties in finding their limit(s). These examples should help us 0 suitably word the definition of *the limit of a function at a point*, covering all possible situations.

Example (9): Let $f_6(x) = \begin{cases} x & \text{if } x < 1 \\ 2 & \text{if } x = 1 \\ x + 2 & \text{if } x > 1 \end{cases}$

Let us consider $\lim_{x \to 1} f_6(x)$ We have the following observations:

(a) As $x \to 1^-, f_6(x) \to 1$ (left-hand limit)
(b) As $x \to 1^+, f_6(x) \to 3$ (right-hand limit)
(c) $f_6(1) = 2$

Thus, $\lim_{x \to 1^2} f_6(x) = 1 \neq \lim_{x \to 1^+} f_6(x) = 3$

Obviously, $\lim_{x \to 1} f_6(x)$ *does not exist.*

(9)
(a) Let $f_4(x) = \begin{cases} 2x + 10, & \text{for } 1 \le x < 3 \\ 7x - 5, & \text{for } 3 < x \le 5 \end{cases}$ (Here, $\lim_{x \to 3} f_4(x) = 16$)

(b) Let $f_5(x) = \begin{cases} 2x^3 - 10, & \text{for } x < 2 \\ 3x^3 - 18, & \text{for } x > 2 \end{cases}$ (Here, $\lim_{x \to 2} f_5(x) = 6$)

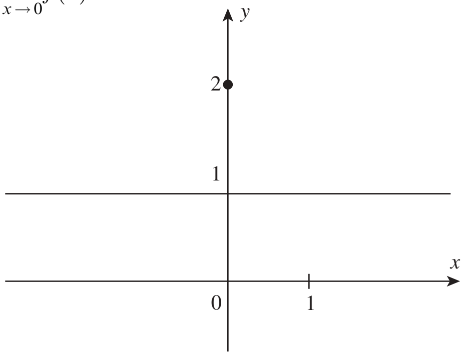
Example (10): Let $f_7(x) = \frac{1}{x-1}$, for all $x \neq 1$

Observe that as $x \to 1^+$ (as x assumes values closer and closer to 1 from the right hand side) $f_7(x)$ *gets larger and larger positive values*. On the other hand, when $x \to 1^-$ (as x assumes values closer and closer to 1 from the left hand side), $f_7(x)$ *gets larger and larger negative values*.

Thus, $\lim_{x \to 1} f_7(x)$ *does not exist*. Here, it may also be noted that $f_7(1)$ *is not defined. In this case, neither the value* $f_7(1)$ *exists nor does the limit of the function, as* $x \to 1$.

Example (11): Consider the function $f(x) = \begin{cases} 1 & \text{for } x \neq 0 \\ 2 & \text{for } x = 0 \end{cases}$

Observe that for all values of x (other than zero), $f(x) = 1$. Since, $\lim_{x \to 0^-} f(x) = 1$ and $\lim_{x \to 0^+} f(x) = 1$, hence $\lim_{x \to 0} f(x) = 1$.

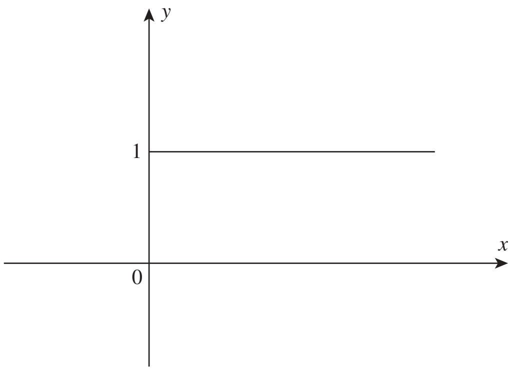


Hiccup Function

Note that though $f(0) = 2$, yet this does not make any difference for the existence of the limit, which is 1. In view of this example, *we would like that the definition of* $\lim_{x \to a} f(x)$ *should be independent of the value f(a).* (This function is known as a "*hiccup function*" due to the appearance of its graph.)

Example (12): Now consider the function defined by

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$$

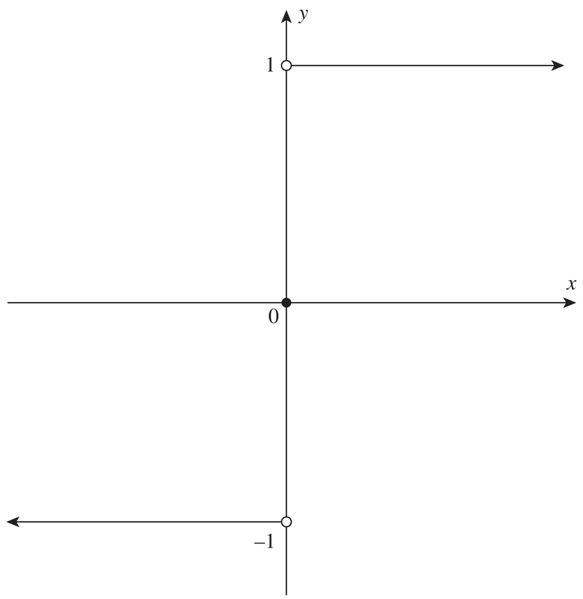


Diving Board function

We ask the question: *Does "f" have a limit as x → 0?* Notice that in any interval about 0, say $(-1/1000,\ 1/1000)$ the function assumes both the values 0 and 1. Observe that $\lim\limits_{x \to 0^-} f(x) = 0$, and $\lim\limits_{x \to 0^+} f(x) = 1$. Here, left-hand limit ≠ right-hand limit, therefore, we conclude that "f" does not have limit. This function is sometimes called "*diving board function*".

Example (13): Consider the graph of the signum function defined by

$$\operatorname{sgn} x = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$



Signum Function

Since, sgn $x = -1$, if $x < 1$ and sgn $x = 1$, if $x > 0$. We have,

$$\lim_{x \to 0^-} \operatorname{sgn} x = \lim_{x \to 0^-} (-1) = -1 \quad \text{and} \quad \lim_{x \to 0^+} \operatorname{sgn} x = \lim_{x \to 0^+} (-1) = 1$$

Because the left-hand limit and the right-hand limit are not equal, the two-sided limit, $\lim\limits_{x \to 0} \operatorname{sgn} x$ does not exist. Hence, we say that, $\lim\limits_{x \to 0} \operatorname{sgn} x$ does not exist.

TABLE 7a.3

x	$\sin x$	$\sin x/x$
−0.10	−0.0998333	0.99833
−0.09	−0.0898785	0.99865
−0.05	−0.0499792	0.99958
−0.03	−0.0299955	0.99985
−0.02	−0.0199987	0.99993
−0.01	−0.00999983	0.999983
0.00	0.00000	?
0.01	0.00999983	0.999983
0.02	0.0199987	0.99993
0.03	0.0299955	0.99985

Example (14): Now, let us evaluate the following limit

$$\lim_{x \to 0} \frac{\sin x}{x}, \quad (x \text{ in radians})^{(10)}$$

Here, *there is no way of canceling terms in the numerator and denominator*. Since $\sin x \to 0$ as $x \to 0$, the quotient $\sin x/x$ might appear to approach 0/0. But, we know that 0/0 is undefined, so if the above limit exists, then we must find it by a different technique. *Since we do not have any other simpler way of rewriting* $\sin x/x$ *to obtain the limit, we use a calculator to find the values of* $\sin x/x$ *for values of x close to 0 and angles x* (in $\sin x$) *in radians.*[(11)]

(Other methods of finding this limit will be discussed later.)

From Table 7a.3, it is obvious that, as $x \to 0$, either from the right or from the left, the value of $\sin x/x$ approaches closer and closer to the number 1. We, therefore, agree to write $\lim_{x\to 0}(\sin x/x) = 1$. This limit is used very often to find the limits of many trigonometric functions (including various functions involving trigonometric functions), and plays a very important role in deriving many useful results. *It must be emphasized that the limiting value of* $\lim_{x\to 0}(\sin x/x)$ *is 1 provided x is measured in radians.* If x is measured in degrees, this limit will be different (and thus, the above does not hold). We will discuss this particular limit (x measured in degrees) later in Chapter 11a.

We have discussed limits informally. In some cases, we were able to deduce limits easily. However, when we tried to ascertain; whether

$$\lim_{x \to 0} \frac{\sin x}{x}$$

exists, we were reduced to calculating $\sin x/x$ for several values of x approaching 0. Using these calculations, we *guessed* that the above limit exists and it should be 1. However, *the uncertainty about this limit leads us to seek a formal definition of limit.*

[(10)] We have so far considered only algebraic functions. The purpose of considering this trigonometric limit is to convey that the concept of limit is applicable to all types of functions.

[(11)] Radian measure of any angle subtended at the center of unit circle equals the length of the circumference, which is taken to have subtended the angle in question. Thus, the measure of an angle "x" radians and the real number "x" representing the length of circumference in the question, both have the same numerical value. In other words, the angle "x" in radians may be looked upon as a real number "x".

7a.4.1 Points of Concern: Formulating the Precise Definition

In formulating the precise definition of $\lim_{x \to a} f(x)$ *we will allow f to be undefined at* "*a*" and ensure that the following requirements are covered in the definition.

(1) Even when *f*(*a*) is not defined (i.e., *f* is not defined at "*a*"), $\lim_{x \to a} f(x)$ may exist.
(2) If "*f*(*a*) happens to be defined at "*a*", *we would like the definition of* $\lim_{x \to a} f(x)$ *to be independent of the value f(a)* [see Examples (5) and (11)].
(3) If $\lim_{x \to a} f(x)$ exists, we would like the *limit to be the same, whether we approach from the left hand side or the right hand side.* For any reason *if the limit is not unique* (i.e., if it is found that left-hand limit $\neq$ right-hand limit) *then we agree to say that the limit does not exist* [see Examples (9), (12), and (13)].

7a.4.2 Rigorous Study of Limits

We gave an informal definition of limit of a function in Section 7a.4.[12]

Here is *a slightly better, reworded definition.*

Definition: To say that $\lim_{x \to a} f(x) = l$, means that the difference between *f(x)* and *l* can be made arbitrarily small (i.e., as small as we please) by demanding that *x* be considered sufficiently close to "*a*", but not exactly "*a*"

We are now ready to formulate a precise *definition of limit.*

7a.4.3 The Formal Definition of Limit

We have said that *l* is the limit of *f*(*x*) as "*x*" approaches "*a*", if *f*(*x*) gets close to *l* as *x* gets close to *a*. But precisely *what does this mean? Does it mean to say that f(x) gets close to l or that x gets close to a*? We begin to answer this question by reinterpreting $\lim_{x \to a} f(x) = l$. We demand that if *x* is considered close to "*a*" (but distinct from "*a*") "*then f(x) must be at least as close to l as we wish*". (This statement is very important.)

In other words, even when *f* is not defined at "*a*", we should be able to obtain the values *f*(*x*) closer and closer to *l* as *x* is assigned values nearer and nearer to "*a*".

In order to put this definition in precise mathematical terms, we shall be using Greek letters ε and δ to stand for *arbitrary positive numbers*. We think of ε and δ as *small positive numbers*, which can be chosen to be as small as we please.[13]

7a.4.4 Making the Definition Precise (ε, δ Definition of Limit)

To say that *f*(*x*) differs from *l* by less than ε is to say that $|f(x) - l| < \varepsilon$. Next, to say that *x is sufficiently close to a, but different from a,* is to say that for some $\delta > 0$, *x* is in the small open interval $(a - \delta, a + \delta)$ with "*a*" deleted. We demand that "*x*" be chosen distinct from "*a*", so that

[12] We reproduce it here, for convenience: Let *f*(*x*) be a function. If "*x*" assumes values nearer and nearer to "*a*", except possibly the value "*a*", *f*(*x*) assumes values nearer and nearer to some finite number *l* then we say that *f*(*x*) tends to the limit *l* as "*x*" tends to "*a*" and we write $\lim_{x \to a} f(x) = l$.

[13] At this point, it must be clearly understood that the arbitrary positive numbers ε and δ are not to be confused with variables. It is because of their arbitrary nature that we can choose their numerical values as per our requirements.

the value of "f" at "a" (if it exists there), *has no influence on the existence or value of the limit*. The best way to say this is to write $0 < |x - a| < \delta$.[14]

We are now in a position to give the ε, δ *definition of limit*.

7a.4.4.1 **ε, δ Definition of Limit** Let f be a function defined at every number of some open interval containing "a", *except possibly at the number* "a" *itself*. We say that *the limit of f(x) as* "x" *approaches* "a" *is l*, if the following statement is true.

For *every number* $\varepsilon > 0$, *there exists a number* $\delta > 0$, such that

$$\text{if } 0 < |x - a| < \delta \quad \text{then} \quad |f(x) - l| < \varepsilon \tag{4}$$

It is to be emphasized that the number ε is chosen first and then the number δ has to be produced. Once we have chosen $\varepsilon > 0$, we must search for a number $\delta > 0$ to ensure that if x is in the interval $(a - \delta, a + \delta)$, with $x \neq a$, then the distance between $f(x)$ and l is less than ε. *If for every* $\varepsilon > 0$, it is possible to get *a corresponding* $\delta > 0$, such that the condition (4) is satisfied, then we say that *the limit at* "a" *exists*, or that f has a limit at "a" or that the limit $\lim_{x \to a} f(x)$ exists.

Using the above ε, δ definition, it can be easily proved that a function, can have *at most one limit at* "a" (we do not prove it here). This justifies calling it *"the" limit* (and not "a" limit) of "f" at "a".

7a.4.5 Geometric Interpretation of the Definition

It is useful to understand carefully the following *geometric interpretation of the definition of the limit of a function f*. Figure 7a.1 shows a portion of the graph of f near the point where $x = a$.

Because f is not necessarily defined at a, there need be no point on the graph with abscissa a. Observe that if x on the horizontal axis, lies between $a - \delta_1$ and $a + \delta_1$, then $f(x)$ on the vertical axis will lie between $l - \varepsilon_1$ and $l + \varepsilon_1$. In other words, by restricting x (on the horizontal axis) to lie between $a - \delta_1$ and $a + \delta_1$, $f(x)$ on the vertical axis can be restricted to lie between $l - \varepsilon_1$ and $l + \varepsilon_1$.

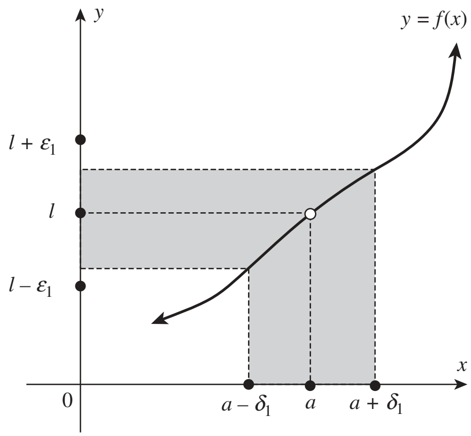


FIGURE 7a.1

[14] Note that $|x - a| < \delta$ describes that $a - \delta < x < a + \delta$, while $0 < |x - a|$ tells that $x \neq a$.

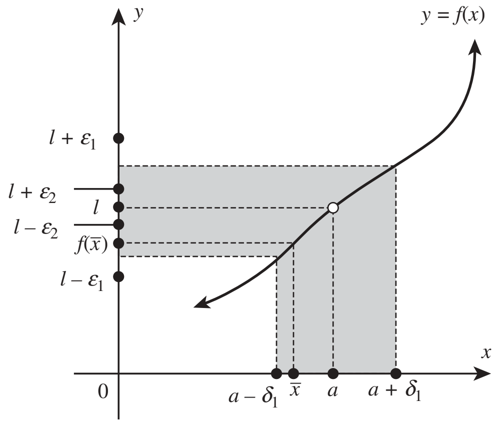


FIGURE 7a.2

Thus, if $0 < |x-a| < \delta_1$ then $|f(x) - l| < \varepsilon_1$.

[In Figure 7a.1, observe that the function values (on the vertical axis), lie *well within the interval* $(l - \varepsilon_1, l + \varepsilon_1)$.]

If smaller value of ε *is chosen, then it can require a different choice for* δ. In Figure 7a.2, it is seen that for $\varepsilon_2 < \varepsilon_1$, the δ_1 value does not serve the purpose since it is too large; so that, there are values of x (like $\bar{x}$) in the open interval $(a - \delta_1, a + \delta_1)$, for which $0 < |\bar{x} - a| < \delta_1$, but $|f(\bar{x}) - l| > \varepsilon_2$.[15]

So we must choose a smaller value δ_2 as shown in Figure 7a.3, such that if $0 < |x - a| < \delta_2$ then $|f(x) - l| < \varepsilon_2$.

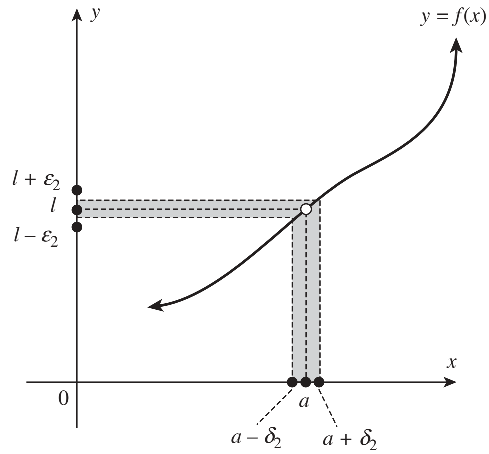


FIGURE 7a.3

[15] In other words, if we choose $\varepsilon_2 < \varepsilon_1$, then to restrict the value of $f(x)$ to lie between $l - \varepsilon_2$ and $l + \varepsilon_2$, we must search for a positive number δ_2, so that whenever x lies between $a - \delta_2$ and $a + \delta_2$, $f(x)$ lies between $l - \varepsilon_2$ and $l + \varepsilon_2$. In general, the closer $f(x)$ is to be to l the nearer x must be to a.

Note (5): Some people call the above ε, δ definition as the most important definition in calculus. Use of symbols "ε" and "δ" in the definition make it look abstract. But the one who has gone through the process of developing this definition, appreciates its wording and the roles of the variables "ε" and "δ". To be able to prove something requires that we should be very clear about the meaning of the words we are using. This is especially true for the word "limit", because all of *Calculus* rests on the meaning of this word.

If "l" is the limit of f as "x" approaches "a", then we write

$$\lim_{x \to a} f(x) = l$$

(Note that, in the definition of limit, nothing is mentioned about the function value at $x = a$.)

Remark: The ε, δ definition of limit does not give any method for evaluating $\lim_{x \to a} f(x)$. It can be used only to verify, whether a given number (or a guessed number) "l" is the limit of the function or not as $x \to a$.

7a.5 TESTING THE DEFINITION [APPLICATIONS OF THE ε, δ DEFINITION OF LIMIT]

It is desirable to test the ε, δ definition against familiar examples to see whether it gives results consistent with our past experience. For instance, our experience tells us that as $x \to 4$, $3x \to 3(4) = 12$, and $(3x - 7) \to 3(4) - 7 = 5$.

Now we give the following examples to show our ε, δ definition gives the kinds of results we want.

Example (15): Use the epsilon, delta definition to prove that $\lim_{x \to 4} (3x - 7) = 5$.

Solution: The first requirement of our definition is that $(3x - 7)$ be defined at every number in some open interval containing 4 except possibly at 4. Here, since $(3x - 7)$ is defined for all real numbers, any open interval containing 4 will satisfy this requirement.

Now, we must show that for any $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$0 < |x - 4| < \delta \quad \text{then } |(3x - 7) - 5| < \varepsilon \tag{5}$$

$$\Leftrightarrow 0 < |x - 4| < \delta \quad \text{then } |(3x - 7) - 5| < \varepsilon$$

$$\Leftrightarrow 0 < |x - 4| < \delta \quad \text{then } 3|x - 4| < \varepsilon$$

$$\Leftrightarrow 0 < |x - 4| < \delta \quad \text{then } |x - 4| < \frac{1}{3}\varepsilon$$

This statement indicates that $(1/3)\varepsilon$ is a satisfactory δ. With this choice of δ, we have the following argument:

$$\begin{aligned} & 0 < |x - 4| < \delta \\ \Rightarrow\ & 3|x - 4| < 3\delta \\ \Rightarrow\ & |3x - 12| < 3\delta \\ \Rightarrow\ & |(3x - 7) - 5| < 3\delta \\ \Rightarrow\ & |(3x - 7) - 5| < \varepsilon \left[\because\ \delta = \frac{1}{3}\varepsilon \quad \therefore \quad 3\delta = \varepsilon \right] \end{aligned}$$

We have, therefore, established that if $\delta = (1/3)\varepsilon$, statement (5) (based on the definition) holds. This proves, that $\lim_{x \to 4}(3x - 7) = 5$.

Again, we discuss the above limit in a slightly different way. We begin with what we call a *preliminary analysis*. It is *not* part of the proof. It is the kind of work, which may be treated as rough work. We include it here, so that our proof will look more logical, systematic and convenient.

To prove that $\lim_{x \to 4}(3x - 7) = 5$[16]

Preliminary Analysis

Let ε be *any* positive number. We must produce a $\delta > 0$, such that $0 < |x - 4| < \delta \Rightarrow |(3x - 7) - 5| < \varepsilon$, where ε is *any arbitrary small positive number*, that we may like to choose.

For this purpose, consider the inequality,

$$\begin{aligned} |(3x - 7) - 5| < \varepsilon &\Leftrightarrow |3x - 12| < \varepsilon \\ &\Leftrightarrow |3(x - 4)| < \varepsilon \\ &\Leftrightarrow |3||x - 4| < \varepsilon \\ &\Leftrightarrow |x - 4| < \frac{\varepsilon}{3} \end{aligned}$$

This suggests the way for choosing δ. Of course, any smaller δ (for example, $\delta = \varepsilon/4$, etc.) would work. Now, we proceed to give the *Formal Proof.*

Formal Proof:

To show that, $\lim_{x \to 4}(3x - 7) = 5$.

Consider,

$$\begin{aligned} &|(3x - 7) - 5| \\ &= |3x - 12| \\ &= |3(x - 4)| \\ &= 3|x - 4| \end{aligned}$$

We know that

$$x \to 4 \Leftrightarrow 0 < |x - 4| < \delta, \quad \text{for any } \delta > 0.$$

Let $\varepsilon > 0$ be given. We choose $\delta = \varepsilon/3$, based on our preliminary analysis.

Now, $0 < |x - 4| < \delta$ means

$$0 < |x - 4| < \frac{\varepsilon}{3} \quad [\because\ \delta = \varepsilon/3]$$

$$\Rightarrow 0 < 3|x - 4| < \varepsilon$$
$$\Rightarrow 0 < |(3x - 7) - 5| < \varepsilon$$

[$\because 3|x - 4|$ is the simplified expression of $|(3x - 7) - 5|$, Thus, for any $\varepsilon > 0$, it is possible to produce $\delta > 0$ (here $\delta = \varepsilon/3$) such that $0 < |x - 4| < \delta \Rightarrow |(3x - 7) - 5| < \varepsilon$

$$\Rightarrow \lim_{x \to 4}(3x - 7) = 5 \quad \text{(Proved)}$$

Note that, here, δ depends on ε (i.e., $\delta = \varepsilon/3$) and this may be the situation in general, however, this may not always be the case [see Example (4)].

[16] Note that, we are not asked to evaluate the limit $\lim_{x \to 4}(3x - 7)$. Also, observe that this limit is given to be 5, and we have to prove that this statement is true.

Example (16): Prove that $\lim_{x\to 2} \frac{2x^2-3x-2}{x-2} = 5,\ \ x \neq 2$

Preliminary Analysis
We are looking for δ such that $0 < |x-2| < \delta \Rightarrow \left|\frac{2x^2-3x-2}{x-2} - 5\right| < \varepsilon$

Now for $x \neq 2$,

$$\begin{aligned}\left|\frac{2x^2-3x-2}{x-2} - 5\right| < \varepsilon &\Leftrightarrow \left|\frac{(2x+1)(x-2)}{(x-2)} - 5\right| < \varepsilon \\ &\Leftrightarrow |(2x+1)-5| < \varepsilon \\ &\Leftrightarrow |2(x-2)| < \varepsilon \\ &\Leftrightarrow |2||x-2| < \varepsilon \\ &\Leftrightarrow |x-2| < \frac{\varepsilon}{2}\end{aligned}$$

This indicates that $\delta = \varepsilon/2$ will work.

Formal Proof: To show that $\lim_{x\to 2} \frac{2x^2-3x-2}{x-2} = 5$
Consider,

$$\begin{aligned}\left|\frac{2x^2-3x-2}{x-2} - 5\right| &= \left|\frac{(2x+1)(x-2)}{(x-2)} - 5\right| \\ &= |(2x+1)-5| = |2(x-2)| \\ &= 2|x-2|\end{aligned}$$

Let $\varepsilon > 0$ be given. We have to search for a δ such that

$$\text{if } 0 < |x-2| < \delta \quad \text{then} \left|\frac{2x^2-3x-2}{x-2} - 5\right| < \varepsilon$$

We know that $x \to 2 \Leftrightarrow 0 < |x-2| < \delta$ for every $\delta > 0$

We choose $\delta = \varepsilon/2$, then

$$\begin{aligned}&0 < |x-2| < \varepsilon/2 \quad [\because\ \delta = \varepsilon/2] \\ &\Rightarrow 2|x-2| < \varepsilon \quad \text{or} \left|\frac{2x^2-3x-2}{x-2} - 5\right| < \varepsilon\end{aligned}$$

It follows that $\lim_{x\to 2} \dfrac{2x^2-3x-2}{x-2} = 5$ (Proved)

Note (6): The cancellation of the factor $(x-2)$ is legitimate because $0 < |x-2| \Rightarrow x \neq 2$. Thus, division by 0 is avoided.

7a.5.1 Simpler and Powerful Rules for Finding Limits (Algebra of Limits)

Limits are extremely important throughout Calculus. Most readers will agree that proving the existence of limit using ε, δ definition is both time consuming and difficult. Also, up to this point, we do not have any general method that can be applied to any function to find its limit at a given point "a".

Of course, as a general method, we can prepare a table listing values of x, closer and closer to "a", and the corresponding values $f(x)$. Such a table may help us *guess a number* to which $f(x)$ approaches, suggesting the limit of f, as $x \to a$. *Once such a number (say "l") is guessed, the ε, δ definition can be used to check whether l is the limit of "f" or not*. However, such a process of finding the values of "f" as $x \to a$ is generally very tedious, as we have seen in the case of computing

$$\lim_{x \to 0} \frac{\sin x}{x} = 1$$

Yet, it is useful to have some experience in computing limit(s) by the above process, in very simple cases.(17)

Fortunately, such a procedure will usually not be necessary because simpler and powerful rules for finding limits are available and we shall discuss about them shortly in Section 7a.5.2.

Now, we shall verify the following two basic limits using ε, δ definition.

- $\lim_{x \to a} c = c$ and
- $\lim_{x \to a} x = a$

These limits will be treated as *standard results, so that they can be freely used in* evaluating the limits of many other functions.(18)

(17) Here, it may also be mentioned that simply by studying the values of a function, it may not be possible to guess the limit of a function, especially when the given function consists of a combination of functions. For example, consider $\lim_{x \to 0} \left[x^2 - \frac{\cos x}{10{,}000}\right]$. Following the procedure used earlier, we have constructed Table 7a.4, of values for the given function. Table 7a.4 suggests that the desired limit is 0. But, that is wrong. If we recall the graph of $y = \cos x$, we realize that $\cos x$ approaches 1 as x approaches 0. Thus, $\lim_{x \to 0} \left[x^2 - \frac{\cos x}{10{,}000}\right] = 0^2 - \frac{1}{10{,}000} = -\frac{1}{10{,}000}$. This situation will be clearer when we study algebra of limits.

TABLE 7a.4

x	$x^2 - \frac{\cos x}{10{,}000}$
$\pm$	0.99995
± 0.5	0.24991
± 0.1	0.00990
± 0.01	0.000000006
$\downarrow$	$\downarrow$
0	?

(18) Later on, when we have studied the properties of trigonometric, exponential, and logarithmic functions, we will be able to establish some other basic limits like, $\lim_{x \to 0} \frac{\sin x}{x} = 1$, ($x$ in radians), $\lim_{x \to 0} \frac{a^x - 1}{x} = \log_e a$, $\lim_{x \to 0} (1 + x)^{1/x} = e$, etc., which will be treated as standard limits.

Example (17): Show that $\lim_{x \to a} c = c$.

Preliminary Analysis
We write $\lim_{x \to a} c = \lim_{x \to a} f(x)$, where $f(x) = c$.
Let ε be *any* positive number. We must produce a $\delta > 0$, such that

$$0 < |x - a| < \delta \Rightarrow |f(x) - c| < \varepsilon \quad \text{where } f(x) = c.$$

Consider, $|f(x) - c| < \varepsilon \Leftrightarrow |c - c| < \varepsilon$

$$\Leftrightarrow 0 < \varepsilon \quad \textit{which is true for any } \varepsilon > 0.$$

Thus, for a *constant function,* $f(x) = c$, we have, for *any* $\delta > 0$,

$$0 < |x - a| < \delta \Rightarrow |f(x) - c| = |c - c| = 0 < \varepsilon.$$

It follows that, $\lim_{x \to a} c = c$.

Remark: In the case of a constant function, *the (positive) number δ does not depend on the arbitrary positive number ε, since any constant function $f(x) = c$ does not change with x.* In other words, x approaching any number "a" does not have any effect on the limit of a constant function. Accordingly,

$$\lim_{x \to 3} 1 = 1, \ \lim_{x \to -\sqrt{2}} \frac{\pi}{3} = \frac{\pi}{3} \quad \text{and} \quad \lim_{x \to 2} (-\pi) = -\pi$$

Example (18): Show that, $\lim_{x \to a} x = a$

Solution: Here, $f(x) = x$ for all x.
Let $\varepsilon > 0$, be an arbitrary number.
We must find a number $\delta > 0$, such that,

$$0 < |x - a| < \delta \Rightarrow |x - a| < \varepsilon$$

Consider, $|f(x) - a| = |x - a|$.
In this case, for $|x - a| < \varepsilon$, we can choose $\delta = \varepsilon$, so that we can write $0 < |x - a| < \varepsilon \Rightarrow |x - a| < \varepsilon$, that is, $0 < |x - a| < \delta \Rightarrow |x - a| < \varepsilon$ [by putting $\varepsilon = \delta$ on the left-hand side].
From the above statement, we conclude that $\lim_{x \to a} x = a$. In view of the above, we can write

$$\lim_{x \to 4\pi} x = 4\pi \quad \text{and} \quad \lim_{x \to -\sqrt{11}} x = -\sqrt{11}$$

Remark: A slight alteration in the situation would show that for any fixed numbers a, b, and c.

$$\lim_{x \to a} (bx + c) = ba + c, \quad \lim_{x \to a} |x| = |a|$$

It follows that

$$\lim_{x \to 2} (-5x + 3) = (-5) \cdot (2) + 3 = -7$$

$$\text{and} \quad \lim_{x \to -3} |x| = |-3| = -(-3) = 3$$

7a.5.2 Algebra of Limits [Limits Theorem]

For computing limits, there are methods which are simpler than using the ε, δ definition. In these methods, *we employ theorems* (called *limit theorems*) *whose proofs are based on the* ε, δ *definition*. In fact, these theorems define the *algebra of limits,* and they are useful in finding the limits of various combinations of functions. *We accept these theorems without proof*, which are given below.

7a.5.3 Theorem (A): Main Limit Theorem

Let n be a *positive integer*, k be a *constant*, and f and g be functions, such that

$$\lim_{x \to a} f(x) \quad \text{and} \quad \lim_{x \to a} g(x) \text{ exist, then}$$

$$\lim_{x \to a} [f(x) + g(x)], \quad \lim_{x \to a} kf(x), \quad \lim_{x \to a} [f(x) - g(x)], \quad \text{and} \quad \lim_{x \to a} [f(x) \cdot g(x)] \text{ exist.}$$

Let

$\lim_{x \to a} f(x) = l$ and $\lim_{x \to a} g(x) = m$, then we have the following theorems (rules).

(1) **Sum Rule:**

$$\begin{aligned} \lim_{x \to a} [f(x) \pm g(x)] &= \lim_{x \to a} f(x) \pm \lim_{x \to a} g(x) \\ &= l \pm m \end{aligned}$$

(This rule is applicable for a finite number of functions.)

(2) **Constant Multiple Rule:**

$$\begin{aligned} \lim_{x \to a} kf(x) &= k \lim_{x \to a} f(x), \quad \text{for any constant } k. \\ &= k \cdot l \end{aligned}$$

(3) **Product Rule:**

$$\lim_{x \to a} [f(x) \cdot g(x)] = \lim_{x \to a} f(x) \cdot \lim_{x \to a} g(x) = l \cdot m$$

(4) **Quotient Rule:**

If $\lim_{x \to a} f(x)$ and $\lim_{x \to a} g(x)$ exist and $\lim_{x \to a} g(x) \neq 0$, then, $\lim_{x \to a} \dfrac{f(x)}{g(x)}$ exists and, we have the following rule

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)} = \frac{l}{m}, \quad (m \neq 0)$$

A special case of this rule is the following.

If a is any real number except zero, then

$\lim_{x \to a} \frac{1}{x} = \frac{1}{a}$. Also, $\lim_{x \to a} \frac{1}{f(x)} = \frac{1}{\lim_{x \to a} f(x)}$ provided $\lim_{x \to a} f(x) \neq 0$.[19]

(5) $\lim_{x \to a} [f(x)]^n = \left[\lim_{x \to a} f(x)\right]^n$

(6) $\lim_{x \to a} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x \to a} f(x)}$, provided $\lim_{x \to a} f(x) > 0$, when n is even.

(7) $\lim_{x \to a} f[g(x)] = f\left(\lim_{x \to a} g(x)\right) = f(m)$.

Now, we also include the following two results (to be treated as theorems), which we have already proved above.

(8) $\lim_{x \to a} c = c$

(9) $\lim_{x \to a} x = a$

Since "*limits are real numbers*", any combination of limits must follow the rules for combining real numbers. This should help us remember the above theorems. Remember that we have accepted the above theorem and rules without proof. Hence, one should not bother about their proofs immediately. For the time being, it is more important to see how all these theorems are applied. The proofs may be referred to in any standard book on the subject.[20]

Exercise: Using ε, δ definition, show that

Q. (1) $\lim_{x \to 2} (3x - 2) = 4$

Q. (2) $\lim_{x \to 4} 2\left[\frac{x^2 - 16}{x - 4}\right] = 16, \quad (x \neq 4)$

Q. (3) $\lim_{x \to 1} \left[\frac{x^2 + 2x - 3}{x - 1}\right] = 4, \quad (x \neq 1)$

Q. (4) $\lim_{x \to 5} \left[\frac{x^2 - 2x - 15}{x - 5}\right] = 24, \quad (x \neq 5)$

Q. (5) $\lim_{x \to 3} (x^2 + x - 5) = 7$

Q. (6) Prove that if $a > 0$, $\lim_{x \to a} \sqrt{x} = \sqrt{a}$

Note: Solutions to Q. (5) and Q. (6) are given below in Examples (19) and (20).

[19] We can deduce rule (5) from rule (4).

$$\left[\text{Note that } \frac{f(x)}{g(x)} = f(x) \cdot \frac{1}{g(x)}\right]$$

[20] *Calculus with Analytic Geometry* (Alternate Edition) by Robert Ellis and Denny Gulick, HBJ Publication.

Example (19): Prove that $\lim_{x \to 3} (x^2 + x - 5) = 7$,

Preliminary Analysis

Our task is to find δ such that $0 < |x-3| < \delta \Rightarrow |(x^2+x-5)-7| < \varepsilon$.

Consider, $|(x^2+x-5)-7| = |x^2+x-12| = |(x+4)(x-3)| = |x+4||x-3|$

Since the second factor $x-3$ *can be made as small as we please, it is enough to bound the factor* $|x+4|$ (i.e., to find the maximum value of this factor as $x \to 3$). To do this, we first agree to make $\delta \leq 1$.

Let us see what happens when we choose $\delta \leq 1$.[21]

We have $|x-3 < 1 \Rightarrow -1 < x-3 < 1$

$$\Rightarrow 2 < x < 4$$
$$\Rightarrow 2+4 < x+4 < 4+4$$
$$\Rightarrow 6 < x+4 < 8$$

(When $\delta \leq 1$, the value of $|x-3| \leq 1$ suggests that maximum value of $|x-3|$ can be 1.)

Then, $|x-3| < \delta$ implies

$$\begin{aligned} |x+4| &= |(x-3)+7| \\ &\leq |x-3| + |7| \quad \text{(Triangle Inequality)} \\ &< 1+7 = 8 \end{aligned}$$

This indicates that *if we also take* $\delta \leq \varepsilon/8$, the product $|x+4|\ |x-3|$ will be less than ε.

Formal Proof:

Let $\varepsilon > 0$ be given.

Choose $\delta = \min\{1, \varepsilon/8\}$; *that is choose* δ *to be smaller of 1 and* $\varepsilon/8$.

Then, $0 < |x-3| < \delta$ implies

$$\begin{aligned} |(x^2+x-5)-7| &= |x^2+x-12| \\ &= |x+4||x-3| < 8. \quad \varepsilon/8 = \varepsilon. \end{aligned}$$

Example (20): Prove that if $a > 0$, $\lim_{x \to a} \sqrt{x} = \sqrt{a}$

Preliminary Analysis

(Note that $\sqrt{x}$ is defined only for $x \geq 0$.)

We must find δ such that $0 < |x-a| < \delta \Rightarrow |\sqrt{x} - \sqrt{a}| < \varepsilon$

Consider,
$$\begin{aligned} |\sqrt{x} - \sqrt{a}| &= \left| \frac{(\sqrt{x}-\sqrt{a}) \cdot (\sqrt{x}+\sqrt{a})}{(\sqrt{x}+\sqrt{a})} \right| \\ &= \left| \frac{x-a}{\sqrt{x}+\sqrt{a}} \right| \\ &= \frac{|x-a|}{\sqrt{x}+\sqrt{a}} \leq \frac{|x-a|}{\sqrt{a}} \end{aligned}$$

Now, to make $\frac{|x-a|}{\sqrt{a}}$ less than ε, requires that we make $|x-a| < \varepsilon\sqrt{a}$.

[21] Note that we may as well choose $\delta \leq 2$ or $\delta \leq 3$, (or $\delta \leq$ any other convenient positive number) and then obtain the relation between that number and ε.

Formal Proof: To prove $\lim_{x \to a} \sqrt{x} = \sqrt{a}$, $(a > 0)$
Let $\varepsilon > 0$ be given.

Choose $\delta = \varepsilon\sqrt{a}$

Then, $0 < |x - a| < \delta$ implies

$$|\sqrt{x} - \sqrt{a}| = \left|\frac{(\sqrt{x} - \sqrt{a}) \cdot (\sqrt{x} + \sqrt{a})}{(\sqrt{x} + \sqrt{a})}\right|$$

$$= \left|\frac{x - a}{\sqrt{x} + \sqrt{a}}\right| \leq \frac{|x - a|}{\sqrt{a}} < \frac{\varepsilon\sqrt{a}}{\sqrt{a}}, \quad [\because |x - a| < \varepsilon\sqrt{a}]$$

i.e.,

$$\frac{|x - a|}{\sqrt{a}} < \varepsilon$$

$$|\sqrt{x} - \sqrt{a}| < \varepsilon$$

$$\lim_{x \to a} \sqrt{x} = \sqrt{a}$$

Remark: *There is one more technical point.* We should insist that $\delta \leq a$, for then $|x - a| < \delta$ implies $x > 0$ *so that* $\sqrt{x}$ *is defined.*[22]

Thus, for absolute rigor, we must *choose* δ *to be smaller than a and* $\varepsilon\sqrt{a}$.

Note (8): In Example (6) given above, we had to rationalize the numerator for the purpose of our demonstration. *Rationalization is a trick frequently useful in calculus.*

7a.5.3.1 Applications of the Main Limit Theorem

Example (21): Evaluate the following limits:

(a) $\lim_{x \to -1} x^2$

(b) $\lim_{x \to -1} (\pi x + x^2)$

(c) $\lim_{x \to -1} \dfrac{x^2}{x + 3}$

Solution:

(a) $\lim_{x \to -1} x^2 = \left(\lim_{x \to -1} x\right)\left(\lim_{x \to -1} x\right) = (-1)(-1) = 1$

(Here, we have applied the *product rule*.)

(b) We have $\lim_{x \to -1} \pi x = \pi(-1) = -\pi$ (*By constant multiple rule*)

and $\lim_{x \to -1} x^2 = 1$ [By part (a) above].

We conclude that

$$\lim_{x \to -1} (\pi x + x^2) = \lim_{x \to -1} \pi x + \lim_{x \to -1} x^2 = -\pi + 1 \qquad \text{(By sum rule)}$$

[22] $|x - a| < \delta \Rightarrow -\delta < x - a < \delta \Rightarrow a - \delta < x < a + \delta$ (i). Now, $\delta \leq a \Rightarrow a - \delta \geq 0$. $\therefore$ From (i) it follows that $0 \leq x < a + \delta$.

(c) We have $\lim_{x\to -1}(x+3) = -1+3 = 2,$ (By sum rule)

and $\lim_{x\to -1} x^2 = 1$ [By part (a)].

We conclude from the quotient rule that

$$\lim_{x\to -1}\frac{x^2}{x+3} = \frac{\lim_{x\to -1} x^2}{\lim_{x\to -1}(x+3)} = \frac{1}{2}$$

Note (9): Rule no. (6) of Theorem A, *demands special attention.*

We have,

$$\lim_{x\to a}\sqrt[n]{f(x)} = \sqrt[n]{\lim_{x\to a} f(x)}, \text{provided } \lim_{x\to a} f(x) > 0, \quad \text{when } n \text{ is even.}$$

Recall that the nth root function, $\sqrt[n]{x}$ is defined for any real number x, if n is *odd*. However, *if n is even then $\sqrt[n]{x}$ is defined only for $x \geq 0$*, with the understanding that only non-negative values of the nth root are accepted. *As a particular case of rule (6)*, we have

$$\lim_{x\to a}\sqrt[n]{x} = \sqrt[n]{a}, \qquad \begin{cases}\text{for all } a \text{ if } n \text{ is odd.}\\ \text{for } a > 0, \text{ if } n \text{ is even.}\end{cases}$$

In particular, $\lim_{x\to a}\sqrt{x} = \sqrt{a}, \quad$ for $a > 0$.

For example, $\lim_{x\to 1/4}\sqrt{x} = \sqrt{\frac{1}{4}} = \frac{1}{2}.$

7a.5.3.2 Substitution Rule

Consider the following limit

$$\lim_{x\to 1}\sqrt{x^5 - 4x^2 + 3x + 2}$$

To evaluate this limit we have to apply the Rule (7), which states that

$$\lim_{x\to a} f[g(x)] = f\left(\lim_{x\to a} g(x)\right)$$

Now, suppose $\lim_{x\to a} g(x) = c$ (i.e., some constant) and we substitute $y = g(x)$.

Then, we can write, $\lim_{x\to a} g(x) = \lim_{y\to c} y.$

This is a valid statement and we can write $\lim_{x\to a} f[g(x)] = \lim_{y\to c} f(y)$, *provided $f(y)$ exists.*

This is known as *Substitution Rule.* Frequently, the process is straight-forward.

This rule might look innocent but it is a very convenient and useful rule for evaluating certain limits. The following examples will convince the reader about its usefulness.

Example (22):

$$\text{Find } \lim_{x\to 1}\sqrt{x^5 - 4x^2 + 3x + 2} \tag{6}$$

In trying to evaluate this limit, we first let $y = x^5 - 4x^2 + 3x + 2$ and notice that as $x \to 1$, y approaches $(1)^5 - 4(1)^2 + 3(1) + 2 = 2$.

This suggests that *by substituting* $y = x^5 - 4x^2 + 3x + 2 \ldots$ (7)

We can easily evaluate the limit of the expression on the right-hand side of Equation (7), as $x \to 1$.

Now, in order to evaluate the limit at (6) above, we substitute $y \to 2$, for $x \to 1$.

Thus, we write

$$\lim_{x \to 1} \sqrt{x^5 - 4x^2 + 3x + 2} = \lim_{y \to 2} \sqrt{y} = \sqrt{2} \quad \text{Ans.}$$

Example (23): Consider $\lim_{x \to 2} \sqrt{x + \frac{1}{x}}$. We first let $y = x + \frac{1}{x}$. Then, we notice that $\lim_{x \to 2} y = \lim_{x \to 2} \left(x + \frac{1}{x}\right) = 2 + \frac{1}{2} = \frac{5}{2}$. $\left[\because\ y = x + \frac{1}{x}\right]$.

$\therefore$ By the substitution rule, we get $\lim_{x \to 2} \sqrt{x + \frac{1}{x}} = \lim_{y \to 5/2} \sqrt{y} = \sqrt{\frac{5}{2}}$.

Example (24): If $\lim_{x \to 3} f(x) = 4$ and $\lim_{x \to 3} g(x) = 8$, find $\lim_{x \to 3} \left[f^2(x) \cdot \sqrt[3]{g(x)}\right]$

Solution:

$$\begin{aligned}
\lim_{x \to 3} \left[f^2(x) \cdot \sqrt[3]{g(x)}\right] &= \lim_{x \to 3} f^2(x) \cdot \lim_{x \to 3} \sqrt[3]{g(x)} \\
&= \left[\lim_{x \to 3} f(x)\right]^2 \cdot \sqrt[3]{\lim_{x \to 3} g(x)} \\
&= [4]^2 \cdot \sqrt[3]{8} \\
&= 16 \cdot 2 = 32 \qquad \text{Ans.}
\end{aligned}$$

Note (10): *Usefulness of the substitution rule* is appreciated when we have to evaluate the following limit.

Example (25):

$$\lim_{x \to 1} \frac{x^{1/4} - 1}{x^{1/3} - 1}$$

Here, we observe that the indices of x are fractions. Hence, it is not possible to factorize both numerator and denominator. Further, we see that the denominator of these indices is 4 and 3, and their L.C.M. is 12.

We substitute $x = y^{12}$. Thus, we get $y = x^{1/12}$ $\therefore$ $x^{1/4} = (y^{12})^{1/4} = y^3$ and $x^{1/3} = (y^{12})^{1/3} = y^4$. Also, we see that as $x \to 1$, $y \to 1$.

$$\therefore \quad \text{Required limit is } \lim_{y \to 1} \frac{y^3 - 1}{y^4 - 1} \text{ that is } \lim_{y \to 1} \frac{y^3 - 1}{y^4 - 1}$$

(Note that, now the numerator and the denominator both can be factorized.)

$$\begin{aligned}
&= \lim_{y \to 1} \frac{(y-1)(y^2 + y + 1)}{(y-1)(y^3 + y^2 + y + 1)} \\
&= \lim_{y \to 1} \frac{(y^2 + y + 1)}{(y^3 + y^2 + y + 1)}, \quad (\because\ y \neq 1) \\
&= \frac{1 + 1 + 1}{1 + 1 + 1 + 1} = \frac{3}{4} \qquad \text{Ans.}
\end{aligned}$$

Note (11): Many such limits are evaluated in the Chapter 7b.
Another example in which *the beauty of the substitution rule* can be enjoyed is the following limit.

Example (26): $\lim_{x \to 0} \frac{\sqrt{1+x}-1}{x}$ [Hint: Put $y=1+x$, then as $x \to 0$, $y \to 1$. Hence, the limit reduces to the form $\lim_{y \to 1} \left((y^{1/2} - 1)/(y - 1)\right)$.]

7a.6 THEOREM (B): SUBSTITUTION THEOREM[23]

If f is a *polynomial function* or a *rational function*, then

$$\lim_{x \to a} f(x) = f(a)$$

provided that, in the case of a rational function, the value of the denominator at "a" is not zero.

Note that Theorem (B) allows us *to find limits for polynomials and those of rational functions by simply substituting "a" for x throughout.* Let us see what happens when in a rational function the limit of the denominator is zero.

Note (12): Suppose, we have to find $\lim_{x \to 1} \frac{x^5+2x+8}{x^2-2x+1} = \lim_{x \to 1} \frac{x^5+2x+8}{(x-1)^2}$.

In this case, neither Theorem (B) nor the "Quotient Rule" of Theorem (A) applies, *since the limit of the denominator is 0.* However, since the limit of the numerator is 11, we see that *as x nears 1, we are dividing a number near 11 by a positive number near 0.* The result is a large positive number. In fact, the resulting number can be made as large as you like by letting x get close enough to 1. Here, we may say that the *limit does not exist, but later on in Chapter 7b we will allow ourselves to say that the limit is* $+\infty$.

(This becomes possible once we accept $+\infty$ and $-\infty$ as limits.)

Note (13): Now suppose that we have to find $\lim_{t \to 2} \frac{t^2-t-2}{t^2+2t-8} = \lim_{t \to 2} \frac{(t-2)(t+1)}{(t-2)(t+4)}$

Again, in this case, *Theorem (B) does not apply*. But this time, *the quotient takes the meaningless form 0/0 at $t=2$. Whenever this happens we should look for an algebraic simplification of the quotient (by factorization), before taking the limit.*

$$\lim_{t \to 2} \frac{t^2 - t - 2}{t^2 + 2t - 8} = \lim_{t \to 2} \frac{(t-2)(t+1)}{(t-2)(t+4)}$$

$$= \lim_{t \to 2} \frac{t+1}{t+4}, \quad [\because t \neq 2]$$

$$= \frac{3}{6} = \frac{1}{2} \quad \text{Ans.}$$

[23] The substitution theorem [i.e., Theorem (B)] discussed here should not be confused with the *substitution rule* discussed in Section 7a.5.3.2.

7a.7 THEOREM (C): SQUEEZE THEOREM OR SANDWICH THEOREM

Let *f*, *g*, and *h* be functions satisfying $f(x) \le g(x) \le h(x)$, *for all x near a, except possibly at a.*

If $\lim_{x \to a} f(x) = \lim_{x \to a} h(x) = l$, then $\lim_{x \to a} g(x) = l$[24]

Proof: Let $\varepsilon > 0$ be given. Choose δ_1 such that, $0 < |x - a| < \delta_1 \Rightarrow l - \varepsilon < f(x) < l + \varepsilon$, and δ_2 such that $0 < |x - a| < \delta_2 \Rightarrow l - \varepsilon < h(x) < l + \varepsilon$. Also, choose δ_3 so that $0 < |x - a| < \delta_3 \Rightarrow l - \varepsilon < g(x) < l + \varepsilon$.

Let $\delta = \min\{\delta_1, \delta_2, \delta_3\}$. Then $0 < |x - a| < \delta \Rightarrow l - \varepsilon < f(x) < g(x) < h(x) < l + \varepsilon \Rightarrow l - \varepsilon < g(x) < l + \varepsilon$.

Hence, we conclude that $\lim_{x \to a} g(x) = l$.

Note (14): This theorem will be found very *useful in evaluating limits of a variety of trigonometric functions, to be studied later.*

Remark: Suppose $\lim_{x \to a} f(x)$ *does not exist*, then, limit rules can help in proving this fact, as the following example illustrates.

Example (27): Show that $\lim_{x \to 0} \frac{1}{x}$ does not exist.

Solution: To prove the above result, we approach *by the indirect method.*

Suppose that $\lim_{x \to 0} (1/x)$ exists, and let $\lim_{x \to 0} (1/x) = l$. Consider, $1 = x \cdot (1/x)$.

$$\therefore \quad \text{We have } \lim_{x \to 0} 1 = \lim_{x \to 0} \left(x \cdot \frac{1}{x} \right) \text{ or } 1 = \left(\lim_{x \to 0} x \right) \cdot \left(\lim_{x \to 0} \frac{1}{x} \right)$$

$$\therefore \quad 1 = 0 \cdot l = 0$$

This is absolutely false (since $1 \neq 0$). Therefore, $\lim_{x \to 0} (1/x)$ *cannot exist.*

7a.8 ONE-SIDED LIMITS (EXTENSION TO THE CONCEPT OF LIMIT)

Now, we are in a position to give the ε, δ definitions for left-hand and right-hand limits of a function.

Definition: Let *f* be defined on some open interval (c, a).[25] A number "l" is the *limit of f(x) as x approaches a from the left*, if, for *every* $\varepsilon > 0$, there is a *corresponding* $\delta > 0$, such that $-\delta < x - a < 0 \Rightarrow |f(x) - l| < \varepsilon$.

In this case we write, $\lim_{x \to a^-} f(x) = l$, and we say that the *left-hand limit* of "f" at "a" exists. Right-hand limits are treated in a completely analogous way. Thus, if "f" is defined on some open interval (a, c), then a number "l" is *the limit of f(x) as x approaches "a" from the right*, if for every $\varepsilon > 0$, there is a *corresponding* $\delta > 0$, such that $0 < x - a < \delta \Rightarrow |f(x) - l| < \varepsilon$.

In this case, we write $\lim_{x \to a^+} f(x) = l$, and we say that the *right-hand limit of "f" at "a" exists.*

[24] Roughly speaking, the theorem tells us that if a function can be "sandwiched" between two other functions, each of which approaches the same limit "l" (say) as x approaches a, then the sandwiched function also approaches the same limit "l" as x approaches a. For obvious reasons, we call it the "Sandwich Theorem".

[25] Observe that "f" is not defined at "a". Besides, "a" can be approached only from the left-hand side of "a".

Remark: One can show that $\lim_{x\to 0^+} \sqrt{x} = 0$, *but it must be clear that neither* $\lim_{x\to 0^-} \sqrt{x}$ nor $\lim_{x\to 0} \sqrt{x}$ *exists* (because $\sqrt{x}$ is not defined to the left of 0). Similarly, for $f(x) = \sqrt{x}, x \in (1, 2)$, $\lim_{x\to 2^-} \sqrt{x} = \sqrt{2}$, but neither $\lim_{x\to 2^+} \sqrt{x}$ nor $\lim_{x\to 2} \sqrt{x}$ exists.

Note (15): Right-hand and left-hand limits are called *one-sided limits*. Ordinary limits are called *two-sided limits*.

Note (16): Sometimes a function "f" is defined by two (or more) different rules. In such cases, one rule may be applicable for the values of x less than "a" and the other for the values of x greater than "a".

We have already given examples, wherein

$\lim_{x\to a^-} f(x) = l_1$ and $\lim_{x\to a^+} f(x) = l_2$. $(l_1 \neq l_2)$ [see Examples (6)–(9)].

If $l_1 = l_2 = l$ *(say), then we say that* $\lim_{x\to a} f(x)$ *exists*, otherwise we say that the limit does not exist. Thus, the statement

$$\lim_{x\to a} f(x) = l \Leftrightarrow \lim_{x\to a^-} f(x) = l = \lim_{x\to a^+} f(x).$$

Note (17): The concept of one-sided limits will be very useful in studying the concept of continuity of a function:

(i) at any point in an interval and
(ii) at the end point of a closed interval.[26]

Example (28): Let us find $\lim_{x\to 1^-} \sqrt{1-x^2}$

Solution: First observe that $\sqrt{1-x^2}$ is not defined for $|x| > 1$. (Why?)
Let $y = 1 - x^2$

For any value of x, such that $|x| \leq 1$ we have $y \geq 0$. Thus, when $x \to 1^-, y \to 0^+$. Therefore, we have, $\lim_{x\to 1^-} \sqrt{1-x^2} = \lim_{x\to 0^+} \sqrt{y} = 0$.

Remark: Although $\lim_{x\to 1^-} \sqrt{1-x^2}$ exists, $\lim_{x\to 1^+} \sqrt{1-x^2}$ *does not exist*, because $1 - x^2$ will be negative whenever x lies to the right of 1, and we know that *square root of a negative number is not defined.*

[26] The concept of "continuity of a function" is discussed in Chapter 8.

7b Methods for Computing Limits of Algebraic Functions

7b.1 INTRODUCTION

In Chapter 7a, we introduced the *notion of limit of a function.* There, we defined the meanings of certain notations (such as $x \to a$, $x \to a^+$, $x \to a^-$, where "a" is a real number), applied intuitive and logical thinking to compute the limit(s) of some *polynomials and rational functions.* In fact, this has been the *simplest and the most practical* way of introducing the concept of limit of a function.

Recall that, in the process of assigning meaning to a *rational* function like $f(x) = ((x^2 - 9)/(x - 3))$, wherein the variable x is *permitted to assume values closer and closer to* 3, we learnt that whereas the value of $f(x)$ at $x = 3$ is *not defined*, we can still give a *logical meaning to the statement* $\lim_{x \to 3} f(x)$, *which matches our intuitive meaning of the statement.* This allows us to assign the number 6 to the statement $\lim_{x \to 3} f(x)$ (i.e., $\lim_{x \to 3} \left((x^2 - 9)/(x - 3)\right)$), which we called the limit of $f(x)$ at $x = 3$.

Such examples help us distinguish between the value f of the function $f(x)$ at $x = a$ and the limit of $f(x)$ as the variable x approaches the number a. We get that the *value of a function and the limit of a function are two different numbers.* Of course, under certain situation both may stand for the same number.

Note: In our study of differential calculus we will be required to compute the derivative of a function, which itself is the limit of a particular kind (this we will understand later in Chapter 9). Hence, *it is necessary to understand the limiting process in full clarity.*

Recall that, the ε, δ definition of the limit introduced in Chapter 7a does not help in evaluating $\lim_{x \to a} f(x)$. It can only be used to verify whether a given number (or a guessed number) is the limit of the given function $f(x)$, as $x \to a$. The method of preparing the tables: one for the values of x closer and closer to "a" and the other for the corresponding values "$f(x)$", can help us *guess the number* to which the values "$f(x)$" approaches. But, this process is not only *tedious* but also *unreliable* (under certain situations) as shown in Chapter 7a.

Further, there is *no general theorem*, which can be applied to a given function to obtain its limit at a desired point. However, there are *limit theorems* (based on ε, δ definition of limit), which offer very simple methods for evaluating limits of (all) functions. Standard limits have

What must you know to learn calculus? 7b-Methods for computing the limits of algebraic functions. Limit at infinity [i.e., $\lim_{x \to \infty} f(x)$, $\lim_{x \to -\infty} f(x)$***] and infinite limits [i.e., meaning of*** $\lim_{x \to a} f(x) = \pm\infty$***].***

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

been established for different functions (through ε, δ definition) and then *by using these standard limits directly, we can easily obtain their limits*, avoiding all practical difficulties associated with ε, δ definition of limit.[(1)]

In Chapter 7a, we have handled various algebraic functions and obtained their limit(s) for developing the ε, δ definition of limit. What remains to be discussed are different methods for *evaluating limits of many other types of algebraic functions*. Accordingly, we now introduce the following methods.

7b.2 METHODS FOR EVALUATING LIMITS OF VARIOUS ALGEBRAIC FUNCTIONS

7b.2.1 Direct Method [or Method of Direct Substitution]

This method is applicable in the case of very simple functions, in which *the value of the function* and *the limit of the function both are the same*. For learning the concept of limit, such functions are neither important nor useful, since they *do not distinguish between the two different ideas involved*. If we replace x by a in the formula defining $f(x)$, we get the value $f(a)$, and the limit $\lim_{x \to 3} f(x)$ both representing the same (finite) number.

Example (1):

$$\lim_{x \to 2}(x^2 + 3) = \lim_{x \to 2} x^2 + \lim_{x \to 2} 3$$
$$= 2^2 + 3 = 7. \quad \text{Ans.}$$

Example (2):

$$\left[\lim_{x \to 5} \frac{\sqrt{x-1}+2}{\sqrt{x+31}}\right] = \frac{\lim_{x \to 5} \sqrt{x-1} + \lim_{x \to 5} 2}{\lim_{x \to 5} \sqrt{x+31}}$$
$$= \frac{\sqrt{5-1}+2}{\sqrt{5+31}} = \frac{4}{6} = \frac{2}{3} \quad \text{Ans.}$$

Example (3):

$$\lim_{x \to 1} \frac{x^3 - 9}{x - 3} x \neq 3$$
$$\frac{\lim_{x \to 1}(x^3 - 9)}{\lim_{x \to 1}(x - 3)} = \frac{(1^3 - 9)}{(1 - 3)} = \frac{-8}{-2} = 4 \quad \text{Ans.}$$

7b.2.2 Factorization Method

For computing limit(s) of the type, $\lim_{x \to a} (f(x)/g(x))$, where $f(a) = 0$ and $g(a) = 0$, the *direct substitution method fails*. In such cases, we search for a common factor $(x - a)$ in $f(x)$ and $g(x)$ by factorizing them and canceling this factor to reduce the quotient to the simplest form

[(1)] In fact, the standard limits of trigonometric functions are established in Chapter 11a and those for exponential and logarithmic functions in Chapter 13a. Accordingly the methods of computing limits of functions involving trigonometric functions are discussed in Chapter 11b and those involving exponential and logarithmic functions are discussed in Chapter 13b.

and then apply the direct method to obtain the limit. [Remember that $x \to a$ means that $x \neq 0$, at any stage. In other words $(x - a) \neq 0$, at any stage. This permits us to cancel the common factor $(x-a)$ from both numerator and denominator.]

Example (4): Evaluate $\lim_{x \to 1} \frac{x^2 - 4x + 3}{x^2 + 2x - 3}$.

Solution:

$$\lim_{x \to 1} \frac{x^2 - 4x + 3}{x^2 + 2x - 3} = \lim_{x \to 1} \frac{(x - 3)(x - 1)}{(x + 3)(x - 1)}$$
$$= \lim_{x \to 1} \frac{x - 3}{x + 3}, \ [(x - 1) \neq 0]$$
$$= \frac{1 - 3}{1 + 3} = \frac{-2}{4} = -\frac{1}{2} \quad \text{Ans.}$$

Note: For evaluating $\lim_{x \to a} (f(x)/g(x))$, we may also follow the following steps:

(i) Put $x = a + h$ ($\therefore$ as $x \to a$, $h \to 0$)

(ii) Simplify numerator and denominator and cancel the common factor h.

(iii) Put $h = 0$, in the remaining expression in h and obtain the limit.

Example (5): Evaluate $\lim_{x \to 4} \frac{x^3 - 8x^2 + 16x}{x^3 - x - 60}$

Solution: Consider $x^3 - 8x^2 + 16x$

$$= x(x^2 - 8x + 16), \quad \{16 = (-4)(-4)\}$$
$$= x(x^2 - 4x - 4x + 16)$$
$$= x[x(x - 4) - 4(x - 4)]$$
$$= x[(x - 4)(x - 4)]$$

Now consider $x^3 - x - 60$

$$= x^3 - 4x^2 + 4x^2 - 16x + 15x - 60$$
$$= x^2(x - 4) + 4x(x - 4) + 15(x - 4)$$
$$= (x - 4)(x^2 + 4x + 15)$$

$$\lim_{x \to 4} \frac{x^3 - 8x^2 + 16x}{x^3 - x - 60} = \lim_{x \to 4} \frac{x(x - 4)(x - 4)}{(x - 4)(x^2 + 4x + 15)}$$
$$\lim_{x \to 4} \frac{x(x - 4)}{(x^2 + 4x + 15)} = \frac{4(4 - 4)}{(4^2 + 4(4) + 15)} = 0 \quad \text{Ans.}$$

An Important Standard Limit

We will prove the following limit.

$$\lim_{x \to a} \frac{x^n - a^n}{x - a} = n \cdot a^{n-1} \tag{1}$$

Let n be a natural number and $a > 0$.

Consider $x^n - a^n$

$$= x^n - x^{n-1} \cdot a + x^{n-1} \cdot a - x^{n-2} \cdot a^2 + x^{n-2}a + x^{n-3}a^{3^2} - x^{n-3} \cdot a^3$$
$$-x^{n-4} \cdot a^4 + \cdots\cdots + xa^{n-1} - x^0 \cdot a^n$$

$$= x^{n-1}(x-a) + x^{n-2} \cdot a(x-a) + x^{n-3} \cdot a^2(x-a) + \cdots\cdots + x^0 \cdot a^{n-1}(x-a)$$

$$= (x-a)[x^{n-1} + x^{n-2} \cdot \leq a + x^{n-3} \cdot a^2 \cdots\cdots + a^{n-1}]$$

$$\frac{x^n - a^n}{x-a} = [x^{n-1} + x^{n-2} \cdot a + x^{n-3} \cdot a^2 + \cdots\cdots + a^{n-1}$$

Therefore, $\lim\limits_{x \to a} \dfrac{x^n - a^n}{x-a} = \lim\limits_{x \to a} (x^{n-1} + x^{n-2} \cdot a + x^{n-3} \cdot a^2 + \cdots\cdots + a^{n-1})$

$$= a^{n-1} + a^{n-1} + a^{n-1} + \cdots\cdots n \text{ terms}$$

$$= n \cdot a^{n-1} \quad \text{(Proved)}$$

(We will use this formula in evaluating the following limits.)

7b.2.3 Applications of the Standard Limit in Solving Special Type of Problems

Example (6): Evaluate $\lim\limits_{x \to 1} \dfrac{x + x^2 + x^3 + \cdots\cdots + x^n - n}{x-1}$

Solution: The given limit

$$= \lim_{x \to 1} \frac{(x + x^2 + x^3 + \cdots\cdots + x^n) - (1 + 1 + 1 + \cdots\cdots n \text{ times})}{x-1}$$

$$= \lim_{x \to 1} \frac{(x-1) + (x^2-1) + (x^3-1) + \cdots\cdots + (x^n-1)}{x-1}$$

$$= \lim_{x \to 1} \frac{(x-1) + (x-1)(x+1) + (x-1)(x^2+x+1) + \cdots\cdots + (x-1)(x^{n-1} + x^{n-2}a + x^{n-1}a^2 \cdots + 1)}{x-1}$$

$$= \lim_{x \to 1} [1 + (x+1) + (x^2+x+1) + \cdots\cdots + (x^{n-1} + x^{n-2} + \cdots + 1)], \quad [\because (x-1) \neq 0]$$

$$= 1 + (1+1) + (1+1+1) + \cdots\cdots n \text{ times})$$

$$= 1 + 1 + 3 + 4 + 5 + 6 \cdots\cdots + n$$

$$= \frac{n(n+1)}{2} \quad \text{Ans.}$$

We have seen above that,

$$x^n - a^n = (x-a)[x^{n-1} + x^{n-2} \cdot a + x^{n-3} \cdot a^2 \cdots\cdots + a^{n-1}]$$

where n is a natural number and $a > 0$.[(2)]

The above formula can be used to evaluate limits of the form $(x^n - a^n)/(x^m - a^m)$, (where n, $m \in N$, and $a > 0$).

[(2)] Note that the expression $x^n - a^n$ can be factorized only if $n \in N$ and $a > 0$.

For this purpose, we write

$$\lim_{x\to a}\frac{x^n-a^n}{x^m-a^m}=\lim_{x\to a}\frac{x^n-a^n}{x-a}\cdot\frac{x-a}{x^m-a^m}=\lim_{x\to a}\frac{x^n-a^n}{x-a}\div\lim_{x\to a}\frac{x^m-a^m}{x-a}$$

and apply the *standard limit* to obtain

$$\lim_{x\to a}\frac{x^n-a^n}{x^m-a^m}=\frac{n}{m}a^{n-m} \tag{2}$$

which is a corollary to the standard limit (1).

Example (7): Evaluate $\lim_{x\to a}\frac{x^5-a^5}{x^3-a^3}$

Solution: $\lim_{x\to a}\frac{x^5-a^5}{x^3-a^3}=\lim_{x\to a}\frac{5\cdot a^{5-1}}{3\cdot a^{3-1}}=\frac{5a^4}{3a^2}=\frac{5}{3}a^2$ Ans.

Remark: Formula (2) has been proved for natural numbers n and m. However, the result is true for rational values of n and m. The following examples tell how this is justified.

Example (8): Evaluate $\lim_{x\to 1}\frac{x^{1/4}-1}{x^{1/3}-1}$

Note (1): In such cases the important point is that the given limit can be converted in the form (2) by substitution as follows.

Here, the indices of x are fractions (i.e., the positive rational numbers) and hence we cannot factorize. The denominators of these indices are 4 and 3. Their L.C.M. is 12. Therefore, we use the substitution $x=t^{12}$, for our purpose.

Solution: Put $x=t^{12}$ $\therefore t=x^{1/12}$

$$\therefore x^{1/4}=(t^{12})^{1/4}=t^3 \quad\text{and}\quad x^{1/3}=(t^{12})^{1/3}=t^4$$

Also we see that as $x\to 1$, $t\to 1$.

$$\therefore \text{Required limit } =\lim_{t\to 1}\frac{t^3-1}{t^4-1}=\lim_{t\to 1}\frac{t^3-1^3}{t^4-1^4}=\frac{3(1)^{3-1}}{4(1)^{4-1}}=\frac{3\cdot 1^2}{4\cdot 1^3}=\frac{3}{4} \quad \text{Ans.}$$

Note (2): We can also apply Corollary (2) directly and obtain the limit as follows:

$$\lim_{x\to 1}\frac{x^{1/4}-1}{x^{1/3}-1}=\lim_{x\to 1}\frac{x^{1/4}-1^{1/4}}{x^{1/3}-1^{1/3}}$$

$$=\frac{(1/4)}{(1/3)}\cdot(1)^{(1/4)-(1/3)}=\frac{3}{4}\cdot(1)^{1/12}=\frac{3}{4} \quad \text{Ans.}$$

Example (9): Evaluate $\lim_{x\to 1}\frac{x^{2/5}-3^{2/5}}{x^{1/2}-3^{1/2}}$

$$=\frac{(2/5)}{(1/2)}\cdot 3^{(2/5)-(1/2)}$$

$$=\frac{4}{5}\cdot(3)^{-(1/10)}$$

$$=\frac{4}{5}\cdot\frac{1}{3^{(1/10)}} \quad \text{Ans.}$$

Example (10): Evaluate $\lim_{x \to 2} \dfrac{x^{-3} - 2^{-3}}{x - 2}$

Solution: $\lim_{x \to 2} \dfrac{x^{-3} - 2^{-3}}{x - 2} = \lim_{x \to 2} \dfrac{x^{-3} - 2^{-3}}{x^1 - 2^1}$

$$= \frac{-3}{1} \cdot 2^{(-3)-(1)} = -3 \cdot 2^{-4}$$

$$= \frac{-3}{16} \quad \text{Ans.}$$

Note (3): To evaluate limits of this type, it is always useful to convert the given limit to the standard form as follows:

$$x^{-3} - 2^{-3} = \frac{1}{x^3} - \frac{1}{2^3} = \frac{2^3 - x^3}{2^3 \cdot x^3} = \frac{-(x^3 - 2^3)}{8x^3}$$

$$\therefore \text{ The given limit is } \lim_{x \to 2} -\frac{1}{8x^3}\left(\frac{x^3 - 2^3}{x - 2}\right)$$

$$= -\left(\frac{1}{8 \cdot 2^3} \cdot \frac{3}{1} 2^{3-1}\right) = -\frac{1}{64} \cdot 3 \cdot 4 = -\frac{3}{16} \quad \text{Ans.}$$

Example (11): Evaluate $\lim_{x \to a} \dfrac{(x+2)^{5/3} - (a+2)^{5/3}}{x - a}$

Solution: $\lim_{x \to a} \dfrac{(x+2)^{5/3} - (a+2)^{5/3}}{x - a}$

$$= \lim_{(x+2) \to (a+2)} \frac{(x+2)^{5/3} - (a+2)^{5/3}}{(x+2) - (a+2)} \quad [\because x - a = (x+2) - (a+2)]$$

$$= \frac{5}{3} \cdot (a+2)^{(5/3)-1}$$

$$= \frac{5}{3} \cdot (a+2)^{2/3} \quad \text{Ans.}$$

Example (12): Evaluate $\lim_{x \to 1} \dfrac{1 - x^{-1/3}}{1 - x^{-2/3}}$

Solution: $\lim_{x \to 1} \dfrac{1 - x^{-1/3}}{1 - x^{-2/3}}$

$$= \lim_{x \to 1} \frac{(x^{1/3} - 1)/x^{1/3}}{(x^{2/3} - 1)/x^{2/3}}$$

$$= \lim_{x \to 1} \frac{x^{1/3} \cdot (x^{1/3} - 1)}{(x^{2/3} - 1)} = 1^{1/3} \lim_{x \to 1} \frac{x^{1/3} - 1}{x - 1} \cdot \frac{x - 1}{(x^{2/3} - 1)}$$

$$= \lim_{x \to 1} \frac{x^{1/3} - 1}{x - 1} \div \lim_{x \to 1} \frac{x^{2/3} - 1}{x - 1}$$

$$= \left(\frac{1}{3}\right) \cdot 1^{(1/3)-1} \div \frac{2}{3}(1)^{(2/3)-1} = \frac{1}{3} \div \frac{2}{3} = \frac{1}{2} \quad \text{Ans.}$$

Exercise [Application of Standard Limits]

(1) Evaluate the following limits:

(i) $\lim_{x\to 1} \frac{x^{n-1}}{x^{m-1}}$

(ii) $\lim_{x\to 0} \frac{(1+x)^n - 1}{x}$ [Hint : $1 + x = y$]

(iii) $\lim_{x\to 3} \frac{x^{-3} - 3^{-3}}{x^{-2} - 3^{-3}}$

(iv) $\lim_{x\to 2} \left(\frac{1}{x^3} - \frac{1}{8}\right)/(x-2)$

(v) $\lim_{x\to -1} \frac{x^3+1}{x+1}$ $\left[\text{Hint} : x^3 + 1 = x^3 - (-1)\right]$

(vi) $\lim_{x\to 0} \frac{\sqrt{1+x} - 1}{x}$

(vii) $\lim_{x\to a} \frac{(x+3)^{7/2} - (a+3)^{7/2}}{x-a}$

(viii) $\lim_{x\to a} \frac{x^3 - 64}{x^2 - 16}$

(ix) $\lim_{x\to 1} \frac{x + x^2 + x^3 + x^4 + x^5 - 5}{x-1}$

(x) If $\lim_{x\to -a} \frac{x^9 + a^9}{x+a} = 9$, find the value of a. [Hint: $x + a = x - (-a)$]

(xi) If $\lim_{x\to 2} \frac{x^n - 2^n}{x-2} = 80$, and n is a positive integer, then find the value of n.

(xii) $\lim_{x\to 2} \frac{(x^2 - x - 2)^{20}}{(x^3 - 12x + 16)^{10}}$ [Hint: factorize N^r and D^r]

Answers

(i) n/m (ii) n (iii) 1/2 (iv) −3/16 (v) 3

(vi) 1/2 (vii) $(7/2)(a+3)^{5/2}$ (viii) 6 (ix) 15 (x) ±1

(xi) 5 (xii) $(3/2)^{10}$.

7b.2.4 Method of Simplification

Sometimes it is required to simplify the given function and then evaluate the limit.

Example (13): Evaluate $\lim_{x\to 3} \left(\frac{1}{x-3} + \frac{1}{3-x}\right)$

Solution: $\lim_{x\to 3} \left(\frac{1}{x-3} + \frac{1}{3-x}\right)$

$$= \lim_{x\to 3} \left(\frac{1}{x-3} - \frac{1}{x-3}\right)$$
$$= \lim_{x\to 3} (0) = 0$$

Example (14): Evaluate $\lim_{x\to 5}\left(\frac{1}{x-5}-\frac{5}{x^2-5x}\right)$

Solution: $\lim_{x\to 5}\left(\frac{1}{x-5}-\frac{5}{x^2-5x}\right)$

$$= \lim_{x\to 5}\left(\frac{1}{x-5}-\frac{5}{x(x-5)}\right), \left\{x\to 5 \therefore x\neq 5 \therefore (x-5)\neq 0\right.$$
$$= \lim_{x\to 5}\left(\frac{x-5}{x(x-5)}\right)$$
$$= \lim_{x\to 5}\left(\frac{1}{x}\right)=\frac{1}{5} \quad \text{Ans.}$$

Example (15): Evaluate $\lim_{x\to -2}\left(\frac{1}{x^2+5x+6}+\frac{1}{x^2+3x+2}\right)$

Solution: We have, x^2+5x+6

$$= (x^2+3x)+(2x+6) \quad [6=3\times 2]$$
$$= x(x+3)+2(x+3)$$
$$= (x+3)\,(x+2)$$

and x^2+3x+2

$$= (x^2+2x)+x+2 \quad [2=2\times 1]$$
$$= (x+2)\,(x+1)$$

$\therefore$ The given limit is

$$= \lim_{x\to -2}\left(\frac{1}{(x+3)(x+2)}+\frac{1}{(x+2)(x+1)}\right)$$
$$= \lim_{x\to -2}\left(\frac{x+1+x+3}{(x+3)\,(x+2)\,(x+1)}\right)$$
$$= \lim_{x\to -2}\left(\frac{2(x+2)}{(x+3)\,(x+2)\,(x+1)}\right)$$
$$= \lim_{x\to -2}\left(\frac{2}{(x+3)\,(x+1)}\right) \quad \left\{\because x\to -2,\ \therefore\ x\neq -2,\ \therefore (x+2)\neq 0\right.$$
$$= \frac{2}{(22+3)(22+1)}=\frac{2}{(1)(21)}=-2. \quad \text{Ans.}$$

Exercise [Method of Simplification]

(2) Evaluate the following limits:

(i) $\lim_{x\to 3}\left(\frac{1}{(x-3)}-\frac{7}{(x^2+x-12)}\right)$

(ii) $\lim_{x\to \frac{b}{a}}\left(\frac{a}{(ax-b)}-\frac{b}{(ax^2-bx)}\right)$

(iii) $\lim_{x \to b} \left(\frac{1}{(y - 3by + 2b^2)} - \frac{7}{(2y^2 - 3by + b^2)} \right)$

(iv) $\lim_{x \to 3} (x^2 - 9)\left(\frac{1}{(x+3)} + \frac{1}{(x-3)} \right)$

(v) $\lim_{x \to 2} \left(\frac{x^8 - 16}{x^4 - 4} + \frac{x^2 - 9}{x^2 - 3} \right)$

Answers

(i) 1/7 (ii) a/b (iii) $-3/b^2$ (iv) 6 (v) 25.

7b.2.5 Method of Rationalization

If the numerator or the denominator or both contain functions of the type $[\sqrt{f(x)} - g(x)]$ or $[\sqrt{f(x)} - \sqrt{g(x)}]$, and the direct method fails to give the limit, we rationalize the given function by multiplying and dividing by $[\sqrt{f(x)} + g(x)]$ or $[\sqrt{f(x)} + \sqrt{g(x)}]$, as the case may be. After simplification of the function, we evaluate the limit by the earlier methods.

Example (16): Evaluate $\lim_{x \to 0} \frac{x}{\sqrt{1+x} - 1}$

Solution: Consider $\frac{x}{\sqrt{1+x} - 1} = \frac{x}{\sqrt{1+x} - 1} \times \frac{\sqrt{1+x} + 1}{\sqrt{1+x} + 1} = \frac{x(\sqrt{1+x} + 1)}{x}$

$\therefore$ Given limit is $\lim_{x \to 0} \frac{x(\sqrt{1+x} + 1)}{x} = \lim_{x \to 0} (\sqrt{1+x} + 1), \quad [\because x \neq 0]$

$$= \sqrt{1+0} + 1 = 1 + 1 = 2 \quad \text{Ans.}$$

Example (17):

$$\lim_{x \to 3} \frac{x-3}{\sqrt{x-2} - \sqrt{4-x}}$$

Solution: Consider

$$\frac{x-3}{\sqrt{x-2} - \sqrt{4-x}} = \frac{x-3}{\sqrt{x-2} - \sqrt{4-x}} \times \frac{\sqrt{x-2} + \sqrt{4-x}}{\sqrt{x-2} + \sqrt{4-x}}$$

$$= \frac{(x-3)(\sqrt{x-2} + \sqrt{4-x})}{(x-2) - (4-x)}$$

$$= \frac{(x-3)(\sqrt{x-2} + \sqrt{4-x})}{2x - 6}$$

$$= \frac{(x-3)(\sqrt{x-2} + \sqrt{4-x})}{2(x-3)}$$

$\therefore$ Required limit is $\lim_{x \to 3} \frac{(x-3)(\sqrt{x-2} + \sqrt{4-x})}{2(x-3)}$

$$= \lim_{x \to 3} \frac{(\sqrt{x-2} + \sqrt{4-x})}{2} \qquad [\because x \to 3, \therefore x \neq 3, (x-3) \neq 0]$$

$$= \frac{(\sqrt{3-2} + \sqrt{4-3})}{2} = \frac{(\sqrt{1} + \sqrt{1})}{2} = \frac{1+1}{2} = 1 \quad \text{Ans.}$$

Example (18): Evaluate $\lim\limits_{x\to 0}\dfrac{\sqrt{a+x}-\sqrt{a-x}}{\sqrt{b+x}-\sqrt{b-x}}$

Solution: Consider $\sqrt{a+x}-\sqrt{a-x}$

$$= (\sqrt{a+x}-\sqrt{a-x})\times\frac{\sqrt{a+x}+\sqrt{a-x}}{\sqrt{a+x}+\sqrt{a-x}}$$

$$= \frac{(a+x)-(a-x)}{\sqrt{a+x}+\sqrt{a-x}} = \frac{2x}{\sqrt{a+x}+\sqrt{a-x}}$$

Consider $\sqrt{b+x}-\sqrt{b-x}$

$$= (\sqrt{b+x}-\sqrt{b-x})\frac{\sqrt{b+x}+\sqrt{b-x}}{\sqrt{b+x}+\sqrt{b-x}}$$

$$= \frac{2x}{\sqrt{b+x}+\sqrt{b-x}}$$

$\therefore$ Given limit is $\lim\limits_{x\to 0}\dfrac{2x}{\sqrt{a+x}+\sqrt{a-x}}\div\dfrac{2x}{\sqrt{b+x}+\sqrt{b-x}}$

$$= \lim_{x\to 0}\frac{\sqrt{b+x}+\sqrt{b-x}}{\sqrt{a+x}+\sqrt{a-x}}$$

$$= \frac{\sqrt{b+0}+\sqrt{b-0}}{\sqrt{a+0}+\sqrt{a-0}} = \frac{2\sqrt{b}}{2\sqrt{a}} = \sqrt{\frac{b}{a}} \quad \text{Ans.}$$

Exercise [Method of Rationalization]

(3) Evaluate the following limits:

(i) $\lim\limits_{x\to 2}\dfrac{(\sqrt{x^2+5}+\sqrt{5x-1})}{x^2-4}$

(ii) $\lim\limits_{x\to 0}\dfrac{x}{1-\sqrt{1-x}}$

(iii) $\lim\limits_{h\to 0}\dfrac{\sqrt{x+h}-\sqrt{x}}{h}$

(iv) $\lim\limits_{x\to 0}\dfrac{\sqrt{1+x}-\sqrt{1-x}}{x}$

(v) $\lim\limits_{x\to 0}\dfrac{\sqrt{a+2x}-\sqrt{3x}}{\sqrt{3a+x}-\sqrt{x}}, (a\neq 0)$

(vi) $\lim\limits_{x\to 1}\dfrac{(\sqrt{x^2-1}+\sqrt{x-1})}{\sqrt{x^3-1}}$

[Hint: Take $\sqrt{x-1}$ common from numerator and denominator.].

Answers

(i) $-1/24$ (ii) 2 (iii) $1/2\sqrt{x}$ (iv) 1 (v) $1/\sqrt{3}$ (vi) $(\sqrt{2}+1)/\sqrt{3}$

7b.3 LIMIT AT INFINITY

Evaluating limit(s) of the form $\lim_{x \to \pm\infty} f(x)$.

The concept of infinity (∞) was introduced in Chapter 2, and the concept of interval involving infinity [i.e., (a, ∞) $[a, \infty)$, $(-\infty, b)$, $(-\infty, b]$ and $(-\infty, \infty)$] was introduced in Chapter 3. Thus, $(3, \infty)$ is *our way of denoting the set of all real numbers greater than* 3, and similarly we denote *the set of all numbers less than or equal to 5* by the interval $(-\infty, 5]$.

We know that infinity (∞) *does not represent a number*. In this section, we will use the symbols ∞ and $-\infty$ in a *new way, maintaining the same clear understanding about the concept.*(3)

Consider the function $f(x) = (x/(1 + x^2))$. We ask the question: *What happens to f(x) as x gets larger and larger?* In symbols, we ask for the value of $\lim_{x \to \infty} f(x)$. We use the symbol $x \to \infty$ as a shorthand way of saying that x gets larger and larger without bound.

(When we write $x \to \infty$, we are *not* implying that somewhere *far, far to the right on the x-axis*, there is a number bigger than all other numbers to which x is approaching. Rather, we use $x \to \infty$ to say that x is permitted to assume larger and larger values endlessly.)

In Table 7b.1, we have listed values of *f*, for larger and larger values of x, for several values of x.

It appears that *f*(x) *gets smaller and smaller as x gets larger and larger*. Therefore, we write $\lim_{x \to \infty} = \left(x/(1 + x^2)\right) = 0$.

Experimenting with *large negative values of x*, would again lead us to write $\lim_{x \to -\infty} = \left(x/(1 + x^2)\right) = 0$. We say that the limit of *f*(x) at infinity is 0.

7b.3.1 Rigorous Definitions of Limits as $x \to \pm\infty$

In analogy with our ε, δ definition for ordinary limits, we make the following definitions.

Definition (a): (Limit as $x \to \infty$)
Let *f* be defined on $[a, \infty)$ *for some number* "*a*". We say that a number l is the *limit of f*(x) *as x approaches* ∞, if for *every* $\varepsilon > 0$ there is a *corresponding* number M, such that

$$x > M \Rightarrow |f(x) - l| < \varepsilon.$$

TABLE 7b.1

x	$f(x) = \frac{x}{(1+x^2)}$
10	0.099
100	0.010
1000	0.001
10,000	0.0001
$\downarrow$	$\downarrow$
∞	?

(3) The concept of "infinity" has inspired and also confused mathematicians from time immemorial. The deepest problems and profoundest paradoxes of mathematics are often intertwined with the use of this word. Yet, mathematical progress can in part be measured in terms of understanding of the role of infinity. [*Calculus with Analytic Geometry* (Fifth Edition) by Edwin J. Purcell and Dale Verberg (p. 184), Prentice-Hall Publication.]

In this case, we write, $\lim_{x \to \infty} f(x) = l$.

We say that *the limit of f(x) exists* as $x \to \infty$ (*or that f has a limit at* ∞).

Definition (b): (Limit as $x \to -\infty$)

Let "f" be defined on $(-\infty, a]$ *for some number* "a". We say that a number l is the limit of $f(x)$ as x approaches $-\infty$, if for every $\varepsilon > 0$ there is a *corresponding* number M such that

$$x < M \Rightarrow f(x) - l < \varepsilon.$$

In this case, we write, $\lim_{x \to -\infty} f(x) = l$.

We say that *the limit of f(x) exists* as $x \to -\infty$ (or that *f has a limit at* $-\infty$).

Remark: Definitions (a) and (b) will remain unchanged even when *f* is defined on the intervals (a, ∞) and $(-\infty, a)$, respectively.

Note (4): The number M in the above definitions corresponds to the number δ as in all other definitions of limit(s) so far. We think of x as "close to" ∞ when $x > M$, just as we say that x is "close to" a when $a - \delta < x < a$. M can depend on ε. In general, *the smaller* ε *is, the larger M will have to be.*

Example (19): To show that if K is a *positive integer*, then

$$\lim_{x \to \infty} \frac{1}{x^K} = 0 \quad \text{and} \quad \lim_{x \to -\infty} \frac{1}{x^K} = 0$$

Solution: Let $\varepsilon > 0$ be given. We have to find a (positive) number M such that

$$x > M \Rightarrow \left| \frac{1}{x^K} - 0 \right| < \varepsilon$$

$$\Rightarrow \left| \frac{1}{x^K} \right| < \varepsilon,$$

$$\Rightarrow \frac{1}{x^K} < \varepsilon, \quad \begin{cases} \text{since we want } x \text{ to be greater than a} \\ \text{positive number } M, \text{ it follows that } x > 0. \\ \text{Hence, } x^K > 0. \end{cases}$$

$$\Rightarrow x^K > \frac{1}{\varepsilon}$$

$$\Rightarrow x > \sqrt[K]{1/\varepsilon}$$

Thus, $M \geq \sqrt[K]{1/\varepsilon}$ implies $|(1/x^K) - 0| < \varepsilon$. It follows that, $\lim_{x \to \infty} 1/x^K = 0$. Similarly, we can prove that $\lim_{x \to -\infty} 1/x^K = 0$. (In particular, $\lim_{x \to \infty} 1/x = 0$ and $\lim_{x \to -\infty} 1/x = 0$.) These *limits* must be treated as *standard limits*.

We must face the question of whether the main limit theorem (i.e., Theorem A) holds for them. The answer is yes. *We accept the corresponding statements of the limit theorem, without proof.*

Thus, the limit as x approaches ∞ (or $-\infty$) is unique, when it exists. Furthermore, if $\lim_{x\to\infty} f(x)$ and $\lim_{x\to\infty} g(x)$ exist, we have

$$\lim_{x\to\infty} [f(x) + g(x)] = \lim_{x\to\infty} f(x) + \lim_{x\to\infty} g(x)$$

$$\lim_{x\to\infty} f(x) \cdot g(x) = \lim_{x\to\infty} f(x) \cdot \lim_{x\to\infty} g(x)$$

(There are corresponding formulas for limits at $-\infty$.)

Example (20): Prove that $\lim_{x\to\infty} \dfrac{x}{1+x^2} = 0.$

Solution: Here we use a *standard trick*: dividing numerator and denominator by *the highest power of x that appears in the denominator*.

$$\lim_{x\to\infty} \frac{x}{1+x^2} = 0 = \lim_{x\to\infty} \frac{x/x^2}{(1+x^2)/x^2} = \lim_{x\to\infty} \frac{1/x}{1/x^2+1}$$

$$= \frac{\lim_{x\to\infty} 1/x}{\lim_{x\to\infty} 1/x^2 + \lim_{x\to\infty} 1} = \frac{0}{0+1} = 0$$

Example (21): To find $\lim_{x\to-\infty} \dfrac{2x^3}{1+x^3}$

Solution: Divide numerator and denominator by x^3, we get

$$\lim_{x\to-\infty} \frac{2x^3}{1+x^3} = \lim_{x\to-\infty} \frac{2}{1/x^3+1} = \frac{2}{0+1} = 2.$$

Remark: We can think of the $\lim_{x\to\infty} f(x)$ as a kind of *left-hand limit*, *because x approaches* ∞ *from the left*. Similarly, we can think of $\lim_{x\to-\infty} f(x)$ as the *right-hand limit*.

Exercise

(4) Evaluate the following limits:[(4)]

(i) $\lim_{x\to\infty} \dfrac{2x^2-4x+5}{3x^3-x+7}$

(ii) $\lim_{x\to\infty} \dfrac{(2x-1)^{20} \cdot (3x-1)^{30}}{(2x+1)^{50}}$

(iii) $\lim_{x\to\infty} \left(\sqrt{x+1} - \sqrt{x}\right)$

(4) [Hint: Consider the highest powers of the terms involving x, in both numerator and denominator, and proceed to compute the limit.]

(iv) $\lim_{x\to\infty} \dfrac{\sqrt{x^2+2}-x}{\sqrt{x^2+3}-x}$

(v) $\lim_{x\to\infty} \dfrac{\sqrt{x^2+5}-\sqrt{x^2+3}}{\sqrt{x^2+3}-\sqrt{x^2+1}}$

Answers

(i) 0 (ii) $(3/2)^{30}$ (iii) 0 (iv) 2/3 (v) 1.

7b.4 INFINITE LIMITS

$\lim_{x\to a} f(x) = \pm\infty$, where "*a*" is finite.[(5)]

So far we have considered the cases where as $x \to a$ (a finite number), $f(x) \to l$, (a finite number). But, it may happen that *as x → a, f(x) increases (or decreases) endlessly*. Symbolically, we express these statements as follows:

$$x \to a \Rightarrow f(x) \to \infty \quad \text{or} \quad \lim_{x\to a} f(x) = \infty$$

$$\text{or} \quad x \to a \Rightarrow f(x) \to -\infty \quad \text{or} \quad \lim_{x\to a} f(x) = -\infty$$

Consider the graph of $f(x) = 1/(x-2)$, as shown in Figure 7b.1.

Note that it makes no sense to ask for $\lim_{x\to 2} 1/(x-2)$ (why?), but we think it is reasonable to write $\lim_{x\to 2^-} 1/(x-2) = -\infty$, and $\lim_{x\to 2^+} 1/(x-2) = \infty$. The following definition relates to this situation.

7b.4.1 Definition (Infinite Limits)

We say that $\lim_{x\to a^+} f(x) = \infty$, if for *each positive number M*, there corresponds a $\delta > 0$, such that $0 < x - a < \delta \Rightarrow f(x) > M$.

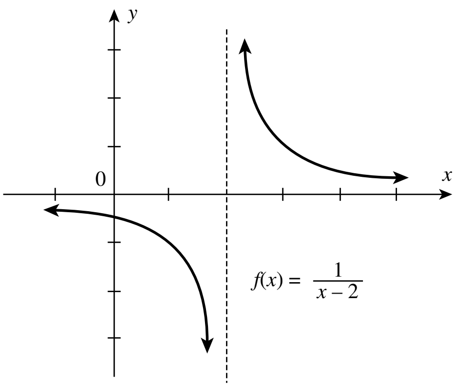


FIGURE 7b.1

[(5)] So far, only finite numbers were considered to be the limit(s) of function(s). Now, we shall consider infinite limits.

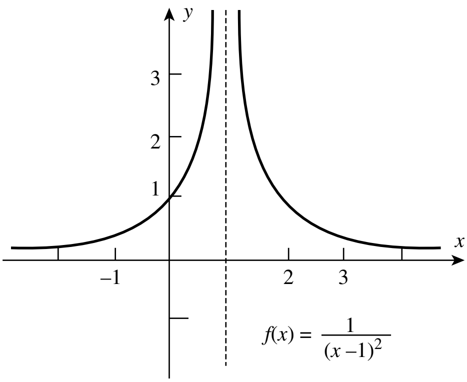


FIGURE 7b.2

There are corresponding definitions of $\lim_{x\to a^+} f(x) = -\infty$, $\lim_{x\to a^-} f(x) = \infty$, and $\lim_{x\to a^-} f(x) = -\infty$.

Example (22): Find $\lim_{x\to 1^-} \dfrac{1}{(x-1)^2}$ and $\lim_{x\to 1^+} \dfrac{1}{(x-1)^2}$

Solution: The graph of $f(x) = \left(1/(x-1)^2\right)$ is shown in Figure 7b.2.

We think it is quite clear that

$$\lim_{x\to 1^-} \frac{1}{(x-1)^2} = \infty,$$
$$\lim_{x\to 1^+} \frac{1}{(x-1)^2} = \infty.$$

Since both limits are ∞*, we could also write*

$$\lim_{x\to 1} \frac{1}{(x-1)^2} = \infty \qquad f(x) = \frac{1}{(x-1)^2}$$

Example (23): Find $\lim_{x\to 2^+} \dfrac{x+1}{x^2+5x+6}$

Solution: $\lim_{x\to 2^+} \dfrac{x+1}{x^2+5x+6} = \lim_{x\to 2^+} \dfrac{x+1}{(x-3)(x-2)}$

As $x \to 2^+$, we see that $x+1 \to 3$, $x-3 \to -1$, and $x-2 \to 0^+$. Thus, the numerator is approaching 3, but the denominator is negative and approaching 0. We conclude that $\lim_{x\to 2^+} \left((x+1)/(x-3)(x-2)\right) = -\infty$.

The concept of limit at infinity [i.e., $\lim_{x\to\infty} f(x)$, $\lim_{x\to -\infty} f(x)$], and the meaning of infinite limits [$\lim_{x\to a} f(x) = \pm\infty$] as discussed above, make it easy to introduce the concept of asymptote(s).

7b.5 ASYMPTOTES

Definition: An asymptote *to a curve* is defined as a *straight line*, which has the *property* that the *distance from a point on the curve to the line tends to zero* as *the distance* of this point to the *origin* increases *without bound.* There are *vertical, horizontal, and oblique asymptotes.*

7b.5.1 Vertical Asymptotes

The *graph of the function* $y = f(x)$ has a *vertical asymptote* for $x \to a$, if $\lim_{x \to a} f(x) = +\infty$ or $\lim_{x \to a} f(x) = -\infty$ (see Figure 7b.3a and b).

Note (5): In the case of a *vertical asymptote* for $x \to a$, the point $x = a$ is a *point of discontinuity*. (In Chapter 8, this is classified under the discontinuity of the *second kind.*) The equation of the vertical asymptote has the form $x = a$. (In Figure 7b.3a, it is $x = 0$, and in Figure 7b.3b it is $x = a$.)

7b.5.2 Horizontal Asymptotes

The graph of the function $y = f(x)$ for $x \to +\infty$ or for $x \to -\infty$, has a horizontal asymptote, if $\lim_{x \to +\infty} f(x) = b$ or $\lim_{x \to -\infty} f(x) = b$, where b is a *finite number*.

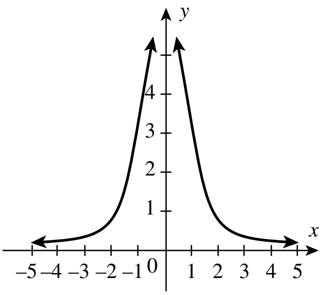
It may happen that *either only one or none* of these limits is finite. Then, the graph has *either one or no horizontal asymptote*. Of course, the graph of a function may have *two horizontal asymptotes.*

The equation of the horizontal asymptote has the form $y = a$.

(In Figure 7b.4a, it is $y = b$, and in Figure 7b.4b the two asymptotes are $y = 1$ and $y = -1$.)

7b.5.3 Oblique Asymptotes

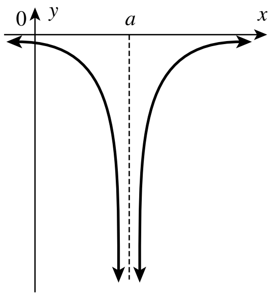
In Figure 7b.5a and b, it is indicated that the graph of the function $y = f(x)$ has an oblique asymptote $y = kx + b$.



(*a*)

$f(x) = \dfrac{3}{x^2}$

x	$f(x) = \dfrac{3}{x^2}$
1	3
0.5	12
0.25	48
0.1	300
0.01	30,000
0.001	3,000,000



(*b*)

$g(x) = -\dfrac{1}{(x-a)^2}$

FIGURE 7b.3 Vertical Asymptotes.

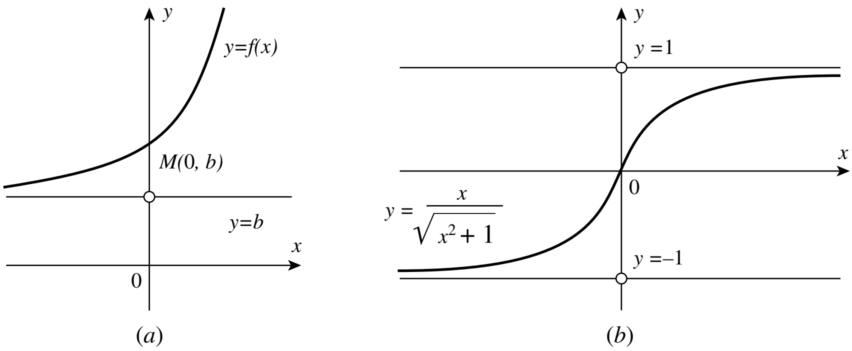


FIGURE 7b.4 Horizontal Asymptotes.

In this case, the following equality holds true.

$$\lim_{x \to \pm\infty} [f(x) - (kx + b)] = 0$$

or
$$\lim_{x \to \pm\infty} [f(x) - kx - b] = 0 \tag{3}$$

Taking out x, as a factor, we get

$$\lim_{x \to \pm\infty} x\left[\frac{f(x)}{x} - k - \frac{b}{x}\right] = 0$$

$$\Rightarrow \frac{f(x)}{x} - k - \frac{b}{x} = 0 \tag{4}$$

Now, observe that $\lim_{x \to \pm\infty} b/x = 0$ always. Thus, we get the formulas for computing the parameters k and b given by

$$\lim_{x \to \pm\infty} [f(x) - kx] = b \quad \text{from (3), and}$$
$$\lim_{x \to \pm\infty} f(x)/x = k \quad \text{from (4)}$$

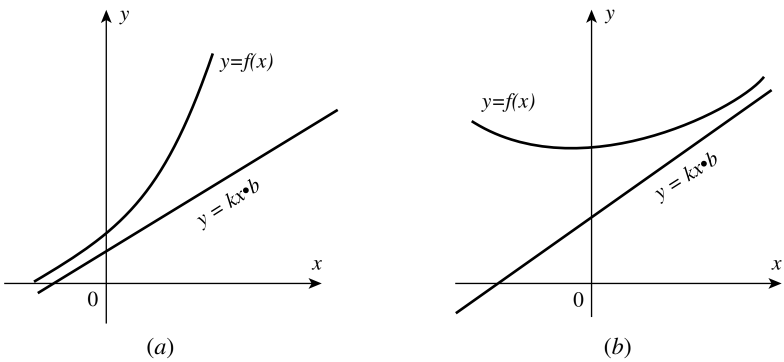


FIGURE 7b.5 Oblique Asymptotes.

Thus, we get the procedure for finding oblique asymptote for the given curve as follows: The constant k is given by the limit $\lim_{x\to\pm\infty} f(x)/x = k$, and the constant b is obtained by computing the limit $\lim_{x\to\pm\infty} [f(x) - kx] = b$.

Having found the values of k and b, we can write down the equation (of straight line) representing the oblique asymptote.

Note (6): For finding the asymptotes to the given curves, both the cases $x \to +\infty$ and $x \to -\infty$, should be considered separately.

Example (24): Find the asymptotes to the curve $y = \frac{1}{x-3}$

Solution: We have $\lim_{x\to\pm\infty} 1/(x-3) = 0$. Therefore, the curve has a horizontal asymptote at $y = 0$. Further, we observe that

$$\lim_{x\to 3^-} \frac{1}{x-3} = -\infty \text{ and } \lim_{x\to 3^+} \frac{1}{x-3} = +\infty.$$

Hence, the curve has a vertical asymptote at $x = 3$ (see Figure 7b.6).

Example (25): Find the oblique asymptotes to the curve $y = \frac{x^2}{x-1}$

Solution: From the given equation, we obtain

$$\begin{aligned} k &= \lim_{x\to\pm\infty} \frac{f(x)}{x} = \lim_{x\to\pm\infty} \frac{x^2}{(x-1)\cdot x} = \lim_{x\to\pm\infty} \frac{x}{x-1} \\ &= \lim_{x\to\pm\infty} \frac{x-1+1}{x-1} = \lim_{x\to\pm\infty} \left(1 + \frac{1}{x-1}\right) = 1 \end{aligned}$$

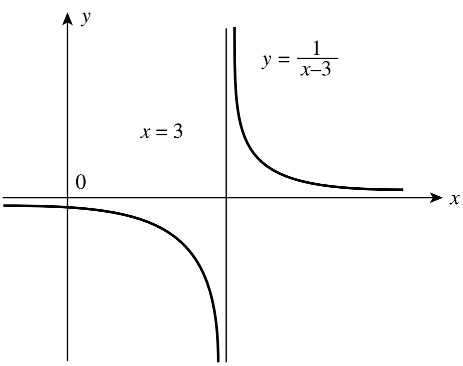


FIGURE 7b.6

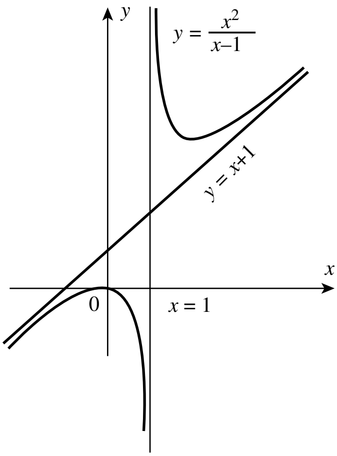


FIGURE 7b.7

and

$$\begin{aligned} b &= \lim_{x\to\pm\infty}[f(x)-kx] = \lim_{x\to\pm\infty}\left[\frac{x^2}{(x-1)}-x\right] \\ &= \lim_{x\to\pm\infty}\left[\frac{x^2-x+x}{(x-1)}\right] = \lim_{x\to\pm\infty}\left[\frac{x}{(x-1)}\right] \\ &= \lim_{x\to\pm\infty}\frac{x-1+1}{x-1} = \lim_{x\to\pm\infty}\left(1+\frac{1}{x-1}\right) = 1 \end{aligned}$$

Thus, $k=1$ and $b=1$. Consequently, for $x \to +\infty$ or for $x \to -\infty$, the graph of the function has an oblique asymptote, $y=kx+b=x+1$ (see Figure 7b.7).

Remark: Observe that the curve shown in Figure 7b.7 also has the vertical asymptote, x=1.

TABLE 7b.2 Good and Bad Uses of Infinity (∞)[a]

Expression	Is it *Right* or *Wrong*? What it Means? Remarks
$\frac{1}{\infty}=0$	Usually *right*, but likely to create confusion. Here, we really mean to say that $\lim_{x\to\infty} 1/x = 0$, which is a right statement
$3\cdot\infty=\infty$	This is *right*, as mathematical shorthand. It means if a quantity increases without bound, so does three times that quantity
$\infty+\infty=\infty$	This is *right* again. It means if two quantities increase without bounds, so does their sum
$\frac{1}{0}=\infty$	This is *wrong*. Division by 0 is not defined for real numbers. Besides, note that $\lim_{x\to 0^+} 1/x=\infty$, but $\lim_{x\to 0^-} 1/x=-\infty$. Hence, it is *worse* to write the expression under consideration
$\infty-\infty=0$	This is *wrong* again. Note that, as $x\to\infty$, $x^3\to\infty$, and $x^2\to\infty$ but $x^3-x^2\to\infty$[b]
$\frac{\infty}{\infty}=1$	This is *wrong*. Note that, as $x\to\infty$, $x^3\to\infty$, and $x^2\to\infty$, but $x^3/x^2=x\to\infty$. Again, $x^2/x^3=1/x\to 0$

[a] The reader may also refer to the algebra of infinity (∞), given at the end of Chapter 2.

[b] $\lim_{x\to\infty}(x^3-x^2) = \lim_{x\to\infty} x^3\left(1-\frac{1}{x}\right)=\lim_{x\to\infty}(x^3)=\infty$.

8 The Concept of Continuity of a Function, and Points of Discontinuity

8.1 INTRODUCTION

The study of calculus begins with the concept of limit introduced and discussed in Chapters 7a and 7b. Of all the many consequences of this concept, one of *the most important* is the concept of a *continuous function*. *One cannot think of the subject of calculus without continuous functions*, which we study now.

The word *continuous* means much the same in mathematics as in everyday language. We can introduce *the concept of continuity proceeding from a graphic representation of a function*. A function is continuous if its graph *is unbroken, i.e., free from sudden jumps or gaps*.

Suppose a function is defined on an interval *I*. We say that the function is *continuous on the interval I*, if its graph consists of *one continuous curve*, *so that it can be drawn without lifting the pencil*. There is no break in any of the graphs of continuous functions (Figure 8.1a–c).

If the graph of a function is *broken at any point* "*a*" of an interval, we say that the function is *not continuous* (or that it *is discontinuous*) at "*a*". We give the following definition:

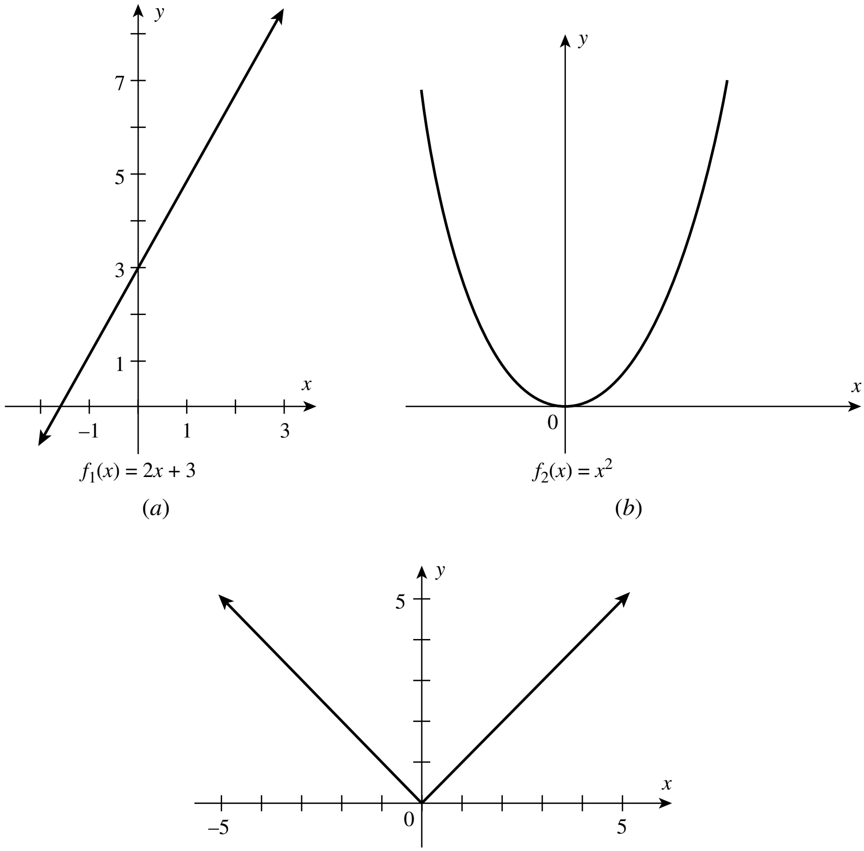
Definition: A function is *discontinuous* at $x = a$, *if and only if* it is *not continuous* at $x = a$.

This point "*a*" is called *the point of discontinuity* of the function. The domain of a function plays an important role in the definition of continuity (and discontinuity) of a function. A function may be continuous on one set but discontinuous on another set. It is useful to recall the definitions of *the domain of definition* and *the natural domain* of a function, from Chapter 6.

- The set of all *those numbers that can be used in the definition of a function* (which we call "*input numbers*") constitute *the domain of definition* (or simply *the domain*) of the function.

What must you know to learn calculus? 8-Continuity of functions and the points of discontinuity. The definition of continuity "at a point" and "in an interval." Types of discontinuities and some theorems defining properties of continuous functions.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.



$f_1(x) = 2x + 3$ (*a*) $f_2(x) = x^2$ (*b*)

$f_3(x) = |x|$

(*c*)

FIGURE 8.1

- If *the domain* of the given function is *not specified*, we take the domain as the *largest set of real numbers for which the rule of the function makes sense and gives real-number values*. This is called *the natural domain* of the function.

Note that, *the natural domain* of a *given* function $f(x)$ is a *fixed set* of points for which $f(x)$ is defined. *It does not include those points* at which $f(x)$ is *not defined*. On the other hand, *the domain of definition* of a given function $f(x)$ is the set of *all input numbers* that can be used in the definition of a function. (It may include even those numbers at which the function is not defined.). *It is not a fixed set of points. The* "*domain of definition*" of a function *can be varied*. Of course, when we change the domain of a function *we define a new function*.

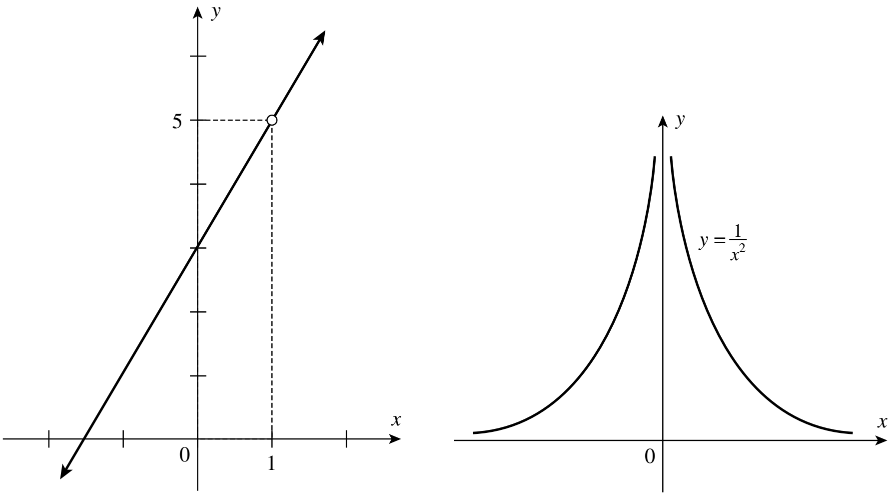
For the purpose of studying the property of continuity (and discontinuity) of a function we shall always take *the domain* of a function as an *interval*. One may also consider a domain, which is *the union of an interval with some isolated points*. Obviously, such a domain is more general (than an interval) for the purpose of discussion. Here, we may state (in advance) that a point of discontinuity can be any point "*a*", *provided there exists some neighborhood of* "*a*" *in which the function is defined. The function may or may not be defined at a.* It follows that an isolated point (if any) of a domain cannot be a point of discontinuity. In fact, a function whose *domain of definition* is a singleton "a" is considered continuous at "a". Intuitively, this statement might appear to create a situation of confusion but it is true as will be seen when we define one-sided continuity of a function. The conclusion follows from the definition of continuity "at a point" (to be discussed later), and the fact that a constant sequence $\{a, a, a,\}$ converges to "a", which means both the one-sided limits defined at "a" are equal. We do not discuss sequences and their properties in this compilation.

For example, the natural domain of the function $f(x) = 1/(x-5)$ is $\{x \in R | x \neq 5\}$. We exclude "5" (from the domain of "f") to avoid *division by zero*. Note that, "f" is defined for each x in its *natural domain*. We can also say that the functions $g(x) = 1/(x-5), x \in (5, \infty)$ and $h(x) = 1/(x-5)$, $x \in (-\infty, 5)$, (wherein the number "5" is excluded from the definitions of these functions) are defined for each x in their respective domains.

We can also include the number "5" in the definition of such a function. For example, consider, $F(x) = 1/(x-5), x \in [4, 7]$, and so on. We say that *the domain (of definition) of F is the interval* [4, 7] in which the function F is *not defined* at $x = 5$. (Shortly, it will be seen that any such point "a" is the point of discontinuity of "f.")

To understand the concept of continuity better, it is useful to study the following graphs of functions, which represent *discontinuous functions*.

The graph of the function $f_1(x)$, appears in Figure 8.2a. It consist of all points on the line $y = 2x + 3$, except $(1, 5)$. The graph *has a break* at the point $(1, 5)$. Here $f_1(x)$ is *not continuous* at $x = 1$ since "1" is not in the domain of $f_1(x)$. We say that $f_1(x)$ is not defined at $x = 1$. We can also say that $f_1(x)$ is continuous for all x, except for $x = 1$. It is also correct to say that $f_1(x)$ is discontinuous at $x = 1$ (or that it is discontinuous in any interval containing "1").



$f_1(x) = 2x + 3, x \neq 1$

(a)

$f_2(x) = \frac{1}{x^2}, x \neq 0$

(b)

FIGURE 8.2

Note (1): Some authors do not prefer to say that $f_1(x)$ is discontinuous at $x = 1$ (or a function like $1/x^2$ is discontinuous at $x = 0$). They are of the opinion that we should not consider the question of continuity (or discontinuity) at a point that is not in the domain of the function. (We shall come back to this discussion, shortly.)

Now consider the function $f_2(x) = 1/x^2$, $x \neq 0$. Its graph appears in the Figure 8.2b. Observe that as $x \to 0$, $1/x^2 \to \infty$, which means that $f_2(x)$ does not exist at $x = 0$ or that $f_2(x) = 1/x^2$ is not defined at $x = 0$. We say that in any interval containing "0", the function $f_2(x)$ is discontinuous at the point $x = 0$. This is an example of infinite discontinuity to be discussed later.

Note (2): We say that a function $f(x)$ is not defined at $x = a$ if either "a" is not in the domain of $f(x)$ or $f(x) \to \infty$ as $x \to a$.

We give below *some more situations* when a function may be discontinuous "at a point", in *the interval* of its definition. The functions $f_3(x)$ and $f_4(x)$ *are defined for all* x. Note that the point (1, 5) is torn out from the graph of $f_3(x)$ and shifted to the location (1, 2). Here, the point (1, 5) of the graph jumps out from the height 5 to 2, creating a break in the graph at $x = 1$ (Figures 8.3 and 8.4).

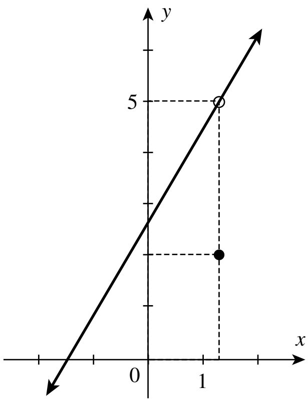
The graph of the function $f_4(x)$, shows a break at the point $x = 1$. Here, a portion of the graph has a finite vertical jump at $x = 1$ making the graph discontinuous at $x = 1$.

Next, consider the graphs of the functions $f_5(x)$ and $f_6(x)$ as indicated in Figures 8.5 and 8.6, respectively.

The function $f_5(x)$ is defined for all x. There is a *finite jump* in the graph *suddenly* at $x = 0$ [as in the case $f_4(x)$] causing a *break*. Thus, $f_5(x)$ is *discontinuous* at $x = 0$.

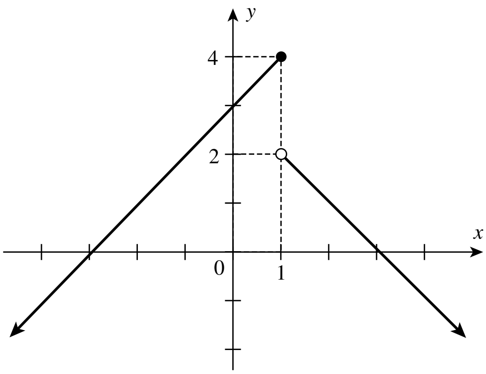
The function $f_6(x)$ is *not defined* at $x = 0$ but it is defined for all other values of x. We observe that as $x \to 0^+$, $1/x \to \infty$, and as $x \to 0^-$, $1/x \to -\infty$. (This is another example of *infinite discontinuity*, to be discussed later.)

From the above discussion (and the graphs), it is clear that the question of continuity must be considered only for those points, which are in the domain of the function. However, a point of discontinuity may or may not be in the domain of the function.



$$f_3(x) = \begin{cases} 2x + 3 & \text{if } x \neq 1 \\ 2 & \text{if } x = 1 \end{cases}$$

FIGURE 8.3



$$f_4(x) = \begin{cases} 3 + x, & \text{if } x \le 1 \\ 3 - x, & \text{if } x > 1 \end{cases}$$

FIGURE 8.4

Note (3): If a function "*f*" is not defined at some point "*a*" (say), then "*a*" may not be a point of discontinuity of "*f*". This will be clear from the Examples (1) and (2) to follow shortly. (The important point to be emphasized is that if "*f*" is defined on an interval containing "*a*", but "*f*" is not defined at "*a*", then "*a*" must be the point of discontinuity.)

Now, we give the *intuitive definition* of continuity of a function "at a point".

8.1.1 Intuitive Definition of Continuity of a Function at Any Point "*a*"

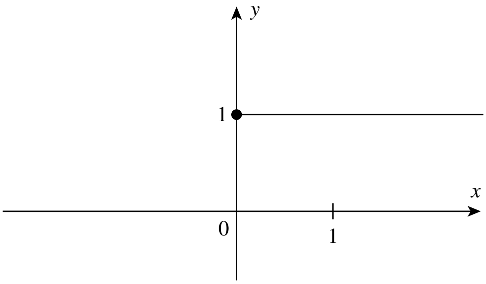
Let $y = f(x)$ be a function defined on an interval I, which contains a point "*a*" in its *interior* or on its *boundary.*

Roughly speaking the function "*f*" is *continuous* at the point $x = a$, provided that its graph does not have a break at $x = a$.[1]

8.1.1.1 Points of Discontinuity of a Function *An elementary function* can have a discontinuity only at *separate points* of a certain interval *but not at all of its points.* [The Dirichlet Function (see Chapter 6, Section 6.10) which is defined throughout the real line is *not continuous at any point*. Of course, it is not an elementary function.] The following Figure 8.7a and b are the *graphs of functions which are discontinuous as indicated there.*

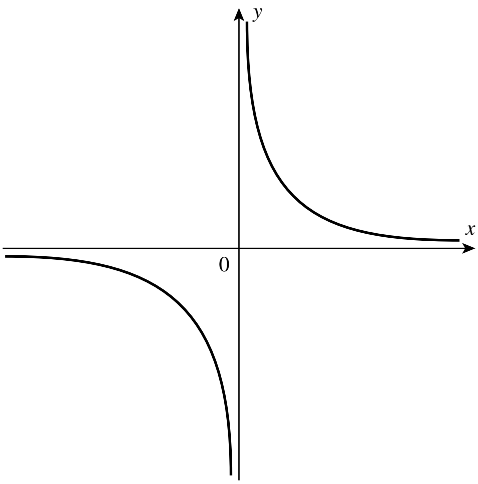
The function represented by the graph in Figure 8.7a is *discontinuous* at $x = 1$ and $x = 2$. It is *continuous* at all other points of its domain. The function graphed in Figure 8.7b is *discontinuous* at $x = 1$, $x = 2$, and $x = 4$. It is *continuous* at all other points of its domain.

[1] If the graph of a function has a break in its interior, then it is not difficult to imagine such a break. However, if the break occurs at the end point of the graph then it is not easy to visualize such a break. It is for this reason that the phrase: "roughly speaking" is used in the definition.



$$f_5(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x \geq 1 \end{cases}$$

FIGURE 8.5



$f_6(x = \frac{1}{x}, x \neq 0)$

FIGURE 8.6

From a graphical point of view, the following definitions are useful in deciding whether any *point* "*a*" *is a point of discontinuity* of the given function or not.[2]

Definition (a): A point at which a function is *not continuous*, but is defined in its neighborhood, is *a point of discontinuity*.

Definition (b): A point at which a *function is not defined* (*but is defined in a neighborhood*), is *a point of discontinuity*.

[2] *Elements of Higher Mathematics for High School Students* by D. K. Faddeev, M. S. Nikulin, and I. F. Sokolovsky, Mir Publisher, Moscow, 1989.

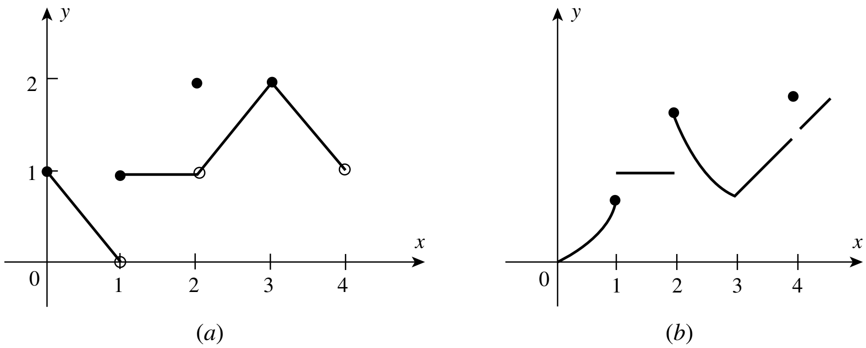


FIGURE 8.7 Graphs of functions with discontinuities: (a) at $x = 1$ and $x = 2$ (b) at $x = 1$, $x = 2$ and $x = 4$.

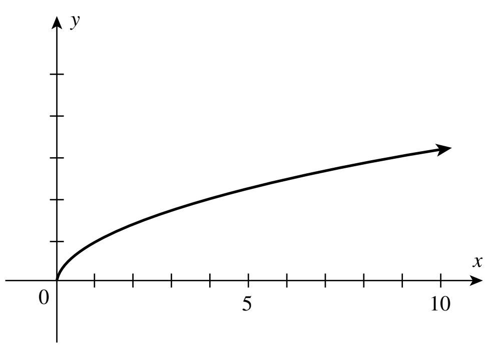
Note (4): We know that the property of discontinuity of a function "f" depends upon the interval I on which the function "f" is defined. Accordingly, an arbitrary point "a" outside the interval I cannot be point of discontinuity of the function (note that "f" may or may not be defined at "a"). The following Examples (1) and (2) make the situation clear.

Example (1): Consider the function $g_1(x) = \sqrt{x}$ which is *defined only for* $x \geq 0$, which means that the domain of $\sqrt{x}$ is $[0, \infty)$. Obviously, "-5" is not in the domain of $\sqrt{x}$ or that $\sqrt{x}$ is not defined at $x = -5$ (see Figure 8.8). Further, observe that there is no neighborhood of "-5" in which the function $\sqrt{x}$ is defined. Hence, the condition of Definition (b) is not satisfied for the number "-5" (or any other negative number). Therefore, the number "-5" cannot be a point of discontinuity of $\sqrt{x}$. Accordingly, it will be wrong to say that $\sqrt{x}$ is discontinuous at "-5," since it is not in the domain of $\sqrt{x}$.

Example (2): Now let us consider the function $g_2(x)$ defined by:

$$g_2(x) = 2x + 3,\ x \in [2,\ 8], \qquad x \neq 3$$

Obviously, this function is *not defined* at $x = 3$, but it is *discontinuous* at $x = 3$ (why?). [Note that there is neighborhood of "3" in which the function $g_2(x)$ is defined, and so "3" is a point of



$g_1(x) = \sqrt{x}$

FIGURE 8.8

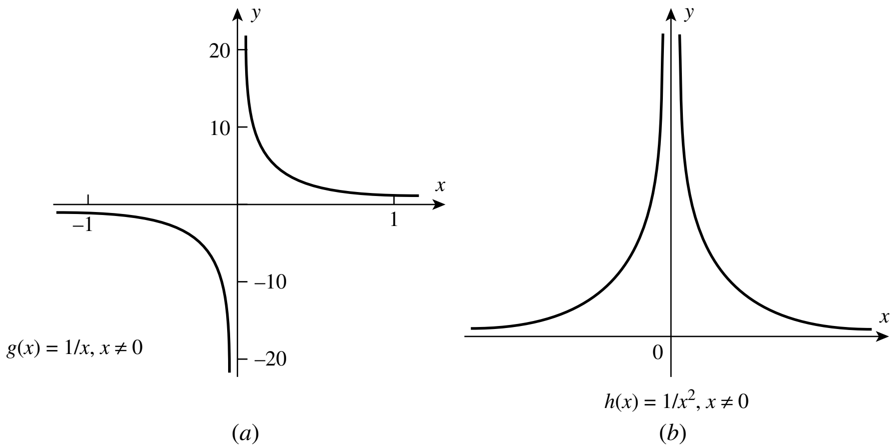


FIGURE 8.9

discontinuity in view of the Definition (b) above.] On the other hand, if we consider any other point outside the interval [2, 8], say "10", then obviously, *it will be wrong to say that* $g_2(x)$ *is discontinuous at* "10".

Now, in view of the Definition (b) it is easy to understand that the functions $G(x) = 1/x$, $x \neq 0$ and $H(x) = 1/x^2$, $x \neq 0$ are both discontinuous at $x = 0$ (see Figure 8.9a and b, respectively). We know that, these functions have *infinite discontinuity* at $x = 0$. However, the function $H(x) = 1/x^2$, $x \neq 0$ is an example of infinite discontinuity with one sign as $x \to 0$. Here, we can also say that both these functions are *continuous on the intervals* $(-\infty, 0)$ and $(0, \infty)$ whose union forms *the natural domain* of these functions.[3]

8.2 DEVELOPING THE DEFINITION OF CONTINUITY "AT A POINT"

Of course, graphical intuition is helpful in understanding the concept of continuity, but a precise definition of continuity cannot depend on pictures.

The notion of continuity can be best expressed *through limits* as will be clear from the following examples.

Example (3): Consider the following functions:

(i) $f(x) = \dfrac{x^2 - 9}{x - 3}, \quad x \neq 3, \quad \text{i.e.,} \ \dfrac{(x-3)(x+3)}{(x-3)}, \quad x \neq 3$

(ii) $g(x) = \begin{cases} x + 3, & \text{if } x \neq 3 \\ 2, & \text{if } x = 3 \end{cases}$

[3] John B. Fraleigh, a world-class author of the book *Calculus with Analytic Geometry* (published by Addison-Wesley, 1979) has expressed (in a footnote on p. 52) that we should not even prefer to consider the question of continuity at a point that is not in the domain of the function. He is of the opinion that whereas we can talk about the continuity of $1/x$ on the intervals $(-\infty, 0)$ and $(0, \infty)$, he is against the statement that $1/x$ is discontinuous at $x = 0$. However, in view of the definition of the domain of a function defined earlier, and the definition (b) given above, we agree to say that $1/x$ is discontinuous at $x = 0$. This is a matter of approach and outlook that helps in accepting the concept of discontinuity with uniformity in our thinking.

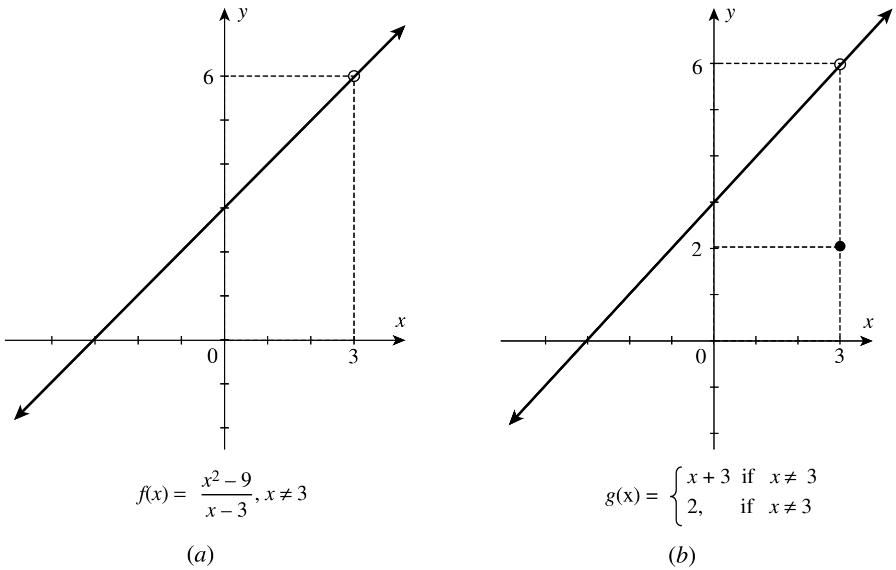


FIGURE 8.10

The graphs of $f(x)$ and $g(x)$ appear in Figure 8.10a and b, respectively. In the graph of $f(x)$ the point (3, 6) is *missing* from the graph. On the other hand, in the graph of $g(x)$, it appears as if the point (3, 6) is torn out from the graph of $g(x)$ and pushed at a new location (3, 2) which is vertically below the point (3, 6) at a lower height. Thus, a *break* is created in these graphs, making both of them *discontinuous* at $x = 3$.

$$(i)\, f(x) = \frac{x^2 - 9}{x - 3}, \quad x \neq 3 \quad g(x) = \begin{cases} x + 3, & \text{if } x \neq 3 \\ 2, & \text{if } x = 3 \end{cases}$$

Stated another way, if we were to trace these graphs with a pencil, we would have to lift the pencil at x = 3. These situations can be technically expressed through *limits* as follows.

(i) The function $f(x) = (x^2 - 9)/(x-3)$ *is not defined* at $x = 3$, or we say that the function $f(x)$ is *meaningless* for $x = 3$ (why?). Here is a preliminary remark: *The concept of limit of a function* $f(x)$, as $x \to a$ *is connected with the behavior of the function in the vicinity of the point "a", except for the point "a" itself.* Note that, the ratio $(x^2-9)/(x-3)$ is *identically equal to the expression* $(x + 3)$ *at all points, except for the point* $x = 3$. Consequently, *in the vicinity of the point* $x = 3$ *as well the functions* $(x^2-9)/(x-3)$ *and* $(x + 3)$ *coincide*, and we have

$$\lim_{x \to 3}(x^2 - 9)/(x - 3) = \lim_{x \to 3}(x + 3) = 6, \quad (x \neq 3)$$

Thus, although the function, $f(x) = (x^2 - 9)/(x - 3) = ((x - 3)(x + 3))/(x - 3)$ is *meaningless, at the point* $x = 3$, *this does not exclude the possibility of the existence of the limit of the function* as $x \to 3$. Also, note that $f(x)$ is *discontinuous* at $x = 3$, for obvious reasons.

Now, if we agree to define $f(x)$ at $x = 3$ by $f(3) = 6$, then it appears as if the missing point (3, 6) is brought back into the gap, making the graph continuous. In other words, the *discontinuity of the function $f(x)$ is removed* by defining $f(x)$ at $x = 3$ *suitably* so that this value equals the limit $f(x)$ as $x \to 3$.

(ii) The function $g(x)$ *has the same function values as the function $f(x)$* when $x \neq 3$. But *it is also given that* $g(3) = 2$. Thus, $g(x)$ is *defined for all values of x, but still there is a break in its graph* at $x = 3$.

If, however, we *redefine* $g(x)$ at $x = 3$, by $g(3) = 6$, *it is equivalent to shifting the point* (3, 2) (of the graph) *to the location* (3, 6), [which is vertically above the point (3, 2)]. *This fills up the gap in the graph* at $x = 3$ and makes it continuous.

The *redefined function* $g(x)$ is given by:

$$g(x) = \begin{cases} x + 3, & \text{if } x \neq 3 \\ 6, & \text{if } x = 3 \end{cases}$$ (4)

Note that by *redefining the function* $g(x)$ at $x = 3$ (which is the point of discontinuity) *the value* $g(3)$ *is numerically made equal to* the $\lim_{x \to 3} g(x)$.

This is *the basis of our definition* of continuity of a function at any point "*a*" in *the domain of definition* of the function. At this state, it is necessary to consider one more *contrasting situation* as indicated in the following example (in which a function is *discontinuous*), before formulating the definition of continuity at any point "*a*" in an interval *I*.

Example (4): Let ϕ be defined by

$$\phi(x) = \begin{cases} 3 + x, & \text{if } x \leq 1 \\ 3 - x, & \text{if } x > 1 \end{cases}$$

Figure 8.11 shows the graph of ϕ. Here, we note that

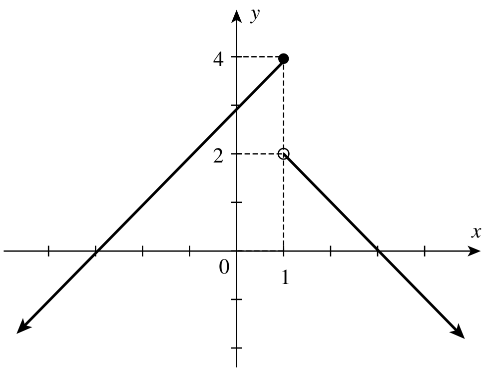


FIGURE 8.11

(4) Here we have retained the same name of the redefined function, to convey that the discontinuity of $g(x)$ can be removed. Technically, this is not correct. The refined function must be denoted by a different notation, say $h(x)$ or $\psi(x)$, and so on.

(i) $\phi(1) = 4$ (as clear from the graph),

(ii) $\lim_{x \to 1^-} \phi(x) = \lim_{x \to 1^-} (3 + x) = 4$, and

(iii) $\lim_{x \to 1^+} \phi(x) = \lim_{x \to 1^+} (3 - x) = 2$.

Here, we know that limit $\lim_{x \to 1^-} \phi(x) \neq \lim_{x \to 1^+} \phi(x)$, *which means that the* $\lim_{x \to 1} \phi(x)$ *does not exist*. We observe that the graph of ϕ has a break at the point $x = 1$, where the $\lim_{x \to 1} \phi(x)$ *does not exist*. From the above discussion, involving the functions $f(x)$, $g(x)$, and $\phi(x)$ we get the following total picture.

Let a function "$f(x)$" be defined on an interval I, and let "a" be an *arbitrary* point in I. Then, there are *three contrasting possibilities* for the behavior of $f(x)$ near "a" as follows:

(i) $\lim_{x \to a} f(x)$ *does not exist* (see Figure 8.12a and b)

(ii) $\lim_{x \to a} f(x)$ *exists*, but $\lim_{x \to a} f(x) \neq f(a)$ (see Figure 8.12c and d)

(iii) $\lim_{x \to a} f(x)$ *exists*, and $\lim_{x \to a} f(x) = f(a)$ (see Figure 8.12e)

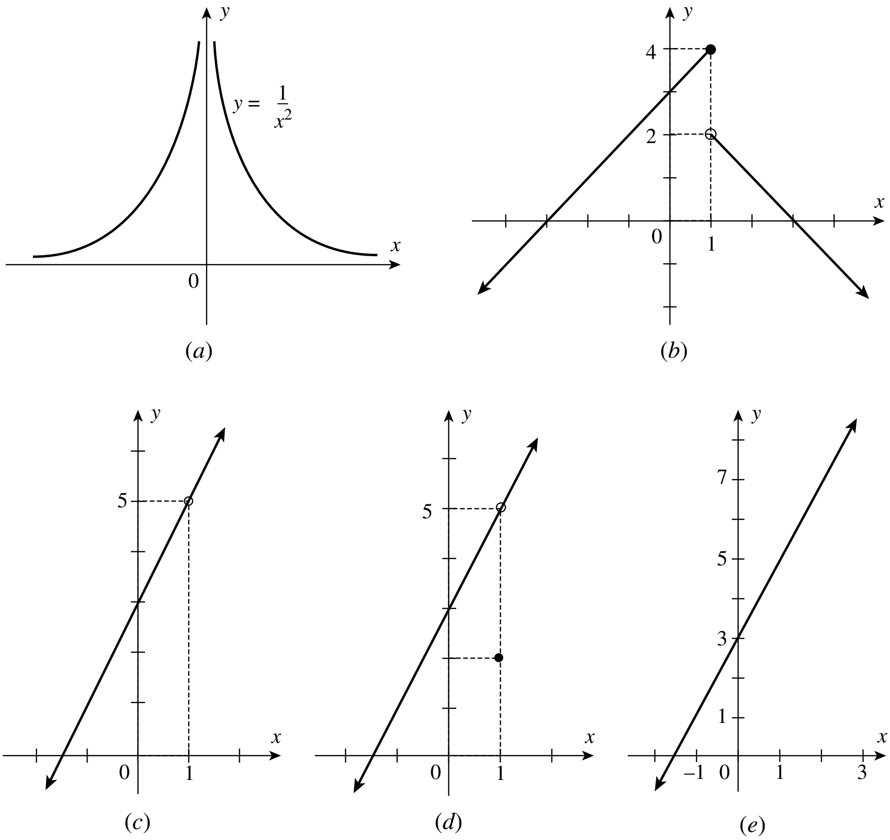


FIGURE 8.12

$\lim_{x \to a} f(x)$ *does not exist* [i.e., $\lim_{x \to a^-} f(x) \neq \lim_{x \to a^+} f(x)$. (Here, "*a*" stands for "0", with reference to Figure 8.12a and it stands for "1" with reference to the Figure 8.12b–e.)

For Figure 8.12c, $\lim_{x \to a} f(x)$ *exists*, but $\lim_{x \to a} f(x) \neq f(a)$ [since, *f*(*a*) is not defined].

For Figure 8.12d, $\lim_{x \to a} f(x)$ *exists*, but $\lim_{x \to a} f(x) \neq f(a)$ [since, *f*(*a*) is different from $\lim_{x \to a} f(x)$].

For Figure 8.12e, $\lim_{x \to a} f(x)$ exists, and $\lim_{x \to a} f(x) = f(a)$.

Notice that in Figures 8.12c and d the graphs appear to be broken at "1". Next observe that in Figure 8.12e the graph appears to be *unbroken* (i.e., continuous) at "1", with *f*(*x*) approaching *f*(1) as *x* approaches "1". *This type of behavior is of great importance in calculus.*

8.2.1 Defining Continuity of a Function at Any Point "*a*"

From the above observations, we can now give the following definition(s) of *continuity at any point* "*a*", *in its domain, using the concept of limit.*

Definition [Continuity]: Let a function "*f*" *be defined in an interval I*, and let "*a*" *be any point in I*. The function "*f*" is said to be *continuous* at the point "*a*", *if and only if* the following three conditions are *met*:

$$\left.\begin{array}{l} (i)\ f(x) \text{ is defined at } x = a \\ (ii)\ \lim_{x \to a} f(x) \text{ exists; and} \\ (iii)\ \lim_{x \to a} f(x) = f(a). \end{array}\right\} \qquad (1)^{(5)}$$

In fact, these three conditions of continuity "at a point", are summed up in the following *short definition*.

A function *f*(*x*) is said to be *continuous at a point* $x = a$, *if the limit of the function* as $x \to a$, *is equal to the value of the function* for $x = a$, which we express by the statement,

$$\lim_{x \to a} f(x) = f(a) \qquad (2)$$

There is *another way to express continuity of a function at a point* "*a*". In the statement (2), if we replace x by $a + h$, then as $x \to a$, we have $h \to 0$ (see Figure 8.13).

Thus, the statement

$$\lim_{h \to 0} f(a + h) = f(a) \qquad (3)$$

defines continuity of the function "*f*" at "*a*".

(5) We give here the meanings of certain statements, which are frequently used in mathematics.

- *f*(*x*) is defined at $x = a$ means, the value *f*(*a*) is a finite number.
- *f*(*x*) is not defined at $x = a$ means, either the point (*a*, *f*(*a*)) is missing from the graph (which also means that "*a*" is not in the domain of "*f*") or *f*(*a*) is not finite [i.e., as $x \to a$, $f(x) \to \infty$].
- $\lim_{x \to a} f(x)$ exists means $\lim_{x \to a^-} f(x) = \lim_{x \to a^+} f(x)$, both being finite.

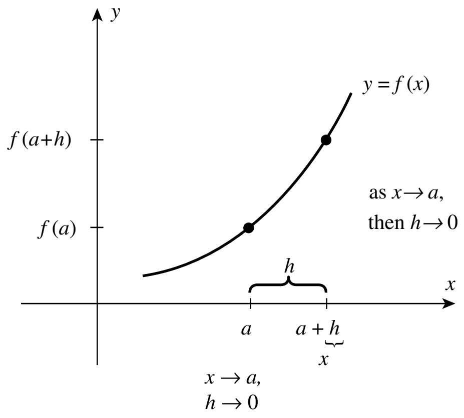


FIGURE 8.13

The statement (3) of *the definition of continuity* is very useful and *convenient* for applying to the *trigonometric*, *exponential*, and *logarithmic functions*, and so on to prove their continuity or otherwise.

Note (5): It is important to remember that *the value f(a) and* $\lim_{x \to a} f(x)$ are *two different concepts* and hence even when both the numbers exist, *they may be different*. The concept of continuity of the function (at any point $x = a$, in its domain) is based on *the existence and equality of these two numbers, at* "*a*".

Remark: In the notion of limit $\lim_{x \to a} f(x)$, *the value f(a) plays no role* [since, $\lim_{x \to a} f(x)$ may exist, even when *f*(*a*) is *not defined*] *but the value f*(*a*) *becomes very important* when we consider the *continuity* of "*f*(*x*)" at "*a*".

8.2.2 A Little More About Continuity

The condition of continuity of a function in an interval can be described as *the property of the function to change gradually* within that interval in the sense that small variations of argument (i.e., the independent variable) *generate small variations of the function itself.*[(6)]

In descriptive geometrical terms, the continuity of a function at a given point signifies that *the difference of the ordinates on the graph* of the function $y = f(x)$ at the points $x_0 + h$ and x_0 will, *in absolute value*, be arbitrarily small, provided $|h|$ is sufficiently small (i.e., if we can choose $|h|$ arbitrarily small, closer and closer to zero).

If the function f(x) is known to be continuous at the point "*a*", then the problem of calculating the limit of the function $f(x)$ as $x \to a$ is *trivial*, since the *calculation* of the limit at the point "*a*" *reduces to the calculation of the value of the function at the point* "*a*".

[(6)] This is a characteristic feature of many phenomena and processes, for instance, expansion in the length of metal rods on heating, the growth of an organism during a period, and variation in air temperature during the day, and so on, are considered continuous processes.

For example, if $h(x) = (x^2-3)/(x-1)$, then we can easily compute

$$\lim_{x \to 2} h(x) = \lim_{x \to 2} \frac{x^2 - 3}{x - 1} = \frac{2^2 - 3}{2 - 1} = 1$$

It can be shown that *all the basic elementary functions are continuous* in the intervals where they are defined. (Of course, the proofs can be seen in the advanced courses in mathematical analysis.)

Besides, continuous functions can be easily investigated and their properties can be studied. Hence, it is often important that a function be continuous wherever possible. We ask the question, *is it possible to remove the discontinuity of a function*? The answer is "only *sometimes*". *If the discontinuity* of a function is not removable it is called an *irremovable* or an *essential discontinuity*. We shall discuss about removable and irremovable discontinuities shortly.

Note (6): The continuity of a function can be expressed either in terms of *the points at which the function is continuous* or in terms of *the points at which the function is discontinuous* or by considering the entire situation covering all the points of interval.

For example, *technically it is correct to say* that the function

$$F(x) = \frac{(x-1)(2x+3)}{(x-1)}, \quad x \neq 1$$

is *continuous throughout its domain where it is defined* or that it is *discontinuous* at $x = 1$ or that *it is continuous for all* x *except for* $x = 1$, *where it is not defined.*

Remark: If we simply say that the function $F(x)$ is *continuous throughout its domain of definition*, then there comes up an element of curiosity (or discomfort) in the reader's mind who is able to visualize the point of discontinuity in the expression defining $F(x)$. Therefore, from this point of view it is more convenient and convincing to say that $F(x)$ is continuous for all x except for $x = 1$, where it is discontinuous. We can understand *the concept of continuity better, if we study its opposite—the concept of discontinuity*.

8.2.3 Definition

A function is discontinuous at $x = a$ if and only if it is not continuous at $x = a$. (Recall that we have already given this definition earlier in Section 8.1.1. Note that, this is an *indirect definition* wherein by *denying the property of continuity to a function*, "at a point", we identify it as a discontinuous *function*.)

When we say that a function is not continuous at $x = a$, we mean that the *condition of continuity is violated* at $x = a$, so that

$$\lim_{x \to a} f(x) \neq f(a) \tag{4}$$

The point "a" is then called a point of discontinuity of the function.

8.2.3.1 A Point of Discontinuity in Terms of Limit(s) Having discussed the definition of continuity in terms of limits (in Sections 8.2.1 and 8.2.2) we now use our knowledge, to discuss and find out what happens at a point of discontinuity (of a function) in terms of limits.

With reference to the *definition of discontinuity* at (4) above, we can say that, *a function defined on an interval I is discontinuous at a point $a \in I$, if at least one of the following conditions occur at the point $x = a$.*

(i) The function $f(x)$ is *not defined* at $x = a$,

(ii) $\lim_{x \to a} f(x)$ *does not exist* [which means that $\lim_{x \to a^-} f(x) \neq \lim_{x \to a^+} f(x)$ *or at least one of the one-sided limits is infinite*],

(iii) $\lim_{x \to a} f(x) \neq f(a)$, in the arbitrary approach of $x \to a$ (which means that the expressions on the *right* and the *left* both exist but they are unequal).

8.2.4 Removable and Irremovable Discontinuities of Functions

If $\lim_{x \to a} f(x)$ *exists* but $f(a)$ is *either not defined*, or *not equal to* $\lim_{x \to a} f(x)$ then, we may redefine "f" (*at the point of discontinuity* "a") such that we assign to $f(a)$, the number which equals the limit $\lim_{x \to a} f(x)$. This makes the function $f(x)$ continuous at $x = a$ (by definition of limit). This we have seen in the process of developing the definition of continuity (see Figure 8.10a and b). Such a discontinuity is called removable discontinuity, for obvious reasons.

It is not always possible to remove the discontinuity of a function. If the discontinuity is not removable it is called an irremovable (or an essential) discontinuity of the function, as mentioned earlier. If $\lim_{x \to a} f(x)$ *does not exist* then $f(x)$ is said to have *an irremovable* (or essential) *discontinuity* at $x = a$. (Note that the graphs of the functions in Figures 8.2a and 8.3 indicate the point of *removable discontinuities* whereas those displayed in Figures 8.2b, 8.4, 8.5, and 8.6 indicate the points of *irremovable discontinuities*.)

The *simplest type of essential discontinuity* occurs at those points at which a function makes a (finite) jump, that is, where the function has a *definite limit* as $x \to a^-$ and a *different definite limit* as $x \to a^+$. Such discontinuities are displayed in Figures 8.4 and 8.5.

Note (7): It must be clear that if the graph of the function has a *finite jump of a point alone,* then the function is said to have removable discontinuity at that point. But, if there is a *finite jump of a portion of the curve*, then such a function has *irremovable* (or essential) *discontinuity* at the point of jump.

Remark: In the case of an irremovable discontinuity it does not matter whether or how the function is defined at the point of discontinuity. This will be clear from the following example, and many more later on.

Example (5): Recall the function $f(x) = 1/x$, $x \neq 0$. Clearly this function is *not continuous* at $x = 0$, and in any interval containing the point "0". The examination of the graph of $1/x$ in the vicinity of the point $x = 0$ clearly shows that *it splits into two separate curves at the point* $x = 0$ (see Figure 8.9a). Further note that, in this case, *we cannot make* "f" *continuous by assigning any value to* $f(0)$. Also observe that neither $\lim_{x \to 0} f(x)$ exists nor $f(0)$ is defined. We say that "f" has *an infinite discontinuity at $x = 0$*. This is an *essential discontinuity* of the function.

The same behavior is observed in the graph of the function $y = \tan x$, in the vicinity of the points $x = (2k + 1)\pi/2$. Even in the case of signum function (denoted by $y = \text{sgn}\, x$), and the

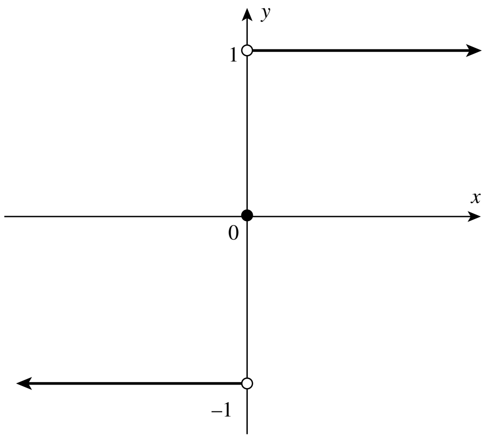


FIGURE 8.14

function $y = |x|/x$ (which are the examples of jump discontinuity), the curve "*splits*" into two *separate curves* [see Examples (6) and (7), given below].

Example (6): Let the function $f(x) = \text{sgn}\, x = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$

The function *f*(*x*) is called *signum* function (or *sign function*) denoted by sgn *x* and read "*signum of x*" (Figure 8.14). (It gives the sign of *x*.) Note that the function sgn *x is defined for all x*. Because sgn $x = -1$, if $x < 0$ and sgn $x = 1$, if $x > 1$, we have

$$\lim_{x \to 0^-} \text{sgn}(x) = \lim_{x \to 0^-} (-1), \quad \text{and} \quad \lim_{x \to 0^+} \text{sgn}(x) = \lim_{x \to 0^+} (1) = 1$$

Thus, the *left-hand limit* and the *right-hand limit* are not equal, which means that $\lim_{x \to 0} \text{sgn}(x)$ does not exist. Accordingly, *f*(*x*) is *discontinuous* at $x = 0$.

Note that *f*(0) exists. Obviously, the function sgn *x* has a *jump discontinuity* at $x = 0$.

Example (7): Consider the function $y = \dfrac{|x|}{x}$, $x \neq 0$ (see Figure 8.15).

The arrows at the ends of the rectilinear portions of the graph mean that for $x = 0$, the function is not defined but for the values of *x* less than zero the value of the function is "– 1", and for the values of *x* exceeding zero, it is equal to "1". Hence, there exists no number to which the value of the function becomes arbitrarily close for all the values of *x*, approaching the point "0". (In other words, this function has no limit as $x \to 0$.)

Note (8): If we add the point $x = 0$ to the domain of this function and put $y = 0$, for $x = 0$, we get the signum function discussed in the previous example.

Remark: We must distinguish between a *jump discontinuity* and an *infinite discontinuity*. Recall that a function has a jump discontinuity at $x = a$, if *both the one-sided limits are finite*

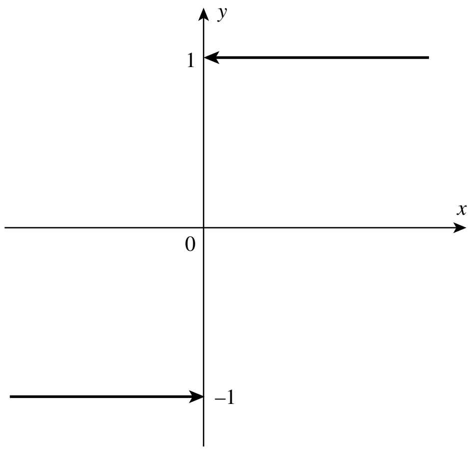


FIGURE 8.15

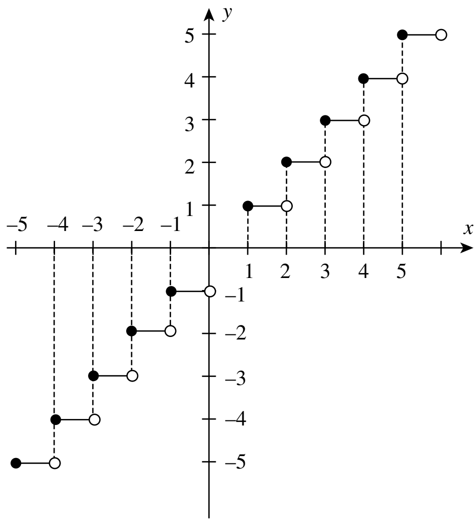


FIGURE 8.16 Greatest integer function

and *unequal*. In the case of an infinite discontinuity, at least one of the one-sided limits is infinity. Of course, both are irremovable discontinuities.

Example (8): The *greatest integer function* of x denoted by $[x]$ is defined as: $[x] =$ the greatest integer less *than* or equal *to x*. Thus, for all numbers *x less than 2 but near 2*, $[x] = 1$, and for all numbers *greater than 2 but near 2,* $[x] = 2$.[7]

The graph of $[x]$ *takes a jump at each integer as* clear from the graph (Figure 8.16).

[7] Obviously, $[3.1] = 3$, $[2.99] = 2$, $[2] = 2$, $[0] = 0$, $[0.9] = 0$, $[-3.1] = -4$, $[-2.99] = -3$, $[7.2] = 7$.

Now we ask the question: is $[x]$ near to a single number 1, when x is near 2? The answer is "No". When $x \to 2^-$, $[x] \to 1$, but when $x \to 2^+$, $[x] \to 2$. Thus, $\lim_{x \to 2}[x]$, *does not exist. This is* as well *true for any other integer*. Thus, $[x]$ is *not continuous for any integer x.*

8.2.4.1 Infinite discontinuities appear at points of discontinuity "*a*" for which $\lim_{x \to a} f(x) = \infty$. We have already seen that for the function $y = 1/x^2$, this point is $x = 0$, and for the function $y = \tan x$, such points are $x = (2k+1)\pi/2$.[8]

Note (9): Because, continuity is defined in terms of limits, we can get information about continuity from the various limit theorems already stated. For example, we know that *if "a" is in the domain of the rational function f*(x) which means that $f(a)$ is defined, then we have, $\lim_{x \to a} f(x) = f(a)$. It follows that, *any rational function is continuous at every point where it is defined (i.e., at the points where the denominator does not become zero).* Similarly, *each of the six trigonometric functions sin x, cos x, tan x, and so on, is continuous at every point where they are defined* (i.e., in their natural domains). (In Chapter 11a, we have shown that $\lim_{x \to a} \sin x = \sin a$, and $\lim_{x \to a} \cos x = \cos a$, which is equivalent to saying that both the functions are continuous at any point "*a*" in their domain.)

Example (9): Find any points of discontinuity for the function $f(x)$ given by

$$f(x) = \frac{x^4 - 3x^3 + 2x - 1}{x^2 - 4}$$

The denominator is zero when $x = \pm 2$. Hence "f" is not defined at ± 2 and accordingly it is *discontinuous* at these points. Otherwise, the function is "well behaved". In fact, any rational function (i.e., any quotient of polynomials) is discontinuous at points where the denominator becomes 0, but it is continuous at all other points.

Earlier, we have broadly identified the discontinuities of functions as: (i) removable discontinuities and (ii) irremovable discontinuities.

Our discussion of discontinuous functions suggests that the following finer classification of the points of discontinuity should help in understanding various types of discontinuities.

8.3 CLASSIFICATION OF THE POINTS OF DISCONTINUITY: TYPES OF DISCONTINUITIES

If a function $y = f(x)$ has a discontinuity for $x = a$, then *to identify the character* (or nature or type) *of the discontinuity*, it is necessary to find the *left-hand* and *right-hand* limits of the function $f(x)$ as $x \to a$.

Depending on the behavior of a function in the vicinity of the point of discontinuity, we distinguish between two basic kinds of discontinuity:

(i) *A Discontinuity of the First Kind.* In this case, *there exist both the one-sided limits*. That is, $\lim_{x \to a^-} f(x)$ and $\lim_{x \to a^+} f(x)$, both are *finite numbers.*

[8] Some other examples are the point $x = 0$ for the function $y = \log x$ (which is defined on the right of that point of discontinuity $x = 0$), and the points $x = -1$ and $x = 1$ for the function $y = 1/\sqrt{1 - x^2}$. In this case, the points of discontinuity are the end points of the interval $[-1, 1]$. Note that this function is defined only on the open interval $(-1, 1)$, where it is continuous.

This is an important class of points of discontinuity. Obviously the *jump discontinuity* belongs to the *first kind.* Also note that *removable discontinuity* is of the *first kind.*

(ii) *A Discontinuity of the Second Kind.* All other discontinuities (which are not of the first kind) are called discontinuities of *second kind. In this case, at least one of the one-sided limits does not exist or is infinite.*

In view of the above classification, note that the discontinuities indicated in the graphs of Figures 8.2a, 8.3, 8.4, and 8.5 are of the *first kind*, whereas those indicated in the graphs of Figures 8.2b and (8.6) are of the second kind.[9]

8.4 CHECKING CONTINUITY OF FUNCTIONS INVOLVING TRIGONOMETRIC, EXPONENTIAL, AND LOGARITHMIC FUNCTIONS

Recall that the concept of continuity of a function $f(x)$, at a point "a" is defined in terms of the equality of both the one-sided limits of the function at "a" with the value $f(a)$. All the statements of the definition of continuity are useful in dealing with different requirements of the problems. One definition which may be useful for checking the continuity of an algebraic function may not be convenient for checking the continuity of a trigonometric or exponential function. Hence, depending on the type of function and the requirement involved, one may have to choose the suitable definition to be applied. Of course, all the three definitions are equivalent.

So far we have discussed about the continuity (and discontinuity) of some algebraic functions only. In almost all the cases, the graphs of the functions were also given, for easy understanding. In fact, this approach has been quite simple (and systematic), since the concept of limit of a function was introduced with the help of simple algebraic functions only. In our further study, it will be found that the continuity of a function can be checked without having an idea about the graph of the function. In fact, in most of the cases, it may not be possible to draw the graph of the function or even imagine its shape.

Now, it is proposed to discuss the continuity of functions involving trigonometric, exponential and logarithmic functions. Accordingly, it is necessary to study their properties and the methods for computing their limit(s). Since, this requirement is met in different chapters, it is necessary that we assume certain results (i.e., the standard limits), since these will be needed to compute the limits involving these functions.

- *Standard Limit of Trigonometric Functions.* We know that *the trigonometric functions are defined for the angle "x", expressed in radians*, which represents *real numbers*, as discussed in Chapter 5.

  The following *trigonometric limits*, are discussed at length in Chapter 11a.

  (i) $\lim_{x \to 0} \cos x = 1,$

  (ii) $\lim_{x \to 0} \sin x = 0$

  (iii) $\lim_{x \to 0} \dfrac{\sin x}{x} = 1$

  (iv) $\lim_{x \to 0} \dfrac{\cos x - 1}{x} = 0$

[9] One must not think that a point of discontinuity of the second kind is necessarily a point of infinite discontinuity. There are bounded functions having neither a left-hand limit nor a right-hand limit as the argument (i.e., a independent variable) approaches a point of discontinuity. Such an example is the function $y = \sin(1/x)$. When $x \to 0$, the function does not tend to any limit, finite or infinite, left-hand or right-hand.

These results are treated as *standard trigonometric limits*, and they are used for computing limits of other functions involving trigonometric functions as discussed in Chapter 11b.

- *Standard Limits of Exponential Functions.* If "a" is a *positive real number*, then the function f defined by $f(x)=a^x$ is called *an exponential function.* The number "e" is introduced later in Chapter 13a and the natural exponential function is denoted by $f(x)=\mathrm{e}^x$. The following results (A) and (B) are also proved there. These are treated as *standard limits.*

1. $\lim_{x\to 0}(1+x)^{1/x}=\mathrm{e}$ (A)

   Further, if $f(x)\to 0$, as $x\to 0$, then

   $$\lim_{x\to 0}(1+kf(x))^{1/kf(x)}=\mathrm{e}, \quad (k\neq 0)$$

   This result is easily obtained by expressing its left-hand side in the form as shown above on the left-hand side of (A). For this purpose, we use *the method of substitution* as follows:

   Put $k\,f(x)=t$, then as $x\to 0$, $t\to 0$ [since, $f(x)\to 0$ as $x\to 0$]

   Therefore, $\lim_{x\to 0}(1+kf(x))^{1/kf(x)}=\lim_{t\to 0}(1+t)^{1/t}=\mathrm{e}$

2. $\lim_{x\to 0}\frac{a^x-1}{x}=\log_e a$, where $a>0$ (B)

   However, if $f(x)\to 0$, as $x\to 0$, and $k\neq 0$ then, $t=k\,f(x)\to 0$, as $x\to 0$.

   Therefore, $\lim_{x\to 0}\frac{a^{kf(x)}-1}{kf(x)}=\lim_{t\to 0}\frac{a^t-1}{t}=\log_e a$

   The methods for computing the limits of exponential and logarithmic functions are discussed in Chapter 13b.

Besides, for computing the limit(s) of certain functions involving exponential functions, the following results will be found very useful. We know that,

$$\lim_{x\to 0^-}1/x=-\infty \quad \text{and} \quad \lim_{x\to 0^+}1/x=\infty$$

Therefore, as $x\to 0^-$, $5^{1/x}\to 0$ (since, $5^{1/x}\to 5^{-\infty}=1/5^{\infty}=0$) and as $x\to 0^+$, $5^{1/x}\to\infty$ (since, $5^{1/x}\to 5^{\infty}=\infty$).

Example (10): Check whether the function $f(x)=\frac{2^{1/x}+2}{2^{1/x}+1}$ is continuous at $x=0$.

Solution: Note that the function $f(x)$ is not defined at $x=0$. To check whether this function is continuous at $x=0$, we compute its one-sided limits. As $x\to 0$ from the left (i.e., as $x\to 0^-$, $1/x\to-\infty$, so that $2^{1/x}\to 0$).

$$\therefore \quad \lim_{x\to 0^-}f(x)=\lim_{x\to 0^-}\frac{2^{1/x}+2}{2^{1/x}+1}=\frac{0+2}{0+1}=2 \quad (5)$$

However, as $x\to 0$ from the right (i.e., as $x\to 0^+$, $1/x\to\infty$, so that $2^{1/x}\to\infty$ and $2^{(-1/x)}\to 0$.)

Note that, here it is not useful to apply the result $\lim_{x\to 0^+}2^{1/x}=\infty$. Hence, we express the given function in a different form, so that its limit can be computed easily. We have,

$$f(x) = \frac{2^{1/x}+2}{2^{1/x}+1} = \frac{2^{1/x}(1+2.2^{-1/x})}{2^{1/x}(1+2^{-1/x})}, \quad x \neq 0$$

$$= \frac{1+2.2^{-1/x}}{1+2^{-1/x}}, \ x \neq 0$$

$$\therefore \quad \lim_{x \to 0^+} f(x) = \frac{1+0}{1+0} = 1 \quad (6)$$

Thus, the function $f(x)$ has a limit 2, as $x \to 0^-$ and a limit 1 as $x \to 0^+$. These limits are unequal (and finite). Therefore, the function in question is discontinuous at $x=0$, and the discontinuity is of *the second kind*. (Note that without having any idea of the graph of this function, we have obtained the above result.)

Example (11): Prove that the function defined by

$$f(x) = \begin{cases} x \sin \frac{1}{x}, & \text{when } x \neq 0 \\ 0, & \text{when } x = 0 \end{cases}$$

is continuous at $x=0$.

Solution: We shall compute the left-hand limit and right-hand limit of this function, at $x=0$. Since we have to find the limit of $f(x)$ at $x=0$, we put $x=0+h$. Therefore, as $x \to 0$, $h \to 0$. We know that, on the left side of "0", each number is *negative* and on the right side of "0", each number is *positive* (by convention), which means

$$\lim_{x \to 0^+} f(x) = \lim_{h \to 0} f(0+h) \text{ and } \lim_{x \to 0^-} f(x) = \lim_{h \to 0} f(0-h)$$

$$\lim_{x \to 0^+} f(x) = \lim_{x \to 0^+} x \cdot \sin \frac{1}{x}$$

$$= \lim_{h \to 0} (0+h) \cdot \sin\left(\frac{1}{0+h}\right) = \lim_{h \to 0} (h) \cdot \sin\left(\frac{1}{h}\right)$$

$$= \lim_{h \to 0} (h) \cdot \sin \frac{1}{h} = 0. \text{ (a finite quantity)} = 0$$

(Since $\sin(1/x)$ is a bounded function, which lies between -1 and 1.)

Now, $\lim_{x \to 0^-} f(x) = \lim_{x \to 0^-} x \cdot \sin \frac{1}{x}$

$$= \lim_{h \to 0} (0-h) \cdot \sin\left(\frac{1}{0-h}\right) = \lim_{h \to 0} (-h) \cdot \sin\left(-\frac{1}{h}\right)$$

$$= \lim_{h \to 0} (h) \cdot \sin \frac{1}{h} = 0 \cdot \text{(a finite quantity)} = 0$$

(Since $\sin(1/x)$ is a bounded function which lies between -1 and 1.)

As $\lim_{x \to 0^-} f(x) = \lim_{x \to 0^+} f(x) = f(0)$, $f(x)$ is continuous at $x=0$.

Example (12): $f(x) = \begin{cases} \sin\dfrac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$

Test the continuity of $f(x)$ at $x=0$.

Solution: Note that $f(x)$ is defined for all x. However, since the part $\sin(1/x)$ is not defined for $x=0$, there is a possibility of discontinuity at $x=0$. The function $f(x)$ is well defined for all other values of x. The value of $f(x)$ in the neighborhood of "0" is given by

$f(0+h) = \sin\dfrac{1}{(0+h)}$, where h is a real number other than 0.

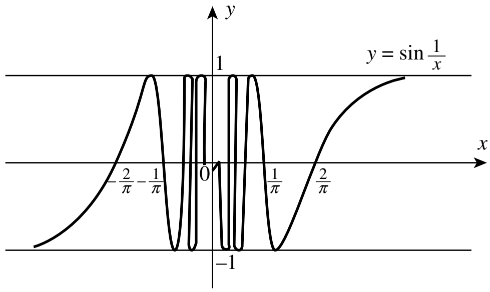
Or $f(h) = \sin\dfrac{1}{h}$, $(h \neq 0)$

$\therefore \lim\limits_{h\to 0} f(h) = \lim\limits_{h\to 0} \sin\dfrac{1}{h}$, which does not exist.

[Indeed, the $\lim\limits_{h\to 0} f(h)$oscillates between -1 and $+1$.]

In other words, the $\lim\limits_{h\to 0} \sin\dfrac{1}{h}$ does not exist at $h=0$. Hence, the given function $f(x)$ is not continuous at $x=0$.

Note (10): The function $\sin(1/x)$ is defined for all values of x except for $x=0$. It does not approach either a finite limit or infinity as $x \to 0$. The graph of this function is shown below.



Note (11): The function $f(x)$, defined in Example (7) is a peculiar function wherein the point of discontinuity does not fit into the first kind, since it is neither a removable discontinuity nor a jump discontinuity. Hence, this is an example of the second kind of discontinuity.

Example (13): $f(x) = \begin{cases} x^2 \cdot \sin\dfrac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$

Test the continuity of $f(x)$ at $x=0$.

Solution: Note that $f(x)$ is defined for all x. However, since the part $x^2 \cdot \sin(1/x)$ is not defined at $x=0$, there is a possibility of discontinuity of $f(x)$ at $x=0$. Therefore, we compute the left-hand and the right-hand limits of $f(x)$, at $x=0$.

Put $x = 0 + h$. Therefore, $x \to 0 \Rightarrow h \to 0$.

Also, $\lim_{x \to 0^+} f(x) = \lim_{h \to 0} f(0+h)$ and $\lim_{x \to 0^-} f(x) = \lim_{h \to 0} f(0+h)$

$$\begin{aligned}\lim_{x \to 0^+} f(x) &= \lim_{x \to 0^+} x^2 \cdot \sin\frac{1}{x} \\ &= \lim_{h \to 0} (0+h)^2 \cdot \sin\frac{1}{(0+h)} \\ &= \lim_{h \to 0} h^2 \cdot \sin\frac{1}{h} = 0 \cdot (\text{a finite quantity}) \\ &= 0.\end{aligned} \tag{7}$$

$$\begin{aligned}\text{Again, } \lim_{x \to 0^-} f(x) &= \lim_{x \to 0^-} x^2 \cdot \sin\frac{1}{x} \\ &= \lim_{h \to 0} (0-h)^2 \cdot \sin\frac{1}{(0-h)} \\ &= \lim_{h \to 0} h^2 \cdot \sin\frac{1}{h} = 0 \cdot (\text{a finite quantity}) \\ &= 0 \text{ (as above)}\end{aligned} \tag{8}$$

$$\text{Also we have } f(0) = 0 \tag{9}$$

In view of the statements at (7), (8), and (9) above, $f(x)$ is continuous at $x = 0$.

Example (14): Test the continuity/discontinuity of the following function at $x = 0$.

$$f(x) = \begin{cases} \dfrac{e^{1/x}}{1 + e^{1/x}}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Solution: Given, $f(x) = \dfrac{e^{1/x}}{1 + e^{1/x}}$

Put $x = (0 + h)$ $\therefore$ As $x \to 0$, $h \to 0$.

$$\therefore \quad \lim_{x \to 0^+} f(x) = \lim_{h \to 0} f(0+h) \tag{10}$$

$$\lim_{x \to 0^-} f(x) = \lim_{h \to 0} f(0-h) \tag{11}$$

In view of (10) above, we have

$$\lim_{x \to 0^+} f(x) = \lim_{h \to 0} \frac{e^{1/(0+h)}}{1 + e^{1/(0+h)}} = \lim_{h \to 0} \frac{e^{1/h}}{1 + e^{1/h}} = l_1 \text{ (say)}$$

Now, dividing the numerator and the denominator by $e^{1/h}$, we express l_1 by

$$l_1 = \lim_{h \to 0} \frac{1}{e^{-1/h} + 1}$$

We know that as $h \to 0$, $-\dfrac{1}{h} \to -\infty$, so that $e^{-1/h} \to 0$.

Therefore, $l_1 = \dfrac{1}{0+1} = 1$, which is the limit from the right. (12)

In view of (11) above, we have

$$\lim_{x\to 0^-} f(x) = \lim_{h\to 0} \frac{e^{1/(0-h)}}{1+e^{1/(0-h)}}$$

$= \lim\limits_{h\to 0} \dfrac{e^{-1/h}}{1-e^{-1/h}} = \dfrac{0}{1-0} = 0$, which is the limit from the left. (13)

Since $\lim\limits_{x\to 0^+} f(x) \neq \lim\limits_{x\to 0^-} f(x)$, we conclude that *f*(*x*) is *discontinuous* at $x = 0$. [Also note that *the point of discontinuity* (at $x = 0$) is of *the second kind*.]

Example (15): $f(x) = \begin{cases} \dfrac{\sin 2x}{x}, & x \neq 0 \\ 1, & x = 0 \end{cases}$. Is $f(x)$ continuous at $x = 0$?

Solution: Note that the function is defined for all x. To find whether *f*(*x*) is continuous at $x = 0$ or not, we check the left-hand and the right-hand limits at $x = 0$.

Put $x = (0+h)$ ∴ As $x \to 0$, $h \to 0$.

$$\therefore \quad \lim_{x\to 0^+} f(x) = \lim_{h\to 0} f(0+h) \tag{14}$$

$$\lim_{x\to 0^-} f(x) = \lim_{h\to 0} f(0-h) \tag{15}$$

Now, in view of (14), we have

$$\lim_{h\to 0} f(x) = \lim_{h\to 0} \frac{\sin 2(0+h)}{(0+h)} = \lim_{h\to 0} \frac{\sin 2h}{h}$$
$$= \lim_{h\to 0} \frac{\sin 2h}{2h} \cdot (2) = 2 \lim_{2h\to 0} \frac{\sin 2h}{2h} = 2 \tag{16}$$

$$\lim_{x\to 0^-} f(x) = \lim_{h\to 0} \frac{\sin 2(0-h)}{(0-h)} = \lim_{h\to 0} \frac{-\sin 2h}{-h}$$
$$= \lim_{h\to 0} \frac{\sin 2h}{2h} \cdot (2) = 2 \tag{17}$$

Here, we have $\lim\limits_{x\to 0^+} f(x) = \lim\limits_{x\to 0^-} f(x) = 2$, which means that $\lim\limits_{x\to 0} f(x)$ exists and it is 2. But, it is given that $f(0) = 1$. Thus, $\lim\limits_{x\to 0} f(x) \neq f(0)$. Hence, the given function is not continuous at $x = 0$.

Note (12): Since the limit of *f*(*x*) and its value both exist at $x = 0$, the given function *f*(*x*) can be made continuous at $x = 0$, if we redefine the function at $x = 0$ by $f(0) = 2$ (instead of 1).

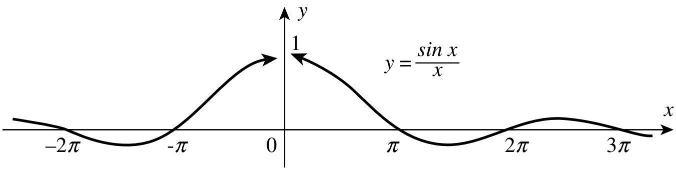
Example (16): Let $f(x) = (\sin x)/x$. Define a function $g(x)$ which is continuous, and $g(x) = f(x)$ for all $x \neq 0$.

Solution: We have, $\lim_{x\to 0} f(x) = \lim_{x\to 0} \frac{\sin x}{x} = 1$

Let $g(x) = \begin{cases} \frac{\sin x}{x}, & \text{for } x \neq 0 \\ 1, & \text{for } x = 0 \end{cases}$

Then, $g(x)$ is continuous at "0". Since $\lim_{x\to 0} g(x) = 1 = g(0)$. Furthermore, $g(x) = f(x)$ for all x, as was desired.

Note (13): The graph of the function (sin x)/x is given below. *It gives a feel of how it becomes continuous when we redefine it at x = 0 as 1.*



Example (17): Discuss the continuity of the function

$$f(x) = \begin{cases} \frac{(3^x - 1)^2}{\sin x \log(1+x)}, & \text{for } x \neq 0 \\ 2\log 3, & \text{for } x = 0 \end{cases}$$

Solution: Given $f(0) = 2\log 3$ (18)

$$\begin{aligned} \lim_{x\to 0} f(x) &= \lim_{x\to 0} \frac{(3^x - 1)^2}{\sin x \log(1+x)} \\ &= \lim_{x\to 0} \frac{((3^x - 1)/x)^2}{(\sin x/x)\cdot(1/x)\log(1+x)} \\ &= \frac{\lim_{x\to 0}((3^x - 1)/x)^2}{\left[\lim_{x\to 0}(\sin x/x)\right]\cdot\log\left[\lim_{x\to 0}(1+x)^{1/x}\right]} = \frac{(\log 3)^2}{(1)\cdot\log_e \mathrm{e}} = (\log 3)^2 \end{aligned} \quad (19)$$

From (18) and (19), we have $\lim_{x\to 0} f(x) \neq f(0)$,

$\therefore$ f is discontinuous at $x = 0$.

Note (14): Since, the $\lim_{x\to 0} f(x)$ and the value of "f" at $x = 0$, both exist, it is possible to remove the discontinuity by redefining f as follows:

$$f(x) = f(x) = \begin{cases} \frac{(3^x - 1)^2}{\sin x \log(1+x)}, & \text{for } x \neq 0 \\ (\log 3)^2, & \text{for } x = 0 \end{cases}$$

Remark: In the definition of $f(x)$, the value $f(0)$ is given to be $2\log 3 = \log 3^2 = \log 9$.

However, for continuity of "f" at $x = 0$, it is found to be $(\log 3)^2$. [Note that $\log 3^2 \neq (\log 3)^2$.]

- *The Problems Related with the Concept of Continuity can be Classified as Follows:*
  - Type (1): Discontinuity of a function at a given point. We have already discussed a good number of such problems.
  - Type (2): To find the value of the unknown, if *$f(x)$ is given to be continuous at a certain point.*
  - Type (3): To find the value $f(a)$ *when f is given to be continuous* at $x = a$. [Such problems demand that we must compute the $\lim_{x\to a} f(x)$. Then by definition, $f(a) = \lim_{x\to a} f(x)$.]
  - Type (4): It is given that $f(x)$ is continuous at $x = a$, and it is required to state either the $\lim_{x\to a} f(x)$ or the value $f(a)$, when anyone of them is given. [Such problems require minimum effort. (Why?)]

Example (18): Find the value of k, if

$$f(x) = \begin{cases} \dfrac{1-\cos kx}{x\sin x}, & \text{for } x \neq 0 \\ 2, & \text{for } x = 0 \end{cases}$$

is continuous.

Solution: Since f is continuous at $x = 0$, $\therefore \lim_{x\to 0} f(x) = f(0)$ (20)

Now, it is given that $f(0) = 2$. (21)

Hence our problem reduces to computing the limit of $f(x)$ as $x \to 0$.

Consider,

$$\lim_{x\to 0} f(x) = \lim_{x\to 0} \frac{1-\cos kx}{x\sin x} = l \quad \text{(say)}$$

(We try to apply some method of expressing the given expression in a convenient form so that the above limit can be easily evaluated.) We have,

$$l = \lim_{x\to 0} \frac{(1-\cos kx)}{x\sin x} \cdot \frac{(1+\cos kx)}{(1+\cos kx)}$$

$$= \lim_{x\to 0} \frac{1-\cos k^2 x}{x\sin x(1+\cos kx)}$$

$$= \lim_{x\to 0} \frac{1-\cos k^2 x}{x\sin x(1+\cos kx)}$$

$$l = \lim_{x\to 0} \frac{(\sin kx/kx)^2 \cdot k^2}{(\sin x/x)(1+\cos kx)}$$

$$= \frac{1^2 \cdot k^2}{1(1+\cos 0)} = \frac{k^2}{2} \quad (\text{as } \cos 0 = 1) \qquad (22)$$

Substituting the values from (21) and (22) in (20), we get

$$\frac{k^2}{2} = 2$$
$$k^2 = 4 \quad \therefore \quad k = \pm 2$$

Example (14): $f(x) = \dfrac{(5^x - 2^x)\cdot x}{\cos 5x - \cos 3x}$, for $x \neq 0$, is continuous at $x = a$. Find $f(0)$.

Solution: It is given that f is continuous at $x = 0$. Therefore, by definition, we have,

$$f(0) = \lim_{x\to 0} f(x). \tag{23}$$

Thus, our problem is reduced to computing the $\lim_{x\to 0} f(x)$.

Now, $\lim_{x\to 0} f(x) = \lim_{x\to 0} \dfrac{(5^x - 2^x)\cdot x}{\cos 5x - \cos 3x} = l$ (say)

Note (15): The standard limits of exponential functions and trigonometric functions suggest that: (i) numerator and denominator must be divided by x^2 and (ii) the denominator must be expressed as a product of "sine functions", using the trigonometric identity.

$$\text{Now, } \cos 5x - \cos 3x = -2\sin\left(\frac{5x+3x}{2}\right)\cdot \sin\left(\frac{5x-3x}{2}\right)$$
$$= -2\sin 4x \cdot \sin x$$

$$l = \lim_{x\to 0} \frac{((5^x - 2^x)/x)}{((-2\sin 4x\cdot \sin x)/x^2)}$$
$$= \lim_{x\to 0} \frac{[((5^x-1)/x) - ((2^x-1)/x)]}{(-8)(\sin 4x/4x)(\sin x/x)}$$
$$= \frac{\log_e 5 - \log_e 2}{(-8)(1)(1)}$$
$$= -\frac{1}{8}\log_e\left(\frac{5}{2}\right)$$

$\therefore$ From (23) we have $f(0) = -\dfrac{1}{8}\log_e\left(\dfrac{5}{2}\right)$

Example (20): The function f is defined by

$$f(x) = \begin{cases} \dfrac{e^x - 1 - x}{x^2}, & \text{for } x \neq 0 \\ \dfrac{1}{2}, & \text{for } x = 0 \end{cases}$$

is continuous at $x = 0$. What is $\lim_{x\to 0} f(x)$?

Solution: [If the problem is read carefully, it must be clear that we do not have to compute $\lim_{x\to 0} f(x)$. On the other hand, we have to simply state the number that gives the limit.]

Since, $f(x)$ is continuous at $x = 0$,

$\lim_{x \to 0} f(x) = f(0)$

But $f(0) = \frac{1}{2}$.

$\therefore \lim_{x \to 0} f(x) = \frac{1}{2}$.

8.5 FROM ONE-SIDED LIMIT TO ONE-SIDED CONTINUITY AND ITS APPLICATIONS

In Chapter 7b, the concept of limit of a function was extended to include *one-sided limits* (and limits involving ∞). The importance of one-sided limits has since been seen *in testing the continuity of a function at any point and in identifying the type of discontinuity at that point*.

Now, we extend *the concept of limit to define the concept of one-sided continuity*, which is useful in defining continuity in a closed interval. For this purpose, we start our discussion with the function $\sqrt{x}$.

We know that the domain of the square root function $\sqrt{x}$ is $[0, \infty)$. Therefore, the $\lim_{x \to 0} \sqrt{x}$ *does not exist*. As a consequence, under the definition of continuity, the square root function $\sqrt{x}$ is not continuous at $x = 0$ (Why?).

However, it has a right-hand limit at 0. We express this fact by saying that the square root function $\sqrt{x}$ is *continuous from the right of* "*0*". We give the following definitions of one-sided continuity.

- **Definition [Continuity from the Right]:** A function $f(x)$ is continuous from the right at a point "a" in its domain, if $\lim_{x \to a^+} f(x) = f(a)$.
- **Definition [Continuity from the Left]:** A function $f(x)$ is continuous from the left at a point "a" in its domain, if $\lim_{x \to a^-} f(x) = f(a)$.

In view of the above definitions a function whose domain is a singleton is considered continuous at that point. See Note (17) on Page 176 (Chapter 7a).

8.6 CONTINUITY ON AN INTERVAL

We say that a function is continuous on an interval if it is continuous at each point there. It must be clear that each point in the interval has to satisfy all the three conditions of *continuity at a point* as stated in the definition (1). This is exactly what it means for continuity on an *open interval*. When we consider a *closed interval* $[a, b]$, *we face a problem* as we have seen in the case of the square root function $\sqrt{x}$.

We overcome this situation by agreeing as follows: we say that "f" is continuous on *closed interval* $[a, b]$, if it is continuous at each point of (a, b) and if the following limits exist:

$$\lim_{x \to a^+} f(x) = f(a), \quad \text{and} \quad \lim_{x \to b^-} f(x) = f(b)$$

(These are *one-sided limits at the end points of a closed interval.*)

Remark: To define the continuity of a *function at any end point of a closed interval, we agree to accept the one-sided limit, as the limit of the function, at that point*. It is a matter of convenience that, we accept the one-sided limit(s), as the limit(s) at the end point(s). We give the following well accepted definition of *continuity on an interval.*

Definition: We say that "f" is *continuous on an open interval* (a, b), if *it is continuous at each point of that interval.*

It is *continuous on the closed interval* $[a, b]$, if it is continuous on (a, b), *right continuous at a*, and *left continuous at b.*

Note (16): If there exists *at least one point in the domain of a function* (assumed to be an interval) where *it is not continuous*, then the function is said to be *discontinuous in its domain.* Thus, if a function is *not continuous even at an endpoint* of a closed interval $[a, b]$, then it is said to be *discontinuous* on $[a, b]$.

Example (21): Given $f(x) = x/(x - 2)$. Test the continuity of the function in the intervals (1, 2), [1, 2], and (1, 3). Note that, $f(x)$ is not defined for $x = 2$. Accordingly, $f(x)$ is *continuous in any interval which does not contain 2*. Thus, "f" is continuous on (1, 2), but it is *discontinuous* on [1, 2] and on (1, 3).

- *Some Theorems on Continuity (Without Proof):*
  1. If f and g are two functions continuous at the number "a", then $f+g$, $f-g$, $f{\cdot}g$, are continuous at "a" and f/g is continuous at "a", provided that $g(a) \neq 0$.
  2. *Continuity of a Composite Function:* If the function g is continuous at "a" and the function f is continuous at $g(a)$, then the composite function f o g is continuous at "a".

Remark: A function continuous in a domain is continuous on any nonempty subset of the domain.

- *Bounded and Unbound Intervals:* Any interval of the form $[a, b]$, $(a, b \in R)$ is said to be *closed and bounded.* Open intervals (a, b) are *bounded*, if a and b are finite numbers. (In fact, in this notation a and b are assumed to be finite.) The interval $(-\infty, \infty)$, which represents the *entire real line*, is both open and closed, and of course *unbounded.* Intervals of the form (a, ∞) and $(-\infty, b)$ are said to be *closed and unbounded.* Remember that the symbol "∞" does not represent a real number.

8.7 PROPERTIES OF CONTINUOUS FUNCTIONS

Continuous functions have many useful properties that discontinuous functions do not have. A function continuous on a closed and bounded interval $[a, b]$ possesses many important properties. Here we state, without proof, one of them, namely the *Intermediate Value Theorem* (IVT) with some of its consequences and applications.

8.7.1 The Intermediate Value Theorem: IVT

If function "f" is continuous on closed interval $[a, b]$, and if $f(a) \neq f(b)$, then *for any number k between f(a) and f(b)*, there exists a number *c between a and b* such that

$$f(c) = k\cdot \tag{24}$$

The *intermediate value theorem* assures us that if the function f is *continuous on the closed (and bounded) interval* $[a, b]$, then $f(x)$ assumes *every value between f(a) and f(b)*, as x assumes all values between a and b. For example, if a function f is continuous throughout the interval [2, 6], and if $f(2) = 1$ and $f(6) = 4$, then *every number between 1 and 4 must be in the range of f.*

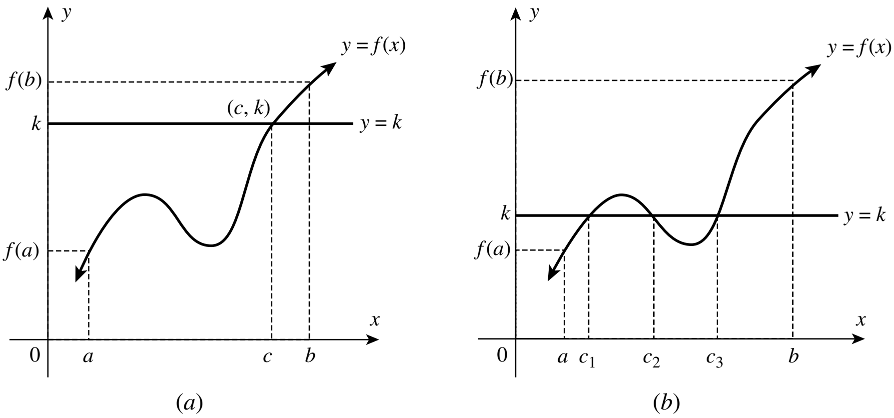


FIGURE 8.17

In terms of geometry, the *intermediate value theorem states that the graph of a function, continuous on a closed interval must intersect every horizontal line* $y = k$*, between the lines* $y = f(a)$ *and* $y = f(b)$*, at least once.* Refer to Figure 8.17a, where (0, k) is a point on the y-axis between (0, $f(a)$) and (0, $f(b)$); the line $y = k$ intersects the graph of f, at the point (c, k), where c lies between a and b.

For some values of k, we may have more than one possible value of c. The theorem states that *at least one value of c exists*, but such a value is *not necessarily unique*. Figure 8.17b shows three possible values of c (c_1, c_2, and c_3) for a particular k.

The following theorem is a *direct consequence* (a corollary) *of the intermediate value theorem.*

8.7.2 The Intermediate Zero Theorem

If the function f is *continuous on a closed interval* $[a, b]$ and if $f(a)$ and $f(b)$ have *opposite signs*, then there exists a number c between a and b such that $f(c) = 0$.

Proof: The hypothesis of the intermediate value theorem is satisfied by the function f, and because $f(a)$ and $f(b)$ have opposite signs, the number "0" qualifies as a number k between $f(a)$ and $f(b)$. Thus, there is a number c between a and b such that

$$f(c) = 0 \tag{25}$$

Such a number "c", is called a *zero* or a *root* of "f". For example, let $f(x) = x^2 - 4x - 5 = (x - 5)(x + 1)$.

Then, we get $f(x) = 0$, for $x = -1$ and $x = 5$.

Accordingly, the zeros of "f" are -1 and 5.

Remark: Zeros of functions of the form $f(x) = ax^2 + bx + c$, where $a \neq 0$ can be located by means of the quadratic formula. But for *higher-degree polynomials* (and *functions in general*) there is no simple formula from which we can determine a zero.

8.7.3 Importance of Zeros of a Function

Let f be continuous on an interval I. If f has both positive and negative values on I, then the *intermediate value theorem implies* that $f(x) = 0$, *for some x in I*, that is, f has a zero in I. Equivalently, if f has no zero in I, then *either $f(x) > 0$*, for all x in I *or $f(x) < 0$*, for all x in I. This fact yields *a procedure for discovering the intervals on which a continuous function f is positive, and those on which f is negative.*[10]

Example (22): Let $f(x) = (x+1)^2(x-2)(x-3)$. We shall determine *the intervals* on which *f is positive* and those on which *f is negative*.

Solution: Note that the zeros of "f" are -1, 2, and 3. Now we can determine the sign of $f(x)$ on the relevant intervals $(-\infty, -1)$, $(-1, 2)$, $(2, 3)$, and $(3, \infty)$, by preparing the following table.

Interval (I)	An Arbitrary but Convenient Point c in (I)	$f(c)$	Sign of $f(x)$ on the Interval
$(-\infty, -1)$	−2	20	+
$(-1, 2)$	0	6	+
$(2, 3)$	5/2	−49/16	−
$(3, \infty)$	4	50	+

From the table, we observe that f is positive on $(-\infty, -1)$, $(-1, 2)$, and $(3, \infty)$, and is negative on $(2, 3)$. Thus, the *intermediate value theorem* can be used to determine the intervals, where continuous functions are positive, where they are negative, and where they are zero.

Remark: The method used in the above example, applies to a function such as a rational function, even if it is not defined on certain points in its domain.

The intermediate value theorem will *not* hold, if the function f is *discontinuous* at a point in $[a, b]$.[11]

The intermediate value theorem can also be used to show that *every non-negative number has a square root*, that is, the domain of the *square root function* consists of *all non-negative numbers*, as asserted in Chapter 6.

Proof: To prove the assertion, we select *any non-negative number p*, and show that p has a square root which is a non-negative number, that is, there is a number $c \geq 0$, such that $c^2 = p$.

[10] Here it may be mentioned (in advance) that once we have introduced the concept of the derivative and studied its properties, we shall apply this technique to determine the zeros of the derivative of f. This will help us in finding the intervals on which the function "f" is increasing and those on which f is decreasing. This technique is very useful for studying many applications of derivatives; for instance, maximum and minimum values of a function, and some other related concepts.

[11] For details refer to the following:

1. *Calculus with Analytical Geometry* (Alternate Edition) by Robert Ellis and Denny Gulick (pp. 97–100), HBJ Publication.
2. *The Calculus 7 of a Single Variable* by Louis Leithold (pp. 87–89), HCC Publishers.

For this purpose, we consider $f(x) = x^2$, for $x \geq 0$. We know that f is continuous on any interval, and observe that

$$f(0) = 0 \leq p, \quad (\because p \geq 0)$$

$$\leq p^2 + 2p + 1 = (p+1)^2 = f(p+1)$$

Thus, $f(0) \leq p \leq f(p+1)$. But the *intermediate value theorem* says that there is a number c in $[0, p+1]$ such that $f(c) = p$ or equivalently, $c^2 = p\ (\because f(c) = c^2)$. Thus, p has a square root.

Since p was an arbitrarily chosen non-negative number, it follows that *the square root is defined for every non-negative number*.

In fact, a *continuous function f* defined on a *closed interval* $[a, b]$, possesses many properties, that we shall be using. Some of these properties are listed below.

If *f is continuous on a closed interval* $[a, b]$, then

- f is bounded on $[a, b]$,
- f has a *maximum* and a *minimum value* on $[a, b]$,
- f is *uniformly continuous* on $[a, b]$.

The importance of *continuity of a function on a closed interval* will become more and more apparent when the reader proceeds through his study of calculus. This *property* is a part of the hypothesis of many key theorems, such as the *mean value theorem*, the *fundamental theorems of calculus* and the *extreme value theorem*.

8.7.3.1 Continuity of Some Elementary Functions It can be shown that

(i) A constant function is continuous for all x.

(ii) A polynomial function $f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$ is continuous for all values of x on $(-\infty, \infty)$.

(iii) x^n; $n > 0$ is continuous for all values of x.

(iv) A rational function is continuous at every point in its domain.

(v) $\dfrac{1}{x^n}$; $n > 0$ is continuous for all values of x, except $x = 0$.

(vi) Trigonometric functions: $f(x) = \sin x$ and $g(x) = \cos x$ are continuous on $(-\infty, \infty)$. Other trigonometric functions (i.e., $\tan x$, $\cot x$, $\sec x$, $\operatorname{cosec} x$) are continuous for all values of x for which they are defined.

(vii) Inverse trigonometric functions are continuous for all values of x for which they are defined.

(viii) The exponential function: $f(x) = a^x$, $(a > 0)$ is continuous on $(-\infty, \infty)$. (In particular, e^x is continuous for all x.)

(ix) The logarithmic function: $f(x) = \log_a x$, $(a > 0)$ is continuous on $(0, \infty)$.

It is now proposed to solve the following problems that may also be treated as an exercise. [This discussion is expected to be useful for a beginner to get a deeper idea of the concept of continuity (and discontinuity). This should prepare him to handle difficult problems presented in various exercises in the textbooks.]

Exercise I

Discuss the continuity of the following functions in the intervals indicated against them. In case a function is discontinuous, state whether the discontinuity is removable or irremovable.

Q. (1): $f(x) = \dfrac{1}{x-2}$ at $x = 2$

Q. (2): $g(x) = \begin{cases} \dfrac{1}{x-2}, & \text{if } x \neq 2 \\ 3, & \text{if } x = 2 \end{cases}$

Q. (3): $\phi(x) = \begin{cases} |x-3|, & \text{if } x \neq 3 \\ 2, & \text{if } x = 3 \end{cases}$

Q. (4): $f(x) = \begin{cases} x^2 + 2, & \text{for } x > 1 \\ 5x - 1, & \text{for } x \leq 1 \end{cases}$

Q. (5): $g(x) = \begin{cases} x + 6, & \text{if } x \geq 3 \\ x^2, & \text{if } x < 3 \end{cases}$

Q. (6): $f(x) = \begin{cases} x + 2, & \text{if } x > 2 \\ x^2, & \text{if } x < 2 \end{cases}$

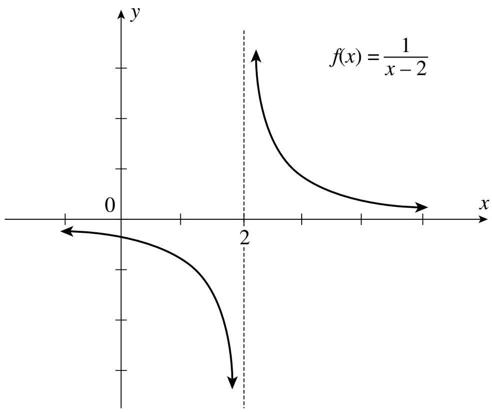
Q. (7): $f(x) = \begin{cases} x^2, & \text{for } x \leq 1 \\ x, & \text{for } x > 1 \end{cases}$

Q. (8): $f(x) = x^2/(1+x^2)$

Q. (9): Show that the function $f(x) = 5$ is continuous for every value of x.

Solutions

(1) Let f be defined by $f(x) = 1/(x-2)$. The graph of f has a break at the point where $x = 2$; so we investigate the conditions of definition (1). Note that "f" is not defined at $x = 2$. Hence, f is discontinuous at 2. Again, $\lim\limits_{x \to 2} f(x)$ does not exist (Why?). This is an example of *infinite discontinuity of second kind.*



(2) Let g be defined by

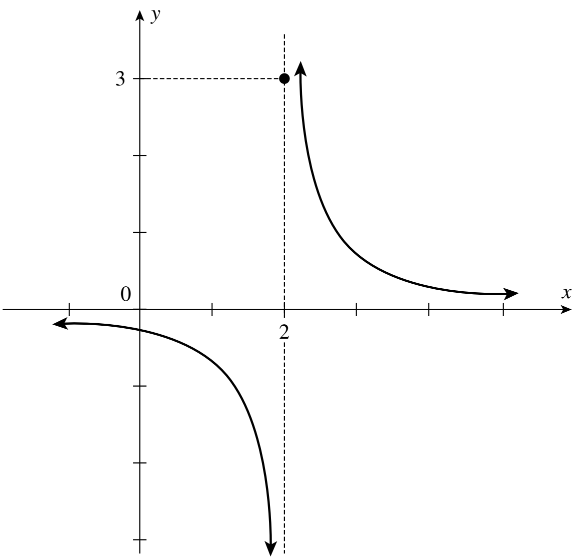
$$g(x) = \begin{cases} \dfrac{1}{x-2}, & \text{if } x \neq 2 \\ 3, & \text{if } x = 2 \end{cases}$$

Note that $g(x)$ is defined for all x. Here again, the graph of g *has a break* at 2. We check the conditions of Definition (1), at $x = 2$. Observe that

(i) $g(2) = 3$

(ii) $\lim_{x \to 2^-} g(x) = \lim_{x \to 2^-} \dfrac{1}{x-2} = -\infty$ and $\lim_{x \to 2^+} g(x) = \lim_{x \to 2^+} \dfrac{1}{x-2} = +\infty$.

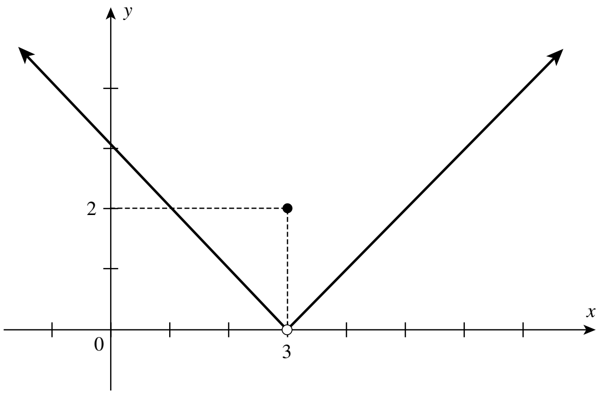
Thus, $\lim_{x \to 2} g(x)$ *does not exist*. Obviously, g is discontinuous at 2. The discontinuity is *infinite* and it is of *second kind.*



(3) Let ϕ be defined by

$$\phi(x) = \begin{cases} |x-3|, & \text{if } x \neq 3 \\ 2, & \text{if } x = 3 \end{cases}$$

The graph of ϕ shows that it has a discontinuity at $x = 3$. We check the three conditions of definition (1) at $x = 3$.



(i) $\phi(3) = 2$

(ii) $\lim_{x \to 3^-} \phi(x) = \lim_{x \to 3^-} (3 - x) = 0$

$\lim_{x \to 3^+} \phi(x) = \lim_{x \to 3^+} (x - 3) = 0$

Hence, the $\lim_{x \to 3} \phi(x)$ exists.

(iii) $\lim_{x \to 3} \phi(x) \neq \phi(3)$

Because *condition (iii) is not satisfied,* ϕ is discontinuous at 3. This *discontinuity is removable* because, if $\phi(3)$ *is redefined to be 0*, then the new function becomes continuous at $x = 3$. This is a discontinuity of *first kind.*

(4) Let us determine whether $f(x) = \begin{cases} x^2 + 2, & \text{for } x > 1 \\ 5x - 1, & \text{for } x \leq 1 \end{cases}$ is continuous at $x = 1$.

Solution: The functions having values $x^2 + 2$ and $5x - 1$ *are polynomials and are therefore continuous everywhere*. Thus, the only number at which continuity is *questionable is* 1. We check the three conditions for continuity at "1".

(i) $f(1) = 4$. Thus, $f(1)$ exists.

(ii) $\lim_{x \to 1^+} f(x) = \lim_{x \to 1^+} (x^2 + 2) = 3$, and $\lim_{x \to 1^-} f(x) = \lim_{x \to 1^-} (5x - 1) = 4$

Thus, $\lim_{x \to 1^+} f(x) \neq \lim_{x \to 1^-} f(x)$.

Therefore, $\lim_{x \to 1} f(x)$ *does not exist*, and so "f" *is discontinuous at* $x = 1$.

This is an example of jump discontinuity, which is of course irremovable. It is of the second kind.

(5) Let $g(x) = \begin{cases} x + 6, & \text{if } x \geq 3 \\ x^2, & \text{if } x < 3 \end{cases}$

The only *possible trouble may occur when* $x = 3$.
We observe that, $g(3) = 3 + 6 = 9$.

$$\text{Further, } \lim_{x \to 3^+} g(x) = \lim_{x \to 3^+} (x + 6) = 3 + 6 = 9$$

$$\text{and } \lim_{x \to 3^-} g(x) = \lim_{x \to 3^-} (x^2) = 9$$

$$\text{Thus, } \lim_{x \to 3} f(x) = f(3). \quad \therefore f \text{ is continuous at } x = 3.$$

(6) Let $f(x) = \begin{cases} x+2, & \text{if } x > 2 \\ x^2, & \text{if } x < 2 \end{cases}$

Since "f" is *not defined* at $x = 2$, it is *discontinuous* there. (It is continuous for all other x.)

This discontinuity can be removed by redefining f. Note that

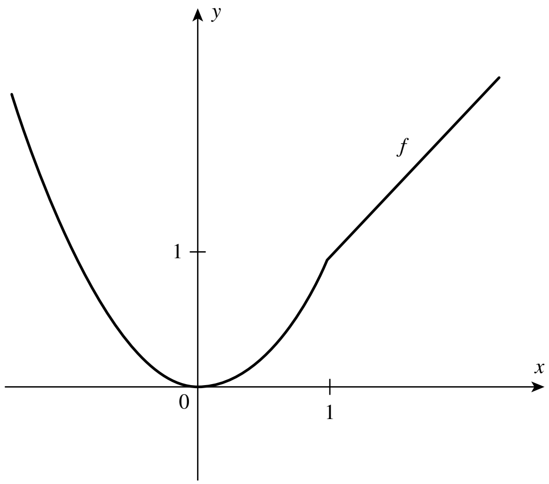
$$\lim_{x \to 2^+} f(x) = \lim_{x \to 2^+} (x+2) = 4 \text{ and } \lim_{x \to 2^-} f(x) = \lim_{x \to 2^-} (x^2) = 4.$$

Thus, $\lim_{x \to 2} f(x)$ exists. Hence this discontinuity can be removed.

Also note that, by including "2" in the domain of "f" [in any part of the formula defining $f(x)$], we get $f(2) = 4$. Thus, f becomes continuous at "2", if 2 is included in the domain of f.

(7) Let $f(x) = \begin{cases} x^2, & \text{for } x \le 1 \\ x, & \text{for } x > 1 \end{cases}$

Show that "f" is continuous at 1.



Solution: We have $\lim_{x \to 1^-} f(x) = \lim_{x \to 1^-} x^2 = 1$ and $\lim_{x \to 1^+} f(x) = \lim_{x \to 1^+} x = 1$.

$$\therefore \lim_{x \to 1} f(x) = 1 \tag{26}$$

Also,

$$f(1) = (1)^2 \tag{27}$$

Thus, $\lim_{x \to 1} f(x) = f(1)$. Therefore, "$f$" is *continuous* at $x = 1$.

(Note that the graph of "f" *has a sharp corner* at $x = 1$.)[12]

[12] Later on, when the concept of differentiable functions is introduced (in Chapter 9), it will be noted that a continuous function, whose graph has a sharp corner at some point $x = a$ (say), is *not differentiable* at that point.

(8) Let $f(x) = x^2/(1+x^2)$. Determine the numbers at which "f" is continuous.

Solution: Here again "f" is a rational function, *but its denominator* $(1+x^2)$ *is never 0*. Thus, "f" is defined for all x and therefore "f" is *continuous for every real value of* x.

(9) Let us show that the function $f(x) = 5$ is continuous at $x = 7$.

Solution: We must verify that the conditions for continuity are satisfied.

(i) "f" is defined at $x = 7$ [Here, we have $f(7) = 5$.]

(ii) $\lim_{x \to 7} f(x) = \lim_{x \to 7} 5 = 5$

Thus, $\lim_{x \to 7} f(x) = f(7)$. Therefore, $f(x)$ is *continuous* at $x = 7$.

Remark: Note that $f(x) = 5$ is a *constant function*. It is easy to show that every constant function is continuous for every value of x.

Exercise II

Discuss the continuity of the following functions:

(a) $f(x) = \begin{cases} \dfrac{e^{3x} - e^{2x}}{\sin 3x}, & \text{when } x \neq 0 \\ 1, & \text{when } x = 0 \end{cases}$

(b) $f(x) = \begin{cases} \dfrac{\sin 5x}{3x}, & \text{for } x \neq 0 \\ \dfrac{3}{5}, & \text{for } x = 0 \end{cases}$

(c) If $f(x) = \begin{cases} \dfrac{e^{x} - e^{-x}}{x}, & \text{when } x \neq 0 \\ k, & \text{when } x = 0 \end{cases}$ is continuous at $x = 0$, find k.

(d) If the function $f(x) = \dfrac{\log(1 + x)}{\sin x}$, for $x \neq 0$, is continuous at $x = 0$, find $f(0)$.

9 The Idea of a Derivative of a Function

9.1 INTRODUCTION

There are certain problems in mathematics, mechanics, physics, and many other branches of science, *which cannot be solved by ordinary methods of geometry or algebra alone*. To solve these problems, we have to use a new branch of mathematics known as *calculus.* It uses not only the ideas and methods from arithmetic, geometry, algebra, coordinate geometry, trigonometry, and so on, but also *the notion of limit*, which is a *new idea* that lies at the *foundation of calculus*. Using this notion as a tool, *the derivative* of a function is defined as the limit of a particular kind.

The idea of derivative of a function is among the most important and powerful concepts in mathematics. This concept distinguishes calculus from other branches of mathematics. It will be found that the derivative of a function is generally a *new function* (derived from the original function). We call it the *rate function* or the *derivative function.*

Calculus is the mathematics of change. The immense practical power of calculus is due to its ability to describe and predict the behavior of changing quantities. We cannot even begin to answer any question related to change unless we know *what changes and how it changes?* Let us discuss.

We know that

- the area of a circle, $A(r) = \pi r^2$, changes with (respect to) its radius "r".
- the volume of a sphere, $V(r) = (4/3)\,\pi r^3 = kr^3$ $(k = (4/3)\pi)$, changes with (respect to) its radius "r".
- the surface area of a cube, $S(l) = 6l^2$, changes with (respect to) the length "l" of its side.

Consider a function $y = h(x)$ whose graph is a *smooth curve* (not a straight line). Then, the inclination "θ" of the tangent line (drawn at any point of the curve) changes from point to point on the curve. (Later on, this observation will be used to define a (new) concept, namely, "*the slope of a curve*" in terms of the slope of the (tangent) line.)

The fact is that all our effort is aimed at defining the *slope of a curve at a point*, which also stands for *instantaneous rate of change of the function* y $[= h(x)]$ at any value of x. To get a better idea of the whole situation, it is useful to study a little more as explained below.

The notion of *dependent variable* introduced in Chapter 6 suggests that if "f" is any function defined by $y = f(x)$, then the *dependent variable* y $[= f(x)]$ changes whenever there is any

9-The concept of derivative function $f'(x)$ (Instantaneous rate of change of f at x, or slope of the graph of f at x) and the process of obtaining it from $y = f(x)$

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

change in the value of *independent variable x. We say that the quantity f(x) changes with (respect to) x.*[(1)]

9.1.1

The *calculus tool* that tells us about the behavior of changing quantities is called *the derivative function (or the rate function)*. For a given function $y\,[= f(x)]$, *the derivative function (or the rate function)* is denoted by $f'(x)$, which tells us the *instantaneous rate of change of f(x) with (respect to) x*.

In this chapter, we will invest a lot of time and effort in studying *how to define derivative functions formally* and *how to calculate them symbolically*. In the process of defining the derivative function (or a rate function), various subtleties and puzzles will inevitably arise. Nevertheless, *it will not be difficult to grasp the concept (of derivatives) with our systematic approach.*

The relationship between $f(x)$ and $f'(x)$ is the main theme. We will study what it means for $f'(x)$ to be *the rate function* (or derivative function) derived from $f(x)$ and what each function says about the other. *The important requirement is to understand clearly the meaning of the instantaneous rate (or the actual rate) of change of f(x) with respect to x.*

For this purpose, it is necessary to distinguish between the *average rate of change* and the *actual (or instantaneous) rate of change* of a varying (dependent) quantity $f(x)$ with respect to another varying quantity "x", considered to be varying independently.[(2)]

We know that *every rate* is the ratio of two changes that may occur in two related quantities.

For example, consider the volume of a sphere, defined by

$$V(r) = \frac{4}{3}\pi r^3 = kr^3 \left(\text{where constant } k = \frac{4}{3}\pi\right)$$

Note that, $V(r)$ will change if "r" is changed. Now consider the situation when "r" is increased by 2 units from 1 unit to 3 units. We get

Average rate of change in $V(r)$ (for increase in "r" by 2 units)

$$= \frac{\text{Change in } V(r)}{\text{Change } \mathit{in}\ r} = \frac{k(3)^3 - k(1)^3}{(3-1)}$$

$$= \frac{k(27-1)}{(3-1)} = \frac{26k}{2} = 13k \tag{1}$$

Again, consider the situation when r is increased by 2 units from 2 units to 4 units. We get

Average rate of change in $V(r)$ (for increase in "r" by 2 units)

$$= \frac{\text{Change } \mathit{in}\ V(r)}{\text{Change } \mathit{in}\ r} = \frac{k(4)^3 - k(2)^3}{(4-2)}$$

$$= \frac{k(64-8)}{(3-1)} = \frac{56k}{2} = 28k \tag{2}$$

[(1)] Here, it may be mentioned that a falling object (dropped from a tower), orbiting spacecraft, growing populations, decaying radioactive material, rising consumer prices, etc., can all be modeled through calculus. – *The Mathematics of Change*

[(2)] Once we have defined *the rate function*, it will be found that the same principle, suitably interpreted, lies behind all our calculations and applications of derivatives. Later on, it will be found quite useful to see how the graphs of $f(x)$ and $f'(x)$ are related.

Also, it can be checked that average rate of change in $V(r)$, for *one unit increase in* "r" varies as follows:

Change in r	Average rate of change (*for* one unit increase in r)
From $r = 0$ to $r = 1$	k
From $r = 1$ to $r = 2$	$7k$
From $r = 2$ to $r = 3$	$19k$
From $r = 3$ to $r = 4$	$37k$

From the above data, we observe that for two units increase in "r", *the average rate of change* in $V(r)$ is not the same as can be seen from (1) and (2) above. Similarly, the average rate of change in $V(r)$ for a unit change in "r", is different for two *different values of* "r". This observation indicates that *the rate at which* $V(r)$ *increases must be different, for different values of* "r".

The rate of change at *any particular value of* r is called the rate of change (or the instantaneous rate of change) for that value of r. Our interest lies in computing the *actual rate of change in* $V(r)$ *at each value of* "r". This statement might look confusing or even useless to a beginner since, so far, we neither know the usefulness of "the actual rate of change of $V(r)$" nor do we know the method of computing it.[3]

9.1.2 From the Average Rate to the Actual Rate (or the Instantaneous Rate) of Change

Consider *an object moving in a straight line*. A parameter of our interest is its speed. Let the moving object be a car, which may be moving with a *constant speed* or varying speed, with or without stoppages in between. In all situations, we can always compute the average speed of the object by noting *the distance traveled in an interval of time*, and using the formula

$$\text{Average speed} = \frac{\text{Distance traveled}}{\text{Time taken}}$$

Note that, the *average speed does not give any information about the variation in speed during any interval of time.* If one plans to travel 160 km by this car, and hopes to make the trip in 4 h, then it suffices for him to know that *he must travel at an average speed of* 40 km/h. Thus, in such cases, what matters is the *average speed.*

Calculus is not meant for computing average speed(s) (or average rate(s)). These can be computed using simple arithmetic. *Differential calculus* is designed to compute *actual rate(s) of change* (*or instantaneous rate(s) of change*) of *varying quantities.*

To emphasize the importance of *actual speed*, imagine the situation when the car strikes a tree. Here, what matters is the *actual speed of the car at the time of strike*. Similarly, as a bullet travels through air, its average velocity may be around 2000 km/h (i.e., 555 m/s, approximately), but what counts when it strikes a person is the *actual velocity at the instant of striking*.

If it is 2 km/h (i.e., 0.55 m/s), the bullet will drop (without causing any harm), but if it is 555 m/s, the person will drop.

The speedometer of a vehicle indicates its actual speed at each instant (to keep the driver alert, so that he could use necessary controls to avoid accidents). Again, we are interested neither in the speeds of vehicles meeting with accidents nor in the velocities of bullets striking

[3] We have observed that "the average rate of change" can be computed using algebra; but it will be seen shortly that, in general, we cannot compute actual rate of change using algebra.

individuals. However, our interest lies in being able to compute the actual rate(s) of change of varying quantities because there are many scientific problems that require the use of instantaneous speed.[(4)]

Consider a function f (*in the form of a formula*) defining the way in which the quantity $y\ [= f(x)]$ changes with x. Then, the differential calculus helps in computing a new function, denoted by $f'(x)$, which *describes the actual rate of change in f(x) with respect to x*. The new function $f'(x)$ is obtained from the given function $f(x)$, through a definite procedure, to be discussed shortly.[(5)]

In practice, the speed of a car (or any other vehicle) is always varying, reasonably close to the desired average speed. For our purpose, let us assume that a car moves in a straight line according to the formula $y = f(x) = 3x^2$, connecting the distance traveled with time (y in meters, x in seconds). Note that, with the passage of time (i.e., for higher values of x) the car can attain a very high speed, and our interest lies in computing actual speed of the car, at any instant of time. In fact, the actual speed (or instantaneous speed) of the car can be read from the speedometer or it can be obtained by substituting the value of the instant "x" in the formula of *the derivative function* to be obtained from the given function $f(x)$.[(6)]

Note (1): It will be found that in general $f'(x)$ depends on x, except when it is a constant function, (that is, $f'(x) = c$). Also, in certain cases *it will be observed that* $f'(x)$, is not defined for certain values of x, for which $f(x)$ is defined. For the time being, we assume that (unless otherwise noted) our functions are well behaved, which means that the given function $f(x)$ and its derivative function $f'(x)$ both have smooth, unbroken graphs.[(7)]

To get *an idea of the actual speed at any instant*, the simplest way is to compute the *average speed over shorter and shorter intervals* of time. This average speed may be considered very close to *the actual speed (i.e., the speedometer speed) at any time during the same small interval*. However, to get a systematic and definite procedure (to define derivative function), we consider a function f, given by $y = f(x)$ and make a very small positive change Δx in the value of x (at $x = x_1$).

Let the corresponding change in the value of $y\ [= f(x)]$ be computed. This change in the value of y may be any real number (positive, negative, or zero). Then *the ratio* of resulting change, that is, the change in the value of y to the change in the value of x, gives an *approximate value of* $f'(x)$ *at* $x = x_1$. *Our interest lies in this ratio and we shall use it in obtaining the desired formula for* $f'(x)$.

(4) For example, an object near the surface of the Earth falls with varying speed according to a known law $s = 16t^2$ (s in feet, t in second(s)). Therefore, to know its speed at any time means to know its instantaneous speed. It is also known that when an object is far from the Earth and falls toward it under gravitational attraction, then not only its velocity but also its acceleration varies from instant to instant.

A deep investigation of all such motions requires understanding of instantaneous speed and instantaneous acceleration. The problems scientists have faced since the seventeenth century are not only that of treating instantaneous speed and acceleration but also instantaneous rates of changes of forces, energies, intensities of light and sound, and hundreds of other instantaneous rates of change.

(5) Note that, while the *function f tells the way in which the value f(x) changes with x*, the (new) function $f'(x)$ is expected to tell the *actual rate at which f(x) changes with x at each value of x*.

(6) Any function f can be used to build new functions derived in one way or another, from f. For example, consider the functions: $f_1(x) = 2f(x)$, $f_2(x) = f(x) + a$, $f_3(x) = [f(x)]^3 + 2f(x)$, and $f_4(x) = (f(x+0.1) - f(x))/0.1$. All these functions may be called "relatives" of f, and the possibilities are endless. Among all the possible functions one might obtain from the given function $f(x)$, the *derivative* function $f'(x)$ is the most important. Our interest lies in establishing the procedure for defining the derivative function of a given function $y = f(x)$.

(7) This assumption is useful to overcome the initial difficulties in understanding the concept of derivatives. As we develop new languages and tools, we will be able to handle complicated functions for computing their derivatives. Of course, all such functions are defined on intervals.

Here is an informal description of derivatives:

9.2 DEFINITION OF THE DERIVATIVE AS A RATE FUNCTION

Let f be any function. The new function f', called the derivative function of f, is defined by the rule:

$$f'(x) = \textit{instantaneous rate of change of } f \textit{ at } x.$$ [8]

This definition tells us that if f is any function defined by the formula $y = f(x)$, then $f'(x)$ *represents* "*the rate at* x" *at which* y *changes* with respect to x. For instance, *the statement* $f'(3) = 5$ *means that if* $x \approx 3$ *then increasing* x *by a small amount* produces about five times as much increase in $f(x)$.

9.3 INSTANTANEOUS RATE OF CHANGE OF y [$=f(x)$] AT $x = x_1$ AND THE SLOPE OF ITS GRAPH AT $x = x_1$

Most functions of our interest can be graphed, *hence it is natural to expect that the graphs of the functions must reveal useful information about their derivatives.* We ask the question: *What does the derivative mean graphically?*

Suppose, a car starting at some point on the x-axis moves (in the positive direction) a distance given by the formula

$$y = g(x) = 2x + 3 \; [x \text{ units of time, } y \text{ units of distance}]$$

(Let us not worry about the units of y and x.] From the above formula, *it can be easily checked that in each unit of time, the car moves 2 units of distance*. In other words, the car moves with a *constant speed* of 2 units. *The graph of this motion is a straight line with slope* 2. Whenever an object moves with *any constant speed*, the graph of distance against time is a straight line with positive slope, which is numerically equal to the constant speed (see Figure 9.1).[9]

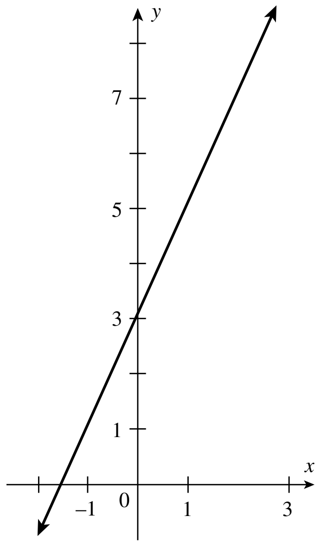
In other words, the slope of a straight line represents the constant speed of the moving object. Note that any constant speed may be looked upon as the instantaneous speed (of the moving object), which represents the derivative of the given function.

Next, suppose the car accelerates *gradually* in the positive direction of x-axis. Let this motion be represented by the graph (Figure 9.2), which we may call the function $h(x)$ (we have not defined $y = h(x)$ by any formula).

Observe from this graph that the value $h(x)$ (i.e., the height of the graph from x-axis) *increases with* x, indicating that the car is *gradually accelerating* (i.e., moving greater and greater distance per unit time as the time x passes). It follows that the slope of the tangent line on each point of the graph increases with x. In other words, the slope at a point on the

[8] The phrase *instantaneous rate of change* is applicable even in the cases where nothing seems to be moving. We say that *a road bends suddenly*. We can discuss how *quickly* the direction of a railway line changes. Words such as "*suddenly*" and "*quickly*," which are originally meant to describe a motion, can also be used to *describe motionless objects*. *Differential calculus* is, therefore, a subject that *can be applied to any thing* that *moves* or *changes*.

[9] The speed of a particle is defined as the absolute value of the instantaneous velocity. Hence, speed is a nonnegative number. The terms speed and instantaneous velocity are often confused. Note that the speed indicates only how fast the particle is moving, whereas the instantaneous velocity also tells the direction of motion.



$y = g(x) = 2x + 3$

FIGURE 9.1 Constant speed.

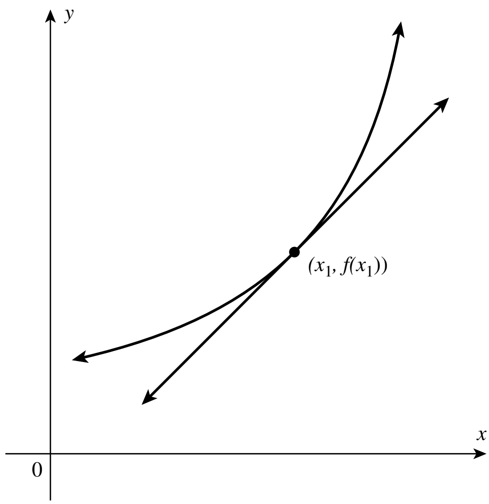


FIGURE 9.2 Varying speed.

curve is the slope of the tangent line at that point. We call it the slope of the curve at that point. *Note that, by using the concept of slope of a line, we have now defined the slope of a curve at a point.*

In view of our observation that the slope of a straight line represents the constant speed (or the instantaneous speed), we conclude that the slopes of the curves (representing functions) can be interpreted to represent instantaneous rate(s) of change or (derivatives) of functions. We give another informal description of derivatives.

9.3.1 Definition (The Derivative as a Slope Function)

Let $f(x)$ be any function given by $y = f(x)$. The derivative function $f'(x)$ is given by the rule

$$f'(x) = \textit{slope of the graph of } y = f(x) \textit{ at any point } x.$$

So far, we have introduced only what it means for $f'(x)$ to be the derivative of $f(x)$. We have neither given its definition nor described the method of obtaining it from the original function $f(x)$.

From the above description, we get that, *to find the instantaneous rate of change of a given function $y = f(x)$ at a desired point x_1, we should compute the slope of the tangent line at the point $(x_1, f(x_1))$ of the graph of f.*

From all that we have discussed so far, to understand the *derivative*, we proceed to consider the following two problems, which are the foundation of *differential calculus*.

(a) *The Problem of the Tangent Line*: To define *the tangent line to a curve at a point* and to find its slope at that point.[10]

(b) *The Problem of Instantaneous Velocity*: An object is moving in a *straight line. We are given a rule* (a function), which *tells where the object is at any time*, and we are asked to find *how fast it is moving at any desired time.*

The two problems, *one geometric* and *the other mechanical*, might appear to be unrelated, but the fact is that they define one and the same problem, as will be clear from the discussion that follows. Let us discuss first *the problem of the tangent line*.

9.3.2 The Problem of the Tangent Line

In our school geometry, we learnt that *the tangent to a circle is a line, which meets the circle, at exactly one point*. To draw a tangent line, to a circle at any given point P, we join "O," the center of the circle, with P. Then, *the line perpendicular to OP at P is the tangent to the circle, at P* (see Figure 9.3).

Using this property of the circle, it is possible to draw a tangent line to a circle, by geometric methods. Euclid's notion of a tangent, as a line touching a curve at one point, is all right for circles, but completely unsatisfactory for most other curves, as will be clear from the following discussion.

Suppose, we want *to draw a tangent line to any other curve, which is not a circle*. The problem is: *How do we get such a line?* Let us try to understand what is meant by a line being a tangent to a curve.

In Figure 9.4a, the lines l_1 and l_2 intersect the curve at exactly one point P. Intuitively, *we would not think of l_2 as the tangent* at this point, but it seems natural to say that l_1 is.

Also, in Figure 9.4b, we would consider l_3 *to be the tangent at P, even though it intersects the curve at other points*. From these examples, it is clear that *we must drop the idea that a tangent line intersects a curve at only one point.* To develop a suitable definition of tangent line, we have to use the *limit concept* as follows.

[10] It is assumed that the given functions are defined on intervals and have smooth and unbroken graphs.

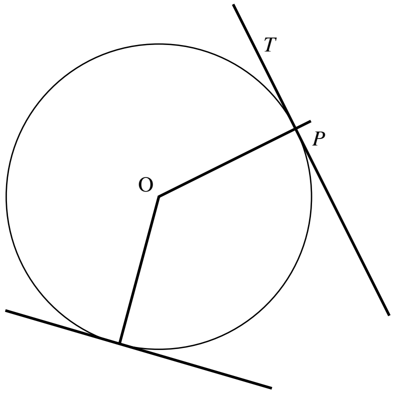


FIGURE 9.3 Tangent line to a circle at any point P.

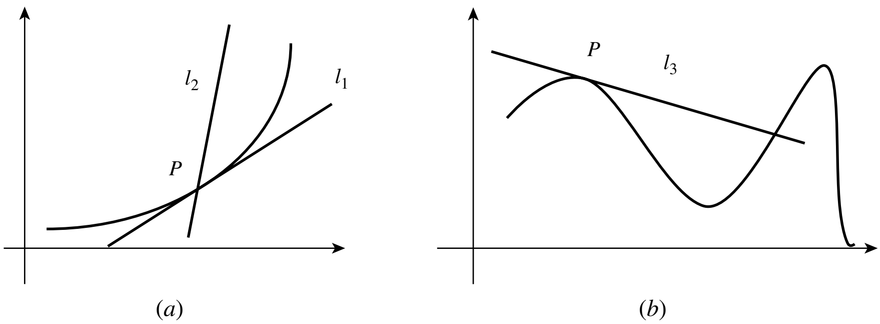


FIGURE 9.4

Consider, a curve that is the graph of a function $y = f(x)$. Let $P(x_1, y_1)$ be a *fixed point* on the curve and $Q(x, y)$ be a nearby *movable point* on that curve. The line through P and Q is called *a secant line*.[11]

Now imagine that the point Q moves *along the curve* approaching closer and closer to P. Then, *the secant PQ is approaching nearer and nearer to a definite line PT*, as shown in Figure 9.5.

While Q approaches P, it *has to pass through an infinite number of positions along the curve* and accordingly *the secant PQ has to pass through an infinite number of positions to approach closer and closer to the definite position PT*. (Note that Q can be considered arbitrarily close to P, but we never allow the point Q to coincide with the point P.) Thus, *the line* PT *is the limiting position of the secant line PQ* and it is the same whether Q approaches P from the left or from the right. This *common limiting position of secant lines is called the tangent line* to the curve at P.

[11] A *fixed point* is identified with coordinates (x_0, y_0), (x_1, y_1), and so on, wherein the coordinates are with subscript 0, 1, 2, and so on. An arbitrary point or a *movable point* is expressed with coordinates (x, y), wherein the coordinates are without subscript.

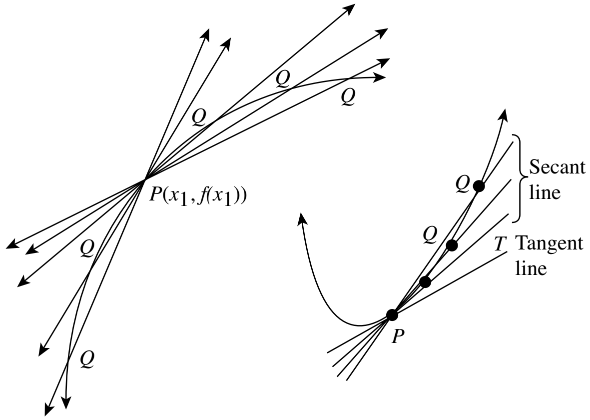


FIGURE 9.5 Limiting position of secant line is defined as the tangent line.

This definition is in agreement with our intuition and avoids the failings previously discussed. We now give the following definition.

9.3.3 Definition (Tangent Line to a Curve at a Point P)

The tangent line PT is the limiting position of the secant lines PQ, as Q approaches P, along the curve.

To draw the tangent line at any given point P of a curve, *it is necessary to know the slope of the tangent line at P. The method of coordinate geometry gives the slope of any secant line* (which passes through any two points on the curve) *but fails to give the slope of the tangent line at any point* of the curve. Let us see why?

To see the actual difficulty, note that *the slope of any secant line denoted by* m_{sec} *passing through two distinct points* $P(x_1, y_1)$ and $Q(x_2, y_2)$ on the curve is given by

$$m_{\text{sec}} = \frac{y_2 - y_1}{x_2 - x_1} \tag{3}$$

Observe that as $Q \to P$ along the curve, the secant line PQ approaches the limiting position PT and hence *the slopes of the secant lines PQ approach the slope of the tangent line PT.* Now consider the expression (3), which gives the slope m_{sec} of the secant line PQ. As $Q \to P$, $x_2 \to x_1$, $y_2 \to y_1$ and $(x_2 - x_1) \to 0$. Therefore, by using (3), we are unable to compute the slope of tangent line. Thus, *although we are able to visualize the existence of the tangent line at P, we are unable to compute its slope at P.*

To find the slope of the tangent line at the point $P(x_1, f(x_1))$, we choose another point $Q(x_2, f(x_2))$ on the curve, distinct from P (see Figure 9.6).

Now we express *the slope of the secant line PQ* as

$$m_{\text{sec}} = \frac{f(x_2) - f(x_1)}{x_2 - x_1} \quad [\text{where } f(x_2) = y_2 \text{ and } f(x_1) = y_1]$$

Since x_2 *can be obtained by adding a nonzero number* h *to* x_1, *we can write* $x_2 = x_1 + h$, *where* $h \neq 0$. Here, h is a *variable nonzero number, positive or negative*. Thus, *the slope of the*

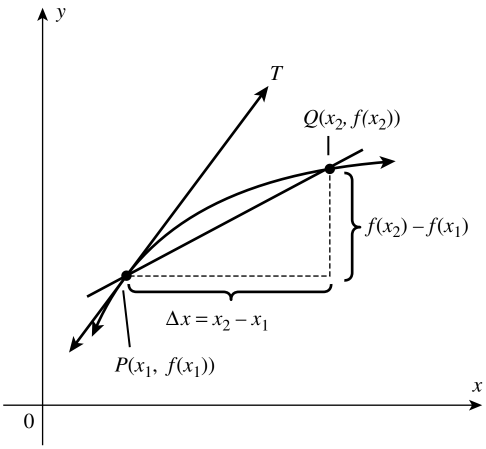


FIGURE 9.6 Tangent line PT at P.

secant line PQ may be expressed as

$$m_{\text{sec}} = \frac{f(x_1+h)-f(x_1)}{(x_1+h)-x_1} = \frac{f(x_1+h)-f(x_1)}{h}$$

Since, the *tangent line is the limiting position of secant lines, the slope of the tangent line at P is the limiting value of the slopes of secant lines PQ as $Q \to P$*. But, as $Q \to P$ along the curve, $x_2 \to x_1$ and so $h \to 0$. (Note that at any stage $h \neq 0$, for if $h = 0$, then $x_2 = x_1$ and then no secant line would exist.)

Therefore, *the slope of the tangent line* at $P(x_1, f(x_1))$ is given by

$$\lim_{h\to 0}\frac{f(x_1+h)-f(x_1)}{h} \tag{4}$$

provided the limit at (4) *exists.*

If the limit at (4) exists, then in view of the definition of derivative as a slope function, we identify the above limit as the derivative function (or the rate function) of f at x_1.

Since, x_1 in (4) can be *any* number (in the domain of f), we may replace it by x *to make the result more general*. Thus, *our problem condenses to evaluating the limit.*

$$\lim_{h\to 0}\frac{f(x+h)-f(x)}{h} \tag{5}$$

which gives the slope of the tangent line at any point $P(x, y)$ of the curve $y = f(x)$, *provided the limit at (5) exists*, and we call it the derivative function of $f(x)$ at x and denote it by the symbol $f'(x)$.

The above discussion suggests that to find the derivative of the given function $f(x)$, we must construct a new function $(f(x+h)-f(x))/h$, $(h \neq 0)$ without bothering to know what this would mean, and take its limit as $h \to 0$. If the limit $\lim_{h\to 0} (f(x+h)-f(x))/h$ exists, we call this limit as the *derivative function* of $f(x)$ and denote it by $f'(x)$. Note that derivative of $f(x)$ can be defined aside from any geometric meaning attached to $f(x)$.

The above discussion also suggests that we can define the *slope of a curve at any point of the curve*, as the slope of the tangent line at that point, obtained from the limit at (5) if it exists.

9.3.4 Definition

The slope of a curve at a point P is the slope of the tangent line at P. (Note that the concept of *slope of a curve* at a point is not to be found anywhere in geometry.)

Not every curve has a definite single tangent at each of its points. For example, if the graph of a function has a sharp corner then *there will be two tangent lines at such a point, one from the left and the other from the right, with different slope. In other words, the slope at any sharp corner of a curve is not unique*. For example, see the graph of $y = |x|$ at the origin.

Besides, there are functions whose graph may have vertical tangent line(s) at certain point(s). We know that the slope of the tangent line is not defined at such points.

For example, this happens in the graph of $y = x^{1/3}$, at the origin. *If the slope of the curve cannot be defined at certain points, we say that the function does not have derivative at those points*. This amounts to saying that the limit at (C) does not exist at such points (see Figures 9.7 and 9.8).

It is now proposed to go back again to *the concept of actual rate of change of a function at a point* (or the actual velocity at any instant), in more details.

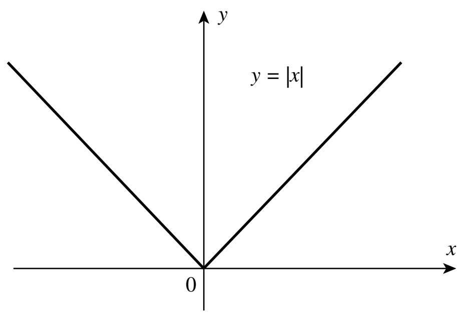


FIGURE 9.7

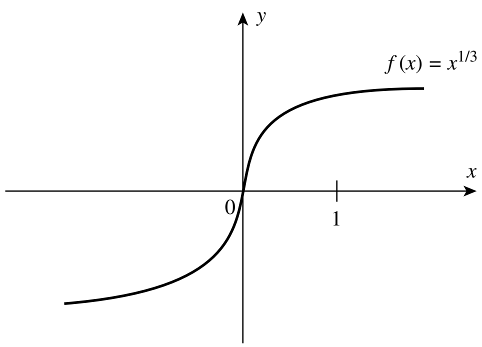


FIGURE 9.8

9.4 A NOTATION FOR INCREMENT(S)

Let $y = f(x)$ be a function of x. The symbol δx (sometimes Δx) is used to denote *an arbitrary nonzero increment in the value of the independent variable* x and the symbol δy (or Δy) is used to denote *the corresponding change in the value of dependent variable* y $[= f(x)]$.[12]

From the relation $f(x) = y$, we write

$$\begin{aligned} f(x+\delta x) &= y+\delta y \\ &= f(x)+\delta y \text{ (since } y = f(x)) \\ \therefore \quad \delta y &= f(x+\delta x)-f(x) \end{aligned}$$

Thus, $[f(x + \delta x) - f(x)] = \delta y$ is the *increment* (or the *resulting change*) in the value of the function, corresponding to the increment, δx in x.[13]

9.4.1 The Increment Ratio (or the Difference Quotient) at x_1

The ratio $(f(x_1+\delta x)-f(x_1))/\delta x = \delta y/\delta x$ is called *the increment ratio* (or the *difference quotient*) *of the function f(x), at the point* x_1. This increment ratio represents the "*average rate of change*", *in the value f(x)*, relative to the change δx at x_1. *Our interest lies in computing the actual rate of change (or the instantaneous rate of change) in the value f(x) relative to the change* δx *at* x_1. Note that for this purpose, the increment δx has no role to play.

9.5 THE PROBLEM OF INSTANTANEOUS VELOCITY

We have seen that in certain situations the instantaneous rate of change of a varying quantity is more significant than its average rate of change—it may be a vehicle hitting a tree or a bullet hitting a person.[14]

The following examples, connecting varying quantities, may be found useful:

(i) As one travels, his distance from the starting point continually changes, as does the time that elapses.[15]

[12] The symbol "δ" is the Greek small "d" and is pronounced "delta". Contrary to the ordinary usage of algebra, δx does not mean a product of δ and x. It is a single symbol and hence the letters should not be separated. A single letter h and k can also be used. An advantage in using the composite symbols δx and δy (instead of single letters h and k) will be noted when we define the derivative of the function $y = f(x)$ as the limit $\lim_{\delta x \to 0} \delta y/\delta x$.

[13] Observe that δx is an arbitrary nonzero increment (positive or negative) in the value of x and δy $[= f(x + \delta x) - f(x)]$ is the corresponding increment in y, which can be any real number (positive, negative, or zero). (Note that for a constant function $y = f(x) = c$, δy will always be zero.) Thus, while x and $x + \delta x$ represent two distinct points on the x-axis, the corresponding values $f(x)$ and $f(x + \delta x)$ need not be distinct on the y-axis.

[14] We may consider another example of a person traveling in a train at a speed of 200 km/h or so. He may hardly be conscious of the speed but a sudden decrease in the speed can throw him out of his seat. In fact, it does not hurt to travel at a high speed such as 200 km/h (or even 1000 km/h). What does hurt is the sudden change in speed.

[15] We have already seen that if the law of motion is a linear function of time, the speed of the object is constant throughout. It means that "average speed" of the object for "any" interval of time is the same. Obviously, then it must also represent the actual speed of the object, at any instant. Furthermore, considering only the algebraic aspect, if the law of motion involves higher powers of t, then "the average speeds are different for different time intervals" and hence the actual speeds are different at different instants. In such cases, computation of instantaneous speed is not simple, even in the case of polynomial functions, when the law of motion involves powers of $t > 2$. On the other hand, if the law of motion involves trigonometric, exponential, or logarithmic functions, or even algebraic functions involving fractional, then differential calculus can help only in computing the actual rates (or the actual speed of moving object at any instant) provided the law of change (or the law of motion) is expressed by a function.

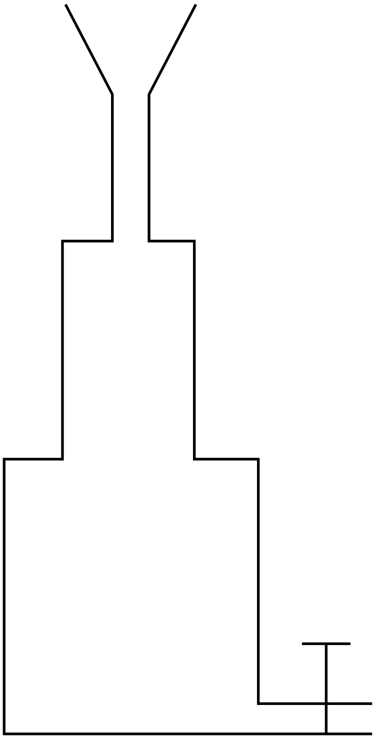


FIGURE 9.9

(ii) A bob on a spring (or on a simple pendulum) moves with constantly varying speed and acceleration.

(iii) For any curve (other than a straight line), the slope of the curve changes from point to point.

(iv) In electrical circuits, as a capacitor is charged (or discharged), the voltage across it changes during the time of charging (or discharging).

(v) In DC circuits, the current takes time to grow to its steady value after the circuit is completed.

Now, we give below some simple experiments by which we can clearly observe the *varying rates of change*.

(a) If water is poured at a constant rate in a glass pot having different diameter in different portions, then we can easily see that water level rises at different rates in different portions (see Figure 9.9). This arrangement also suggests that if water is poured (at a constant rate) in a conical pot, then *water level must rise at different rates at different heights.*

(b) If we walk toward a street light bulb (or go away from the pole), then the rate at which the length of our shadow changes at different distances from the pole is not the same.

(c) It is easy to check that if the radius of a sphere changes, the rate at which the volume of the sphere changes is different for different values of the radius. (This, we have already discussed.).

Having realized the importance of the fact of instantaneous rate of change, we would like to be able to compute the instantaneous rate of change of varying quantities.

But there are certain difficulties in computing instantaneous rates. The first question is: *What is an instant?* It may be difficult to give a good physical definition of an instant, but *the notion of an instant does have some physical meaning*.

For example, when two objects collide, we think of this happening at an instant. A lightning flash is practically instantaneous. We speak of an event happening at 6 o'clock and refer thereby to an instant. Thus, even in common situations, we think of and utilize the notion of an instant. Let us discuss about this notion in details.

9.5.1 The Notion of an Instant

Mathematically, we have less trouble with the concept of an instant. *A mathematician thinks of time as a measurable quantity, measured, say, in seconds*. Then, the passage of time is recorded by the number of seconds measured from some event that is represented as happening at zero time. *Thus $t = 2$ is an instant, 2s after the event that the mathematician has selected as happening at zero time*.

Having understood the notion of an instant, let us try to understand the notion of instantaneous speed. It is true that a person traveling in an automobile has a speed at each instant. *But there is difficulty in stating just what we mean by instantaneous speed, and if we do not know precisely what it means, then we certainly shall have trouble in calculating it.*(16)

9.5.2 From Average Speed to Instantaneous Speed

We know that speed is *the rate of change of displacement compared to time*. Therefore, *the average speed, which applies over an interval of time (rather than at an instant),* is the *distance traveled during any time interval* divided by the time taken. Let the distance traveled by an object in a time interval of "δt" units be "δs" units of distance. Then, we can write

$$\text{Average speed} = \frac{\text{Distance travelled}}{\text{Time taken}} = \frac{\delta s}{\delta t} \text{ units of speed}$$

The definition of average speed permits us to calculate it very easily. Hence, *we are tempted to define and calculate instantaneous speed in the same way*. But at an instant, *zero distance is traveled* and *zero time elapses*. Hence, to define instantaneous speed *as distance divided by time* leads to the expression $0/0$, which is meaningless from a mathematical point of view. Here then lies the problem.

Physically we have every reason to believe that there is such a thing as an instantaneous speed, yet we face difficulty in defining it and calculating it mathematically.(17)

(16) *Note* (2): From the above discussion, one might think that "*differential calculus*" is difficult to learn, but this is not true. Once the basic ideas of *differential calculus* have been grasped, a whole world of problems can be tackled without great difficulty. It is a subject well worth learning and this book is compiled to achieve this goal systematically, maintaining the interest and enthusiasm of the reader.

(17) Of course, now we know that this difficulty can be overcome only by applying the method(s) of evaluating the limit $\lim_{\delta t \to 0} \delta s/\delta t$ and check if the limit exists.

9.5.3 Approaches by Newton and Leibniz

Let us consider *how Newton and Leibniz approached the problem of defining and calculating instantaneous rates*. Though there were differences in their approaches, we shall ignore them and examine the subject in the form in which it has been standardized in recent years.

To start with, let us consider the formula

$$s = 16t^2, \ (t \ \text{seconds}, \ \text{s feet}) \qquad (6)^{(18)}$$

that governs the *free fall motion of a ball*, relating the distance the ball falls to the time it falls.

Note (3): The formula (6) is strictly correct only if the object falls in vacuum. The factor 16 is approximate. Also, *note that the ball falls vertically in a straight line and thus we are considering the motion in a straight line.*

Suppose it takes exactly 4 s for the ball to hit the ground, after it is dropped from a tower, and suppose *it is required to compute the instantaneous speed of the ball at the end of third second.* We prepare the following table:

From this table, we observe that *the average speed of the ball keeps on increasing with time* and therefore its *instantaneous speed* is increasing as the time passes. *What can we say about its speed at the end of the third second?*

Observe that the ball started with no velocity at all, and increased its speed under gravitational attraction. *In first 3 s, the ball falls by 144 ft and so the average speed of the ball during this period is 48 ft/*s. Obviously, then its actual speed at the end of third second must be greater than 48 ft/s, to balance its slow initial speed. Next, we observe that the distance the ball fell by during the third second is 80 ft. Hence, *its actual speed at the end of third second must be greater than 80 ft/s.*

It is reasonable to say that the actual speed at any instant will not differ very much from the average speed during the previous 10th of a second. Furthermore, if we compute the average speed for the previous 1000th of a second, then it will still be closer to the *actual speed*, at the instant under consideration. In other words, *if we take the average speed for smaller and smaller intervals of time around the instant under consideration, then we shall get nearer and nearer to the true speed at the instant in question.*

For many practical purposes, *the average speed during a 1000th of a second may be regarded as the exact speed, but in reality it is still different from the actual speed.* It is important that *we should not agree to accept any approximate value of the average speed howsoever close to the actual speed it might be.*

(Here, we introduce the area of logical thinking, leading to the concept of limit.)

The distance "*s*" traveled by the ball in 3 s is given by

$$s = 16t^2 = 16(3)^2 = 144 \text{ ft} \qquad (7)$$

(18) Formulas for free fall near the Earth's surface:

1. $s = (1/2)gt^2$; s = distance, t = time, g = gravitational constant.
2. $s = 16gt^2$; s = feet, t = seconds, $g = 32\,\text{ft/s}^2$.
3. $s = 490t^2$; s = centimeters, t = seconds, $g = 980\,\text{cm/s}^2$.
4. $s = 4.9t^2$; s = meters, t = seconds, $g = 9.8\,\text{m/s}^2$.

Also, using the formula (6), we can find where the ball will be at the end of $(3 + \delta t)$ s, δt being an *arbitrarily small additional time interval*, after third second. Then, we have

$$\begin{aligned} s+\delta s &= 16(t+\delta t)^2 = 16(t^2+2t\delta t+\delta t^2) \\ &= 16(9+6\delta t+\delta t^2) \end{aligned}$$

$$144+\delta s = 144+96\delta t+16\delta t^2 \tag{8}$$

$$\therefore \quad \delta s = 96\delta t+16\delta t^2 \tag{9}$$

But, we know that *the average speed during the additional time interval* δt *is given by* $\delta s/\delta t$. Therefore, we divide both sides of formula (9) by $\delta t(\delta t > 0)$ and obtain

$$\frac{\delta s}{\delta t} = \frac{96\delta t+16\delta t^2}{\delta t} \tag{10}$$

From formula (10), we observe that *the average speed* $\delta s/\delta t$*(over the time interval* δt*) is a function of* δt. Furthermore, *since* $\delta t \neq 0$*, we can divide the numerator and denominator on the right side of* (10) *by* δt and obtain the simplified expression for $\delta s/\delta t$. Thus, we get

$$\frac{\delta s}{\delta t} = 96+16\delta t, \quad (\delta t \neq 0) \tag{11}$$

Up to this point, Newton and Leibniz had calculated *the average speed of the falling body in the time interval* δt*, after the third second of the fall.* Moreover, since δt can be chosen as small as we please and the above algebra still holds; they had obtained the formula for average speed over any small interval, just after the third second.

But, the problem they set out to solve was to calculate the speed just at the end of the third second, that is, when $\delta t = 0$. One is tempted to put $\delta t = 0$ in (11) and obtain the answer 96. *Unfortunately, the answer happens to be correct, but the reasoning is incorrect* (Why?).[(19)]

To determine the value of $\delta s/\delta t$ (when $\delta t = 0$), we should use formula (10). But if we substitute $\delta t = 0$ in (10), we obtain $\delta s/\delta t = 0/0$, which poses the same difficulty (in obtaining instantaneous speed) as we mentioned at the outset. The situation is exasperating. *The answer we seek is obviously at hand in formula* (11)*, but we cannot use formula* (11).

One is tempted to cheat a little by putting $\delta t = 0$ in formula (11) and get the answer, but it is not correct as discussed above. (The new idea that Newton and Leibniz contributed comes in at this point.) They operated on the expression $96 + 16\delta t$ in the way *we would treat it today for computing its limit as* $\delta t \to 0$.[(20)]

Let us examine formula (11) when δt is not 0, and see what happens to it as δt approaches closer and closer to 0 in value. *For all nonzero values of* δt*, formula* (11) *is valid, and we see that as* $\delta t \to 0$*, the right side of* (11) (i.e., $96 + 16\delta t$*) approaches 96*. We therefore take 96 to be the actual speed at the end of third second.

[(19)] Note that (11) is derived from (10) with the condition that $\delta t \neq 0$. Thus, (11) is not the correct expression for the value of $\delta s/\delta t$ when $\delta t = 0$.

[(20)] No one can read the details of their writings on calculus without being amazed by the number of times they changed their explanations of the limit concept and still failed to get it right. Some of these explanations contained outright contradictions of earlier ones. It is fair to say that though both men had their hands on a sound idea, they could not grasp it securely. The concept of a limit, as we know it today, was not known to either Leibniz or Newton. (*The Calculus of a Single Variable* by Louis Leithold (p. 115), Harper Collins).

In this calculation, we observed the behavior of formula (6) when $\delta t \to 0$, but did not permit δt to assume the value 0. Thus, we *did use formula* (11)*, but the manner in which we used it, is all important*.

In other words, what we do is as follows: *Consider the formula at (6) and try to guess the number to which the expression* $(96 + 16\delta t)$ *approaches as* δt *approaches 0*. This number is called the limit of $(96 + 16\delta t)$ as $\delta t \to 0$, and *we take it as the actual speed at the end of the third second*.[21]

Observe that 96 *is also the value of the expression* $(96 + 16\delta t)$ *for* $\delta t = 0$. *This is equivalent to saying that the limit of the function* $(96 + 16\delta t)$ *as* $\delta t \to 0$ *and value of the function at* $\delta t = 0$ *both are same*. This is due to the fact that *we had a very simple function* "$16t^2$", *which is continuous*. This may not be the situation always, that is, the expression representing the difference quotient $\delta s/\delta t$ may not be as simple as the one in (10).

In other words, it may not be possible to simplify the function $\delta s/\delta t$, to the form that is so convenient for finding its limit.[22]

Since our requirement is to find the limit of the function $\delta s/\delta t$ as $\delta t \to 0$, we must *understand and respect the distinction between the limit of the function as* $\delta t \to 0$, and the value of the function at $\delta t = 0$. We have discussed about this distinction at length, in the process of formulating the δ, ε definition of limit in Chapter 7a. The general fact, *about speed at an instant*, is expressed as follows:

The speed at an instant is the limit approached by the average speed $\delta s/\delta t$ *as* δt *approaches* 0. In our problem, we applied this fact in computing the speed at the end of the third second, *by considering the average speed over smaller and smaller intervals, just exceeding the third second*.[23]

To appreciate the full generality of the process of computing instantaneous rates, we must go a step further. Let us consider the function

$$y = f(x) = 16x^2$$

where $y\ [= f(x)]$ is the *dependent variable* and x is an *independent variable*, representing *any quantity*, and let us ask for *instantaneous rate of change of y with respect to x, at any value of x* (say at $x = x_1$).

(21) At this stage, it is important to consider one more situation that could create confusion in computing the actual speed. To understand it, let us go back to Table 9.1, which gives average speeds of the ball during various intervals. There we observed that during the period of 1 s before $t = 3$ (i.e., from $t = 2$ to $t = 3$), the distance covered by the ball is 80 ft, and during the subsequent period of 1 s after this instant (i.e., from $t = 3$ to $t = 4$), the distance covered is 112 ft. It is therefore reasonable to guess that the velocity at the instant $t = 3$ must lie between 80 and 112 ft/s. Accordingly, one might take the average of 80 and 112, and conclude that the velocity of the ball is 96 ft/s. Unfortunately, this answer is correct. We say "unfortunately" because as a rule taking the average does not give the correct velocity. In fact, it hardly gives the correct velocity. It is only when the law of variable is of the type $s = at^2 + bt + c$ will taking the average work. It is easy to understand why this happens. It may be checked that averaging gives a wrong result for the law, $v(r) = (4/3)\pi r^3$, as we had discussed earlier in this chapter.

(22) For example, it will be found that in the process of computing the instantaneous rate of change for the function $y = \sin x$ (to be discussed later in Chapter 11a), we have to use the result $\lim_{x \to 0} (\sin x)/x = 1$ and similarly for the function $y = a^x$ (in Chapter 13a), we have to use the result $\lim_{x \to 0} (a^x - 1)/x = \log_e a$. In both these cases, there is no way of canceling terms in the numerator and the denominator. Each quotient appears to approach 0/0, which is not defined. However, these limits exist and are evaluated by different techniques. Therefore, to be able to compute the derivatives of certain functions, it is important to learn method(s) of evaluating such limits.

(23) *Note* (4): Now, we propose to treat $y = 16x^2$ to represent any function and try to obtain its (actual) rate of change. It might appear as if we are repeating the whole thing that we have already discussed, but it is not so. If y is the distance s, x is time t, and x_1 is 3, then the above relation reduces to the earlier problem.

TABLE 9.1 Guessing the Actual Speed (of a Freely Falling Ball) From its Average Speed

No of second(s) the ball falls	Total distance fallen in feet ($s = 16t^2$)	Time interval during which the ball falls	Actual period of fall (δt)	Actual distance fallen (δs) during the actual period of fall (δt)	Average speed during the period ($\delta s/\delta t$)
0	0				
1	16	0 to 1 s	1 s	(16–0) = 16 ft	16 ft/s
2	64	1–2 s	1 s	(64–16) = 48 ft	48 ft/s
3	144	2–3 s	1 s	(144 − 64) = 80 ft	80 ft/s
4	256	3–4 s	1 s	(256 − 144) = 112 ft	112 ft/s

From the above relation, we can write the value of "f" at $x = x_1$, and denote it by y_1. Thus, we get

$$y_1 = f(x_1) = 16{x_1}^2 \tag{12}$$

Now, let us give *an arbitrary, nonzero, increment* δx to x_1 and let the corresponding increment in y_1 be denoted by δy. Then we have

$$\begin{aligned} y_1 + \delta y &= f(x_1 + \delta x) \\ &= 16(x_1 + \delta x)^2 \quad ^{24} \\ &= 16{x_1}^2 + 32x_1\delta x + 16\delta x^2 \end{aligned} \tag{13}$$

Therefore, we get from (13) – (12)

$$\delta y = f(x_1 + \delta x) - f(x_1) = 32x_1\delta x + 16\delta x^2 \tag{14}$$

To get *the average rate of change of y with respect to x, in the interval* δx, we divide both sides of (14) by δx and obtain

$$\frac{\delta y}{\delta x} = \frac{f(x_1 + \delta x) - f(x_1)}{\delta x} = \frac{32x_1\delta x + 16\delta x^2}{\delta x} \tag{15}$$

Observe that *right-hand side of* (15) *is a function of* δx. Fortunately, in this case, it is possible to simplify the RHS of equation (15) by dividing both numerator and denominator by δx, which is a *nonzero common factor in both*. Thus, we obtain from (15)

$$\frac{\delta y}{\delta x} = \frac{f(x_1 + \delta x) - f(x_1)}{\delta x} = 32x_1 + 16\delta x, \quad \text{where } \delta x \neq 0 \tag{16}$$

At this stage the *crucial step* is to see *what happens on the right side of* (16) *when* $\delta x \to 0$? In this, case *the answer is obvious*. As $\delta x \to 0$ in value, the quantity $16\delta x \to 0$ and so the limit

[24] Here, x_1 is a particular point on the x-axis and $(x_1 + \delta x)$ is another neighboring point, which is obtained by giving an arbitrary nonzero increment δx to x_1. The increment "δx" given to x_1 is arbitrary; hence, it is expressed without any subscript. Similarly, the resulting increment δy in y_1 is expressed without subscript.

of the function $(32x_1 + 16\delta x)$ is $32x_1$. Thus, the *instantaneous rate of change* of y $[= f(x)]$ with respect to x, at $x = x_1$, is $32x_1$. We write

$$\lim_{\delta x \to 0} \frac{\delta y}{\delta x} = 32x_1 \tag{17}$$

9.5.4 Formula (17) Tells Us Several Valuable Things

(a) The quantity x_1 was *any value of* x. Hence, in steps (12)–(17), we obtained the instantaneous rate of change of y with respect to x, for *any* value of x. *We may emphasize this fact by dropping the subscript* and writing

$$\lim_{\delta x \to 0} \frac{\delta y}{\delta x} = 32x \tag{18}$$

Notation: If the $\lim\limits_{\delta x \to 0} \delta y/\delta x$ exists, we use the notation $\mathrm{d}y/\mathrm{d}x$ to express this limit. We write

$$\lim_{\delta x \to 0} \frac{\delta y}{\delta x} = 32x = \frac{\mathrm{d}y}{\mathrm{d}x} \tag{19}$$

and call it the derivative of the function $y = f(x)$, and it is true for any value of x at which it is defined.

Remark: Since $\mathrm{d}y/\mathrm{d}x$ is a notation for a limit, *it must be treated as a single symbol,* though its appearance is that of a quotient.[25]
Thus, we have calculated the rate of change of y with respect to x, for an infinite number of values of x, in one operation. In fact, the relation (19) is a *new formula* (and *we look at it as a new function of x), which is derived from the given function* $y = 16x^2$. We say that *the function 32x is the derivative of the function* $y = 16x^2$. We can link the formula at (19) with that at (11). As a check on (19), let us note that at $x = 3$, $\mathrm{d}y/\mathrm{d}x = 32(x) = 32(3) = 96$, and this result agrees with the conclusion derived from equation (11).

(b) The second valuable implication of (18) or (19) is that *the result holds regardless of the physical meaning of y or x*. Remember that mathematics treats only pure numbers or pure special relationships. *Hence, we can apply the result to thousands of physical situations in which the original function,* $y = 16x^2$, *applies*. Moreover, the process that we used to obtain the result (19) can be applied to *any function*.

We can calculate *the rate of change of one variable with respect to the other at a value of the second variable*, by the same mathematical procedure that we used for calculating instantaneous rate of change of y with respect to x when $y = 16x^2$. For example, if y represents

[25] Leibniz used the suggestive but misleading notation $\mathrm{d}y/\mathrm{d}x$ for the instantaneous rate. It suggests that the instantaneous rate is obtained by considering an average rate, which is indeed a quotient. On the other hand, this notation is misleading in the sense that it represents instantaneous rate in the form of a quotient, whereas instantaneous rate is not a quotient but the limit approached by a quotient. Besides, the symbols $\mathrm{d}y$ and $\mathrm{d}x$ have not been given independent meaning. They are called differentials of dependent and independent variables, respectively, and their ratio $\mathrm{d}y/\mathrm{d}x$ can be interpreted as the derivative of y w.r.t x. Details are discussed in Chapter 16.

velocity and x time, we can calculate the rate of change of velocity compared to time at an instant. This instantaneous rate of change of velocity is called instantaneous *acceleration*.

As another example, the pressure of the atmosphere varies with height above the surface of the Earth. Given the formula that relates pressure and height, we can calculate the rate of change of pressure compared to height at any given height and the rate of change of surface area of a cube, with respect to the length of its edge.[(26)]

Remark: The original calculus problems of speed and acceleration did involve time and were concerned with rates at an instant of time. Our interest lies in computing the rate of change of the dependent variable y [$= f(x)$] with respect to the independent variable x at any value of x. All such rates are referred to as *instantaneous rates*, despite the fact that time may not be one of the variables involved.

9.5.5

From the above discussion, we note the following:

(i) If y is a function of x denoted by $y = h(x)$, whose graph is *a curve, then the slope of the curve at any point* $P(x, y)$ on the curve, is given by the limit $\lim_{\delta x \to 0} (h(x+\delta x) - h(x))/\delta x$, *provided this limit exists*. We denote it by dy/dx.

(ii) Consider a particle "P" moving in a straight line. Suppose *the position of the particle at any instant* "t" is expressed by function $y = g(t)$, then *the velocity of the particle at any instant t* is given by the limit $\lim_{\delta t \to 0} (g(t+\delta t) - g(t))/\delta t$, *provided this limit exists*. We denote it by dy/dt.

(iii) Let the velocity of a particle at any instant t be given by the function $v = \phi(t)$, then *the instantaneous rate of change of velocity at any instant t*, is given by the limit $\lim_{\delta t \to 0} (\phi(t+\delta t) - \phi(t))/\delta t$, *provided this limit exists*. We denote it by dv/dt. It is called the *instantaneous acceleration of the particle*.

Thus, if $y = f(x)$ is a given function, which may define a *curve*, *the position of a moving particle at time x,* or *the velocity of a particle at time x*, then the limit $\lim\limits_{\delta x \to 0} (f(x+\delta x) - f(x))/\delta x$, *if it exists*, will define, respectively, the *slope of the curve at a point, the velocity of the particle at an instant,* or *the acceleration of the particle at an instant*.

This limit also appears in many other contexts in economics, physics, and chemistry. Since it has various interpretations, it is treated as an abstract mathematical entity called a derivative, and its properties are studied in detail.

In view of the above, it is reasonable and natural to give the following useful definition of the derivative of a function at a point in its domain.

Now, we give the following formal definitions:

- *Derivative of a Function*:
  Let $y = f(x)$ be a given function defined in an open interval (a, b). Let the points x and $(x + \delta x)$ both belong to the domain of function $f(x)$, where δx is an *arbitrary nonzero number*.

[(26)] At the end of this chapter, we have discussed some interesting applications of the process of finding the derivatives or rates of change. There, we have computed the rate of change of the area of a circle with respect to radius and the rate of change of volume of a sphere with respect to its radius.

From the function $f(x)$, we form a new function

$$\phi(x) = \frac{f(x+\delta x) - f(x)}{\delta x}$$

The limit of this ratio, as $\delta x \to 0$, *may or may not exist*.
If

$$\lim_{\delta x \to 0} \frac{f(x+\delta x) - f(x)}{\delta x} = \lim_{\delta x \to 0} \frac{\delta y}{\delta x} \tag{20}$$

exists, then we call it the derivative of f with respect to x.

- *Derivative of a Function at a Particular Point*:
  The derivative of a function $y = f(x)$ *at a particular point* $x = x_1$ *in the domain of* f is given by the limit

$$\lim_{\delta x \to 0} \frac{f(x_1 + \delta x) - f(x_1)}{\delta x} \tag{21}$$

if this limit exists. It is denoted by $f'(x)$ or dy/dx.

If we replace $(x_1 + \delta x)$ by x, and accordingly δx by $x - x_1$, then the derivative of f at x_1 is given by

$$f'(x_1) = \lim_{x \to x_1} \frac{f(x) - f(x_1)}{x - x_1} \qquad (22)^{(27)}$$

if this limit exists.

In all cases, *the number x_1 at which f' is evaluated is held fixed during the limit operation*. Here, *x is the variable and x_1 is regarded as a constant*.

Note (5): Observe that if $f'(a)$ exists, then the letter x in (C) can be replaced by any other letter. For example, we can write

$$f'(a) = \lim_{t \to a} \frac{f(t) - f(a)}{t - a}$$

Note (6): The quotients $(f(x_1 + \delta x) - f(x_1))/\delta x$ and $(f(x) - f(x_1))/(x - x_1)$, both are called *standard difference quotients* of the function f, at the number x_1. If it is desired to

(27) Derivatives can be regarded as a rate measure. It measures the rate at which a function is changing its value with that of the variable upon which it depends. Thus, for a function $y = x^2$ since $(dy/dx) = 2x$, when $x = 1$, $y[= x^2]$ changes its value at two times the rate at which x is changing. Similarly, when $x = 3$, y is changing its value six times the rate at which x is changing.

compute the derivative of a function at a particular point $x = x_1$, and if $f'(x_1)$ exists, *then it is more convenient to evaluate the limit*

$$\lim_{x \to x_1} \frac{f(x) - f(x_1)}{x - x_1}$$

Consider the following example.

Example (1): Let $f(x) = (1/4)x^2 + 1$. Find $f'(-1)$ and $f'(3)$, and draw the line tangent to the graph of f at the corresponding points.

Solution: Using (22), we obtain

$$\begin{aligned} f'(-1) &= \lim_{x \to -1} \frac{((1/4)x^2 + 1) - (5/4)}{x - (-1)} \\ &= \lim_{x \to -1} \frac{(1/4)x^2 - (1/4)}{x + 1} = \lim_{x \to -1} \frac{(1/4)(x^2 - 1)}{x + 1} \\ &= \lim_{x \to -1} \frac{(1/4)(x - 1)(x + 1)}{x + 1} \\ &= \lim_{x \to -1} \frac{1}{4}(x - 1) \ [\because x \neq -1] \\ &= \frac{-1}{2} \end{aligned}$$

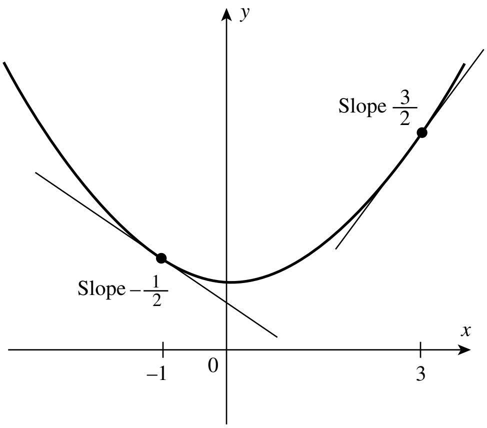
We also obtain

$$\begin{aligned} f'(3) &= \lim_{x \to 3} \frac{((1/4)x^2 + 1) - (13/4)}{x - 3} \\ &= \lim_{x \to 3} \frac{(1/4)x^2 - (9/4)}{x - 3} \\ &= \lim_{x \to 3} \frac{(1/4)(x^2 - 9)}{x - 3} = \lim_{x \to 3} \frac{1}{4}(x + 3) \\ &= \frac{3}{2} \end{aligned}$$

The lines, tangent to the graph at the corresponding points, are shown in Figure 9.10.

Next, we give the following formal definitions.

- *The Natural Domain of Derivative*: Let a set D be the domain of $f(x)$. The question is whether D is also the domain of $f'(x)$? In any case, *the domain of $f'(x)$ cannot be wider than the domain of $f(x)$* because to compute $f'(x)$ we use $f(x)$. *In general, the domain of $f'(x)$ is a subset of D*. It is obtained from D by elimination of those points x for which $f'(x)$ does not exist. It is called *the domain of differentiability of $f(x)$*.
- *Differentiation*: The process of *computing the derivative* of a function is called *differentiation*.



$f(x) = \frac{1}{4}x^2 + 1$

FIGURE 9.10

- *Differentiability of Functions*:
  (i) *Functions differentiable at a point*. If a function has a derivative at x_1 of its domain, then it is said to be differentiable at x_1.
  (ii) *Functions differentiable in an open interval.* A function is *differentiable in an open interval* (a, b) if it is *differentiable at every number* in the open interval.[28]
  (iii) *Functions differentiable in a closed interval.* If $f(x)$ is defined in a closed interval $[a, b]$, then *the definitions of the derivatives at the end points are modified so that the point* $(x + \delta x)$ *lies in the interval* $[a, b]$.

For example, if $x = b$ and $\delta x > 0$, then the point $(x + \delta x)$, that is, $(b + \delta x)$ *will not lie in the interval [a, b]*. Similarly, if $\delta x < 0$, then the point $(a + \delta x)$ will not lie in [a, b]. Hence, *we define the derivative at the end points as follows:*

$$f'(a) = \lim_{\delta x \to 0} \frac{f(a+\delta x) - f(a)}{\delta x}, \quad (\delta x > 0) \text{ and}$$

$$f'(b) = \lim_{\delta x \to 0} \frac{f(b) - f(b-\delta x)}{\delta x}, \quad (\delta x > 0)$$

- *Differentiable Function*: If a function is *differentiable at every number in its domain,* it is called *a differentiable function.*

Note (7): The above definition appears to be quite simple, but certain situations might create confusion. Hence, to get a clear idea of a *differentiable function*, it is useful to consider Examples (2) and (3) as follows:

[28] In this case, at every point of the open interval, the two-sided limit of the difference quotient exists.

Note (8): *It can be proved* that the derivative of x^α is given by

$$\frac{\mathrm{d}(x^\alpha)}{\mathrm{d}x} = \alpha x^{\alpha-1}, \quad (\alpha \in R)$$

However, for our purpose, let us consider (without proof)

$$\frac{\mathrm{d}(x^r)}{\mathrm{d}x} = r x^{r-1}, \quad (r \in Q)$$

Example (2): Let $f(x) = 3/x$, then $f'(x) = -(3/x^2)$.

Note that the domain of f is the set of all real numbers *except the number 0*. Also, $f'(x)$ exists at every real number except "0". Thus, *f is differentiable at every number in its domain*. Hence, *f is a differentiable function*.

Example (3): Let $g(x) = \sqrt{x} = x^{1/2}$, then $g'(x) = (1/2)(x)^{-1/2} = (1/2)\sqrt{x}$

Here, the domain of g is $[0, +\infty)$, but $g'(x)$ does not exist at $x = 0$. Thus, g is *not differentiable* at "0", which is *in the domain of g*. Therefore, we will say that g is *not a differentiable function*.

However, *if we define the function* $\sqrt{x}$ *in the open interval (0, ∞), then it becomes a differentiable function.*

In view of the above, we agree to say that *if the domain of f′ is the same as that of f, then f is a differentiable function.*

Nearly every function we will encounter is differentiable at all numbers or *all but finitely many numbers* in its domain.

Note (9): *The derivative of a function, at a given point (irrespective of its physical meaning) has the same numerical value.*

Note (10): To obtain the derivative of a function, by using the *definition* of the derivative, is known as *the method of finding the derivative from the first principle*.

Notation for Derivative:

We know that differentiation of $y = f(x)$ by *the first principle* involves two steps: *first,* the *formation of the difference quotient* and *second*, the *evaluation of its limit*.

If the limit, $\lim_{\delta x \to 0} (f(x+\delta x) - f(x))/\delta x = \lim_{\delta x \to 0} \delta y/\delta x$ *exists*, then we denote it by the symbol $f'(x)$ or $\mathrm{d}y/\mathrm{d}x$ and call it the *derivative of the function f(x).*

Note (11): We can look at the process of differentiation as an operation. The operation of obtaining $f'(x)$, from $f(x)$, is called differentiation of $f(x)$. The symbol $\mathrm{d}/\mathrm{d}x$ is assigned for this operation. We call it the *operator of differentiation.*(29)

- *The Operator of Differentiation* $\mathrm{d}/\mathrm{d}x$:

(29) The "operator of differentiation" is a new term that we have introduced here. This operator may be looked upon as a machine, which generates a numerical function at the output, in response to a "numerical function at the input".

Numerical function → [operator] → (new) Numerical function

In view of the above discussion, we can say that *the symbol* $\mathrm{d}/\mathrm{d}x$ *stands for the operation of computing the derivative of a given function by the first principle*. In other words, we agree to say that $\mathrm{d}/\mathrm{d}x$ *constructs* from $f(x)$, the difference quotient $(f(x+\delta x)-f(x))/\ \delta x$, and determines its limit as $\delta x \to 0$ (treating the difference quotient as a function of variable δx).[30]

Note (12): The notation $\mathrm{d}/\mathrm{d}x$ *should be interpreted as a single entity and not as a ratio*. (It reads "d over dx"). In Chapter 10, it will be seen that the symbol $\mathrm{d}/\mathrm{d}x$ is also used in a formula to stand for the phrase "*the derivative of*". Thus, the symbol $\mathrm{d}/\mathrm{d}x$ is used to *define the derivatives of combinations of functions*.[31]

9.6 DERIVATIVE OF SIMPLE ALGEBRAIC FUNCTIONS

Now, we proceed to evaluate the derivatives of some simple algebraic functions by definition.

- *The Derivative of the Power Function*: Let us find the derivative of some simple (algebraic) functions.

We begin with

Example (4): Let $y = f(x) = x$. Then, we have

$$f'(x) = \lim_{\delta x \to 0} \frac{f(x+\delta x)-f(x)}{\delta x} = \lim_{\delta x \to 0} \frac{(x+\delta x)-x}{\delta x}$$
$$= \lim_{\delta x \to 0} \frac{\delta x}{\delta x} = 1.$$

That is, the derivative of $f(x) = x$ is a constant equal to 1. We write $(\mathrm{d}/\mathrm{d}x)(\mathrm{x}) = 1$

(This is obvious since $y = x$ is a function whose *graph is a straight line with a constant slope*.)

Example (5): Let $y = f(x) = x^2$, then

$$f'(x) = \lim_{\delta x \to 0} \frac{f(x+\delta x)^2 - x^2}{\delta x} = \lim_{\delta x \to 0} \frac{2x\,\delta x + (\delta x)^2}{\delta x}$$
$$= \lim_{\delta x \to 0}(2x+\delta x)$$
$$= 2x$$

We write $(\mathrm{d}/\mathrm{d}x)(x^2) = 2x$.

[30] We can also say that the operator $\mathrm{d}/\mathrm{d}x$ constructs from $f(x)$ the difference quotient $(f(x)-f(a))/(x-a)$ and determines its limit as $x \to a$ treating it as a function of x.

[31] For example, $\mathrm{d}[f(x)g(x)]/\mathrm{d}x = f(x)(\mathrm{d}g(x)/\mathrm{d}x) + g(x)(\mathrm{d}f(x)/\mathrm{d}x)$.

Example (6): Let $y = f(x) = x^3$, then

$$f'(x) = \lim_{\delta x \to 0} \frac{f(x+\delta x)^3 - x^3}{\delta x}$$

$$= \lim_{\delta x \to 0} \frac{x^3 + 3x^2\,\delta x + 3x(\delta x)^2 + (\delta x)^3 - x^3}{\delta x}$$

$$= \lim_{\delta x \to 0} 3x^2 + 3x\,\delta x + (\delta x)^2 \quad (\because \delta x \neq 0)$$

$$= 3x^2$$

We write $(\mathrm{d}/\mathrm{d}x)(x^3) = 3x^2$.

Observe the general feature of the structure of the derivatives of the power function $y = x^n$ for $n = 1, 2, 3$.

Now, we shall prove that

$$\frac{\mathrm{d}}{\mathrm{d}x}(x^n) = n\,x^{n-1}, \ (n \in N)$$

Proof: Let $y = f(x) = x^n$

$$\therefore y + \delta y = f(x+\delta x) = (x+\delta x)^n$$

$$\therefore \frac{\delta y}{\delta x} = \frac{f(x+\delta x) - f(x)}{\delta x} = \frac{(x+\delta x)^n - x^n}{\delta x}$$

We have

$$\frac{\mathrm{d}}{\mathrm{d}x}(x^n) = \lim_{\delta x \to 0} \frac{\delta y}{\delta x} = \lim_{\delta x \to 0} \frac{(x+\delta x)^n - x^n}{\delta x}$$

Now,

$$\begin{aligned}(x+\delta x)^n &= x^n + {}^nC_1 x^{n-1}\,\delta x + {}^nC_2 x^{n-2}\,(\delta x)^2 + \cdots + (\delta x)^n \\ &= x^n + nx^{n-1}\,\delta x + \frac{n(n-1)}{2} x^{n-2}(\delta x)^2 + \cdots + (\delta x)^n\end{aligned}$$

$$\therefore \frac{\delta y}{\delta x} = nx^{n-1} + \frac{n(n-1)}{2!} x^{n-2}\,\delta x + \cdots + (\delta x)^{n-1} \quad [\because \delta x \neq 0]$$

The expression on the RHS is a sum of n terms; *the first term is independent of* δx and the others tend to zero as $\delta x \to 0$. Therefore,

$$\lim_{\delta x \to 0} \frac{\delta y}{\delta x} = nx^{n-1}$$

Thus, for every *positive integral exponent n*, the power function $y = x^n$ has the derivative $n\, x^{n-1}$. We write

$$\frac{d}{dx}(x^n) = n\, x^{n-1},\ (n \in N).^{(32)}$$

Note (13): In Chapter 15a, where the method of logarithmic differentiation is discussed, we shall show that the above formula remains valid for any (real) exponent n. Thus, we can write

$$\frac{d}{dx}(\sqrt{x}) = \frac{d}{dx}(x^{1/2}) = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}}$$

Similarly,

$$\frac{d}{dx}\left(\frac{1}{\sqrt{x}}\right) = \frac{d}{dx}(x^{-1/2}) = \frac{-1}{2x\sqrt{x}}^{(33)}$$

- *Now, Let Us Consider the Derivative of a Constant, $y = f(x) = c$*. Since, the value of the function does not change, as the independent variable x changes, we have

$$f(x+\delta x) = f(x)$$

$$f(x+\delta x) - f(x) = 0$$

$$\therefore \delta y = f(x+\delta x) - f(x) = 0$$

$$\frac{\delta y}{\delta x} = \frac{0}{\delta x} = 0$$

Consequently, $\lim\limits_{\delta x \to 0} \dfrac{\delta y}{\delta x} = \lim\limits_{\delta x \to 0}(0) = 0.$

Thus, the *derivative of a constant is equal to zero*. It is reasonable to say that *the rate of change of any constant is zero*.

Example (7): Find the derivative of $\sqrt{3x+7}$

Solution: Let $f(x) = \sqrt{3x+7}$

$$f(x+h) = \sqrt{3(x+h)+7}$$

$$f'(x) = \lim_{h \to 0} \frac{\sqrt{3(x+h)+7} - \sqrt{3x+7}}{h}$$

[32] We know that $d(x^2)/dx = 2x$, $d(x)/dx = 1$, $d(x^{-1})/dx = -x^{-2}$. Note that we have not yet encountered any function whose derivative is x^{-1}. This problem is dealt with in Part II of the book.

[33] For writing these results, we have used the formula $d(x^n)/dx = nx^{n-1}$, $n \in R$.

By rationalizing the numerator, we get

$$f'(x) = \lim_{h\to 0} \frac{\sqrt{3(x+h)+7} - \sqrt{3x+7}}{h}$$

$$= \lim_{h\to 0} \frac{\sqrt{3(x+h)+7} - \sqrt{3x+7}}{h} \frac{\sqrt{3(x+h)+7} + \sqrt{3x+7}}{\sqrt{3(x+h)+7} + \sqrt{3x+7}}$$

$$= \lim_{h\to 0} \frac{(3+3h+7)-(3x+7)}{h(\sqrt{3(x+h)+7} + \sqrt{3x+7})}$$

$$= \lim_{h\to 0} \frac{3h}{h(\sqrt{3(x+h)+7} + \sqrt{3x+7})}$$

$$= \lim_{h\to 0} \frac{3}{\sqrt{3(x+h)+7} + \sqrt{3x+7}}$$

$$= \frac{3}{\sqrt{3(x+0)+7} + \sqrt{3x+7}}$$

$$f'(x) = \frac{3}{2\sqrt{3x+7}} \qquad \textbf{Ans}.$$

Example (8): Find the derivative of $1/\sqrt{x}$.

Solution: Let $f(x) = \dfrac{1}{\sqrt{x}}$, $\therefore f(x+h) = \dfrac{1}{\sqrt{x+h}}$

Now consider

$$f(x+h) - f(x) = \frac{1}{\sqrt{x+h}} - \frac{1}{\sqrt{x}} = \frac{\sqrt{x} - \sqrt{x+h}}{\sqrt{x}\sqrt{x+h}}$$

By rationalizing the numerator, we get

$$f(x+h) - f(x) = \frac{\sqrt{x} - \sqrt{x+h}}{\sqrt{x+h}\sqrt{x}} \frac{\sqrt{x} + \sqrt{x+h}}{\sqrt{x} + \sqrt{x+h}}$$

$$= \frac{x-(x+h)}{\sqrt{x+h}\sqrt{x}} \frac{1}{\sqrt{x} + \sqrt{x+h}}$$

$$= \frac{-h}{\sqrt{x+h}\sqrt{x}(\sqrt{x} + \sqrt{x+h})}$$

Now,

$$
\begin{aligned}
f'(x) &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \\
&= \lim_{h \to 0} \frac{-h}{\left[h\sqrt{x+h}\sqrt{x}(\sqrt{x} + \sqrt{x+h})\right]} \\
&= \frac{-1}{(\sqrt{x}\sqrt{x})(\sqrt{x} + \sqrt{x})} \quad [\because h \to 0,\ h \neq 0] \\
&= \frac{-1}{x2\sqrt{x}} \\
&= \frac{-1}{2x\sqrt{x}}
\end{aligned}
$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{\sqrt{x}}\right) = \frac{\mathrm{d}}{\mathrm{d}x}(x)^{-1/2} = \frac{-1}{2x\sqrt{x}} \quad \textbf{Ans.}$$

9.7 DERIVATIVES OF TRIGONOMETRIC FUNCTIONS

To find the derivatives of trigonometric functions by the first principle (i.e., by definition) we have to use the following standard limits:

(i) $\lim_{x \to 0} \frac{\sin x}{x} = 1$ where x is an angle expressed in radians.

(ii) $\lim_{x \to 0} \frac{\cos x - 1}{x} = 0$ [or equivalently x stands for a real variable].

Then, the following results can be proved:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}(\sin x) &= \cos x \\
\frac{\mathrm{d}}{\mathrm{d}x}(\cos x) &= -\sin x \\
\frac{\mathrm{d}}{\mathrm{d}x}(\tan x) &= \sec^2 x \\
\frac{\mathrm{d}}{\mathrm{d}x}(\cot x) &= -\operatorname{cosec}^2 x \\
\frac{\mathrm{d}}{\mathrm{d}x}(\sec x) &= \sec x \cdot \tan x \\
\frac{\mathrm{d}}{\mathrm{d}x}(\operatorname{cosec} x) &= -\operatorname{cosec} x \cdot \cot x
\end{aligned}
$$

(All necessary details about the proof of limits (i) and (ii) are available in Chapter 11.)

9.8 DERIVATIVES OF EXPONENTIAL AND LOGARITHMIC FUNCTIONS

To find the derivatives of exponential and logarithmic functions, by the first principle, we have to use the following limits:

(i) $\lim_{x \to 0} (1+x)^{1/x} = \text{e}.$

(ii) $\lim_{x \to 0} \dfrac{a^x - 1}{x} = \log_e a$, where $a > 0$.

Then, the following results can be proved:

$$\frac{\text{d}}{\text{d}x}(a^x) = a^x \log_e a,\ a > 0,\ (a \neq 0)$$

$$\frac{\text{d}}{\text{d}x}(e^x) = e^x \log_e \text{e} = \text{e}^x$$

$$\frac{\text{d}}{\text{d}x}(\log_e x) = \frac{1}{x},\ x > 0$$

$$\frac{\text{d}}{\text{d}x}(\log_a x) = \frac{1}{x}\frac{1}{\log_e a},\ (x > 0,\ a > 0,\ a \neq 1)$$

(All necessary details about the proof of limits (i) and (ii) are available in Chapter 13.)

Note (14): So far, we have seen the evaluation of derivative(s) of some simple functions by the first principle (i.e., by definition). The direct evaluation of the limit

$$\lim_{\delta x \to 0} \frac{f(x+\delta x) - f(x)}{\delta x}\text{(34)}$$

is most often connected with *lengthy and complicated calculations*. But, it turns out that for basic elementary functions (i.e., basic trigonometric, exponential, and logarithmic functions), it is possible to derive *general formulas expressing their derivatives analytically,* as in the case of the power function $y = x^n$.

Furthermore, the rules for *differentiating combinations of functions* resulting from arithmetical operations (i.e., sums, products, and quotients) and the rules for computing *derivatives of composite functions are readily established, in terms of the derivatives of constituent functions.*

Accordingly, we can always find analytically the derivative of *any combination of finite number of basic elementary functions*, without resorting to the computation of the limit indicated above. (The rules for differentiating combinations of functions are discussed in Chapter 10).

9.9 DIFFERENTIABILITY AND CONTINUITY

There is an important relationship between *differentiability* of a function and *continuity* of that function, as stated in the following theorem.

Theorem: If a function *f* is *differentiable at* x_1, then *f* is *continuous at* x_1.

(34) Evaluation of this limit means applying the operator $\text{d}/\text{d}x$ to the function $f(x)$.

Proof: Suppose f is differentiable at x_1.

$$\therefore \lim_{h\to 0}(f(x_1+h)-f(x_1))/h = f'(x_1) \text{ exists.}$$

Now, consider

$$\lim_{h\to 0}[f(x_1+h)-f(x_1)] = \lim_{h\to 0}\left[h\frac{f(x_1+h)-f(x_1)}{h}\right]$$

$$= \lim_{h\to 0}(h)\lim_{h\to 0}\left[\frac{f(x_1+h)-f(x_1)}{h}\right]$$

$$= 0f'(x_1) = 0$$

$$\therefore \lim_{h\to 0}[f(x_1+h)-f(x_1)] = 0$$

$$\therefore \lim_{h\to 0}f(x_1+h) = \lim_{h\to 0}f(x_1)$$

It means that f is continuous at x. This theorem tells us that if a function f is given (or proved) to be differentiable at $x = x_1$, then it is definitely continuous at $x = x_1$. It also tells us that a function cannot have a derivative at points of discontinuity.

Remark: *From the fact that at some point $x = x_1$ the function $y = f(x)$ is continuous, it does not follow that it is differentiable at that point.* In other words, *if a function is continuous at a point, it is not necessarily differentiable at that point,* as must be clear from the following examples.

Example (9): Let $f(x) = |x|$. It is easy to show that this function is continuous at all points, in particular continuous at 0 (see Figure 9.11a). We can show that it is not differentiable at 0. To find the derivative at $x = 0$, consider the difference quotient:

$$\frac{f(0+h)-f(0)}{h} = \frac{|0+h|-|0|}{h} = \frac{|h|}{h}$$

Let us consider the limit of the above difference quotient as $h\to 0$.
If $h\to 0$ *from the right, then the limit of this ratio is* $+1$ and if $h\to 0$ *from the left, then the limit is* -1. Since $\lim_{h\to 0^+}|h|/h \neq \lim_{h\to 0^-}|h|/h$, it follows that the two-sided limit $\lim_{h\to 0}|h|/h$ does not exist. In other words, $|x|$ *is not differentiable at 0.*

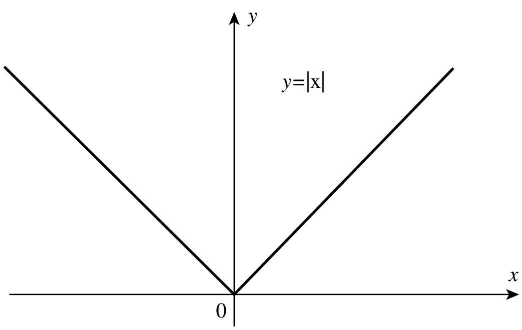


FIGURE 9.11 a

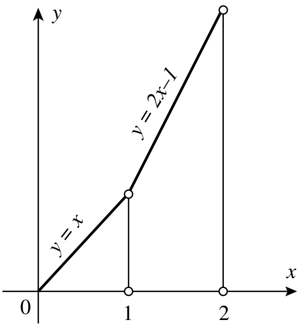


FIGURE 9.11 b

Example (10): A function *f*(*x*) is defined on an interval [0,2] as follows (see Figure 9.11b):

$$f(x) = x, \text{ when } 0 \leq x \leq 1$$
$$f(x) = 2x - 1, \text{ when } 1 < x \leq 2$$

At $x = 1$, *the function has no derivative,* although *it is continuous at this* point as shown below. Consider a *nonzero variable h*. (Note that h stands for an increment that can be either positive or negative, but not zero.) Thus, at $x = 0$, when $h > 0$, we have

$$\lim_{h \to 0} \frac{f(1+h) - f(1)}{h} = \lim_{h \to 0} \frac{[2(1+h) - 1] - [2(1) - 1]}{h}$$
$$= \lim_{h \to 0} \frac{(1+2h) - 1}{h} = \lim_{\delta h \to 0} \frac{2h}{h} = 2, \ [\text{since} f(x) = 2x - 1]$$

Again when $h < 0$, we get

$$\lim_{h \to 0} \frac{f(1+h) - f(1)}{h} = \lim_{h \to 0} \frac{(1+h) - 1}{h} \lim_{\delta h \to 0} \frac{h}{h} = 1.$$

(The definition of a derivative requires that the ratio $\delta y / \delta x$ (as $\delta x \to 0$) should approach one and the same limit regardless of the way in which $\delta x \to 0$.)

Since, the above *limit depends on the sign of* (the increment) *h,* it follows that the function has no derivative at the point $x = 1$. Geometrically, this is in accordance with the fact that at $x = 1$, *the "curve" does not have a definite tangent line.*

Note (15): In Example (9), there is a *sharp corner* at $x = 0$ and in Example (10), such a corner exists at $x = 1$. At such points, the graph is continuous, but *there are possible two tangent lines with different slopes. In other words,*

$$\lim_{\delta x \to 0^-} \frac{\delta y}{\delta x} \neq \lim_{\delta x \to 0^+} \frac{\delta y}{\delta x}$$

It is easy to show that *this function is continuous at* $x = 1$.

We have $f(x) = x$ for $0 \leq x \leq 1$, and $f(x) = 2x - 1$ when $1 < x \leq 2$

$$\therefore f(1) = 1 \tag{23}$$

$$\text{Next,} \quad \lim_{x \to 1^-} f(x) = \lim_{x \to 1^-} x = 1 \tag{24}$$

$$\lim_{x \to 1^+} f(x) = \lim_{x \to 1^+} (2x - 1) = 2 - 1 = 1 \tag{25}$$

From (24) and (25), we have $\lim_{x \to 1^-} f(x) = \lim_{x \to 1^+} f(x)$

Also, we have $f(1) = 1$ [from (23)]

$$\therefore \lim_{x \to 1} f(x) = f(1)$$

$$\therefore f \text{ is continuous at } x = 1.$$

Example (11): The function $y = \sqrt[3]{x}$ *is not differentiable at 0,* though it is continuous for all values of x.

Let us find out whether this function has a derivative at $x = 0$.

We have, at $x = 0$,

$$f'(x) = \lim_{x \to 0} \frac{f(x) - f(0)}{x - 0}$$

$$= \lim_{x \to 0} \frac{x^{1/3} - 0}{x}$$

$$= \lim_{x \to 0} \frac{1}{x^{2/3}}$$

But this limit does not exist. Thus, f is not differentiable at 0. However, f is continuous at 0, because $\lim_{x \to 0} f(x) = \lim_{x \to 0} x^{1/3} = 0 = f(0)$.

Note (16): A function f *can fail to be differentiable at a number* x_1 for one of the following reasons:

- Function f *is not continuous at* x_1.
- Function f is continuous at x_1, but *the graph of f does not have a (unique) tangent line at point* $x = x_1$. Figure 9.11a and b shows the graph of functions satisfying this condition. Observe a "sharp turn" (or corner) in these graphs (see Figure 9.11a at $x = 0$ and Figure 9.11b at $x = 1$).
- Function f is continuous at x_1, but *the graph of f has a vertical tangent line at the point* $x = x_1$. Remember that the slope of a vertical line is not defined. This situation occurs in Example (11) (see Figure 9.12). In such cases, we say (for generality) that the function has *an infinite derivative*, which also means that the function is not *differentiable* at the point in question.

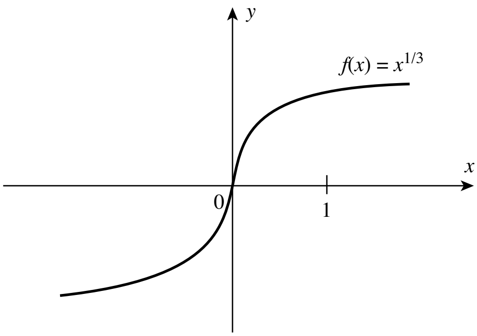


FIGURE 9.12

Note (17): Before giving another example of a function that is continuous but not differentiable at a point of the domain, it is useful to recall the fact that the values of the trigonometric functions sin x and cos x lie between -1 and 1, for all x. Obviously, this is also true for the functions $\sin(1/x), x \neq 0$, and $\cos(1/x), x \neq 0$. In other words, $\lim_{x\to 0} \sin(1/x)$ and $\lim_{x\to 0} \cos(1/x)$ oscillate between -1 and 1, which means that these limits do not exist.
Now, we give an example of a function (involving trigonometric functions) that is continuous at $x = 0$, but not differentiable.

Example (12): Prove that the function defined as follows is continuous at $x = 0$, but not differentiable at $x = 0$.

$$f(x) = \begin{cases} x\cos\frac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Solution:
$$\begin{aligned}\lim_{x\to 0^-} f(x) &= \lim_{x\to 0^-} x\cos\left(\frac{1}{x}\right) \\ &= \lim_{h\to 0}(0-h)\cos\left(\frac{1}{0-h}\right) \\ &= \lim_{h\to 0}(-h)\cos\left(-\frac{1}{h}\right) \\ &= \lim_{h\to 0}(-h)\cos\left(\frac{1}{h}\right) \quad [\because \cos(-\theta) = \cos\theta] \\ &= 0 \times \text{a finite quantity} \left[\because \cos\frac{1}{h} \text{ lies between } -1 \text{ and } 1\right] \\ &= 0 \qquad \text{(i)}\end{aligned}$$

Similarly, $$\lim_{x\to 0^+} f(x) = \lim_{h\to 0} h\cos\frac{1}{h} = 0 \qquad \text{(ii)}$$

Also, $f(0) = 0$, (given) (iii)

In view of (i), (ii), and (iii), we conclude that $f(x)$ is continuous at $x = 0$. Now to find the derivative of $f(x)$, at $x = 0$, we compute the following limit:

$$\lim_{x\to 0}\frac{f(x)-f(0)}{x-0} = \lim_{x\to 0}\frac{f(x)}{x} = \lim_{x\to 0}\frac{x\cos(1/x)}{x}$$

$$= \lim_{x\to 0}\cos\frac{1}{x}, \quad [\because x \neq 0]$$

But, this limit does not exist.
$\therefore f'(0)$ does not exist. That is, $f(x)$ is not differentiable at $x = 0$.

Remark: It can be shown that the function

$$f(x) = \begin{cases} x^2\cos\dfrac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is *continuous and differentiable at* $x = 0$, and obviously at all other points.
Consider the limit (of difference quotient)

$$\lim_{x\to 0}\frac{g(x)-g(0)}{x-0}$$

$$= \lim_{x\to 0}\frac{x^2\cos(1/x)-0}{x-0}$$

$$= \lim_{x\to 0}\frac{x^2\cos(1/x)}{x}$$

$$= \lim_{x\to 0} x\cos\frac{1}{x}, \ (\because x \neq 0)$$

$$= \left(\lim_{x\to 0} x\right)\left(\lim_{x\to 0}\cos\frac{1}{x}\right)$$

$$= (0)(\text{a finite quantity}) = 0$$

Thus, the limit of the difference quotient exists at $x = 0$. This proves that the function $g(x)$ is differentiable at $x = 0$, which also tells us that *$g(x)$ is continuous at $x = 0$*. (The continuity, at $x = 0$, can also be proved independently.) Note that in the definition of this function, *the component x^2* (in $x^2\cos(1/x)$) *plays an important role*. (This also suggests that we can define any number of such differentiable functions.)

Example (13): Prove that *the greatest integer* function $y = [x]$ is not differentiable at $x = 1$.

Solution: Here, we will show that the function $f(x) = [x]$ is not continuous at $x = 1$.

$$\text{Consider } \lim_{x \to 1^-} f(x) = \lim_{x \to 1^-} [x]$$

$$= \lim_{h \to 0} [1 - h], \ (h > 0)$$

$$= 0 \quad \text{(by definition of } [x])$$

$$\text{Again } \lim_{x \to 1^+} f(x) = \lim_{x \to 1^+} [x]$$

$$= \lim_{h \to 0} [1 + h], \ (h > 0)$$

$$= 1 \quad \text{(by definition of } [x])$$

Thus left-hand limit $\neq$ Right-hand limit
$\therefore \lim_{x \to 1} [x]$ does not exist.
$\Rightarrow$ the function $[x]$ *is not continuous at* $x = 1$.
$\Rightarrow$ $[x]$ *is not differentiable at* $x = 1$.

Note (18): The greatest integer function $y = [x]$ is a step function.
Recall that $[x]$ = the greatest integer less than or equal to x.
Selected values of $y = [x]$
Positive values: $[3] = 3$, $[3.1] = 3$, but $[2.9] = 2$
The value 0: $[0] = 0$, but $[0.5] = 0$
Negative values: $[-2] = -2$, $[-1.8] = -2$, but $[-2.1] = -3$
Note that if x is negative, $[x]$ may have a *larger absolute value* than x does.
Furthermore, observe that

$$[5] = 5, \ \lim_{x \to 5^+} [x] = 5, \quad (\because [5.01] = 5)$$

$$\text{but } \lim_{x \to 5^-} [x] = 4, \quad (\because [4.99] = 4)$$

$$\text{Next, } [-4] = -4, \ \lim_{x \to -4^+} [x] = -4, \quad (\because [-3.81] = -4)$$

$$\text{but } \lim_{x \to 4^-} [x] = -5, \quad (\because [-4.12] = -5)$$

From the above, we observe that the function $y = [x]$ is *right continuous for each integral value of* x. Furthermore, $[x]$ is not differentiable for any integral value of x.

9.10 PHYSICAL MEANING OF DERIVATIVES

We know that

$$\text{If } f(x) = 5x + C, \text{ then } f'(x) = 5 \tag{26}$$

$$\text{If } g(x) = x^3, \text{ then } g'(x) = 3x^2 \tag{27}$$

The relation (26) tells us that for function "*f*" the (actual) rate of change is 5 (a constant), which means that for any (small or big) increase (or decrease) in x, the value $f(x)$ must increase (or decrease) five times the change in x, anywhere in the domain of $f(x)$. In other words, the rate of change of $f(x)$ (being constant) does not depend on the value of x.

The relation (27) tells that, for the function "g" *the (actual) rate of change of* $g(x)$ *is* $3x^2$, *which depends on* x. It means that at $x = 1$, *any small change in* x causes *nearly* 3 times the change in the value $g(x)$, and at $x = 2$, any small change in x causes nearly 12 times the change in the value $g(x)$ (since for $x = 2$, $3x^2 = 12$).

In other words, the derivative of a function gives an idea about the variation in y-coordinate of a point on the graph (for a very small variation in the x-coordinate of the point).[35]

9.11 SOME INTERESTING OBSERVATIONS [36]

(a) We know that area A of any circle is given by the formula

$$A(r) = \pi r^2 \tag{28}$$

where r is radius of the circle. Here A is a function of r. *Let us find the instantaneous rate of change of A, with respect to r.*

(We may carry out the process as done in steps (12)–(18) for function $y = 16x^2$. There is, however, no need to repeat all the details. The two functions are practically alike, the only difference being that π occurs in formula (28), whereas the number 16 occurs in $y = 16x^2$. Of course, δr in the present case is an increase in the length of the radius (see Figure 9.13) and δA is the corresponding increase in the area that results due to the increase δr in r.)

We get

$$\frac{dA}{dr} = 2\pi r \tag{29}$$

The result (29) is of interest since it tells us that *the rate of change of area of a circle with respect to radius is the circumference of the circle*. This result is intuitively clear, for as the radius increases, one might say that successive circumferences are added to the area.

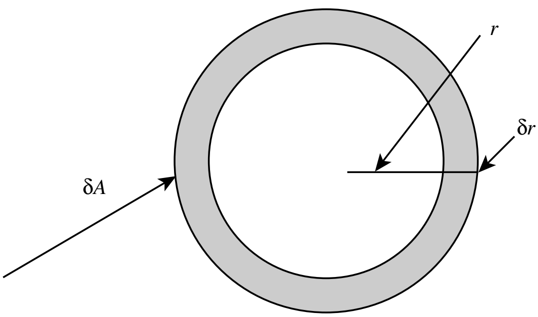


FIGURE 9.13

[35] We have mentioned about the derivative as a rate measurer in an earlier footnote.

[36] The following examples are selected from the book *What Calculus Is About* by W.W. Sawyer, Random House.

(b) With the preceding result at hand, one might guess that *the rate of change of volume of a sphere with respect to the radius could be the area of the surface of the sphere*. In fact, this is true.

The volume of a sphere is given by

$$v(r) = \frac{4}{3}\pi r^3$$

Here v is a function of r. We get

$$\begin{aligned}\frac{dv}{dr} &= \frac{4}{3}3\pi r^2 \\ &= 4\pi r^2\end{aligned}$$

which is the formula for the area of the surface of a sphere.

This can hardly be a coincidence. In fact, *it is easy to see why it occurs*. Suppose we have a sphere and we want to make it a little larger. We might spray an even coating of paint all over its surface, thus giving it an extra skin. It is not at all surprising that *the amount by which the volume has increased during this operation should be closely related to the area of the surface,* on which the skin has been placed (or paint has been sprayed).

In this argument, it is absolutely essential that *the coating should be even* (i.e., the skin must have the same thickness everywhere). In effect, we are estimating the increase δv in the volume by multiplying *the surface area to the thickness* δr *of the skin. This estimate is reasonable only if the coating is thin.*

(c) The idea that objects grow by forming an extra skin can also be illustrated without using circles and spheres. Imagine a cube placed in the corner of a room as shown in Figure 9.14. This cube will grow if someone continually sprays paint onto the exposed faces of the cube.

This is done in such a way that the points A, B, and C move outward at some constant speed (say, 1 mm/s). *At any time, let the side of the cube be x mm. Its volume will be* $v = x^3$. This, we know, grows at the rate $(dv/dr) = 3x^2$. *The picture shows why* $3x^2$ *should come into focus. Observe that the exposed surface consists of three squares.*

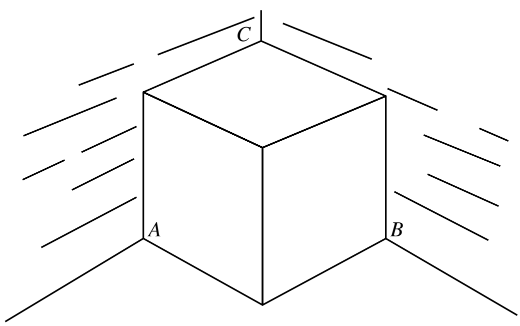


FIGURE 9.14

9.12 HISTORICAL NOTES

Who invented derivatives?

No one person invented the derivative. Related ideas and methods appear throughout mathematical history, spanning at least 2000 years. The modern European development happened largely in the seventeenth and eighteenth centuries as follows:

Pierre de Fermat (1601–1665, French)
Isaac Newton (1642–1727, English)
Gottfried Leibniz (1646 – 1716, German)
Leonard Euler (1707–1783, Swiss)
Joseph-Louis Lagrange (1736–1813, French)

Isaac Newton and Gottfried Leibniz are generally considered the cofounders of modern calculus. Building on Euler's idea of a function, Lagrange may have been first to use the phrase *"derivative function*", and the *prime symbol*, to denote it.

Because of the difficulties in the very foundation of calculus, conflicts and doubts on the soundness of the entire subject were prolonged. Among many contemporaries of Newton were the following:

- Michel Rolle (who then contributed a famous theorem) taught that calculus was a collection of ingenious fallacies.
- Colin Maclaurin (after whom another famous theorem was named) decided that he would give a proper foundation to calculus and published a book on the subject in 1742. The book was undoubtedly profound but also unintelligible.
- One hundred years after Newton and Leibniz, Joseph Louis Lagrange, one of the greatest mathematicians of all times, still believed that calculus was unsound and gave correct results only because errors were offsetting each other. He too formulated his own foundation for calculus, but it was incorrect.
- Near the end of eighteenth century, d'Alembert had to advise students of calculus to keep on with their study: faith would eventually come to them.
- Some of the strongest criticisms came from religious leaders. Of these, the most famous is the highly original philosopher Bishop George Berkely.

  Since the fundamental concept of calculus was not clearly understood and therefore, not well presented by either Newton or Leibniz, Berkely was able to enter the fray with justification and conviction.

  In "The Analyst" (1734), addressed to an infidel mathematician, he condemned instantaneous rates of change of functions as "neither finite quantities nor quantities infinitely small, nor yet nothing". These rates of change were but "the ghosts of the departed quantities".

  To account for the fact that calculus gave correct results, Berkely, like Lagrange, argued that somewhere errors were compensating for each other.

The problem of calculating instantaneous rates, of which speed and acceleration were the most pressing, attracted almost all the mathematicians of the seventeenth century, and the roster of those who contributed to the subject and achieved limited success is extensive. Newton and Leibniz took decisive steps in applying their ideas, which involved both intuition and imagination. They had an idea that made physical sense, and, since mathematics and physical science were closely intertwined, they were not greatly concerned about the lack of mathematical rigor. One might say that in their minds the end justified the means.

- Both Newton and Leibniz had a good "intuitive understanding" of the idea involved. They applied it in a way we may call today, "*the process of computing the limit of a function*". However, in concluding the result(s), the explanations given by them were inconsistent and many a time contradictory. The concept of limit, as we know it today, was not known to either Newton or Leibniz. Since they were not very clear about the idea (of a limit) they applied, they could not define it. Also, nobody before them defined the concept. *The "merely intuitive quality of the idea" (of a limit) hampered progress in the development of calculus, for a century, after Newton and Leibniz.*
- In 1754, the French mathematician d'Alembert (1717–1783) suggested that the logical basis of calculus would reside in the concept of limit. It was French mathematician Augustin-Louis Cauchy (1789–1857), who gave the definition of limit that removed doubts as to the soundness of the subject and made it free from all the confusion. *With the availability of systematic and refined material on the concept of limit, the reader today, can easily grasp the concept*. However, if the reader still finds some difficulty in grasping it, he may be less discouraged when it is told that the concept of limit eluded even Newton and Leibniz.

The history of controversy surrounding calculus is most illuminating. The soundness of calculus was doubted by the greatest mathematicians of the eighteenth century (as mentioned above), yet it was not only applied freely but some of the greatest developments in mathematics-differential equations, the calculus of variation, differential geometry, potential theory- and a host of other subjects comprising what is now called analysis were developed and explored by means of calculus.

Calculus might have been lost to us forever had the mathematicians of that age been too concerned with rigor. We know now that even in mathematics, intuition and physical thinking produce big ideas and that logical perfection must come afterward. We also see more clearly today that the pursuit of absolute rigor in mathematics is an unending endeavor, calling for patience. The understanding and mastery of nature must be sought, with the best tools available.

(Most of these notes are taken from *Mathematics and the Physical World*, by Morris Kline, and *Calculus with Analytic Geometry* (Alternate Edition), by Robert Ellis and Denny Gulick, HBJ Publishers.)

10 Algebra of Derivatives: Rules for Computing Derivatives of Various Combinations of Differentiable Functions

10.1 INTRODUCTION

In Chapter 7a and b, we have studied the concept of limit of a function and used the notion (of limit) as a tool to define the derivative of a function. We know that while the notion of limit is a general notion for functions, *the derivative of a function f(x) is defined by the limit*

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

If this limit exists, we denote it by $f'(x)$ and call it the derivative of f. In other words, the derivative of a function is a *limit of a particular kind*.[1]

Since there are limit rules for sums, differences, products, and quotients of functions, it is natural to ask whether there are corresponding rules for derivatives. Of course, there are rules for computing the derivatives of such combinations of functions, *but some of these rules (or formulas) are quite different from their counterparts for limits.* Also, there are rules governing the derivative(s) of *composite functions* and those of *inverse functions*. All these rules constitute the *algebra of derivatives*. We will also see how these rules are used in applications and in the further development of calculus itself.

The necessity of such rules can be shown by means of the following example. Suppose we have to find the derivative of the function defined by

$$f(x) = x^5 - 4x^3 + 7x + 8$$

This is a simple combination of algebraic functions, but still a complicated formula defining a function. To find its derivative by applying the definition (i.e., by forming its difference quotient

10-Algebra of derivatives (Derivatives of combinations of functions)

[1] It must be clear that the difference quotient $\lim_{h \to 0} f(x+h) - f(x)/h$ is constructed from the given function $f(x)$ and looked upon as a new function of "h". If the limit of this difference quotient as $h \to 0$ exists, only then is $f(x)$ said to be a differentiable function, the limit being denoted by $f'(x)$.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

and evaluating its limit) will be naturally *time consuming and tedious*. But, *if we know how to find the derivative of a combination of functions from the derivatives of the individual functions, then obtaining the derivative $f'(x)$ would be much simpler*.

For example, if $f(x)$ and $g(x)$ are *differentiable functions* of x, then the following results can be proved:

$$\frac{d}{dx}[kf(x)] = k\frac{d}{dx}f(x), \quad (k \in R) \tag{1}$$

$$\frac{d}{dx}[f(x)g(x)] = f(x)\frac{d}{dx}g(x) + g(x)\frac{d}{dx}f(x) \tag{2}$$

These relations define *the rules* (or *formulas*) expressing the *derivative(s) of certain combinations of functions in terms of the derivatives of individual functions*. The functions $f(x)$, $g(x)$, $h(x)$, and so on may be *basic elementary functions* (like x^n, $\sin x$, e^x, $\log_e x$, etc.) or *their (simple) combinations* (like $k \sin x$, $\cos x^2$, e^{3x}, $\log_a(x + 5)$, etc.) that are called *elementary functions*. We distinguish between the terms: *basic elementary functions* and *the elementary functions*.

10.1.1 Definition (A)

Basic elementary functions are the following analytically represented functions:

(i) *Power Function*: $y = x^\alpha$, $\alpha \in R$, $x > 0$[(2)]

(ii) *General Exponential Function*: $y = a^x$, $(a > 0,\ a \neq 1,\ x \in R)$

(iii) *Logarithmic Function*: $y = \log_a x$, $(a > 0,\ a \neq 1,\ x > 0)$

(iv) *Trigonometric Functions*: $y = \sin x$, $y = \cos x$, $y = \tan x$,

$$y = \cot x,\ y = \sec x,\ y = \operatorname{cosec} x.$$

(v) *Inverse Trigonometric Functions*: $y = \sin^{-1} x$, $y = \cos^{-1} x$, $y = \tan^{-1} x$,

$$y = \cot^{-1} x,\ y = \sec^{-1} x,\ y = \operatorname{cosec}^{-1} x.$$

Observe that *certain basic elementary functions are combinations of other basic elementary functions*, for example, $\tan x = \sin x/\cos x$.

10.1.2 Definition (B)

An elementary function is a function that may be represented by a *single formula of the type*

$$y = f(x)$$

where the expression on the right-hand side is made up of *basic elementary functions and constants*, by means of a *finite number of operations of addition, subtraction, multiplication, division*, and *taking the function of a function*.[(3)]

(2) If a is irrational, this function is evaluated by taking logarithms and antilogarithms. Thus, we can write $\log y = \alpha \log x$, which is defined for $x > 0$.

(3) For more details, see *Differential and Integral Calculus* by N. Piskunov (vol. I, pp. 20–24), Mir Publishers, Moscow.

Examples of elementary functions:

$$y = |x| = \sqrt{x^2},\ y = \sqrt{1 + 4\sin^2 x}$$

$$y = \frac{\log x + 2\tan x + 4\sqrt[3]{x}}{10^x - x + 10}$$

10.1.3 An Example of a Nonelementary Function

The function $y = 1.2.3.4.5\ldots n$ $[y = n! = f(n)]$ is *not elementary* because the number of operations that must be performed to obtain y increases with n. In other words, *the number of operations is not bounded.*

10.2 RECALLING THE OPERATOR OF DIFFERENTIATION

We introduced the symbol $\mathrm{d}/\mathrm{d}x$ in Chapter 9 and named it *the operator of differentiation*. Recall that when it is applied to a differentiable function $y = f(x)$, *it carries out the entire operation of computing the derivative* of $f(x)$, in the following two steps:

(a) From the function $f(x)$, it constructs a new function (called the difference quotient)

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{3}$$

where h is a *nonzero variable number*, and

(b) Treating the difference quotient as a function of the variable h, it determines the limit

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{4}$$

If this limit exists, we call it *the derivative of the function* $y = f(x)$, and denote it by the symbol $(\mathrm{d}y/\mathrm{d}x)$[or $(\mathrm{d}/\mathrm{d}x)(y)$]or $f'(x)$ or y'.

10.2.1 Operator of Differentiation

Here a very important question arises.

Why should we introduce the operator of differentiation $\mathrm{d}/\mathrm{d}x$, if $\mathrm{d}f/\mathrm{d}x$ *represents nothing else but the limit transition operation described at* (*4*)?

To answer this, note that the operator $\mathrm{d}/\mathrm{d}x$ stands for *the entire process of computing the derivative of a function*. (This includes the method(s) required for evaluating the limit of the difference quotient.) Accordingly, for certain functions *(which are complicated combinations of functions)*, the process of differentiation by applying the operator will be obviously quite tedious.

Now, suppose *it is proved separately* that for a given function $f(x)$, $\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f'(x)$ exists, then *we can always write* (the notation)

$$\frac{\mathrm{d}}{\mathrm{d}x}[f(x)] = f'(x) \tag{5}$$

without evaluating the limit of the difference quotient of f(x). Thus, we can use the symbol $\mathrm{d}/\mathrm{d}x$ in two useful ways.

First, as an operator, when it defines the entire process of computing the derivative of a given function, and *second*, it can be used to stand for the phrase, *the derivative of f(x)* as indicated at (5) above.

(Note that in the formulas at (1) and (2) above, we have used the result (5), in expressing the derivatives of certain combinations of functions without forming the difference quotient of the given functions and evaluating their limits.)

To continue the discussion smoothly, we give below *one more formula* that expresses the derivative of the ratio of differentiable functions:

Let $f(x)$ and $g(x)$ be *differentiable functions*. Then the derivative of the ratio

$$u(x) = f(x)/g(x) \text{ is given by}$$

$$\mathrm{d}/\mathrm{d}x[u(x)] = \mathrm{d}/\mathrm{d}x\left[\frac{f(x)}{g(x)}\right] = \frac{g(x)(\frac{\mathrm{d}}{\mathrm{d}x})f(x) - f(x)(\frac{\mathrm{d}}{\mathrm{d}x})g(x)}{[g(x)]^2}, \quad g(x) \neq 0 \qquad (6)^{(4)}$$

If the derivatives of $f(x)$ and $g(x)$ are known, then we can easily write down the derivative of the quotient $u(x) = f(x)/g(x)$ using the formula (6), avoiding the direct use of limits. (Note that, here, the symbol $\mathrm{d}/\mathrm{d}x$ *does not demand the formation of difference quotient and evaluation of its limit* since it does not act as an operator of differentiation.) *This simplifies the procedure for evaluating the derivative* of the given combination of functions.

But to use such rules, *we must know the derivatives of the basic elementary functions, appearing in the formulas*. It is therefore necessary *to compute the derivatives of basic elementary functions by some method* and prepare a table for using them in the formulas to be established.

One way is *to obtain the derivatives of basic elementary functions by the first principle*. It is a good exercise but time consuming. There is a simpler way. *We can obtain the derivatives of some selected basic elementary functions by the first principle and then by using these derivatives in the formulas*, we can obtain the derivatives of other basic elementary functions, by using properties of the functions, as will be clear from the following example.

Example (1): We know that $\mathrm{d}/\mathrm{d}x\ (\sin x) = \cos x$. Now, by using the relation $\cos x = \sin((\pi/2) - x)$ and the chain rule for differentiation (to be studied shortly), we can compute $(\mathrm{d}/\mathrm{d}x)(\cos x)$. Furthermore, we know that $\tan x = \sin x/\cos x$, $(\cos x \neq 0)$. Hence, by applying the formula (6), *we can obtain the derivative of tan x, using the derivatives of sin x and cos x*.[5]

10.2.2 Rules of Differentiation of Functions

The above discussion suggests that *our first step should be to establish the rules for differentiation of functions*. For this purpose, first, we find the result of applying the operator $\mathrm{d}/\mathrm{d}x$ to certain combinations of differentiable functions, namely, sums, products, and ratios. (It turns out that the rules for differentiating such combinations of functions are easily *established in terms of the derivatives of the constituent functions*.).

[4] The proof of this formula is discussed later under Rule (4).

[5] This gives an idea about the applications of the rules of differentiation. Later on, in the process of computing derivatives of implicit functions and parametric functions, it will be noted how the rules of differentiation contribute to the further development of calculus.

Second, we find out the result of applying the operator $\mathrm{d}/\mathrm{d}x$ to some *selected basic elementary functions,* namely, functions like $y = x^n$, $y = \sin x$, $y = a^x$, $y = \log_a x$. It is found that the derivatives of these functions are easily computed by applying the operator $\mathrm{d}/\mathrm{d}x$ and using the properties of the functions. It will be seen that using the formulas (or rules) for differentiation and the derivatives of these functions, we can obtain the derivatives of many other basic elementary functions. (We can then prepare a table of these basic elementary functions with their derivatives.)

After these two steps are completed, we may practically forget about the relations of the type (4). *In order to differentiate a function*, it is sufficient to express the given function (via basic elementary functions) and apply the rules of differentiation. Using the differentiation rules and *the table of derivatives for the basic elementary functions*, we are in a position to forget about the relations of the type (4) and compute the *derivatives of elementary functions using the language of the relations of type (5).*

10.2.3 Formal Differentiation

By using the rules of differentiation, we can compute the derivatives (of functions) *without applying the operator* $\mathrm{d}/\mathrm{d}x$ (i.e., without applying the definition of derivatives). Hence, *this method of obtaining derivative(s) is called formal differentiation*. Note that, in obtaining the derivatives of functions *by applying the formal rules of differentiation, the definition of derivatives is indirectly used.*[6]

In a *formal course of differential calculus*, the approach could be to skip the relations of the type (4). These rules allow us to compute the derivatives of most complicated combinations of functions, almost instantly, avoiding evaluation of limit(s). For the time being, we shall accept the following standard results. These results are established later, in different chapters, as indicated below:

10.2.4 Derivatives of Some Basic Elementary Function

1. $\dfrac{d}{dx}(c) = 0, \quad (c \text{ being constant})$
2. $\dfrac{d}{dx}(x^n) = nx^{n-1}, \quad (n \in N)$[7]
3. $\dfrac{d}{dx}(\sin x) = \cos x$
4. $\dfrac{d}{dx}(\cos x) = -\sin x$
5. $\dfrac{d}{dx}(\tan x) = \sec^2 x$
6. $\dfrac{d}{dx}(\cot x) = -\mathrm{cosec}^2 x$
7. $\dfrac{d}{dx}(\sec x) = \sec x \cdot \tan x$
8. $\dfrac{d}{dx}(\mathrm{cosec}\ x) = -\mathrm{cosec}\ x \cdot \cot x$
9. $\dfrac{d}{dx}(e^x) = e^x$
10. $\dfrac{d}{dx}(a^x) = a^x \cdot \log_e a, \quad (a > 0)$
11. $\dfrac{d}{dx}(\log_e x) = \dfrac{1}{x}, \quad (x > 0)$
12. $\dfrac{d}{dx}(\log_a x) = \dfrac{1}{x}\dfrac{1}{\log_e a}, \quad (a > 0,\ x > 0)$

[6] If differential calculus were formulated in terms of limits, using relations of type (4), all books on calculus would have been increased in their volume several folds and become unreadable. The use of the relations of type (5), instead of (4), makes it possible to avoid this.

[7] This rule is true for any real n, as will be clear later on.

Of these results, those at (1) and (2) are already proved in Chapter 9.

- Derivatives of trigonometric functions are established in Chapter 11.
- Derivatives of exponential and logarithmic functions are established in Chapter 13.

Having accepted the standard results stated in Section 10.2.4, our next step is to establish the rules for differentiation.

Our approach will be to state the differentiation rules and discuss their application(s). *Also, we will prove some of these rules, preferably those that demand special care in proof(s), leaving the rest as exercises*. It is assumed that each function under consideration is a differentiable function of a real variable.

Rule (1): Derivative of a sum (or difference) of functions

Let f_1 and f_2 be differentiable functions of x, with the same domain, and let

$$f(x) = f_1(x) + f_2(x)$$

then

$$\frac{d}{dx}f(x) = \frac{d}{dx}f_1(x) + \frac{d}{dx}f_2(x)$$

This rule tells us that *the derivative of a sum (or difference) of functions is the sum (or difference) of their derivatives*. (This rule is similar to the corresponding *rule for limit of a sum (or difference) of functions.*)

Note (1): This rule can be extended to the derivative of the sum (or difference) of any *finite number of differentiable functions*, with the *same domain*. Thus, if

$$f(x) = f_1(x) \pm f_2(x) \pm \ldots \pm f_n(x)$$

then

$$f'(x) = f_1'(x) \pm f_2'(x) \pm \cdots \pm f_n'(x)$$

Example (2):

- $\frac{d}{dx}(\sin x - \cos x) = \cos x + \sin x$
- $\frac{d}{dx}(x^3 + 7x - 5) = 3x^2 + 7$
- $\frac{d}{dx}(a^x - \tan x + \log_e x) = a^x \log_e a - \sec^2 x + \frac{1}{x}$
- $\frac{d}{dx}(x^5 + e^x - \sec x) = 5x^4 + e^x - \sec x \cdot \tan x$

Rule (2): The Constant Rule for Derivatives
If k is any constant, f is any differentiable function, and $g(x) = k \cdot f(x)$, then

$$\frac{d}{dx}g(x) = \frac{d}{dx}[k \cdot f(x)] = k\frac{d}{dx}f(x)$$

Example (3):

- $\frac{d}{dx}(5 \sin x) = 5\frac{d}{dx}(\sin x) = 5 \cos x$
- $\frac{d}{dx}(7x^3) = 7\left[\frac{d}{dx}(x^3)\right] = 7(3x^2) = 21x^2$

Note (2): This rule reminds us of the formula for the *limit of a function multiplied by a constant*. Besides, the difference rule for derivatives is obtained by combining the *addition rule* and the *constant multiple rule* for derivatives.

Remark: The *constant rule* can be interpreted geometrically. The graph of $k \cdot f(x)$ is obtained *by stretching the graph of "f "vertically, with factor k.*

How does such a stretch affect a tangent line to the graph of "f" at $x = a$?
The constant rule says that a *vertical k stretch multiplies slopes of every thing – both the graph and the tangent line – by the same factor k*. For example, *the slope of the* line $y = 3x$ *is three times that for* $y = x$. *Similarly, the slope at each point* of $y = 5x^2$ *is five times that for* $y = x^2$.

Note (3): By combining *rule(2), with rule(1),* we can write

$$\frac{d}{dx}(5 \sin x - 7e^x + 2 \log_e x + 2x^3 - 8x^2 + 3) = 5 \cos x - 7e^x + \frac{2}{x} + 6x^2 - 16x$$

Rule (3): The derivative of product of two functions
Let $f_1(x)$ and $f_2(x)$ be differentiable functions of x and let

$$f(x) = f_1(x)f_2(x)$$

Then

$$\frac{d}{dx}f(x) = \frac{d}{dx}[f_1(x)f_2(x)] = f_1(x) = \frac{d}{dx}[f_2(x)] + f_2(x)\frac{d}{dx}[f_1(x)]^{(8)}$$

Proof: We have

$$f(x) = f_1(x)f_2(x)$$
$$f(x+h) = f_1(x+h)f_2(x+h)$$

(8) To remember this formula, we can read it as follows:

$$\frac{d[f_1(x)f_2(x)]}{dx} = \text{First function} \cdot \frac{d(\text{Second function})}{dx} + \text{Second function} \cdot \frac{d(\text{First function})}{dx}.$$

By the definition of derivative (i.e., by applying the operator d/dx), we have

$$\begin{aligned}
\frac{d}{dx}[f_1(x)f_2(x)] &= \lim_{h\to 0}\frac{f_1(x+h)f_2(x+h)-f_1(x)f_2(x)}{h}\\
&= \lim_{h\to 0}\frac{f_1(x+h)f_2(x+h)-f_1(x+h)f_2(x)+f_1(x+h)f_2(x)-f_1(x)f_2(x)}{h}\\
&= \lim_{h\to 0}\frac{f_1(x+h)\{f_2(x+h)-f_2(x)\}+f_2(x)\{f_1(x+h)-f_1(x)\}}{h}\\
&= \lim_{h\to 0}\left[\frac{f_1(x+h)\{f_2(x+h)-f_2(x)\}}{h}+\frac{f_2(x)\{f_1(x+h)-f_1(x)\}}{h}\right]\\
&= f_1(x)\frac{d}{dx}[f_2(x)]+f_2(x)\frac{d}{dx}[f_1(x)]
\end{aligned}$$

$$\therefore\quad \frac{d}{dx}[f_1(x)f_2(x)] = f_1(x)\frac{d}{dx}[f_2(x)]+f_2(x)\frac{d}{dx}[f_1(x)]$$

Note (4): This rule can also be proved, without using the trick of adding the number 0 (i.e., adding $f_1(x)f_2(x)$ and subtracting the same). Another convenient notation for stating this rule is

$$\frac{d}{dx}[u\,v] = u\frac{\mathrm{d}v}{\mathrm{d}x}+v\frac{\mathrm{d}u}{\mathrm{d}x},$$ where u and v are *differentiable* functions of x.

Note (5): This rule *can be extended to the product of more than two functions* (and in general for a product of finite number of differentiable functions).

Thus,

$$\begin{aligned}
\frac{d}{dx}[u\,v\,w] = \frac{d}{dx}[(u\,v)\,w] &= u\,v\frac{\mathrm{d}w}{\mathrm{d}x}+w\frac{\mathrm{d}(uv)}{\mathrm{d}x}\\
&= uv\frac{\mathrm{d}w}{\mathrm{d}x}+w\left[u\frac{\mathrm{d}v}{\mathrm{d}x}+v\frac{\mathrm{d}u}{\mathrm{d}x}\right]\\
&= uv\frac{\mathrm{d}w}{\mathrm{d}x}+vw\frac{\mathrm{d}u}{\mathrm{d}x}+wu\frac{\mathrm{d}v}{\mathrm{d}x}
\end{aligned}$$

Example (4):

$$\begin{aligned}
\frac{d}{dx}(x^2\log_a x) &= x^2\frac{d}{dx}(\log_a x)+\log_a x\frac{d}{dx}(x^2)\\
&= x^2\frac{1}{x\log_e a}+\log_a x^2 x\\
&= \frac{x}{\log_e a}+2x\log_a x \quad \text{Ans.}
\end{aligned}$$

Remark (1): We have seen that the derivative of a sum (or difference) of functions is a sum (or difference) of their derivatives. *By analogy, it is tempting to assume that the derivative of a* *product of functions is the product of their derivatives.* But this is not correct as can be seen from the example $\frac{d}{dx}(x^2)$[9]

(The correct formula as discussed under Rule (3) was discovered by Leibniz. Hence, it is often called the *Leibniz rule.*)

Remark (2): If k is a *constant* and $f(x)$ is a differentiable function of x, then $d/dx[kf(x)] = kf(x)$. This we have stated as rule (2). It can also be proved by applying the definition of derivative (i.e., by the first principle). It can also be proved by *applying Rule (3)* as follows:

Proof:
$$\frac{d}{dx}[kf(x)] = k\frac{d}{dx}f(x) + f(x)\frac{dk}{dx}$$
$$= k\frac{d}{dx}f(x) + f(x)\,0$$
$$= k\frac{d}{dx}f(x)$$

Example (5): Find the slope of the graph of $h(x) = (7x^3 - 5x + 2)(2x^4 + x + 7)$, at $x = 1$.

Solution: Let $f(x) = 7x^3 - 5x + 2$ and $g(x) = 2x^4 + x + 7$

Then,
$$h(x) = f(x)g(x)$$
$$\therefore h'(x) = f(x)g'(x) + g(x)f'(x) \tag{7}$$
$$= (7x^3 - 5x + 2)(8x^3 + 1) + (2x^4 + x + 7)(21x^2 - 5)$$

Now, by evaluating $h'(x)$ at $x = 1$, we get the slope of the graph of $h(x)$ at that point. We have

$$h'(1) = (7 - 5 + 2)\,(8 + 1) + (2 + 1 + 7)\,(21 - 5)$$
$$= 4(9) + 10\,(16) = 36 + 160 = 196$$

Another approach could be that we expand the right side of (7) and differentiate the resulting polynomial. Besides, note that in this example $h(x)$ is a polynomial, whereas *we will be applying the product rule to many functions other than polynomials.*

Example (6): Differentiate $(x^3 + 5x^2)\sin x$.

Solution: Let $y = (x^3 + 5x^2)\sin x$
$$\therefore \frac{dy}{dx} = (x^3 + 5x^2)\cos x + \sin x(3x^2 + 10x)$$
$$= (x^3 + 5x^2)\cos x + (3x^2 + 10x)\sin x \quad \text{Ans.}$$

[9] Note that $dx^2/dx = d(x \cdot x)/dx$. If the above assumption were true, we would conclude that $dx^2/dx = (dx/dx) \cdot (dx/dx) = 1 \times 1 = 1$, which is not correct.

Example (7): Differentiate $3^x \log_5 x$

$$\begin{aligned}\text{Let } y &= 3^x \log_5 x\\ &= 3^x \frac{d}{dx}(\log_5 x) + (\log_5 x)\frac{d}{dx}(3^x)\\ &= 3^x \frac{1}{x \log_e 5} + (\log_5 x)3^x \log_e 3\\ &= \frac{3^x}{x \log_e 5} + \log_e 3 \cdot 3^x(\log_5 x)\\ &= 3^x \frac{1}{x \log_e 5} + 3^x \log_e 3(\log_5 x) \quad \text{Ans.}\end{aligned}$$

Exercise	Answer
(1) Differentiate $x \log_e x$	$1 + \log_e x$
(2) If $y = (x^2 + 2x)\,3^x$, find dy/dx at $x = 2$	$18(\log 3^4 + 3)$
(3) If $y = 6x \tan x$, find dy/dx at $x = 0$	0

The rules defining the derivatives of *product(s)* and *quotient(s) of functions* are not as straightforward as those of sums and constant multiples. Just as the derivative of the product of two functions is *not* the product of their derivative, *the derivative of the quotient of two functions is not the quotient of their derivatives*, as you see in the next rule.

Rule (4): The derivative of quotient of two functions
Let f_1 and f_2 be differentiable functions of x and let

$$f(x) = \frac{f_1(x)}{f_2(x)}$$

Then,

$$\frac{d}{dx}\left[\frac{f_1(x)}{f_2(x)}\right] = \frac{f_2(x)(\frac{d}{dx})[f_1(x)] - f_1(x)(\frac{d}{dx})[f_2(x)]}{[f_2(x)]^2}$$

Proof: We have $f(x) = \dfrac{f_1(x)}{f_2(x)}$

$$\therefore\ f(x+h) = \frac{f_1(x+h)}{f_2(x+h)}$$

$$\begin{aligned}\frac{d}{dx}\left[\frac{f_1(x)}{f_2(x)}\right] &= \lim_{h\to 0}\frac{(f_1(x+h))/(f_2(x+h)) - (f_1(x))/(f_2(x))}{h}\\ &= \lim_{h\to 0}\frac{f_1(x+h)f_2(x) - f_1(x)f_2(x+h)}{hf_2(x+h)f_2(x)}\end{aligned}$$

As we did in the proof of the product rule, we perform *another clever manipulation*. Again, we add the number 0 in the numerator, but this time the expression (that we add) is $[-f_1(x)f_2(x) + f_1(x)f_2(x)]$.

We get

$$\frac{d}{dx}\left[\frac{f_1(x)}{f_2(x)}\right] = \lim_{h\to 0}\frac{f_1(x+h)f_2(x) - f_1(x)f_2(x) + f_1(x)f_2(x) - f_1(x)f_2(x+h)}{hf_2(x+h)f_2(x)}$$

$$= \lim_{h\to 0}\frac{f_2(x)[(f_1(x+h) - f_1(x))/h] - f_1(x)[(f_2(x+h) - f_2(x))/h]}{f_2(x+h)f_2(x)}$$

Now, taking the limit, we 2 get

$$\frac{d}{dx}\left[\frac{f_1(x)}{f_2(x)}\right] = \frac{f_2(x)(\frac{d}{dx})[f_1(x)] - f_1(x)(\frac{d}{dx})[f_2(x)]}{[f_2(x)]^2}, \text{ where } f_2(x) \neq 0$$

This formula can be remembered as follows.

The derivative of the quotient of two functions:

$$= \frac{\mathrm{Dr}(\frac{d}{dx})\mathrm{Nr} - \mathrm{Nr}(\frac{d}{dx})\mathrm{Dr}}{[\mathrm{Dr}]^2}, \quad \mathrm{Dr} \neq 0$$

where, Nr = Numerator and Dr = Denominator.

Another convenient notation used to state this rule is given below.

If u and v are differentiable functions of x, then

$$\frac{d}{dx}\left[\frac{u}{v}\right] = \frac{v(\frac{\mathrm{d}u}{\mathrm{d}x}) - u(\frac{\mathrm{d}v}{\mathrm{d}x})}{v^2}, \quad \text{where } v \neq 0$$

Note (6): The formula for the derivative of a quotient becomes more concise when the numerator $u = 1$, for all x. In this case, the formula is

$$\frac{d}{dx}\left[\frac{1}{v}\right] = \frac{v(\frac{d}{dx})(1) - 1(\frac{d}{dx})(v)}{v^2} = \frac{-1}{v^2}\frac{\mathrm{d}v}{\mathrm{d}x}$$

Example (8): Show that $(\mathrm{d}/\mathrm{d}x)\tan x = \sec^2 x$.

Solution: We have, $\tan x = \sin x/\cos x$.

$$\therefore \quad \frac{d}{dx}\tan x = \frac{\cos x(\frac{d}{dx})\sin x - \sin x(\frac{d}{dx})\cos x}{\cos^2 x}$$

$$= \frac{\cos x \cdot \cos x - \sin x(-\sin x)}{\cos^2 x}$$

$$= \frac{\cos^2 x + \sin^2 x}{\cos^2 x}$$

$$= \frac{1}{\cos^2 x} = \sec^2 x$$

Similarly,

$$\frac{d}{dx}\cot x = -\text{cosec}^2 x$$

Example (9): Show that $(\mathrm{d}/\mathrm{d}x)\sec x = \sec x \cdot \tan x$.

Solution: $\frac{d}{dx}\sec x = \frac{d}{dx}\left(\frac{1}{\cos x}\right)$

$$= \frac{\cos x(\frac{d}{dx})(1) - 1(\frac{d}{dx})(\cos x)}{(\cos x)^2}$$

$$= \frac{0 - (-\sin x)}{(\cos x)^2} = \frac{1}{\cos x}\frac{\sin x}{\cos x}$$

$$= \sec x \cdot \tan x$$

Similarly,

$$\frac{d}{dx}(\text{cosec}\, x) = -\text{cosec}\, x \cdot \cot x$$

Rule (5): The power rule of differentiation for negative powers
(As an application of Rule (4))

Show that $(\mathrm{d}/\mathrm{d}x)(x^{-n}) = -nx^{-n-1}$, *for any positive integer n.*

Proof: Since $-n$ is a negative integer, it means that n is a positive integer.
We, therefore, express f(x) as a quotient and apply the quotient rule.
We have

$$f(x) = x^{-n} = \frac{1}{x^n}$$

$$f(x) = \frac{(x^n)0 - 1 \cdot n(x^{n-1})}{(x^n)^2}$$

$$= \frac{-nx^{n-1}}{x^{2n}} = -nx^{n-1-2n}$$

$$= -nx^{-n-1}$$

In particular,

$$\frac{d}{dx}\left(\frac{1}{x}\right) = \frac{d}{dx}(x^{-1}) = -1x^{-2} = \frac{-1}{x^2} \quad \text{and} \quad \frac{d}{dx}(x^{-13}) = -13x^{-14}$$

Note (7): The function x^{-2} appears in Newton's law of gravitation and in the formula for the electric force between charges. In addition, x^{-4} appears in the formula for the flow of blood through arteries. Thus, functions of the form x^{-n}, where n is positive, arise in the real world.[10]

[10] *Calculus with Analytic Geometry* (Alternate Edition) by Robert Ellis and Denny Gulick (p. 128), HBJ Publication.

Remark: We have

$$\frac{d}{dx}(x^{-n}) = -nx^{-n-1}$$
$$\text{and } \frac{d}{dx}(x^{n}) = nx^{n-1}$$

Combining these two functions, we get

$$\begin{aligned}\frac{d}{dx}(x^n x^{-n}) &= x^n(-nx^{-n-1}) + x^{-n}(nx^{n-1})\\ &= -nx^{-1} + nx^{-1}\\ &= (-n+n)x^{-1}\\ &= 0x^{-1}\end{aligned}$$

If we assume that $0\,x^{-1} = 0$, then we can write

$$\frac{d}{dx}(x^n x^{-n}) = \frac{d}{dx}(x^{n-n}) = \frac{d}{dx}(x^{\circ}) = 0x^{-1} = 0$$

which means that the result

$\frac{d}{dx}(xn) = nx^{n-1}$ holds *even when* $n = 0$

In particular, if $n = 1$, then $\frac{d}{dx}(x^1) = 1x^{1-1} = x^{\circ} = 1$

Thus, we conclude that if n is any integer (positive, zero, or negative), then the power rule holds.

Remark: Note that, there is *no power function* $y = x^n$, which can give $\mathrm{d}y/\mathrm{d}x = x^{-1} = 1/x$. Later on, we will discover a *new function*, namely, the *logarithmic function* to the base e (denoted by $y = \log_e x$, $(x > 0)$] that gives $\mathrm{d}y/\mathrm{d}x = \mathrm{d}(y)/\mathrm{d}x = \mathrm{d}(\log_e x)/\mathrm{d}x = 1/x = x^{-1}$. (This is discussed in Chapter 13 of *Differential Calculus* and Chapter 6b of *Integral Calculus*.)

Example (10): If $y = \sqrt{\dfrac{1-\sin 2x}{1+\sin 2x}}$, find $\mathrm{d}y/\mathrm{d}x$.

Solution: Consider $1 - \sin 2x$

$$\begin{aligned}&= \sin^2 x + \cos^2 x - 2\sin x\cos x\\ &= (\sin x - \cos x)^2\\ \therefore\quad y &= \sqrt{\frac{1-\sin 2x}{1+\sin 2x}} = \frac{\sin x - \cos x}{\sin x + \cos x}\end{aligned}$$

We have

$$\frac{d}{dx}(\sin x - \cos x) = \cos x + \sin x$$

and

$$\frac{d}{dx}(\sin x + \cos x) = \cos x - \sin x$$

$$\therefore \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{Dr}(\frac{d}{dx})\mathrm{Nr} - \mathrm{Nr}(\frac{d}{dx})\mathrm{Dr}}{[\mathrm{Dr}]^2}, \quad \mathrm{Dr} \neq 0$$

$$= \frac{(\sin x + \cos x)(\cos x + \sin x) - (\sin x - \cos x)(\cos x - \sin x)}{(\sin x + \cos x)^2}$$

$$= \frac{(\sin x + \cos x)^2 + (\sin x - \cos x)^2}{(\sin x + \cos x)^2}$$

$$= \frac{(1 + 2\sin x \cos x) + (1 - 2\sin x \cos x)}{(\sin x + \cos x)^2}$$

$$\therefore \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{2}{(\sin x + \cos x)^2} = \frac{2}{(1 + \sin 2x)}$$ [11]

Example (11): If $y = (\tan x + \sec x)/(\tan x - \sec x)$, find $\mathrm{d}y/\mathrm{d}x$.

Solution: Consider

$$\tan x + \sec x = \frac{\sin x}{\cos x} + \frac{1}{\cos x} = \frac{\sin x + 1}{\cos x}$$

Similarly

$$\tan x - \sec x = \frac{\sin x - 1}{\cos x}$$

$$\therefore y = \frac{\sin x + 1}{\sin x - 1}$$

Furthermore, $\frac{d}{dx}(\sin x + 1) = \cos x$

and $\frac{d}{dx}(\sin x - 1) = \cos x$

$$\therefore \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{d}{dx}\left[\frac{\sin x + 1}{\sin x - 1}\right]$$

$$= \frac{(\sin x - 1)\frac{d}{dx}(\sin x + 1) - (\sin x + 1)\frac{d}{dx}(\sin x - 1)}{(\sin x - 1)^2}$$

$$= \frac{(\sin x - 1)\cos x - (\sin x + 1)\cos x}{(\sin x - 1)^2}$$

$$= \frac{-\cos x - \cos x}{(\sin x - 1)^2}$$

$$\therefore \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{-2\cos x}{(\sin x - 1)^2}$$

[11] *Note* (8): If we express the given function as $y = (\cos x - \sin x)/(\cos x + \sin x)$, then we will get $\mathrm{d}y/\mathrm{d}x = -2/(1 + \sin 2x)$.

Example (12): Differentiate with respect to x, the function, $y = \log_x a$

Solution: We have $\log_x a = \log_e a \cdot \log_x \mathrm{e}$

$$= \frac{\log_e a}{\log_e x} = \frac{k}{\log_e x}$$

Where k (constant) $= \log_e a$

Thus, we have

$$y = \frac{k}{\log_e x}$$

$$\therefore \quad \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\log_e x(0) - k(1/x)}{(\log_e x)^2}$$

$$= \frac{-k}{x\,(\log_e x)^2}$$

$$\therefore \quad \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{-\log_e a}{x(\log_e x)^2} \qquad \textbf{Ans.}$$

Example (13): If $y = \frac{\left(\sqrt{x+1}+\sqrt{x-1}\right)}{\left(\sqrt{x+1}-\sqrt{x-1}\right),}$ find $\mathrm{d}y/\mathrm{d}x$.

Solution: Consider

$$y = \frac{\sqrt{x+1}+\sqrt{x-1}}{\sqrt{x+1}-\sqrt{x-1}} \cdot \frac{\sqrt{x+1}+\sqrt{x-1}}{\sqrt{x+1}+\sqrt{x-1}}$$

$$= \frac{(x+1)+(x-1)+2\sqrt{x+1}\sqrt{x-1}}{(x+1)-(x-1)} = \frac{2x+2\sqrt{x^2-1}}{2} = x+\sqrt{x^2-1}$$

$$\therefore \quad \frac{\mathrm{d}y}{\mathrm{d}x} = 1 + \frac{1}{2\sqrt{x^2-1}} 2x \frac{\mathrm{d}y}{\mathrm{d}x} = 1 + \frac{1}{\sqrt{x^2-1}} x \qquad \textbf{Ans.}$$

Example (14): If $y = \frac{(\sqrt{a}+\sqrt{x})}{(\sqrt{a}-\sqrt{x})}$, find $\mathrm{d}y/\mathrm{d}x$.

Solution: $\frac{d}{dx}\left(\sqrt{a}+\sqrt{x}\right) = 0 + \frac{1}{2\sqrt{x}} = \frac{1}{2\sqrt{x}}$

$$\text{and} \quad \frac{d}{dx}\left(\sqrt{a}-\sqrt{x}\right) = -\frac{1}{2\sqrt{x}}$$

Now,

$$\frac{dy}{dx} = \frac{(\sqrt{a}-\sqrt{x})(1/2\sqrt{x}) - (\sqrt{a}+\sqrt{x})(-1/2\sqrt{x})}{(\sqrt{a}-\sqrt{x})^2}$$

$$= \frac{(\sqrt{a}/2\sqrt{x}) - (1/2) + (\sqrt{a}/2\sqrt{x}) + (1/2)}{(\sqrt{a}-\sqrt{x})^2} = \frac{\sqrt{a}/\sqrt{x}}{(\sqrt{a}-\sqrt{x})^2}$$

$$\therefore \frac{dy}{dx} = \frac{\sqrt{a}}{\sqrt{x}(\sqrt{a}-\sqrt{x})^2} \qquad \textbf{Ans.}$$

Exercise (1):

Find the derivative of the following functions with respect to x:	Answers
1. $\frac{e^x}{\sin x}$	$\frac{e^x(\sin x - \cos x)}{\sin^2 x}$
2. $\frac{a^x}{x^n}$	$\frac{a^x}{x^n}\left(\log_e a - \frac{n}{x}\right)$
3. $\frac{x\cos x}{\log_e x}$	$\frac{(-x\sin x + \cos)\log x - \cos x}{(\log_e x)^2}$
4. $\frac{\log_e x}{\cos x}$	$\frac{\cos x + x\log_e x \sin x}{x\cos^2 x}$
5. $\frac{e^x + e^{-x}}{e^x - e^{-x}}$	$\frac{-4}{(e^x - e^{-x})^2}$
6. $\sqrt{\frac{1+x}{1-x}}$	$\frac{1}{(1-x)\sqrt{1-x^2}}$ or $\frac{1}{(1-x)^{3/2}(1+x)}$
7. $\frac{\sqrt{a+x}-\sqrt{a-x}}{\sqrt{a+x}+\sqrt{a+x}}$	$\frac{a^2 - a\sqrt{a^2-x^2}}{x^2\sqrt{a^2-x^2}}$
8. $\log\sqrt{\frac{a+x}{a-x}}$	$\frac{a}{a^2-x^2}$
9. $\log\sqrt{\frac{x+ab}{x-ab}}$	$\frac{ab}{x^2-a^2b^2}$
10. $\frac{1}{\log_{10} x}$	$\frac{-\log_e 10}{x(\log_e x)^2}$

10.3 THE DERIVATIVE OF A COMPOSITE FUNCTION

We have already introduced the *concept of composite functions* in Chapter 6. Many of the functions we encounter in mathematics and in applications are composite functions. Consider the following examples:

(i) $\sin x^3$ is a function of x^3, and x^3 is a function of x.

(ii) $\log_e x^4$ is a function of x^4, and x^4 is a function of x.

(iii) $e^{\cos 2x}$ is a function of $\cos 2x$, $\cos 2x$ is a function of $2x$, and $2x$ is a function of x.

(iv) $\log(\tan(x/2))$ is a function of $\tan(x/2)$, $\tan(x/2)$ is a function of $x/2$, and $x/2$ is a function of x.

Thus, $\sin x^3$, $\log_e x^4$, $e^{\cos 2x}$, $\log(\tan(x/2))$, and so on are examples of composite functions of x.

If we could discover a general rule for *the derivative of a composite function in terms of the component functions*, then we would be able to find its derivative without resorting to the definition of the derivative.

To find the derivative of a composite function, we apply the *chain rule*, which is one of the important computational theorems in calculus. It assumes a very suggestive form in the Leibniz notation and can be stated as follows:

If y is a function of u, defined by $y = f(u)$ and $\mathrm{d}y/\mathrm{d}u$ *exists*, and if u is a function of x, defined by $u = g(x)$ and $\mathrm{d}u/\mathrm{d}x$ *exists*, then y is a function of x and $\mathrm{d}y/\mathrm{d}x$ exists, and is given by

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}u}\frac{\mathrm{d}u}{\mathrm{d}x} \tag{8}$$

Note (9): The resemblance between (8) and *an algebraic identity makes it easy to remember this rule*. Here, it is important to note that in the product of derivatives on RHS, *there are two separate operators of differentiation*, namely, $\mathrm{d}/\mathrm{d}u$ and d/dx. Hence, $\mathrm{d}y/\mathrm{d}x$ *is not obtained by canceling* $\mathrm{d}u$ *from the numerator and the denominator*.[(12)]

Note (10): The proof of the *chain rule* for all differentiable functions is *sophisticated and appears in advanced texts. A simplified proof* (pertaining to functions satisfying an additional hypothesis) is given below.

Rule (5): The Chain Rule

If $y = f(u)$ *is a differentiable function of* u and $u = g(x)$ is a *differentiable functions of* x, *such that the composite function* $y = f(g(x))$ is defined then

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}u}\frac{\mathrm{d}u}{\mathrm{d}x} \tag{9}$$

Proof: It is given that y is a differentiable function of u and u is a differentiable function of x such that $f(g(x))$ is defined. Thus, y is a function of x.

As x changes to $(x + \delta x)$, let u change to $(u + \delta u)$ and in turn y to $(y + \delta y)$.

$$\therefore \text{ As } \delta x \to 0,\ \delta u \to 0$$

[(12)] When we introduced the Leibniz notation $\mathrm{d}y/\mathrm{d}x$, we emphasized that it should be treated as a single symbol. We did not give independent meanings to $\mathrm{d}y$ and $\mathrm{d}x$. We should, therefore, consider the statement (9) as an equation involving formal differentiation. Later on, we will see the separate meanings attached to $\mathrm{d}y$ and $\mathrm{d}x$ (in Chapter 16), so that the meaning of $\mathrm{d}y/\mathrm{d}x$ is retained.

Now, consider the *algebraic identity*

$$\frac{\delta y}{\delta x} = \frac{\delta y}{\delta u}\frac{\delta u}{\delta x} \quad (\text{where } \delta u \neq 0,\ \delta x \neq 0)$$

Taking limit as $\delta x \to 0$

$$\begin{aligned}\lim_{\delta x\to 0}\frac{\delta y}{\delta x} &= \lim_{\delta x\to 0}\left(\frac{\delta y}{\delta u}\frac{\delta u}{\delta x}\right)\\ &= \left(\lim_{\delta u\to 0}\frac{\delta y}{\delta u}\right)\left(\lim_{\delta x\to 0}\frac{\delta u}{\delta x}\right)\end{aligned} \quad (10)$$

Now,

$$\lim_{\delta u\to 0}\frac{\delta y}{\delta u} = \frac{dy}{du}, \quad [\because y \text{ is a differentiable function of } u]$$

and

$$\lim_{\delta x\to 0}\frac{\delta u}{\delta x} = \frac{du}{dx}, \quad [\because u \text{ is a differentiable function of } x]$$

$\therefore$ RHS of (10) exists.

$\therefore$ LHS of (10) exists, that is, $\lim_{\delta x\to 0}\delta y/\delta x$, which is equal to $\mathrm{d}y/\mathrm{d}x$.

$$\therefore \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}u}\frac{\mathrm{d}u}{\mathrm{d}x} \quad \text{Proved}$$

Rule (5.1): Extension of Chain Rule (i.e. The Compound Chain Rule)
In general, if $y = f(t)$, $t = g(u)$, and $u = h(x)$, where $\mathrm{d}y/\mathrm{d}t$, $\mathrm{d}t/\mathrm{d}u$, and $\mathrm{d}u/\mathrm{d}x$ exist, then y is a function of x and $\mathrm{d}y/\mathrm{d}x$ exists, given by

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}u}\frac{\mathrm{d}u}{\mathrm{d}x}$$

Thus, the derivative of y is obtained in a *chain-like fashion*. In practice, it is convenient to identify the functions t, u, and so on at different stages of differentiation, as indicated in the solved examples.

Remark: In formula (R), y is represented *in two different ways:once as a function of* x and *once as a function of* u. The expression $\mathrm{d}y/\mathrm{d}x$ is the derivative of y, when y is regarded as a function of x.

In the same way, $\mathrm{d}y/\mathrm{d}u$ is the derivative of y, when y is regarded as a function of u.[13]

Formula (9) is especially useful when y is not given explicitly in terms of x, but is given in terms of an intermediate variable (*see solved examples on related rates*.)

Example (15): If $y = \log(\log(\sin x))$, find $\mathrm{d}y/\mathrm{d}x$.

[13] It can be shown that $\mathrm{d}y/\mathrm{d}x$ and $\mathrm{d}y/\mathrm{d}u$ may be different. For example, consider a simple function. Suppose $y = u^2$ and $u = (1/x)$. Then $y = (1/x)^2 = 1/x^2 = x^{-2}$, so that $\mathrm{d}y/\mathrm{d}x = -2/x^3$, whereas $\mathrm{d}y/\mathrm{d}u = 2u = 2/x$. Thus, $\mathrm{d}y/\mathrm{d}x \neq \mathrm{d}y/\mathrm{d}u$.

Solution: We have $y = \log(\log(\sin x))$, differentiating w. r. to x, we get

$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}\log(\log(\sin x)) \\
&= \frac{d}{dx}\log t, \quad [\text{where } t = \log(\sin x)] \\
&= \frac{d}{dt}\log t\frac{dt}{dx}, \quad \left[\because \frac{dy}{dx} = \frac{dy}{dt}\frac{dt}{dx}\right] \\
&= \frac{1}{t}\frac{dt}{dx} \\
&= \frac{1}{\log(\sin x)}\frac{d}{dx}\log(\sin x), \quad [\text{putting the value of } t] \\
&= \frac{1}{\log(\sin x)}\frac{d}{dx}\log u, \quad [\text{where } u = \sin x] \\
&= \frac{1}{\log(\sin x)}\frac{d}{du}(\log u)\frac{du}{dx} \\
&= \frac{1}{\log(\sin x)}\frac{1}{u}\frac{d}{dx}\sin x \\
&= \frac{1}{\log(\sin x)}\frac{1}{\sin x}\cos x \\
\therefore \frac{dy}{dx} &= \frac{\cot x}{\log(\sin x)} \qquad \textbf{Ans.}
\end{aligned}$$

Example (16): If $y = \sqrt{\sec\sqrt{x}}$, find dy/dx.

Solution: We have

$$y = \sqrt{\sec\sqrt{x}}$$

$$\begin{aligned}
\therefore \frac{dy}{dx} &= \frac{d}{dx}\sqrt{\sec\sqrt{x}} = \frac{d}{dx}t^{1/2}, \quad [\text{where } \sec\sqrt{x} = t] \\
&= \frac{d}{dt}t^{1/2}\frac{dt}{dx}, \quad \left[\because \frac{dy}{dx} = \frac{dy}{dt}\frac{dt}{dx}\right] \\
&= \frac{1}{2\sqrt{t}}\frac{d}{dx}\sec\sqrt{x} \\
&= \frac{1}{2\sqrt{\sec\sqrt{x}}}\frac{d}{dx}\sec u, \quad [\text{where } u = \sqrt{x}] \\
&= \frac{1}{2\sqrt{\sec\sqrt{x}}}\frac{d}{du}\sec u\frac{du}{dx} \\
&= \frac{1}{2\sqrt{\sec\sqrt{x}}}\sec u\tan u\frac{d}{dx}x^{1/2} \\
&= \frac{1}{2\sqrt{\sec\sqrt{x}}}\sec\sqrt{x}\tan\sqrt{x}\frac{1}{2\sqrt{x}} \\
&= \frac{\sec\sqrt{x}\tan\sqrt{x}}{4\sqrt{x}\sqrt{\sec\sqrt{x}}} \qquad \textbf{Ans.}
\end{aligned}$$

Example (17): If $y = \log\sqrt{\frac{(1+\sin mx)}{(1-\sin mx)}}$, find dy/dx.

Solution: Given, $y = \log\left(\dfrac{1+\sin mx}{1-\sin mx}\right)^{1/2}$

$$y = \frac{1}{2}\log\left(\frac{1+\sin mx}{1-\sin mx}\right)$$

$$= \frac{1}{2}[\log(1+\sin mx) - \log(1-\sin mx)]$$

$$\therefore \quad \frac{dy}{dx} = \frac{1}{2}\left[\frac{d}{dx}\log(1+\sin mx) - \frac{d}{dx}\log(1-\sin mx)\right]^{(14)}$$

$$= \frac{1}{2}\left[\frac{d}{dx}\log t - \frac{d}{dx}\log u\right] \quad \{\text{where } t = 1+\sin mx, \text{ and } u = 1-\sin mx\}x$$

$$= \frac{1}{2}\left[\frac{d}{dt}\log t\frac{dt}{dx} - \frac{d}{du}\log u\frac{du}{dx}\right]$$

$$= \frac{1}{2}\left[\frac{1}{t}\frac{d}{dx}(1+\sin mx) - \frac{1}{u}\frac{d}{dx}(1-\sin mx)\right]$$

$$= \frac{1}{2}\left[\frac{m\cos mx}{1+\sin mx} + \frac{m\cos mx}{1-\sin mx}\right]$$

$$= \frac{1}{2}m\cos mx\left[\frac{1-\sin mx+1+\sin mx}{1-\sin^2 mx}\right]$$

$$= \frac{1}{2}\left[\frac{2m\cos mx}{\cos^2 mx}\right] = \frac{m}{\cos mx}$$

$$= m\sec mx$$ **Ans.**

Simpler method for Example (17) and other similar problems:

Given $y = \dfrac{1}{2}\log\left(\dfrac{1+\sin mx}{1-\sin mx}\right) = \dfrac{1}{2}[\log(1+\sin mx) - \log(1+\sin mx)]$

$$\therefore \quad \frac{dy}{dx} = \frac{1}{2}\left[\frac{d}{dx}\log(1+\sin mx) - \frac{d}{dx}\log(1+\sin mx)\right]$$

Consider $\dfrac{d}{dx}\log(1+\sin mx)$

$$= \frac{1}{1+\sin mx}(m\cos mx) = \frac{m\cos mx}{1+\sin mx}$$

Similarly $\dfrac{d}{dx}\log(1-\sin mx) = \dfrac{-m\cos mx}{1-\sin mx}$

(14) *Note* (11): From this step onward, we can adopt a simpler approach, as given below, instead of the one that follows in continuation.

$$\therefore \quad \frac{dy}{dx} = \frac{1}{2}\left[\frac{m\cos mx}{1+\sin mx} + \frac{m\cos mx}{1-\sin mx}\right]$$

$$= \frac{1}{2} m\cos mx\left[\frac{1-\sin mx + 1 + \sin mx}{1-\sin^2 mx}\right]$$

$$= \frac{1}{2}\left[\frac{2m\cos mx}{\cos^2 mx}\right] = \frac{m}{\cos mx} = m\sec mx \qquad \textbf{Ans.}$$

Note (12): When computing derivatives by the chain rule, we *do not actually write the functions t, u, and so on, but bear them in mind, and keep on obtaining the derivatives of the component functions, stepwise,* as shown in the following solved examples.

Example (18): If $y = \log(\sin x^2)$, find dy/dx.

Solution: Given, $y = \log(\sin x^2)$.

Using the comments given in the above note, we write

$$\frac{dy}{dx} = \frac{d}{dx}\left[\log(\sin x^2)\right]$$

$$= \frac{1}{\sin x^2}\cos x^2 \frac{d}{dx}(x^2)$$

$$= \cot^2 x \cdot 2x$$

$$= 2x\cot x^2 \qquad \textbf{Ans.}$$

Note (13): Observe that *when we differentiate a function by using the chain rule, we differentiate from the outside inward.* Thus, to differentiate $\sin(3x + 5)$, we first differentiate the outer function $\sin x$ (at $3x + 5$) and then differentiate the inner function $(3x + 5)$, at (x). Similarly, to differentiate $\cos x^7$, we first differentiate the outer function $\cos x$ (at x^7) and then differentiate the inner function x^7, at x.

The chain rule can be applied to even longer composites. The procedure is always the same: *Differentiate from outside inward and multiply the resulting derivatives* (evaluated at the appropriate numbers).

For example,

$$\frac{d}{dx}\left[\sin(\cos(\tan^5 x))\right] = \left[\cos(\cos(\tan^5 x))\right]\left[-\sin(\tan^5 x)\right](5\tan^4 x)\sec^2 x$$

Example (19): If $y = \log\log(\log x)$, find dy/dx.

Solution: We have

$$y = \log\log(\log x)$$

$$\frac{dy}{dx} = \frac{d}{dx}[\log\log(\log x)]$$

$$= \frac{1}{\log(\log x)}\frac{d}{dx}[\log(\log x)]$$

$$= \frac{1}{\log(\log x)}\frac{1}{\log x}\frac{d}{dx}(\log x)$$

$$= \frac{1}{\log(\log x)}\frac{1}{\log x}\frac{1}{x}$$

$$= \frac{1}{x\log x\log(\log x)} \qquad \textbf{Ans.}$$

Example (20): If $y = \log\log\log x^3$, find $\mathrm{d}y/\mathrm{d}x$.

Solution: We have

$$
\begin{aligned}
y &= \log\log\log x^3 \\
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{d}{dx}\left[\log\log\log x^3\right] \\
&= \frac{1}{\log\log x^3}\frac{d}{dx}\left[\log\log x^3\right] \\
&= \frac{1}{\log\log x^3}\frac{1}{\log x^3}\frac{d}{dx}\left[\log x^3\right] \\
&= \frac{1}{\log\log x^3}\frac{1}{\log x^3}\frac{1}{x^3}\frac{d}{dx}(x^3) \\
&= \frac{3x^2}{x^3\log x^3\log\log x^3} \\
&= \frac{3}{x\log x^3\log\log x^3} \qquad \textbf{Ans}.
\end{aligned}
$$

Example (21): If $y = \mathrm{e}^{x^3}$, find $\mathrm{d}y/\mathrm{d}x$.

Solution: We have, $y = \mathrm{e}^{x^3}$

$$
\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{d}{dx}(\mathrm{e}^{x^3}) \\
&= e^{x^3}\frac{d}{dx}(x3) \\
&= e^{x^3}3x^2 \\
&= 3x^2e^{x^3} \qquad \textbf{Ans}.
\end{aligned}
$$

Example (22): If $y = \sqrt{\cos\sqrt{x}}$, find $\mathrm{d}y/\mathrm{d}x$.

Solution: We have, $y = \sqrt{\cos\sqrt{x}}$

$$
\begin{aligned}
\therefore \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{d}{dx}\left[\sqrt{\cos\sqrt{x}}\right] \\
&= \frac{1}{2\sqrt{\cos\sqrt{x}}}\left(-\sin\sqrt{x}\right)\frac{d}{dx}(\sqrt{x}) \\
&= \frac{-\sin\sqrt{x}}{2\sqrt{\cos\sqrt{x}}}\frac{1}{2\sqrt{x}}\frac{d}{dx}(x) \\
&= \frac{-\sin\sqrt{x}}{4\sqrt{x}\sqrt{\cos\sqrt{x}}} \qquad \textbf{Ans}.
\end{aligned}
$$

Example (23): If $y = \sin(\log_{10} x)$, find dy/dx.

We have, $y = \sin(\log_{10} x)$

$$\frac{dy}{dx} = \frac{d}{dx}[\sin(\log_{10} x)]$$
$$= \cos(\log_{10} x)\frac{d}{dx}(\log_{10} x)$$
$$= \cos(\log_{10} x)\frac{1}{x\log_e 10}$$
$$\therefore \quad \frac{dy}{dx} = \frac{\cos(\log_{10} x)}{x\log_e 10} \qquad \textbf{Ans.}$$

Example (24): If $y = \log[\sin x^\circ + \cos x^\circ]$, find dy/dx.

$$\therefore \frac{dy}{dx} = \frac{1}{(\sin x^\circ + \cos x^\circ)}\frac{d}{dx}[\sin x^\circ + \cos x^\circ]$$
$$= \frac{1}{(\sin x^\circ + \cos x^\circ)}\frac{d}{dx}\left[\sin\frac{\pi x}{180} + \cos\frac{\pi x}{180}\right]$$
$$= \frac{1}{(\sin x^\circ + \cos x^\circ)}\left[\cos\left(\frac{\pi x}{180}\right)\frac{\pi}{180} - \sin\left(\frac{\pi x}{180}\right)\frac{\pi}{180}\right]$$
$$\therefore \frac{dy}{dx} = \frac{\pi}{180}\left[\frac{\cos x^\circ - \sin x^\circ}{\cos x^\circ + \sin x^\circ}\right]\left(\because \frac{\pi x}{180} = x^\circ\right) \qquad \textbf{Ans.}$$

Example (25): If $y = 2^x \cos(3x - 2)$, find dy/dx.

Solution: We have

$$y = 2^x \cos(3x - 2)$$
$$\therefore \frac{dy}{dx} = 2^x \frac{d}{dx}\cos(3x - 2) + \cos(3x - 2)\frac{d}{dx}2^x$$
$$= 2^x[-\sin(3x - 2)]3 + \cos(3x - 2)2^x\log_e 2$$
$$= 2^x[\log_e 2\cos(3x - 2) - 3\sin(3x - 2)] \qquad \textbf{Ans.}$$

Example (26): If $y = 1/(x\log_e x)$, find dy/dx.

Solution: We have

$$y = \frac{1}{x\log_e x} = [x\log_e x]^{-1}$$
$$\therefore \frac{dy}{dx} = \frac{d}{dx}[x\log_e x]^{-1}$$
$$= -1[x\log_e x]^{-2}\frac{d}{dx}(x\log_e x)$$
$$= \frac{-1}{(x\log_e x)^2}\left[x\frac{1}{x} + \log_e x(1)\right]$$
$$= \frac{-(1 + \log_e x)}{(x\log_e x)^2} \qquad \textbf{Ans.}$$

Note (14): Some important Observations about the Chain Rule

Suppose we have to differentiate the function: $y = (x^2+2)^2$. Then, we may write it as

$$y = x^4 + 4x^2 + 4$$

and differentiate it easily. *But this method is impractical for a function such as*

$$y = (x^2+2)^{1000} \quad \text{or} \quad y = (x^2+2)^{5/3}$$

Note that, since $y = (x^2+2)^{1000}$ is like $y = u^{1000}$, where, $u = x^2+2$, we can write (using the chain rule),

$$\begin{aligned}\frac{\mathrm{d}y}{\mathrm{d}x} &= 1000u^{999}\,(2x)\\ &= 2000x(x^2+2)^{999} \qquad \textbf{Ans.}\end{aligned}$$

Example (27): If $y = \left(\frac{x}{x+3}\right)^5$, find $\mathrm{d}y/\mathrm{d}x$.

Solution: We have

$$\begin{aligned} y &= \left(\frac{x}{x+3}\right)^5\\ \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{d}{dx}\left(\frac{x}{x+3}\right)^5\\ &= 5\left(\frac{x}{x+3}\right)^4 \frac{d}{dx}\left(\frac{x}{x+3}\right)\\ &= 5\left(\frac{x}{x+3}\right)^4 \left[\frac{(x+3)(1)-x(1)}{(x+3)^2}\right]\\ &= 5\left(\frac{x}{x+3}\right)^4 \left[\frac{3}{(x+3)^2}\right]\\ &= \frac{5x^4 3}{(x+3)^4(x+3)^2}\\ &= \frac{15x^4}{(x+3)^6} \qquad \textbf{Ans.}\end{aligned}$$

(i) *When we apply the chain rule we use the power rule first and then the quotient rule.*

(ii) *The power rule is a special case of the chain rule.*

Let us prove the following result:

$$\frac{d}{dx}[f(x)]^n = n[f(x)]^{n-1} f'(x)$$

Proof: Let $y = [f(x)]^n$

$$y = u^n, \text{ where } u = f(x)$$

$$\frac{\mathrm{d}y}{\mathrm{d}u} = nu^{n-1} \quad \text{and} \quad \frac{\mathrm{d}u}{\mathrm{d}x} = f'(x)$$

Now,

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$
$$= nu^{n-1} f'(x) = n[f(x)]^{n-1} f'(x)$$
$$\therefore \quad \frac{d}{dx}[f(x)]^n = n[f(x)]^{n-1} f'(x)$$

In particular, we have,

$$\frac{d}{dx}\frac{1}{[f(x)]^n} = \frac{d}{dx}[f(x)]^{-n}$$
$$= -n[f(x)]^{-n-1} f'(x)$$
$$= \frac{-n}{[f(x)]^{n+1}} f'(x)$$

and

$$\frac{d}{dx}\sqrt{f(x)} = \frac{d}{dx}[f(x)]^{1/2} = \frac{1}{2}[f'(x)]^{-1/2} = \frac{1}{2\sqrt{f'(x)}}$$

Similarly, we can prove the following results, using the *chain rule*.

An important requirement is that we must remember the derivatives of basic elementary functions involved.

y	dy/dx
$\sin[f(x)]$	$\cos[f(x)]f'(x)$
$\cos[f(x)]$	$-\sin[f(x)]f'(x)$
$\tan[f(x)]$	$\sec^2[f(x)]f'(x)$
$\cot[f(x)\}$	$-\text{cosec}^2[f(x)]f'(x)$
$\sec[f(x)]$	$\sec[f(x)]\tan[f(x)]f'(x)$
$\text{cosec}[f(x)]$	$-\text{cosec}[f(x)]\cot[f(x)]f'(x)$

y	dy/dx
$a^{[f(x)]}$	$a^{[f(x)]} \log_e a f'(x)$
$e^{[f(x)]}$	$e^{[f(x)]} f'(x)$
$\log_e[f(x)]$	$1/(f(x))f'(x)$
$\log_a[f(x)]$	$1/([f(x)]\log_e a)f'(x)$
If $(d/dx)f(x) = \phi(x)$, then $(d/dx)f(ax+b) = a\,\phi(ax+b)$	

(iii) All the above results are the corollaries to the chain rule. *They should not be used as formulas*. In other words, to write the derivative of a composite function, we must write all the steps before reaching the final answer, as shown in the solved examples, (4)–(13).

10.4 USEFULNESS OF TRIGONOMETRIC IDENTITIES IN COMPUTING DERIVATIVES

The following examples indicate that trigonometric identities can be used in expressing *certain combinations of functions* (in suitable forms), convenient for computing their derivatives in a simple form.

Example (28): If $y = \frac{(a\cos x - b\sin x)}{(a\sin x + b\cos x)}$, find $\mathrm{d}y/\mathrm{d}x$.

Solution: If this function is considered as a quotient, then its differentiation by the rule (5) will be very complicated. Hence, we simplify the given function by changing the constants, as given below.[15]

Put $a = r\sin t$ and $b = r\cos t$

$$\therefore \quad \tan t = \frac{a}{b}, \quad \text{and} \quad r^2 = a^2 + b^2 \quad (\text{where } r \text{ and } t \text{ are obviously constants})$$

$$\therefore = \frac{r\sin t\cos x - r\cos t\sin x}{r\sin t\sin x + r\cos t\cos x}$$

$$= \frac{r\sin(t-x)}{r\cos(t-x)} = \tan(t-x)$$

$$\therefore \frac{\mathrm{d}y}{\mathrm{d}x} = -\sec^2(t-x) = \frac{-1}{[\cos(t-x)]^2}$$

$$= \frac{-1}{[\cos t\cos x + \sin t\sin x]^2}$$

$$= \frac{-1}{\left[\frac{b}{r}\cos x + \frac{a}{r}\sin x\right]^2}$$

$$= \frac{-r^2}{(b\cos x + a\sin x)^2}$$

$$\therefore \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{-(a^2+b^2)}{(b\cos x + a\sin x)^2} \qquad \textbf{Ans.}$$

Example (29): If $y = \frac{\sin x}{(1+\cos x)}$, find $\mathrm{d}y/\mathrm{d}x$.

We have $(d/dx)(\sin x) = \cos x$ and $(d/dx)(1+\cos x) = -\sin x$.

[15] A good number of examples of this type are discussed in Part II (Chapter 3) of this book.

Now consider,

$$\frac{dy}{dx} = \frac{(1+\cos x)(\frac{d}{dx})(\sin x) - \sin x(\frac{d}{dx})(1+\cos x)}{(1+\cos x)^2}$$

$$= \frac{(1+\cos x)(\cos x) - \sin x(-\sin x)}{(1+\cos x)^2}$$

$$= \frac{1+\cos^2 x + \sin^2 x}{(1+\cos x)^2}$$

$$\therefore \quad \frac{dy}{dx} = \frac{2}{(1+\cos x)^2} \qquad \textbf{Ans}.$$

Also, it is easy to show that,

$$y = \frac{\sin x}{1+\cos x} = \tan\frac{x}{2}^{(16)}$$

$$\therefore \frac{dy}{dx} = \sec^2\frac{x}{2}\left(\frac{1}{2}\right) = \frac{1}{2}\sec^2\frac{x}{2} \qquad \textbf{Ans}.$$

Example (30): If $y = \log\tan((\pi/4)+(x/2))$, find dy/dx.

Solution: We have, $y = \log\tan((\pi/4)+(x/2))$.

Using trigonometric identities and algebraic operations, we can show that

$$\sqrt{\frac{1+\sin x}{1-\sin x}} = \tan\left(\frac{\pi}{4}+\frac{x}{2}\right)^{(17)}$$

(16) $\frac{\sin x}{1+\cos x} = \frac{2\sin(x/2)\cos(x/2)}{1+2\cos^2(x/2)-1} = \frac{2\sin(x/2)\cos(x/2)}{2\cos^2(x/2)} = \frac{\sin(x/2)}{\cos(x/2)} = \tan\frac{x}{2}$

(17) Similarly, we can easily prove the following results:

$$\sqrt{\frac{1+\sin 2x}{1-\sin 2x}} = \tan\left(\frac{\pi}{4}+x\right), \sqrt{\frac{1-\sin x}{1+\sin x}} = \tan\left(\frac{\pi}{4}-\frac{x}{2}\right)$$

$$\sqrt{\frac{1-\cos 2x}{1+\cos 2x}} = \tan x, \sqrt{\frac{1+\cos x}{1-\cos x}} = \cot\left(\frac{x}{2}\right)$$

$$\frac{\sin x}{1+\cos x} = \tan\frac{x}{2}, \frac{\cos x}{1+\sin x} = \tan\left(\frac{\pi}{4}-\frac{x}{2}\right).$$

We should be able to obtain these results and they need not be remembered. For necessary details, refer to Part II of this book, Chapter 2

Assuming this result, we have,

$$y = \log\left(\frac{1+\sin x}{1-\sin x}\right)^{1/2} = \frac{1}{2}[\log(1+\sin x) - \log(1-\sin x)]$$

$$\frac{dy}{dx} = \frac{1}{2}\left[\frac{1}{(1+\sin x)}\cos x - \frac{1}{(1-\sin x)}(-\cos x)\right]$$

$$= \frac{1}{2}\left[\frac{\cos x}{(1+\sin x)} - \frac{-\cos x}{(1-\sin x)}\right]$$

$$= \frac{1}{2}\left[\frac{\cos x - \cos x \sin x + \cos x + \cos x \sin x}{1-\sin^2 x}\right]$$

$$= \frac{1}{2}\left[\frac{2\cos x}{\cos^2 x}\right] = \frac{1}{\cos x} = \sec x$$

$$\therefore \quad \frac{dy}{dx} = \sec x \qquad \textbf{Ans.}$$

Exercise (2): Differentiate the following functions w.r.t. x:

(1) $\log(\log \sin x)$ (2) $[\log(\log(\log x))]^4$ (3) $\sqrt{\sin\sqrt{x}}$ (4) $\dfrac{\sin\sqrt{x}}{\sqrt{x}}$

(5) $\cos(x^3 e^x)$ (6) $e^{e^{e^x}}$ (7) 2^{2^x} (8) $\log_7(\log_7 x)$ (9) $\dfrac{\sin x^\circ}{x}$

(10) $\dfrac{e^x + e^{-x}}{e^x - e^{-x}}$ (11) $\log\sqrt{\dfrac{1+\sin 3x}{1-\sin 3x}}$ (12) $\log\sqrt{\dfrac{1+\cos x}{1-\cos 3x}}$

Answers:

(1) $\dfrac{\cot x}{\log \sin x}$ (2) $\dfrac{4[\log(\log(\log x))]^3}{x \log x \log(\log x)}$ (3) $\dfrac{\cos\sqrt{x}}{4\sqrt{x}\sqrt{\cos\sqrt{x}}}$

(4) $\dfrac{\sqrt{x}\cos\sqrt{x} - \sin\sqrt{x}}{2x^{3/2}}$ (5) $-x^2 e^x(x+3)\sin(x^3 e^x)$ (6) $e^{e^{e^x}} e^{e^x} e^x$

(7) $2^{2^x} 2^x \log_e 2$ (8) $\dfrac{1}{x(\log 7)^2 \log_7 x}$ (9) $\dfrac{1}{x^2}\left[\dfrac{\pi x}{180}\cos x^\circ - \sin x^\circ\right]$

(10) $\dfrac{-4e^{2x}}{(e^{2x}-1)^2}$ (11) $3\sec 3x$ (12) $-\operatorname{cosec} x$

10.5 DERIVATIVES OF INVERSE FUNCTIONS

We have seen (in Chapter 2) that if a function $y = f(x)$ is one–one and onto, from A to B, then the inverse of f exists, and is denoted by f^{-1}. Also, f^{-1} is a one–one and onto function from B to A. The inverse function consists of the same pairs of elements but in reverse order.[18]

Now the question is*: If f is differentiable, will f^{-1} be differentiable? If so, at what points, and what is the rule of differentiation*? If this information is available to us, it will help us to obtain the derivatives of $\log_e x$, $\sin^{-1} x$, $\cos^{-1} x$, *and so on whenever they are defined.*

Rule (6)

Theorem: If $y = f(x)$ is a *differentiable* function of x such that the inverse function $x = f^{-1}(y)$ is defined and dy/dx, dx/dy both exist, then

$dx/dy = \frac{1}{(dy/dx)}$, provided $dy/dx \neq 0$.[19]

Proof: Suppose $y = f(x)$ be a one–one mapping of A onto B, where A and B are subset of Real numbers.

Let $x = f^{-1}(y)$ be the inverse mapping of B onto A.

Then, the composite mapping (ff^{-1}) is the identity mapping of B onto B.

That is, $f[f^{-1}(y)] = y$

Differentiating both sides of the above equation w.r.t. y, we get

$$\frac{d}{dy} f\left[f^{-1}(y)\right] = \frac{dy}{dy} = 1 \tag{11}$$

By the chain rule for composite functions (and remembering that $f^{-1}(y) = x$), we get LHS of (11)

$$\frac{d}{dx} f(x) \frac{dx}{dy} = 1$$

using $f(x) = y$, we get

$$\frac{dy}{dx} \frac{dx}{dy} = 1$$

We have thus shown that if y is a function of x and x is the inverse function of y, then $dx/dy = \frac{1}{(dy/dx)}$, provided $dy/dx \neq 0$.

Corollary: If $x = g(y)$ is a differentiable function of y such that the inverse function $y = g^{-1}(x)$ exists, then $dy/dx = \frac{1}{(dx/dy)}$, provided $dx/dy \neq 0$.

We can also prove the above results as follows:

[18] It means that the domain of f^1 is the range of f, and range of f^1 is the domain of f.

[19] Suppose $y = f(x)$ has an a differentiable function, which has an inverse. Then, we can express the inverse function by the equation $x = f^{-1}(y)$. Accordingly, the derivative of f is expressed by dy/dx, whereas the derivative of f^{-1} by dx/dy.

Theorem: (Method II)
If $y = f(x)$ is a derivable function of x, such that the inverse function $x = f^{-1}(y)$ is defined and $\mathrm{d}y/\mathrm{d}x$, $\mathrm{d}x/\mathrm{d}y$ both exist, then $\mathrm{d}x/\mathrm{d}y = 1/(\mathrm{d}y/\mathrm{d}x)$, provided $\mathrm{d}y/\mathrm{d}x \neq 0$.

Proof: As x changes to $x + \delta x$, let y change to $y + \delta y$.

$\therefore$ As $\delta x \to 0$, $\delta y \to 0$

Now, consider the *algebraic identity*,

$$\frac{\delta y}{\delta x} = 1 \quad (\delta x \neq 0,\ \delta y \neq 0)$$

$$\therefore \quad \frac{\delta x}{\delta y} = \frac{1}{\delta y/\delta x}$$

Taking limit as $\delta x \to 0$

$$\lim_{\delta x \to 0} \frac{\delta x}{\delta y} = \lim_{\delta x \to 0} \left[\frac{1}{\left(\frac{\delta y}{\delta x}\right)}\right] = \left[\frac{1}{\lim\limits_{\delta x \to 0} \left(\frac{\delta y}{\delta x}\right)}\right] \tag{12}$$

$\because$ y is a differentiable function of x,

$$\therefore \lim_{\delta x \to 0} \frac{\delta y}{\delta x} = \frac{\mathrm{d}y}{\mathrm{d}x} \tag{13}$$

$\therefore \lim\limits_{\delta x \to 0} \frac{\delta x}{\delta y}$ exists, provided $\frac{\mathrm{d}y}{\mathrm{d}x} \neq 0$

Again, as $\delta x \to 0$, $\delta y \to 0$

$$\therefore \lim_{\delta x \to 0} \frac{\delta x}{\delta y} = \lim_{\delta y \to 0} \frac{\delta x}{\delta y} = \frac{\mathrm{d}x}{\mathrm{d}y} \tag{14}$$

using (13) and (14) in (12), we get

$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{1}{\mathrm{d}y/\mathrm{d}x}$, provided $\frac{\mathrm{d}y}{\mathrm{d}x} \neq 0$

Note: Since f and f^{-1} are mutually inverse functions, we also have

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{\mathrm{d}x/\mathrm{d}y}, \quad \text{provided } \frac{\mathrm{d}x}{\mathrm{d}y} = 0$$

Summary of Differentiation Rules

Rule (1) (Derivative of a *sum* of functions)	$\frac{d}{dx}[f_1(x)+f_2(x)] = \frac{d}{dx}[f_1(x)] + \frac{d}{dx}[f_2(x)]$
Rule (2) (Derivative of a *constant multiple* of a function)	$\frac{d}{dx}kf(x) = k\frac{d}{dx}f(x)$ [k = constant)
Rule (3) (Derivative of a *product* of functions)	$\frac{d}{dx}[f_1(x)f_2(x)] = f_1(x)\frac{d}{dx}[f_2(x)] + f_2(x)\frac{d}{dx}[f_1(x)]$
Rule (4) (Derivative of *ratio* of functions)	$\frac{d}{dx}\left[\frac{f_1(x)}{f_2(x)}\right] = \frac{f_2(x)(\frac{d}{dx})[f_1(x)] - f_1(x)(\frac{d}{dx})[f_2(x)]}{[f_2(x)]^2}$
Rule (5) (Derivative of *composite* functions): the chain rule	$\frac{d}{dx}g[f(x)] = \frac{\mathrm{d}}{\mathrm{d}f}[g(f)] \times \frac{d}{dx}[f(x)]$
Rule (6) (Derivative of *inverse* of functions)	$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{1}{\mathrm{d}y/\mathrm{d}x}$ that is, $\frac{\mathrm{d}}{\mathrm{d}y}x(y) = \frac{1}{(d/dx)y(x)}$

Remark: One may get an impression that by using differentiation rules (1)–(6), we should be able to compute the derivative of any function. However, there are still some functions whose derivatives cannot be computed with these rules. On the other hand, *the derivative of such a function can sometimes be computed directly from the definition*. For instance, consider the function

$$f(x) = x\,|x|$$

We cannot apply any of the rules to obtain $f'(0)$ because |x| is not differentiable at 0. Nevertheless, using the definition of derivative, we find that

$$f'(0) = \lim_{x\to 0}\frac{f(x)-f(0)}{x-0} = \lim_{x\to 0}\frac{x|x|-0}{x-0} = \lim_{x\to 0}|x| = 0$$

However, a great majority of the differentiable functions that we will encounter can be differentiated by rules (1)–(6).

11a Basic Trigonometric Limits and Their Applications in Computing Derivatives of Trigonometric Functions

11a.1 INTRODUCTION

Every time we come across new functions, we would like to find if they are differentiable, and if so, we would like to find their derivatives. In Chapter 9, we have seen that the *derivative of a function is the limit of the particular kind.* To compute the derivative(s) of basic trigonometric functions, we shall be using the following *basic trigonometric limits*:

(i) $\lim_{x \to 0} \cos x = 1$
(ii) $\lim_{x \to 0} \sin x = 0$
(iii) $\lim_{x \to 0} \frac{\sin x}{x} = 1$ and
(iv) $\lim_{x \to 0} \frac{\cos x - 1}{x} = 0$

But how do we get these limits? We shall obtain the above results shortly.

In Chapter 5, we extended the definitions of trigonometric ratios (of an acute angle) to the *trigonometric functions of real variable.* [1]

In Chapter 7, we introduced the concept of limit of a function, and gave illustrative examples, for evaluating the limits of some simple algebraic functions, involving polynomials (including rational functions).[2]

Also, we stated the *main limit theorem* (without proof), introduced the substitution rule (with its usefulness), and proved the *sandwich theorem* (or the *squeezing theorem*), which is very useful in evaluating limits of a variety of trigonometric functions.

11a-Basic trigonometric limits and their applications in computing derivatives of trigonometric functions

[1] For this purpose, the concept of *directed angles* and their *radian measure was introduced with a logical understanding that angles of any magnitude and sign* could be generated.
[2] Starting from intuitive meaning of limit, we entered into the rigorous study of the concept and developed ε, δ definition of limit by considering a good number of suitable examples to cover various possible situations, so as to make the definition complete in all respects.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

If the reader has gone through the concept of limit (as discussed in Chapters 7a and 7b), then he will be able to appreciate that *our way of introducing the concept of limit has been the simplest and the most practical one*, and that it would not be so simple if it were introduced by considering trigonometric or any other functions.

Now, since we have some idea about the limit concept, we are in a position to discuss and establish the basic trigonometric limits mentioned above (i–iv). It is important that we are very clear about the definition of trigonometric functions of an arbitrary angle whose measure "θ" is expressed in radians. (Recall that radian measure "x" in $\sin x$ stands for a real number.) Accordingly, we can use, instead of "θ", any other symbol (like x, y, or t) that is used to represent real numbers, and identify it, as the measure of an angle in radians. It is in this sense that expressions such as $x + \sin x$, $x \cos x$, $(\sin x)/x$, and other similar ones are understood.

Also, recall (from Chapter 7a) that for evaluating the limit $\lim_{x\to 0} (\sin x)/x$, we listed the values of the ratio $(\sin x)/x$, for several values of x closer and closer to the number "0", and observed that the value of this ratio approaches nearer and nearer to 1, as x tends to "0". We therefore guessed (and, in fact, agreed) that

$$\lim_{x\to 0} \frac{\sin x}{x} = 1, \quad (x \text{ in radians})^{(3)}$$

Fortunately, our guess happens to be correct but the feeling of uncertainty (and incompleteness) remains in our mind. This situation demands that we should prove the above result in a more systematic way. Such a proof is available in the text.

11a.2 BASIC TRIGONOMETRIC LIMITS

To prove the basic trigonometric limits (i) and (ii), we recall the definitions of the sine and cosine functions (with reference to a circle of radius "r" *centered at the origin, and an angle of θ radians, placed in standard position at the center of the circle,* as shown in Figure 11a.1).

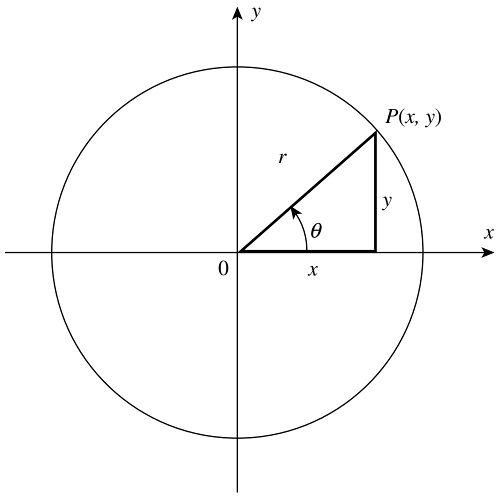


FIGURE 11a.1 Angle θ in standard position.

(3) It is due to this limit (and many other problems creating various situations in the way of guessing limits of certain functions) that compelled mathematicians to seek a suitable definition of limit of a function. This is how the ε, δ definition of limit came into existence. This definition might look abstract to someone, but in fact it is a beautiful definition of limit, complete in all respects.

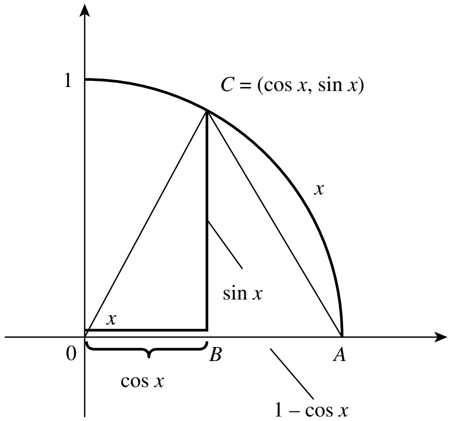


FIGURE 11a.2 Unit circle centered at the origin.

The terminal side of the angle intersects the circle at a unique point $P(x, y)$. We define the *sine function* and *cosine function* by

$$\sin\theta = \frac{y}{r} \quad \text{and} \quad \cos\theta = \frac{x}{r} \text{ (4)}$$

If $r = 1$, then sin $\theta =$ y and cos $\theta = x$(5)

Since *we normally use "x" to represent points in the domain of a function, we will usually follow that convention* for the sine and cosine functions and replace θ by x (see Figure 11a.2).

In Figure 11a.2, let C be any point on the unit circle (placed in the standard position) such that it is at the end of the arc length x. Since this arc length subtends an angle of x radians at the center, we identify the point C as a function of the angle x and define cosine and sine functions of this angle as follows:

$$\sin x = y\text{–coordinate of } C$$

$$\cos x = x\text{–coordinate of } C$$

Since *C (cos x, sin x) can move endlessly around the unit circle* (with positive or negative arc length), the *domain of both sine and cosine functions* is $(-\infty, +\infty)$. The *largest value* either function may have is 1 and the *smallest value* is -1. Also, observe that *both these functions assume all values between −1 and 1*. Hence, the *range of both the functions is* $[-1,\ 1]$.

Note that as $x \to 0$, the point $P(\cos x, \sin x)$ moves toward (1, 0) so that we get

$$\lim_{x\to 0} \cos x = 1 \quad \text{and} \quad \lim_{x\to 0} \sin x = 0$$

Thus, we have shown *the correctness of the results* (i) and (ii).

(4) The properties of similar triangles imply that sin θ and cos θ depend only on θ, not on the value of r.

(5) We repeat that in the expression sinθ, "θ" represents a number. Thus, we write sin2 to mean sin(2 radian).

(Though we have concluded results (i) and (ii) in a very simple way, their rigorous proof uses *sandwich theorem.*)[6]

Note (1): Now, onward, we shall be using results (i) and (ii) freely in solving problems and obtaining other results.

Now, our next goal is to show that for any real number "a",

$$\lim_{x \to a} \sin x = \sin a \quad \text{and} \quad \lim_{x \to a} \cos x = \cos a \tag{1}$$

We know that, if "a" is a fixed number and $x = a + h$, then

$$\lim_{x \to a} f(x) = l \quad \text{if and only if} \quad \lim_{h \to 0} f(a + h) = l$$

Therefore, in order to prove the result(s) at (1) above, we can instead show that

$$\lim_{h \to 0} \sin(a + h) = \sin a \quad \text{and} \quad \lim_{h \to 0} \cos(a + h) = \cos a^{(7)}$$

Solution: Let "a" be a *fixed number*. To prove that $\lim_{h \to 0} \sin(a + h) = \sin a$ and hence that $\lim_{x \to a} \sin x = \sin a$, we *use the trigonometric identity:*

$$\sin\,(a + h) = \sin\,a \cdot \cos\,h + \cos\,a \cdot \sin\,h$$

Since "a" is *fixed*, $\sin\,a$ and $\cos\,a$ are *constants.*

Now,

$$\begin{aligned}\lim_{h \to 0} \sin(a + h) &= \lim_{h \to 0} (\sin a \cos h + \cos a \sin h) \\ &= \sin a\Big(\lim_{h \to 0} \cos h\Big) + \cos a\Big(\lim_{h \to 0} \sin h\Big)\end{aligned}$$

(Here, we have applied the sum and constant multiple rules for limits.)

$$\begin{aligned}&= (\sin a)1 + (\cos a)0 \quad \text{[Applying the results (i) and (ii)]} \\ &= \sin a\end{aligned}$$

Similarly, to prove that $\lim_{h \to 0} \cos(a + h) = \cos a$, *we use the trigonometric identity:* $\cos(a + h) = \cos\,a \cdot \cos\,h - \sin\,a \cdot \sin\,h$ and conclude the result, $\lim_{x \to 0} \cos x = \cos a$.

[6] For this proof refer to *Calculus with Analytic Geometry* (Alternate Edition) by Robert Ellis and Denny Gulick, HBJ Publishers.

[7] In fact, these results tell us that both sine and cosine functions are continuous at any point "a" in their domain (see Section 8.2.1, Statement (3) of continuity of a function in Chapter 8). It is important to remember that continuity of a function at a point is a higher concept than the existence of the limit at that point.

Remark: Note that the proofs of the results at (1) depend on the limits (i) and (ii).

The other trigonometric functions have similar properties, as can be verified from (1), by using the limit rules. For example,

$$\lim_{x\to a} \tan x = \lim_{x\to a} \frac{\sin x}{\cos x} = \frac{\lim_{x\to a} \sin x}{\lim_{x\to a} \cos x} = \frac{\sin a}{\cos a} = \tan a$$

for any number a in the domain of the tangent function. In particular, $\lim_{x\to 0} \tan x = 0$.

A very useful trigonometric limit is the result (iii), that is,

$$\lim_{x\to a} \frac{\sin x}{x} = 1, \quad (x \text{ in radians})$$

This result is proved using the *sandwich theorem* (also called *squeezing theorem*). We have already stated and proved this theorem in Chapter 7a. Here, we give the geometrical view of this theorem.

11a.2.1 Geometrical View of Squeezing Theorem (the Sandwich Theorem)

The squeezing theorem says (in effect) that if the graphs of *f* and *h* converge at a point *P* in the plane and if the graph of *g* is "squeezed" (or sandwiched) between the graphs of *g* and *h*, then the graph of *g* converges with the (graphs of) *f* and *h* at P (Figure 11a.3).

Theorem: Show that $\lim_{x\to 0} \sin \frac{x}{x} = 1, \quad (x \text{ in radians})$.

Proof: Consider a unit circle with center "O", placed at the origin, and let the radian measure of angle AOC be *x* radians (Figure 11a.4).

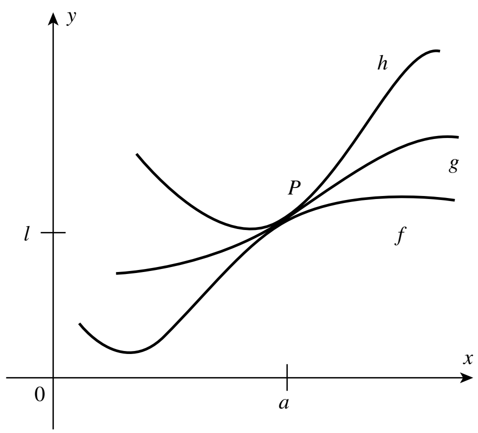


FIGURE 11a.3 Geometrical View of the Sandwich Theorem.

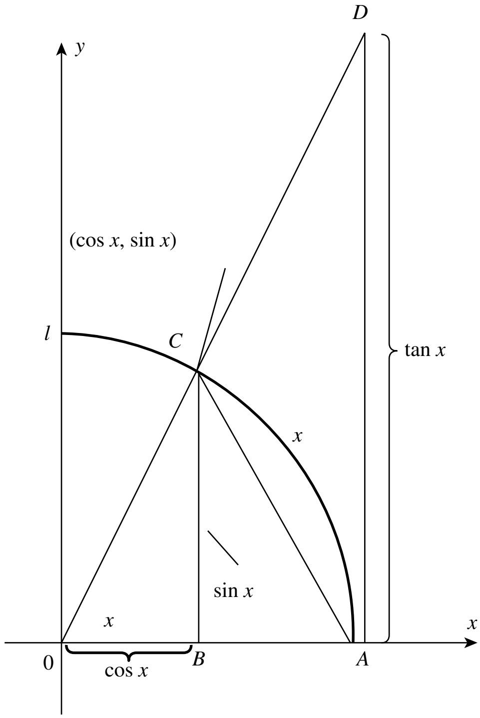


FIGURE 11a.4 Applying Geometric Considerations and Sandwich Theorem.

Using Figure 11a.4, we obtain the following equations, which are valid for $0 < x < \dfrac{\pi}{2}$

$$\text{Area of triangle OAC} = \frac{1}{2}|OA||BC| = \frac{1}{2}1 \cdot \sin x = \frac{\sin x}{2}$$

$$\text{Area of sector OAC} = \frac{x}{2\pi}(\text{Area of circle}) = \frac{x}{2}\,^{(8)}$$

$$\text{Area of triangle OAD} = \frac{1}{2}|OA||AD| = \frac{1}{2}1 \cdot \tan x = \frac{1}{2}\frac{\sin x}{\cos x}$$

It is *geometrically clear* that

Area of ΔOAC < area of sector OAC < area of ΔOAD, so that

$$\frac{\sin x}{2} \le \frac{x}{2} \le \frac{1}{2}\frac{\sin x}{\cos x}$$

(8) $\dfrac{\text{Area Sector OAC}}{\text{Area of Circle}} = \dfrac{x}{2\pi}$ $\therefore$ Area of sector OAC $= \dfrac{x}{2\pi}\pi = \dfrac{x}{2}$. Note that the area of the unit circle $= \pi(1)^2 = \pi$.

Separately, the first and second inequalities yield

$$\frac{\sin x}{x} \le 1 \quad \text{and} \quad \cos x \le \frac{\sin x}{x} \quad (\alpha)$$

Combining the inequalities in (α), we get

$$\cos x \le \frac{\sin x}{x} \le 1, \quad \text{for} \quad 0 < x < \frac{\pi}{2}$$

Furthermore, using the fact that

$$\cos(-x) = \cos x \quad \text{and} \quad \frac{\sin(-x)}{-x} = \frac{-\sin x}{-x} = \frac{\sin x}{x}$$

we obtain

$$\cos x \le \frac{\sin x}{x} \le 1, \quad \text{for } 0 < |x| < \frac{\pi}{2}$$

But $\lim\limits_{x\to 0} \cos x = 1$, and $\lim\limits_{x\to 0} 1 = 1$, it follows from the sandwich theorem that $\lim\limits_{x\to 0} \dfrac{\sin x}{x} = 1$ Proved.

Remark: We emphasize that the result $\lim\limits_{x\to 0} \sin \frac{x}{x} = 1$ is *valid only if the angle x in sinx is expressed in radians*. In case, angle x in sin$x$ is expressed in degrees, then the limit in question does not hold. *Let us see why*?

If "x" in sin$x$ is in degrees, then the limit to be evaluated is $\lim\limits_{x\to 0} \sin x^\circ/x$, where $x^\circ = (\pi x/180)^c$radians. Note that the *degree measure of an angle is a linear function of the radian measure x.*

Note (2): To evaluate $\lim\limits_{x\to 0} \sin x^\circ/x$, we must express the numerator $\sin x^\circ$ *as a function of real variable x*. Hence, we replace x° by the number $\pi x/180$ and then adjust the denominator suitably so that we can apply the result $\lim\limits_{t\to 0} \sin \dfrac{t}{t} = 1$.

Let us evaluate this limit.

We have

$$\lim_{x\to 0} \frac{\sin x^\circ}{x} = \lim_{x\to 0} \frac{\sin(\pi x/180)}{x}$$

$$\therefore \qquad = \lim_{x\to 0} \frac{\sin(\pi x/180)(\pi x/180)}{(\pi x/180)}$$

$$= 1 \frac{\pi}{180} \quad \left\{ \because \text{as } x \to 0, \frac{\pi x}{180} \to 0 \right\}$$

$$= \frac{\pi}{180} \ne 1$$

Now, using the result $\lim\limits_{x\to 0} \sin \dfrac{x}{x} = 1$, we can easily prove the result $\lim\limits_{x\to 0} \dfrac{x}{(\cos x - 1)} = 0.$

11a.2.2 Important Observation

To evaluate the limit, $\lim_{x\to 0} \dfrac{x}{(\cos x - 1)}$, notice that $\lim_{x\to 0} x = 0$, *so we cannot apply the quotient rule (for limits) directly*. However, we can evaluate this limit through the following *mathematical manipulations*:

$$\frac{\cos x - 1}{x} = \left(\frac{\cos x - 1}{x}\right)\left(\frac{\cos x + 1}{\cos x + 1}\right) = \frac{\cos^2 x - 1}{x(\cos x + 1)}$$

$$= \frac{-\sin^2 x}{x(\cos x + 1)} = \left(\frac{\sin x}{x}\right)\left(\frac{-\sin x}{\cos x + 1}\right)$$

$$\therefore \quad \lim_{x\to 0} \frac{\cos x - 1}{x} = \left[\lim_{x\to 0}\left(\frac{\sin x}{x}\right)\right]\left[\lim_{x\to 0}\frac{-\sin x}{\cos x + 1}\right]^{(9)}$$

$$= (1)\left(\frac{0}{1+1}\right) = 0$$

11a.3 DERIVATIVES OF TRIGONOMETRIC FUNCTIONS

By using the basic trigonometric limits (listed in the beginning at (i)–(iv)) and applying the definition of the derivative, we can compute the derivatives of all basic trigonometric functions.

11a.3.1 The Derivatives of sin *x* and cos *x* (From the First Principle)

To find the derivative of $f(x) = \sin x$, using *the definition of the derivative*.

We have,

$$\frac{d}{dx} f(x) = \lim_{\delta x\to 0} \frac{f(x+\delta x) - f(x)}{\delta x}$$

provided the limit on the RHS exists.

$$\therefore \frac{d}{dx} \sin x = \lim_{\delta x\to 0} \frac{\sin(x+\delta x) - \sin x}{\delta x}$$

$$= \lim_{\delta x\to 0} \frac{(\sin x \cos \delta x + \cos x \sin \delta x) - \sin x}{\delta x}$$

$$[\because \sin(x+y) = \sin x \cos y + \cos x \sin y]$$

(9) To compute the limit of a function, which is in the form of a ratio, some trick like algebraic manipulation or the use of some algebraic/trigonometric identity is almost always needed to eradicate the troublesome denominator.

$$= \lim_{\delta x \to 0} \left[\frac{\sin x(\cos \delta x - 1)}{\delta x} + \frac{\cos x \sin \delta x}{\delta x}\right]$$

$$= \lim_{\delta x \to 0} \left[\frac{\sin x(\cos \delta x - 1)}{\delta x} + \frac{\cos x \sin \delta x}{\delta x}\right]$$

$$= \sin x\left[\lim_{\delta x \to 0} \frac{\cos \delta x - 1}{\delta x}\right] + \cos x \lim_{\delta x \to 0} \frac{\sin \delta x}{\delta x}$$

$$= (\sin x)0 + (\cos x)1 \quad \left[\because \lim_{\delta x \to 0} \frac{\cos \delta x - 1}{\delta x} = 0\right]$$

$$= \cos x, \quad \text{for all } x$$

Thus,

$$\frac{\mathrm{d}}{\mathrm{d}x} \sin x = \cos x, \quad \text{for all } x \tag{1a}$$

We can also prove the above result by using the *trigonometric identity*, $\sin 2x = 2 \sin x \cos x$, as follows:

To prove

$$\frac{\mathrm{d}}{\mathrm{d}x} \sin x = \cos x, \quad \text{for all } x$$

Proof: Let $f(x) = \sin x$

$$\therefore \; f(x + \delta x) = \sin(x + \delta x)$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) = \lim_{\delta x \to 0} f(x + \delta x) - f(x) \delta x$$

$$\therefore \; \frac{\mathrm{d}}{\mathrm{d}x}(\sin x) = \lim_{\delta x \to 0} \frac{\sin(x + \delta x) - \sin x}{\delta x}$$

$$= \lim_{\delta x \to 0} \frac{\sin(x + \delta x) + \sin(-x)}{\delta x} \quad [\because -\sin x = \sin(-x)]$$

$$\therefore \; \frac{\mathrm{d}}{\mathrm{d}x}(\sin x) = \lim_{\delta x \to 0} \frac{\left[2\sin\left(\frac{x + \delta x - x}{2}\right) \cos\left(\frac{x + \delta x + x}{2}\right)\right]^{(10)}}{\delta x}$$

(10) Using the trigonometric identity, $\sin A + \sin B = 2 \sin \frac{(A+B)}{2} \cos \frac{(A-B)}{2}$.

$$\left[\because \quad \sin A + \sin B = 2\sin\left(\frac{A+B}{2}\right)\cos\left(\frac{A-B}{2}\right)\right]$$

$$= \left[\lim_{\delta x \to 0} \frac{2\sin\left(\frac{\delta x}{2}\right)\cos\left(\frac{2x+\delta x}{2}\right)}{\delta x}\right]$$

$$= \left[\lim_{\delta x \to 0} \frac{\sin(\delta x/2)}{\delta x/2}\right] \lim_{\delta x \to 0} \cos\left(\frac{2x+\delta x}{2}\right) = 1\cos x$$

$$\therefore \quad \frac{\mathrm{d}}{\mathrm{d}x}(\sin x) = \cos x$$

Similarly, we can show that

$$\frac{\mathrm{d}}{\mathrm{d}x}(\cos x) = -\sin x \tag{2a}$$

Note: It is *convenient* to use the symbol "h" instead of the composite symbol δx.

11a.3.2 Derivative of tan x (from the First Principle)

To prove $\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)(\tan x) = \sec^2 x$.

Proof: Let $f(x) = \tan x$

$$f(x+h) = \tan(x+h)$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(\tan x) = \lim_{h \to 0} \frac{\tan(x+h) - \tan x}{h}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(\tan x) = \lim_{h \to 0} \frac{1}{h}\left[\frac{\sin(x+h)}{\cos(x+h)} - \frac{\sin x}{\cos x}\right]$$

$$= \lim_{h \to 0} \frac{1}{h}\left[\frac{\sin(x+h)\cos x - \cos(x+h)\sin x}{\cos(x+h)\cos x}\right]$$

$$= \lim_{h \to 0} \frac{1}{h}\left[\frac{\sin(x+h-x)}{\cos(x+h)\cos x}\right]$$

$$= \lim_{h \to 0} \frac{1}{h}\left[\frac{\sin h}{\cos(x+h)\cos x}\right]$$

$$= \lim_{h\to 0} \frac{\sin h}{h} \cdot \lim_{h\to 0} \frac{1}{\cos(x+h)\cos x}$$

$$= (1)\frac{1}{\cos x \cos x} = \frac{1}{\cos^2 x} = \sec^2 x$$

$$\frac{d}{dx}(\tan x) = \sec^2 x \tag{3a}$$

Similarly, *we can prove the following results*:

$$\frac{d}{dx}(\cot x) = -\text{cosec}^2\, x \tag{4a}$$

$$\frac{d}{dx}(\sec x) = \sec x \tan x \tag{5a}$$

$$\frac{d}{dx}(\text{cosec}\, x) = \text{cosec}\, x \cot x \tag{6a}$$

11a.3.3 Alternative Simpler Methods (for Finding Derivatives of Basic Trigonometric Functions)

Here, we use *formal rules of differentiation, trigonometric identities, and derivatives of sin x and cos x* (i.e., $\left(\frac{d}{dx}\right)(\sin x) = \cos x$, $\left(\frac{d}{dx}\right)(\cos x) = -\sin x$), which we have obtained by applying the definition of the derivative.

To prove $\left(\frac{d}{dx}\right)(\cos x) = -\sin x$.

We know that,

$$\cos x = \sin\left(x + \frac{\pi}{2}\right)$$

$$\therefore \quad \frac{d}{dx}(\cos x) = \frac{d}{dx}\sin\left(x+\frac{\pi}{2}\right) = \cos\left(x+\frac{\pi}{2}\right) \quad \left[\because \frac{d}{dt}(\sin t) = \cos t\right]$$

$$\therefore \quad \frac{d}{dx}(\cos x) = -\sin x \quad \left[\because \cos\left(x+\frac{\pi}{2}\right) = -\sin x\right]$$

Therefore,

$$\frac{d}{dx}(\cos x) = -\sin x$$

Similarly, we can show that

$$\frac{d}{dx}(\sin x) = \cos x$$

11a.3.4 Derivatives of tan *x*, cot *x*, sec *x*, and cosec *x* (Alternative Simpler Methods)

These trigonometric functions are *quotients involving only sin x and cos x, so their derivatives can be found using the quotient rule for differentiation.*

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}(\tan x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\sin x}{\cos x}\right) &= \frac{\cos x\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)(\sin x) - \sin x\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)\cos x}{\cos^2 x}\\ &= \frac{(\cos x)(\cos x) - (\sin x)(-\sin x)}{\cos^2 x}\\ &= \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x\end{aligned}$$

In exactly the same way we can show that,

$$\frac{\mathrm{d}}{\mathrm{d}x}(\cot x) = -\operatorname{cosec}^2 x$$

Note: *The formulas for the derivative of a quotient becomes more concise when the quotient is of the form* $1/g(x)$ for all x.

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\text{Numerator (Nr)}}{\text{Denominator (Dr)}}\right) = \frac{\mathrm{Dr}\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)\mathrm{Nr} - \mathrm{Nr}\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)\mathrm{Dr}}{(\mathrm{Dr})^2}.$$

When the Numerator = 1, the RHS reduces to $-\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)(\mathrm{Dr})/(Dr)^2$.
In this case, the formulas is

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{g(x)}\right) = \frac{-\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)g(x)}{[g(x)]^2} \tag{7}$$

Example (1): Show that $\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)(\sec x) = \sec x \tan x$.

Solution: From the formulas (7), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}x}(\sec x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{\cos x}\right) = \frac{-(-\sin x)}{\cos^2 x} = \frac{1}{\cos x}\frac{\sin x}{\cos x} = \sec x \tan x$$

Similarly,

$$\mathrm{d}(\operatorname{cosec} x) = -\operatorname{cosec} x \cot x$$

11a.3.5 A Question For Consideration

Now the *next question is*: How can we find the derivative of $\sin x^3$, or in general, that of $\sin u$, where *u is a differentiable function of x?*

(To find the derivatives of such functions, we apply the *chain rule*.)

We have

$$\frac{d}{dx}(\sin u) = \frac{d}{du}(\sin u)\frac{du}{dx} = \cos u \,\frac{du}{dx}$$

Thus, for the function $y = \sin(x^3)$, we have

$$\frac{dy}{dx} = \cos(x^3)\frac{d}{dx}(x^3) = 3x^2 \cos x^3$$

Similarly, for the function $y = \cos(2x^5)$, we may put $2x^5 = u$, so that we have

$$y = \cos u$$

$$\frac{dy}{dx} = \frac{d}{du}(\cos u)\frac{du}{dx}$$

$$\frac{dy}{dx} = -\sin u\frac{du}{dx}$$

$$= -\sin(2x^5)10x^4, \left[\text{because} \frac{du}{dx} = 10x^4\right]$$

$$= -10x^4 \sin(2x^5) = -10x^4 \sin(2x^5)$$

We list below for convenience the *formulas for derivatives of basic trigonometric functions, proved above, in three sets.*

Set (1)

1(a) $\frac{d}{dx}(\sin x) = \cos x$

2(a) $\frac{d}{dx}(\cos x) = -\sin x$

3(a) $\frac{d}{dx}(\tan x) = \sec^2 x$

4(a) $\frac{d}{dx}(\cot x) = -\operatorname{cosec}^2 x$

5(a) $\frac{d}{dx}(\sec x) = \sec x \tan x$

6(a) $\frac{d}{dx}(\operatorname{cosec} x) = -\operatorname{cosec} x \cot x$

Corresponding to the formulas for derivatives of basic trigonometric functions, we list their chain rule formulas.

Set (2)

1(b) $\frac{d}{dx}(\sin u) = \cos u \frac{du}{dx}$

2(b) $\frac{d}{dx}(\cos u) = -\sin u \frac{du}{dx}$

3(b) $\frac{d}{dx}(\tan u) = \sec^2 u \frac{du}{dx}$

4(b) $\frac{d}{dx}(\cot u) = -\text{cosec}^2 u \frac{du}{dx}$

5(b) $\frac{d}{dx}(\sec u) = \sec u \tan u \frac{du}{dx}$

6(b) $\frac{d}{dx}(\text{cosec}\, u) = -\text{cosec}\, u \cot u \frac{du}{dx}$

Set (3)

In the trigonometric functions sin x, cos x, and so on, *if x is replaced by the linear function $(ax+b)$* then we have the following *standard results*, known as the derivatives of *extended forms of basic trigonometric functions.*

1(c) $\frac{d}{dx}\sin(ax+b) = a\cos(ax+b)$

2(c) $\frac{d}{dx}\cos(ax+b) = -a\sin(ax+b)$

3(c) $\frac{d}{dx}\tan(ax+b) = a\sec^2(ax+b)$

4(c) $\frac{d}{dx}\cot(ax+b) = -a\,\text{cosec}^2(ax+b)$

5(c) $\frac{d}{dx}\sec(ax+b) = a\sec(ax+b)\tan(ax+b)$

6(c) $\frac{d}{dx}\text{cosec}(ax+b) = -a\,\text{cosec}(ax+b)\cot(ax+b)$

Note (5): The functions cos x, cot x, and cosec x (starting with "co") are called *cofunctions* of sin x, tan x, and sec x, respectively. Note that the *derivatives of cofunctions are with negative sign.*

Note (6): Importance of the Radian Measure

The radian measure of an angle is convenient for calculus on trigonometric functions. We know that $\left(\frac{d}{dx}\right)(\sin x) = \cos x$, provided "$x$", represents a real variable (or equivalently, angle x is expressed in radians).

On the other hand, if angle x is expressed in degrees and we have to compute the derivative of $\sin x^\circ$, then we proceed as follows:

$$\therefore \quad \frac{d}{dx}\sin x^\circ = \frac{d}{dx}\sin\left(\frac{\pi}{180}x\right) \quad \left[\text{since } x^\circ = \frac{\pi x}{180}\text{radians}\right]$$

$$= \cos\left(\frac{\pi}{180}x\right)\frac{d}{dx}\left(\frac{\pi}{180}x\right)$$

$$= (\cos x^{\circ}) \frac{\pi}{180}$$
$$= \frac{\pi}{180} \cos x^{\circ}$$

Remark: *Observe that if we use degree measure, we would have the factor π/180 in our differentiation formulas for basic trigonometric functions.* On the other hand, *if the angles are expressed in radians, then derivatives of trigonometric functions are in their simplest form. It is for this reason that radian measure is considered convenient* for calculus.

11a.3.6 More Uses of Basic Trigonometric Limits

The limits $\lim_{x \to 0} (\sin x/x) = 1$ and $\lim_{x \to 0} (\cos x - 1/x) = 0$ are mainly useful to prove the derivative formulas. However, we can also use them for evaluating other trigonometric limits. Some applications of the result $\lim_{x \to 0} (\sin x/x) = 1$ are given below through examples.

Example (2): To evaluate $\lim_{x \to 0} (\sin x/x^{2/3})$.

Solution: We rewrite $(\sin x)/x^{2/3}$ as follows:

$$\frac{\sin x}{x^{2/3}} = \left(\frac{\sin x}{x}\right) x\, x^{-2/3} = \left(\frac{\sin x}{x}\right) x^{1/3}$$

Since $\lim_{x \to 0} x^{1/3} = 0$, *it follows from product rule (for limits) that*

$$\lim_{x \to 0} \frac{\sin x}{x^{2/3}} = \lim_{x \to 0} \left[\left(\frac{\sin x}{x}\right) x^{1/3}\right]$$

$$= \lim_{x \to 0} \left(\frac{\sin x}{x}\right) \lim_{x \to 0} x^{1/3} = 1.0 = 0$$

Example (3): To evaluate $\lim_{x \to 0} (\sin 5x)/x$.

Solution: Because of the appearance of $5x$ in the numerator, we write

$$\frac{\sin 5x}{x} = 5\left(\frac{\sin 5x}{5x}\right) \tag{8}$$

Furthermore, notice that as $x \to 0$, $5x \to 0$. Now, if we put $5x = y$, we can write

$$\therefore \quad \lim_{x \to 0} \frac{\sin 5x}{x} = \lim_{5x \to 0} \left(\frac{\sin 5x}{5x}\right) 5$$
$$= \lim_{y \to 0} \left(\frac{\sin y}{y}\right) 5$$
$$= 5 \lim_{y \to 0} \frac{\sin y}{y} = (5)(1) = 5$$

Remark: In evaluating the above limit, we have used "*the substitution rule*". (We have already introduced this rule in Chapter 7. However, it is useful to repeat it again.)

11a.3.7 The Substitution Rule

Using the *limit rules* and the *sandwich theorem*, we can evaluate limits of *rational functions* and a variety of *trigonometric functions*. But, as yet we have no convenient method for evaluating limits such as

$$\lim_{x\to 1} \sqrt{(2x^3 + x^2 - 5x + 8)}\,^{(11)}$$

To evaluate this limit, suppose we first let $y = 2x^3 + x^2 - 5x + 8$ and notice that as $x \to 1$, y approaches $2(1)^3 + (1)^2 - 5(1) + 8 = 6$. *It is then suggesting that if we substitute y for $2x^3 + x^2 - 5x + 8$, and substitute $y \to 6$ for $x \to 1$,* then we can write

$$\lim_{x\to 1} \sqrt{(2x^3 + x^2 - 5x + 8)} = \lim_{y\to 6} \sqrt{y}$$

Since $\lim_{y\to 6} \sqrt{y} = \sqrt{6}$, it would follow that

$$\lim_{x\to 1} \sqrt{(2x^3 + x^2 - 5x + 8)} = \sqrt{6}$$

More generally, if $\lim_{x\to a} f(x) = c$ and if $\lim_{y\to c} g(y)$ *exists,* then we have the following result, known as the *substitution rule.*

$$\lim_{x\to a} g(f(x)) = \lim_{y\to c} g(y)^{(12)} \tag{9}$$

To find $\lim_{x\to a} g(f(x))$, by using the substitution rule (9), we approach as follows:

(i) We substitute y for $f(x)$
(ii) find $c = \lim_{x\to a} y$, and then
(iii) compute $\lim_{y\to c} g(y)$

The process is straightforward. Let us consider the following example:

Example (4): To evaluate $\lim_{x\to \pi/3} \cos(x + \pi/6)$.

Solution: Let $y = x + \pi/6$ and notice that

$$\lim_{x\to \pi/3} y = \lim_{x\to \pi/3} \left(x + \frac{\pi}{6}\right) = \frac{\pi}{3} + \frac{\pi}{6} = \frac{\pi}{2}$$

(11) Surely, resorting to ε's and δ's has no appeal in this case because the proof will be very tedious.
(12) *Calculus with Analytic Geometry* (Alternate Edition) by Robert Ellis and Denny Gulick (p. 78), HBJ Publishers.

Then, by substitution rule

$$\lim_{x \to \pi/3} \cos\left(x + \frac{\pi}{6}\right) = \lim_{y \to \pi/2} \cos y = \cos\frac{\pi}{2} = 0$$

Remark: The substitution rule tells us that we can write

$$\lim_{x \to a} g(f(x)) = g\left(\lim_{x \to a} f(x)\right)$$

provided the limit on RHS exists. (See Chapter 7, theorem (A), rule (8).)

11b Methods of Computing Limits of Trigonometric Functions

11b.1 INTRODUCTION

To evaluate limits of the functions involving trigonometric functions, the following points must be remembered:

1. It is assumed that all the trigonometric functions are defined for real variable x. Thus, x in $\sin x$, $\cos x$, $\tan x$, and so on stands for a real variable (or equivalently the angles contained in the trigonometric functions are expressed in radians).

Note (1): In any problem, if the degree measure of an angle is given, then it must be converted into radian measure using the relation, $1^\circ = \pi/180$ radians.

2. We shall apply the following *basic trigonometric limits*, proved in Chapter 11a:

   (i) $\lim_{x \to 0} \sin x = \sin 0 = 0$ **(ii)** $\lim_{x \to 0} \cos x = \cos 0 = 1$

   (iii) $\lim_{x \to 0} \frac{\sin x}{x} = 1$ **(iv)** $\lim_{x \to 0} \frac{\cos x - 1}{x} = 0$

   These are all *standard limits* and *can be directly used for evaluating the required limits.*

3. Using the standard limits given at (2) above, we can easily prove the following results:

Corollary (i)

$$\lim_{x \to 0} \frac{x}{\sin x} = 1$$

Corollary (ii)

$$\lim_{x \to 0} \frac{\tan x}{x} = 1$$

Corollary (iii)

$$\lim_{x \to 0} \frac{x}{\tan x} = 1$$

11b-Methods of computing limits of trigonometric functions using basic trigonometric limits, sandwich theorem, trigonometric identities, and algebraic manipulations methods

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

Corollary (iv)

$$\lim_{x \to 0} \frac{1 - \cos x}{x^2} = \frac{1}{2}^{(1)}$$

Corollary (v)

$$\lim_{x \to 0} \frac{x^2}{1 - \cos x} = 2$$

Remark: Observe that

$$\lim_{x \to 0} \frac{1 - \cos 5x}{x^2} = \lim_{x \to 0} \frac{1 - \cos 5x}{(5x)^2} 25 = \frac{25}{2}$$

Note (2): While the *standard limits* given in (2) above can be used freely for evaluating the required limits, the five corollaries listed above should not be used freely, since they are *not considered* standard limits.

4. We know that $\tan(\pi/2)$ is *not* defined because

(i) as $x \to \pi/2$, from left, $\tan x \to \infty$, while
(ii) as $x \to \pi/2$, from right, $\tan x \to -\infty$.
(iii) $\therefore \lim_{x \to \pi/2} \tan x$ *does not exist.*

Here, the important point to be remembered is that for the function $f(x) = \tan x$, *neither does the value $f(\pi/2)$ exist nor does the* $\lim_{x \to \pi/2} f(x)$ exist. In fact, $\lim_{x \to a} \tan x = \tan a$ provided "a" is *not an odd multiple of* $\pi/2$.

Similarly, we have

- $\lim_{x \to a} \sec x = \sec a$, provided $\sec a$ is defined.
- $\lim_{x \to a} \operatorname{cosec} x = \operatorname{cosec} a$, provided $\operatorname{cosec} a$ is defined.
- $\lim_{x \to a} \cot x = \cot a$, provided $\cot a$ is defined.

Proposition: If $f(x)$ is a *bounded function*, and if $\lim_{x \to a} g(x) = 0$.

$$\text{Then, } \lim_{x \to a} f(x)g(x) = 0.$$

Proof: Since $f(x)$ is a bounded function,

$f(x)$ has both *the lower and the upper bounds.*

Suppose, l is the lower bound, and u is the upper bound of $f(x)$

$$\therefore l \le f(x) \le u, \text{ for every } x \in \text{domain of } f.$$

(1) To prove $\lim_{x \to 0} (1 - \cos x)/x^2 = 1/2$, consider $(1 - \cos x)/x^2 = (1 - \cos x^2)/(x^2(1 + \cos x)) = (\sin^2 x)/(x^2(1 + \cos x))$

$$\therefore \lim_{x \to 0} \left[\frac{1 - \cos x}{x^2} = \lim_{x \to 0} \left(\frac{\sin x}{x}\right)^2 \frac{1}{1 + \cos x}\right] = (1)^2 \frac{1}{(1+1)} = \frac{1}{2}$$

Multiplying by $g(x)$, throughout. Then, for every x such that $g(x) \geq 0$, we have $l \cdot g(x) \leq f(x)g(x)u \cdot g(x)$, and for every x such that $g(x) \leq 0$, we have $l \cdot g(x) \geq f(x)g(x) \geq u \cdot g(x)$.

Thus, in any case, the product $f(x) \cdot g(x)$ lies in between $l \cdot g(x)$ and $u \cdot g(x)$. But, $\lim_{x \to a} l \cdot g(x) = l \cdot \lim_{x \to a} g(x) = l \cdot 0 = 0$ and $\lim_{x \to a} u \cdot g(x) = u \cdot \lim_{x \to a} g(x) = u \cdot 0 = 0$.

$\therefore$ By the *sandwich theorem*

$$\lim_{x \to a} f(x) \cdot g(x) = 0.$$

Remark (1): In the above proposition, the condition that "$f(x)$ is bounded" is *necessary. The following example justifies this remark.*

Example (1): Let $f(x) = 1/x$ and $g(x) = x$.

(Note that $f(x)$ is *not bounded* for $x \to 0$.)
Now,

$$\lim_{x \to 0} [f(x)g(x)] = \lim_{x \to 0} \left(\frac{1}{x} x\right)$$

$$= \lim_{x \to 0} 1 = 1 \ (\because x \neq 0)$$

Thus,

$$\lim_{x \to 0} f(x)g(x) = 1 \neq 0, \quad \text{through } \lim_{x \to 0} g(x) = 0$$

Therefore, the above proposition *may not hold* if $f(x)$ is *not bounded.*

Remark (2): The *above proposition remains valid if we replace* "$x \to a$" by "$x \to \infty$".

Example (2): Evaluate $\lim_{x \to \infty} (\sin x)/x = \lim_{x \to \infty} [(1/x)\sin x]$.

Solution: Note that $-1 \leq \sin x \leq 1, \quad$ for all x

$\therefore$ $\sin x$ is a *bounded function.*

$$\text{Also, } \lim_{x \to \infty} 1/x = 0.$$

$$\therefore \lim_{x \to \infty} \frac{1}{x} \sin x = 0$$

$$\therefore \lim_{x \to \infty} \frac{\sin x}{x} = 0 \quad \text{(by remark (1))}$$

Example (3): Evaluate $\lim_{x\to 0} \sin(1/x)/1/x$.

Solution: $\lim_{x\to 0} \sin(1/x)/1/x = \lim_{x\to 0} x \cdot \sin 1/x$.

We know that $-1 \le \sin\frac{1}{x} \le 1, \quad$ for all x.

$\therefore$ $\sin 1/x$ is a *bounded function.*

$$\text{Next, } \lim_{x\to 0} x = 0.$$

$$\therefore \lim_{x\to 0}\left[x \sin\frac{1}{x}\right] = 0$$

$$\therefore \lim_{x\to 0}\frac{\sin\frac{1}{x}}{\frac{1}{x}} = 0 \quad \text{(by remark (1))}$$

Note (3): We shall be dealing with two types of limits of trigonometric functions:

1. Limits of the type: $\lim_{x\to 0} f(x)$, (type (I)).
2. Limits of the type: $\lim_{x\to a} f(x)$, where $a \neq 0$ (type (II)).

In dealing with limits of the type $\lim_{x\to a} f(x)$, where $a \neq 0$, we *first substitute $t = x - a$*, so that as $x \to a$, $t \to 0$. Thus, *we convert the limits of the type (II) into the form of limits of type (I).*

If the given limit is in the form of a ratio and both the numerator and the denominator are trigonometric functions, then *it is possible to evaluate the limit more easily by canceling a common factor*. Certain *algebraic manipulations and/or use of trigonometric identities may be needed.*

11b.2 LIMITS OF TYPE (I)

Example (4): Evaluate $\lim_{x\to 0} \sin 3x/x$.

Solution: $\lim_{x\to 0}\frac{\sin 3x}{x} = \lim_{x\to 0}\left(\frac{\sin 3x}{3x}\right)3$

Note that as $x \to 0$, $3x \to 0$. If we put $3x = t$, *we get the given limit as*

$$\lim_{t\to 0}\left[\left(\frac{\sin t}{t}\right)3\right] = \lim_{t\to 0}\left(\frac{\sin t}{t}\right)\lim_{t\to 0} 3$$

$$= (1)(3) = 3 \qquad \textbf{Ans.}$$

Note (4): In solving problems, we need not indicate the above substitution. Thus, we can directly write $\lim_{x\to 0}(\sin 3x/3x) = 1$. However, one must remember that *the standard limit is* $\lim_{x\to 0} \sin x/x = 1$, *where x is expressed in radians.*

Example (5): Evaluate $\lim_{x\to 0}(\tan x)/x$.

Solution: $\lim_{x\to 0}\frac{\tan x}{x} = \lim_{x\to 0}\frac{\sin x}{x\cos x}$

$$= \lim_{x\to 0}\left[\frac{\sin x}{x}\frac{1}{\cos x}\right] = \lim_{x\to 0}\left(\frac{\sin x}{x}\right)\left(\frac{1}{\lim_{x\to 0}\cos x}\right)$$

$$= 1 \cdot \frac{1}{1} = 1$$

$$\text{Thus, } \lim_{x\to 0}\frac{\tan x}{x} = 1 \qquad \textbf{Ans.}$$

Remark: This limit is *not* used as a standard limit.

Example (6): Evaluate $\lim_{x\to 0}(x\cos x+\sin x)/(x+\tan x)$.

Solution: Let us denote the above limit by l.

Consider
$$\frac{x\cos x+\sin x}{x+\tan x}=\frac{\cos x+(\sin x/x)}{1+(\tan x/x)}$$
$$=\frac{\cos x+(\sin x/x)}{1+(\sin x/x)(1/\cos x)}$$
$$\therefore\ l=\frac{\lim_{x\to 0}(\cos x+(\sin x/x))}{\lim_{x\to 0}1+\lim_{x\to 0}[(\sin x/x)(1/\cos x)]}=\frac{\cos 0+1}{1+(1)(1/\cos 0)}$$
$$=\frac{1+1}{1+1}=1\quad \textbf{Ans.}$$

Note (5): Observe that *we have not directly used the results* $\lim_{x\to 0}\tan x/x=1$ simply because it is *not considered* a standard limit.

Example (7): Evaluate $\lim_{x\to 0}(\mathrm{cosec}\,2x-\cot 2x)/\sin x$.

Solution: Let us denote the given limit by l.

Consider, $\mathrm{cosec}\,2x-\cot 2x$
$$=\frac{1}{\sin 2x}-\frac{\cos 2x}{\sin 2x}=\frac{1-\cos 2x}{\sin 2x}$$
$$=\frac{1-(1-2\sin^2 x)}{2\sin x\cos x}=\frac{\sin x}{\cos x}$$
$$\therefore\ l=\lim_{x\to 0}\frac{1}{\sin x}\left[\frac{\sin x}{\cos x}\right]=\lim_{x\to 0}\frac{1}{\cos x}=\frac{1}{1}=1\quad \textbf{Ans.}$$

Example (8): Evaluate $\lim_{x\to 0}(\sqrt{2}-\sqrt{1+\cos 2x})/\sin^2 x$.

Solution: Let us denote the above limit by l.

Consider
$$\frac{\sqrt{2}-\sqrt{1+\cos 2x}}{\sin^2 x}\cdot\frac{\sqrt{2}+\sqrt{1+\cos 2x}}{\sqrt{2}+\sqrt{1+\cos 2x}}$$
$$=\frac{2-(1+\cos 2x)}{\sin^2 x(\sqrt{2}+\sqrt{1+\cos 2x})}=\frac{1-\cos 2x}{\sin^2 x(\sqrt{2}+\sqrt{1+\cos 2x})}$$
$$\therefore\ l=\lim_{x\to 0}\frac{2\sin^2 x}{\sin^2 x(\sqrt{2}+\sqrt{1+\cos 2x})}\ [\because \cos 2x=1-2\sin^2 x]$$
$$=\lim_{x\to 0}\frac{2}{\sqrt{2}+\sqrt{1+\cos 2x}},\quad [\text{As } x\to 0,\ \sin x\to 0\quad \therefore \sin x\neq 0]$$
$$=\frac{2}{\sqrt{2}+\sqrt{1+1}}=\frac{2}{2\sqrt{2}}=\frac{1}{\sqrt{2}}=\frac{\sqrt{2}}{2}\quad \textbf{Ans.}$$

Example (9): Method (II)

Evaluate $\lim\limits_{x \to 0} (\sqrt{2} - \sqrt{1 + \cos 2x})/\sin^2 x$.

Let us denote the above limit by l.

Consider $\sqrt{1 + \cos 2x} = \sqrt{1 + 2\cos^2 x - 1} = \sqrt{2} \cos x$.

$$l = \lim_{x \to 0} \frac{\sqrt{2} - \sqrt{2}\cos x}{\sin^2 x}$$

$$= \lim_{x \to 0} \frac{\sqrt{2}(1 - \cos x)}{\sin^2 x} \times \frac{(1 + \cos x)}{(1 + \cos x)}$$

$$= \lim_{x \to 0} \frac{\sqrt{2}(1 - \cos^2 x)}{\sin^2 x(1 + \cos x)}$$

$$= \lim_{x \to 0} \frac{\sqrt{2}\sin^2 x}{\sin^2 x(1 + \cos x)}$$

$$= \lim_{x \to 0} \frac{\sqrt{2}}{1 + \cos x} \quad [\sin x \neq 0]$$

$$= \frac{\sqrt{2}}{1 + 1} = \frac{\sqrt{2}}{2} \quad \textbf{Ans}.$$

Example (10): Evaluate $\lim\limits_{x \to 0} (1 - \cos 4x/x^2) = l$, say.

Solution: Consider, $\dfrac{1 - \cos 4x}{x^2} \, \dfrac{1 + \cos 4x}{1 + \cos 4x}$

$$= \frac{1 - \cos^2 4x}{x^2(1 + \cos 4x)} = \frac{\sin^2 4x}{x^2(1 + \cos 4x)}$$

$$= \frac{\sin^2 4x}{(4x)^2} 16 \frac{1}{(1 + \cos 4x)}$$

$$= \left(\frac{\sin 4x}{4x}\right)^2 16 \frac{1}{(1 + \cos 4x)}$$

$$\therefore l = \lim_{x \to 0} \left[\left(\frac{\sin 4x}{4x}\right)^2 16 \frac{1}{(1 + \cos 4x)}\right]$$

$$= 1^2 \cdot 16 \frac{1}{1 + 1} = 16 \frac{1}{2} = 8 \quad \textbf{Ans}.$$

Note (6): It is useful to prove the following results and compare them.

(i) $\lim_{x\to 0} \frac{1-\cos x}{x} = 0$ **(ii)** $\lim_{x\to 0} \frac{1-\cos x}{x^2} = \frac{1}{2}$

(iii) $\lim_{x\to 0} \frac{1-\cos 2x}{x} = 0$ **(iv)** $\lim_{x\to 0} \frac{1-\cos 5x}{x^2} = \frac{25}{2}$

(v) $\lim_{x\to 0} \frac{1-\cos x}{x^2} = \frac{1}{2}$ **(vi)** $\lim_{x\to 0} \frac{x^2}{1-\cos x} = 2$

Example (11): Evaluate $\lim_{x\to 0}(3\sin x^\circ - \sin 3x^\circ)/x^3 = l$, say.

Solution: Consider $3\sin x^\circ - \sin 3x^\circ$

$$= 3\sin x^\circ - [3\sin x^\circ - 4\sin^3 x^\circ] \quad \because [\sin 3x = 3\sin x - 4\sin^3 x]$$
$$= 4\sin^3 x^\circ$$
$$l = \lim_{x\to 0} \frac{4\sin^3 x^\circ}{x^3}$$
$$= \lim_{x\to 0} 4\left(\frac{\sin x^\circ}{x}\right)^3$$
$$= 4\left[\lim_{x\to 0} \frac{\sin(\pi x/180)(\pi/180)}{\pi x/180}\right]^3$$
$$= 4\left[1\frac{\pi}{180}\right]^3 = 4\left(\frac{\pi}{180}\right)^3 \quad \textbf{Ans.}$$

Example (12): Evaluate $\lim_{x\to 0}(\cos ax - \cos bx)/(\cos cx - \cos dx) = l$, say.[2]

Solution: Consider, $\cos ax - \cos bx = -2\sin\frac{(a+b)x}{2}\sin\frac{(a-b)x}{2}$

$$= \frac{-2\sin((a+b)x/2)\sin((a-b)x/2)}{(a+b)x/2} \; \frac{((a+b)x/2)((a-b)x/2)}{(a-b)x/2}$$

$$\lim_{x\to 0}(\cos ax - \cos bx) = \lim_{x\to 0}\left[-2\frac{(a+b)x}{2}\frac{(a-b)x}{2}\right]$$
$$= \lim_{x\to 0}\left[-\frac{(a^2-b^2)x^2}{2}\right]$$

Similarly, $\lim_{x\to 0}(\cos cx - \cos \mathrm{d}x) = \lim_{x\to 0} - \frac{(c^2-d^2)x^2}{2}$

$$\therefore l = \lim_{x\to 0}\left[\frac{(a^2-b^2)x^2}{(c^2-d^2)x^2}\right]$$
$$= \lim_{x\to 0}\left[\frac{(a^2-b^2)}{(c^2-d^2)}\right] \quad \left[\begin{array}{l}\because x\to 0, x \neq 0 \\ \therefore x^2 \neq 0\end{array}\right]$$
$$= \frac{(a^2-b^2)}{(c^2-d^2)} \quad \textbf{Ans.}$$

[2] Recall that $\cos A + \cos B = 2\cos(A+B/2)\cos(A-B/2)$ and $\cos A - \cos B = -2\sin(A+B/2)\sin(A-B/2)$.

Exercise (1):

(1) $\lim\limits_{x \to 0} \dfrac{\sin ax}{\sin bx}$ **(2)** $\lim\limits_{x \to 0} \dfrac{\sin^2 x}{2x}$

(3) $\lim\limits_{x \to 0} \dfrac{1 - \cos 2x}{x}$ **(4)** $\lim\limits_{x \to 0} \dfrac{x \cos x - \sin x}{x^2 \sin x}$

(5) $\lim\limits_{x \to 0} \dfrac{x^3 \cot x}{1 - \cos 2x}$ **(6)** $\lim\limits_{x \to 0} \dfrac{2 \sin^2 3x}{x^2}$

(7) $\lim\limits_{x \to 0} \dfrac{1 - \cos 5x}{\tan^2 x}$ **(8)** $\lim\limits_{x \to 0} \dfrac{3x^3 - 2x^2 + 1 - \cos 4x}{x \sin x}$

(Hint: Divide numerator and denominator by x^2.)

(9) $\lim\limits_{x \to 0} \dfrac{2 \sin x^\circ - \sin 2x^\circ}{x^3}$ **(10)** $\lim\limits_{x \to 0} \dfrac{\sqrt{1+x} - \sqrt{1-x}}{\sin x}$

(11) $\lim\limits_{x \to \infty} \sin x \tan \dfrac{1}{x}$ $\left[\begin{array}{l} \text{Hint : } x \to \infty, \dfrac{1}{x} \to 0 \\ \therefore \tan \dfrac{1}{x} \to 0. \text{ Also } -1 \le \sin x \le 1 \end{array}\right]$

(12) $\lim\limits_{x \to \infty} \dfrac{\sin\sqrt{x}}{x}$ **(13)** $\lim\limits_{x \to 0} \dfrac{\sin(\alpha + x) + \sin(\alpha - x) - 2 \sin \alpha}{x^2}$

(14) $\lim\limits_{x \to 0} \dfrac{\cos 8x - \cos 2x}{\cos 12x - \cos 4x}$ **(15)** Show that $\lim\limits_{x \to 0} x^2 \cos \dfrac{1}{x} = 0$

(16) Show that $\lim\limits_{x \to 0} x^2 \sin \dfrac{1}{x} = 0$

Answers.

(1) $\dfrac{a}{b}$ **(2)** 0 **(3)** 0 **(4)** $-\dfrac{1}{3}$ **(5)** 2

(6) 18 **(7)** $\dfrac{25}{2}$ **(8)** –16 **(9)** $\left(\dfrac{\pi}{180}\right)^3$ **(10)** 1

(11) 0 **(12)** 0 **(13)** $-\sin \propto$ **(14)** $\dfrac{15}{32}$.

11b.3 LIMITS OF THE TYPE (II) [$\lim\limits_{x \to a} f(x)$, WHERE $a \neq 0$]

Example (13): Evaluate $\lim\limits_{x \to a} (\sin x - \sin a)/(\sqrt{x} - \sqrt{a})$.

Solution: Let the given limit be denoted by l.

Consider $\lim\limits_{x \to a} \dfrac{\sin x - \sin a}{\sqrt{x} - \sqrt{a}} = \lim\limits_{x \to a} \dfrac{\sin x - \sin a}{\sqrt{x} - \sqrt{a}} \times \dfrac{\sqrt{x} + \sqrt{a}}{\sqrt{x} + \sqrt{a}}$[3]

$$= \frac{\sin x - \sin a(\sqrt{x} + \sqrt{a})}{(x - a)}$$

put $t = x - a$. $\therefore$ $x = t + a$

[3] Note (7): The method of rationalization introduced for algebraic functions is also applicable here.

As $x \to a$, $t \to 0$.

$$\therefore\ l = \lim_{t \to 0} \frac{[\sin(t+a) - \sin a][\sqrt{t+a} + \sqrt{a}]}{t}$$

$$= \lim_{t \to 0} \frac{\left[2\cos\left(a + \frac{t}{2}\right)\sin\frac{t}{2}\right][\sqrt{t+a} + \sqrt{a}]}{t}$$

$$= \lim_{t \to 0} \cos\left(a + \frac{t}{2}\right)\left[\frac{\sin t/2}{t/2}\right][\sqrt{t+a} + \sqrt{a}]$$

$$= \cos(a+0)[1][\sqrt{0+a} + \sqrt{a}]$$

$$= 2\sqrt{a}\cos a \quad \textbf{Ans.}$$

Example (14): Evaluate $\lim_{x \to \pi} (\sqrt{2+\cos x} - 1)/(\pi - x)^2 = l$, say.

Solution: Put $x - \pi = t \quad \therefore x = \pi + t$.
Note that $x \to \pi, t \to 0$.

$$l = \lim_{t \to 0} \frac{\sqrt{2+\cos(\pi+t)} - 1}{t^2}$$

$$= \lim_{t \to 0} \frac{\sqrt{2-\cos t} - 1}{t^2}, \quad [\because \cos(\pi+t) = -\cos t]$$

$$= \lim_{t \to 0} \frac{\sqrt{2-\cos t} - 1}{t^2} \frac{\sqrt{2-\cos t} + 1}{\sqrt{2-\cos t} + 1}$$

$$= \lim_{t \to 0} \frac{(2-\cos t) - 1}{t^2(\sqrt{(2-\cos t)} + 1)}$$

$$= \lim_{t \to 0} \frac{1-\cos t}{t^2(\sqrt{(2-\cos t)} + 1)}$$

$$= \lim_{t \to 0} \frac{1-\cos t}{t^2} \lim_{t \to 0} \frac{1}{(\sqrt{(2-\cos t)} + 1)}$$

We have shown earlier that,

$$= \lim_{t \to 0} \frac{1-\cos t}{t^2} = \frac{1}{2}$$

(*We must prove it here again* since it is *not a standard limit. However, we use this result to save time*.)

$$\therefore\ l = \frac{1}{2}\frac{1}{(\sqrt{2-1}+1)} = \frac{1}{2}\frac{1}{2} = \frac{1}{4} \quad \textbf{Ans.}$$

Example (15): Evaluate $\lim_{x\to\pi/4}(1-\tan x)/(1-\sqrt{2}\sin x) = l$, say.

Solution: Consider $\dfrac{1-\tan x}{1-\sqrt{2}\sin x}$

$$= \frac{\cos x-\sin x}{\cos x}\frac{1}{1-\sqrt{2}\sin x}\frac{1+\sqrt{2}\sin x}{1+\sqrt{2}\sin x}$$

$$= \frac{\cos x-\sin x}{\cos x}\frac{1+\sqrt{2}\sin x}{1-2\sin^2 x}$$

$$= \frac{\cos x-\sin x}{\cos x}\frac{1+\sqrt{2}\sin x}{\cos^2 x-\sin^2 x}$$

$$\therefore l = \lim_{x\to\pi/4}\left[\frac{\cos x-\sin x}{\cos x}\frac{1+\sqrt{2}\sin x}{(\cos x-\sin x)(\cos x+\sin x)}\right]$$

$$= \lim_{x\to\pi/4}\left[\frac{1+\sqrt{2}\sin x}{\cos x(\cos x+\sin x)}\right], \quad \left[\begin{array}{l}\because \cos x \neq \cos\frac{\pi}{4} \neq \frac{1}{\sqrt{2}} \\ \sin x \neq \sin\frac{\pi}{4} \neq \frac{1}{\sqrt{2}} \\ \therefore \cos x-\sin x \neq 0\end{array}\right]$$

$$= \frac{1+\sqrt{2}\cdot\frac{1}{\sqrt{2}}}{\frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{2}}\right)} = \frac{1+1}{\frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{2}}} = \frac{2}{1} = 2.$$

Example (16): Evaluate $\lim_{x\to 1}(x^2-3x+2)/[x^2-x+\sin(x-1)] = l$, say.

Solution: Consider x^2-3x+2

$$= x^2-2x-x+2$$
$$= x(x-2)-1(x-2)$$
$$= (x-2)(x-1)$$

$$\therefore l = \lim_{x\to 1}\frac{(x-1)(x-2)}{x(x-1)+\sin(x-1)}, \quad \left(\begin{array}{ll}\text{put } x-1 = t, & \therefore x = t+1 \\ \text{As } x\to 1, & t\to 0\end{array}\right)$$

$$= \lim_{t\to 0}\frac{t(t-1)}{(t+1)t+\sin t}$$

$$= \lim_{t\to 0}\frac{(t-1)}{(t+1)+\sin t/t} \quad [\because t\to 0, \quad \therefore t \neq 0]$$

$$= \frac{0-1}{(0+1)+1} = -\frac{1}{2}$$ **Ans.**

Exercise (2):

Evaluate the following limits (type (II)):

(1) $\lim_{x \to 1} \dfrac{1 + \cos \pi x}{(x-1)^2}$ **(2)** $\lim_{x \to a} \dfrac{\sin x - \sin a}{(x-a)}$

(3) $\lim_{x \to \pi/4} \dfrac{\sin x - \cos x}{(x - \pi/4)}$ **(4)** $\lim_{x \to \pi/2} \dfrac{2x - \pi}{\cos x}$

(5) $\lim_{x \to \pi/2} \dfrac{\sqrt{2} - \sqrt{1 + \sin x}}{\cos^2 x}$ **(6)** $\lim_{x \to \pi} \dfrac{\sqrt{5 + \cos x} - 2}{(\pi - x)^2}$

(7) $\lim_{x \to \pi/4} \dfrac{1 - \tan x}{1 - \sqrt{2} \sin x}$ **(8)** $\lim_{x \to a} \dfrac{\sin(x+a) - \sin(a-x) - 2 \sin a}{x \sin x}$

(9) $\lim_{x \to \pi/6} \dfrac{2 \sin x - 1}{\sqrt{3} \tan x - 1}$ **(10)** $\lim_{x \to a} \dfrac{\cos x - \cos a}{\sqrt{x} - \sqrt{a}}$

Answers:

(1) $\frac{1}{2}\pi^2$ **(2)** $\cos a$ **(3)** $\sqrt{2}$ **(4)** -2 **(5)** $\frac{1}{4\sqrt{2}}$

(6) $\frac{1}{8}$ **(7)** 2 **(8)** $-\sin a$ **(9)** $\frac{3}{4}$ **(10)** $-2\sqrt{2} \sin a$

11b.4 LIMITS OF EXPONENTIAL AND LOGARITHMIC FUNCTIONS

(Basic exponential and logarithmic functions and the related standard limits are discussed in Chapter 13.)

Definition: If $a > 0$, then the function f defined by

$$y = f(x) = a^x$$

is called an exponential function.

Note (8): (i) $a^x = y \Leftrightarrow x = \log_a y$

Thus, we can write

If $a^x = y$, (1)

Then, $\log_a y = x$, (2)

and vice versa.

It is easy to obtain the following results:

$$\log_a a^x = x, \qquad (3)^{(4)}$$

and

$$a^{\log_a y} = y, \qquad (4)^{(5)}$$

(4) This result is obtained if we consider (2) and substitute for y from (1).

(5) This result is obtained if we consider (1) and substitute for x from (2).

(ii) $a^1 = a \quad \therefore \log_a a = 1$
(iii) $a^0 = a \quad \therefore \log_1 a = 0$

$$\text{(i)}\ \log_a xy = \log_a x + \log_a y$$
$$\text{(ii)}\ \log_a \frac{x}{y} = \log_a x - \log_a y$$
$$\text{(iii)}\ \log_a x^m = m \log_a x$$

Rules for Change of Base

$$\log_a x = \frac{\log_b x}{\log_a b} \quad (A)^{(6)}$$

By writing $x = b$, in the above statement, we get

$$\log_a b = \frac{\log_b b}{\log_b a}$$
$$= \frac{1}{\log_b a} \quad (\because \log_b b = 1)$$
$$\therefore \log_a b = \frac{1}{\log_b a}$$

11b.4.1 Common Logarithms and Natural Logarithms

Logarithms to the base e (e ≈ 2.7182) are called natural logarithms (or Napierian logarithms).

(i) When we are considering natural logarithms, the convention is not to write the base e.

In our study of calculus, we are going to use, in general, natural logarithms only. Therefore, we need not write the base when it is to the base e. Thus, log x will mean "$\log_e x$"[7]

Note (9):

$$\text{(ii)}\ \log \mathrm{e} = \log_e \mathrm{e} = 1$$
$$\text{(iii)}\ \log \mathrm{e}^x = \log_e \mathrm{e}^x = x$$
$$\text{(iv)}\ \mathrm{e}^{\log x} = \mathrm{e}^{\log_e y} = y$$

[6] It is very easy to remember this rule. Write the algebraic identity, $x/a = (x/b)/(a/b)$, and it helps to write this rule. One may also write down the identity $x/a = (x/b)(b/a)$, and then write down $\log_a x = \log_b x$, which gives (A).

[7] By not writing the base "*e*" repeatedly we save time and effort. However, one can still write the base "e" for clarity, if and when needed. (At the school level where the base "e" is not introduced, and only base 10 is used, some authors insist that log x should be read to mean $\log_{10} x$.)

We shall assume the following results:

(i) $\lim_{x \to \infty} \left(1 + \frac{1}{x}\right)^x = \mathrm{e}$

(ii) $\lim_{x \to 0} (1 + x)^{1/x} = \mathrm{e}$

(iii) $\lim_{x \to 0} \left(1 + \frac{x}{a}\right)^{a/x} = \mathrm{e}$

(iv) If $f(x) \to 0$, as $x \to 0$, then $\lim_{x \to 0} (1 + kf(x))^{\frac{1}{kf(x)}} = \mathrm{e}$, where $k \neq 0$
It follows that

$$\lim_{x \to 0} (1 - x)^{-1/x} = \mathrm{e}, \quad \text{and} \quad \lim_{x \to 0} \left(1 - \frac{1}{x}\right)^{-x} = \mathrm{e}$$

We also assume the following limit[(8)]:

(i) $\lim_{x \to 0} \frac{a^x - 1}{x} = \log_e a \quad (\mathrm{B})$

(Recall that for proving this result, we first prove $\lim_{x \to 0} \log_a(1 + x)/x = \log_a e = 1/\log_e a$ by the change of base.)

(ii) By replacing a with e in (B), we get

$$\lim_{x \to 0} \frac{e^x - 1}{x} = \log_e \mathrm{e} = 1$$

In particular, $\lim_{x \to 0} \frac{5^x - 1}{x} = \log_e 5$, and $\lim_{t \to 0} \frac{2^t - 1}{t} = \log_e 2$

Let $f(x) \to 0$ as $x \to 0$. If $k \neq 0$, then any number $t = k \cdot f(x) \to 0$ as $x \to 0$.
We have

$$\lim_{x \to 0} \frac{a^{k \cdot f(x)} - 1}{k \cdot f(x)} = \lim_{x \to 0} \frac{a^t - 1}{t} = \log_e a = \log a$$

(8) We have proved these results in Chapter 13.

12 Exponential Form(s) of a Positive Real Number and its Logarithm(s): Pre-Requisite for Understanding Exponential and Logarithmic Functions

12.1 INTRODUCTION

The product $2 \times 2 \times 2 \times 2 \times 2 \times 2 = 64$, is conveniently written in the form $2^6 = 64$, to mean that the number is multiplied by itself, six times. In the expression 2^6, the number "2" is called the *base* and "6" is called the *exponent*. We say that the number 64 is expressed in the *exponential form* as 2^6. Similarly, we can write $4^3 = 64$ and $64^1 = 64$, which are two other exponential forms for 64.

In fact, *any positive number* can be expressed in any number of exponential form(s), by choosing a *positive base* and an appropriate exponent.[(1)]

12.2 CONCEPT OF LOGARITHM

At this stage, we introduce the *concept of logarithm* of a positive real number. If three numbers a, b, and c are so related that

$$a^b = c, \quad (a > 0,\ a \neq 1) \tag{1}$$

then the *exponent* "b" is called the logarithm of "c" to the base "a."

We write

$$\log_a c = b \tag{2}$$

It may be noted that the logarithm of a number can be different for different bases. Detailed discussion about logarithm(s) and their applications will follow later.

What must you know to learn calculus? 12-Logarithms [Exponential form(s) of a positive real number and its logarithm(s)]

[(1)] This statement is true from a mathematical point of view. However, it should not create any fear or confusion in the reader's mind by visualizing the practical difficulties. Later on, it will be clear that our interest lies in only two bases, namely "10" and "e", and tables for exponents are readily available.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

Both (1) and (2) given above express the relations between the three numbers a, b, and c. The relation (1) is in the index form and the relation (2) expresses the same thing in the log arithm (log) form. Let us discuss the role of the conditions $a > 0$ and $a \neq 1$. reflected at (1) above.

(i) By definition, $0^n = 0$, $(n \in N)$.
In general, $0^K = 0$, $(k \in R,\ k \neq 0)$.[(2)]
Note that in the relation $0^k = 0$, the exponent k loses its role and identity due to the base "0." Moreover "0^k" represents the number "0," only. Hence, in order to express a positive number in the exponential form, we cannot consider the number "0" as the base.

(ii) By definition $1^n = 1$ $(n \in N)$.
In general $1^k = 1$ $(k \in R)$
Further, $1^k \div 1^k = 1$ $(k \in R)$[(3)]
Also, $1^k \div 1^k = 1^{k-k} = 1^0$
$\therefore 1^0 = 1$.
Note that in the expression 1^k $(k \in R)$ the exponent k looses its role and identity due to the base "1". Moreover, 1^k always represents the number 1. Hence, in order to express any positive number (other than "1") in the exponential form, we cannot consider the base to be one.

(iii) Now, let us see what happens if the base "a" is taken as a negative number. We know that,

$$(-3)^2 = 9 \text{ and } (3)^2 = 9. \text{ On the other hand, } (-3)^3 = -27 \text{ and } (3)^3 = 27.$$

From the above examples, it is clear that if a negative base is raised to an *even* power we get a positive number, but if it is raised to an odd power, we get a negative number. On the other hand, *if the base is positive, then any power raised to it represents a positive number*. Therefore, to represent a *positive number* in the exponential form, the base is always taken to be a *positive number*, other than 1.

12.3 THE LAWS OF EXPONENT

The laws of exponents are initially defined for *natural numbers* and then extended to *integers and rational numbers*. Let us revise the following definitions and laws of exponents:

(i) $a \times a \times a \ldots\ldots (n \text{ factors}) = a^n$, $(a \in R,\ n \in N)$. In particular, $0^n = 0$.
(ii) $a^{-n} = 1/a^n$, $(a \neq 0)$
(iii) $a^0 = 1$ $(a \neq 0)$
(iv) *The nth root of a positive number* "a". If the exponent of a *positive number* "a" is a rational number of the form $1/n$ $(n \in N)$, then we call it the nth root of "a". Thus, $16^{1/4}$ is called the fourth root of 16 and $125^{1/3}$ is called the third root of 125. The root of a number is also written using a *radical symbol* ($\sqrt{\ }$). An expression for a root is called radical. We write $a^{1/q}$ as $\sqrt[q]{a}$ and read it as the qth root of "a". Here $\sqrt[q]{a}$ is called a radical and q is called the *index* of the radical.

(2) It is assumed that the reader is familiar with the basic laws of exponents, which are used for combining exponents. Further, since the expression 0^0 cannot be assigned any value, we do not define it.
(3) Note that, $a^k \div a^k = 1 = a^{k-k} = a^0$.
$\therefore a^0 = 1 (a \in R,\ a \neq 0)$.
This follows from the laws of exponents, valid for real numbers. Thus, $7^0 = 1$,
$(-5)^0 = 1$, $(5/7)^0 = 1$, $(\sqrt{3})^0 = 1$, and so on.

Definition: The *n*th root of a positive number "*a*" is the *positive number* $\sqrt[n]{a}$ (or $a^{1/n}$), whose *n*th power is "*a*". The above definition tells us that, for $a > 0$,

$$\left(\sqrt[n]{a}\right)^{\text{n}} = ((a^{1/n})^n = a > 0^{(4)}$$

Remark: *A negative number does not have a square root* (since the square of any number is never negative). On the other hand, a positive number has *two square roots*, of which one is positive and the other is negative. For example,

$$5^2 = 25 \text{ and } (-5)^2 = 25, \text{ so that, } (25)^{1/2} = \sqrt{25} = \pm 5.$$

Thus, if n is even, then the *n*th root of a positive number is *not unique*. If we agree to exclude negative value(s) of this *n*th root when n is even, *then nth root of "a" (a > 0) is uniquely defined, whether n is even or odd.* Thus, the *n*th root operation on $a > 0$, *becomes a function, if we discard negative values of* $\sqrt[n]{a}$, whenever n is even.

(v) *Positive Rational Numbers as Exponents*: Let "*a*" be any *positive* real number and p/q be a positive rational number (where p and q both are positive integers).

Note (1): We have already given a meaning to $a^{1/q}$ and now we are in a position to give a meaning to $a^{p/q}$. $a^{p/q}$ is defined as $(a^p)^{1/q}$. Thus, $a^{p/q}$ is defined as *q*th root of a^p. Notice, however, that *if q is an even integer*, then a^p must be positive, so that $(a^p)^{1/q}$ is defined.[5] (This is definitely achieved, if $a > 0$.)

Note (2): We assume that in the expression $a^{p/q}$, the base $a > 0$.

(vi) *Negative Rational Numbers as Exponents*: A negative rational number is generally written in such a form that its *denominator is always positive*. If the denominator is negative, we can multiply both, the numerator and the denominator by "(-1)" and thus make the denominator positive. For example,

$$4^{3/-2} = 4^{-3/2} = (4^{-3})^{1/2} = \left(\frac{1}{4^3}\right)^{1/2} = \frac{1}{(64)^{1/2}} = \frac{1}{8}.$$

Remark: Conventionally, *an exponential expression* is written in such a way that its *exponent is a positive number*.

12.4 LAWS OF EXPONENTS (OR LAWS OF INDICES)

Now, we clearly know the meaning of a^x, where *x is a rational number*. At this stage, *we assume that* a^x *is defined, when x is an irrational number*. Now, we can give the "laws of exponents" (or laws of indices) *valid for real exponents*.

12.4.1 Laws of Exponents (or Laws of Indices) for real exponents

For any real numbers a, b, m, and n, the following laws are valid:

(i) $a^m \times a^n = a^{m+n}$
(ii) $(a^m)^n = a^{m \times n}$

[4] We agree that $\sqrt[n]{0} = 0^{1/\text{n}} = 0$.

[5] We have remarked earlier that the square root of a negative number does not exist. For the same reason, if q is even then, *q*th root of a negative number does not exist.

(iii) $(a \times b)^n = a^n \times b^n$

(iv) $\left(\frac{a^m}{a^n}\right) = a^m \times a^{-n} = a^{m-n}$, if $a \neq 0$.

Remark: If $m = n$, then $m - n = 0$.
It follows that $a^0 = 1$, provided $a \neq 0$.

(v) $\left(\frac{a}{b}\right)^n = \frac{a^n}{b^n}$, if $b \neq 0$.

12.4.2 Applications of the Laws of Exponents

It is interesting to know that the above laws of exponents can be used to multiply and divide any given numbers (however, large or small, they might be) using *addition* and *subtraction*, which are simpler operations. The main ideas of the method were developed and given by John Napier in 1614, as explained below. Let us consider the following two sets, A and B. Set A contains some positive integers which are powers of 2, written in ascending order. Set B consists of corresponding exponents of 2.

12.4.3 Multiplication of Numbers in Set A

Suppose, we have to multiply two numbers in the Set A.

For example, 32×512, we locate the exponents of "2", corresponding to 32 and 512. They are 5 and 9, respectively. Add these exponents, that is, $5 + 9 = 14$. Then we look for the number in the Set A corresponding to the exponent 14. It is 16384.

$$\therefore 32 \times 512 = 16384.$$

Note that, *we have used the operation of addition to calculate the product*. This becomes clear if we look at law (i) above.

$$32 \times 512 = 2^5 \times 2^9 = 2^{5+9} = 2^{14} = 16384$$

Set A	Set B
$2 = 2^1$	1
$4 = 2^2$	2
$8 = 2^3$	3
$16 = 2^4$	4
$32 = 2^5$	5
$64 = 2^6$	6
$128 = 2^7$	7
$256 = 2^8$	8
$512 = 2^9$	9
$1024 = 2^{10}$	10
$2048 = 2^{11}$	11
$4096 = 2^{12}$	12
$8192 = 2^{13}$	13
$16384 = 2^{14}$	14
$32768 = 2^{15}$	15

12.4.4 Division of Numbers in Set A

Now, suppose we wish to divide 8192 by 128. The corresponding exponents of "2" are 13 and 7, respectively. We subtract the exponents, that is, $13 - 7 = 6$. The number in Set A corresponding to exponent 6 is 64.

$$\therefore \frac{8192}{128} = \frac{2^{13}}{2^7} = 2^{13-7} = 2^6 = 64$$

[Note that, we have used law (iv) in the above computation.]

The Sets A and B considered above are quite simple. Similar sets can be designed using other bases such as 3, 4, 5, ..., and so on. Obviously, it will not be convenient to choose the (positive) rational numbers as bases. Recall, that in the statement $a^b = c$, the exponent (or the power) "b" raised to *the positive base* "a" is called *the logarithm of the number* "c".

12.5 TWO IMPORTANT BASES: "10" AND "e"

In the system of logarithms, which we use in our day-to-day calculations (such as those in the field of engineering, etc.), the base 10 is found to be most useful. Logarithms to the base 10 are called *common logarithms*. Once the base "10" is chosen, it has to be raised with a suitable real number "b" (positive, zero, or negative) so that, it represents the given (positive) number c, *exactly or very close to it*. Thus, we write,

$$10^b = c \text{ or } 10^b \approx c$$

where the symbol "≈" stands for "very close to". For example,

$\left.\begin{array}{l}\log_{10}100 = 2.0000\\ \log_{10}1000 = 3.0000\end{array}\right\}$	These values of logarithms are *exact*, since $10^2 = 100$ and $10^3 = 1000$
$\log_{10}5 \approx 0.6990$	These values of logarithms are *not exact*, but they are *very close to the numbers* in question, since $(10)^{0.6990} \approx 5$ and $(10)^{1.4453} \approx 27.8$
$\log_{10}27.8 \approx 1.445$	

Now, the question is, *How do we find these exponents (i.e., logarithms of the given positive numbers) to the base "10"*? For our purpose, the answer is that the logarithms can be found out by using suitable tables.[6]

12.5.1 Notations

In *common logarithms*, the base is always 10, so that, if no base is mentioned, the base 10 is always understood. However, it is useful only while dealing with arithmetical calculations.

[6] Detailed methods, for preparing the tables (of logarithms) are available in many books on algebra and trigonometry. Study of these methods is quite interesting, but here our interest lies in concentrating more on logarithms and their properties.

Important in calculus are logarithms to the base "e", called *natural logarithms* (or Naperian logarithms). The number "e", (which is the base for natural logarithms) is a *typical irrational number*, lying between 2 and 3 ($e \approx 2.71828\ldots$).[7]

The notation for "*natural logarithm*" is "*ln*", but we shall be using $\log_e x$ to mean *ln x*. *Throughout this book, we are going to use natural logarithms only.*

(*Once we get used to it, we will start identifying* log *x to stand for* $\log_e x$.) To avoid, confusion in the notation, whether log *x* should mean $\log_{10} x$ or $\log_e x$, we agree that in dealing with arithmetical calculations it will stand for $\log_{10} x$. On the other hand, *while solving problems in calculus*, it will stand for $\log_e x$. Besides, this notation will be implemented only after a suitable note.

12.6 DEFINITION: LOGARITHM

The logarithm of any number *to a given base*, is equal to the power to which, the base should be raised to get the given number.[8]

We Know That	Therefore, we say That	We Write
$2^6 = 64$	log of 64 to the base 2 = 6	$\log_2 64 = 6$
$4^3 = 64$	log of 64 to the base 4 = 3	$\log_4 64 = 3$
$64^1 = 64$	log of 64 to the base 64 = 1	$\log_{64} 64 = 1$
$5^2 = 25$	log of 25 to the base 5 = 2	$\log_5 25 = 2$
$5^{-3} = 1/125$	log of 1/125 to the base 5 = −3	$\log_5 1/125 = -3$
$a^0 = 1, (a \neq 0)$	log of 1 to the base $a = 0$	$\log_a 1 = 0$
$a^1 = a$	log of a to the base $a = 1$	$\log_a a = 1$

Note (1):

(i) From the first three illustrations, we observe, that *the logarithm of a (positive) number is different for different bases.*

(ii) From the last two illustrations, we get the following two results:

(a) The logarithm of 1 to any base is zero.

(b) The logarithm of any number to the same base (as the number itself) is (i.e., $\log_a a = 1$, $\log_{10} 10 = 1$, $\log_e e = 1$.)

[7] It might look odd to choose "e" as a base. Later, it will be found that choosing "e" as a base, provides many advantages in analysis. It arises quite naturally in calculus (similar to π appearing in geometry) as a basic property of mathematics. A detailed discussion about "e", its origin and properties along with the exponential function e^x and its properties are discussed in the next chapter.

[8] Note that, we shall be considering logarithms of "positive real numbers" only. However, it may be mentioned that logarithms of negative numbers (and those of complex numbers) are also defined and handled, when we deal with the algebra of complex numbers.

Recall the *following three laws of exponents*,

(i) $a^m \cdot a^n = a^{m+n}$,
(ii) $a^m \div a^n = a^{m-n}$,
(iii) $(a^m)^n = a^{mn}$.

Corresponding to the above laws (of exponents), we have the following *three fundamental laws of logarithms:*

(i) $\log_a(mn) = \log_a m + \log_a n$
(ii) $\log_a(m/n) = \log_a m - \log_a n$
(iii) $\log_a m^n = n \log_a m$

Let us prove these laws (or properties) of logarithms.

(I) To prove, $\log_a(mn) = \log_a m + \log_a n$

$$\left.\begin{aligned}\text{Let, } x &= \log_a m, \text{ so that } a^x = m\\ \text{and } y &= \log_a n, \text{ so that } a^y = n\end{aligned}\right\} \tag{3}$$

Now, consider,

$$mn = a^x \cdot a^y = a^{x+y} \quad \text{(by law of exponents)}$$
$$\therefore \log_a mn = x + y, \quad \text{(by definition of logarithm)}$$
$$= \log_a m + \log_a n, \quad \text{[using (3)]}$$

(II) To prove, $\log_a (m/n) = \log_a m - \log_a n$

$$\left.\begin{aligned}\text{Let, } x &= \log_a m, \text{ so that } a^x = m\\ \text{and } y &= \log_a n, \text{ so that } a^y = n\end{aligned}\right\} \tag{4}$$

Now consider,

$$\frac{m}{n} = \frac{a^x}{a^y} = a^x \div a^y = a^{x-y}$$
$$\therefore \log_a(m/n) = x - y \quad \text{(by definition of logarithm)}$$
$$= \log_a m - \log_a n$$

(III) To prove, $\log_a m^n = n \log_a m$

Let, $x = \log_a m$ so that $a^x = m$. Now, consider, $m^n = (a^x)^n = a^{nx}$

$$\therefore \log_a(m^n) = nx \text{ (by definition of logarithm)}$$
$$= n \log_a x$$

It is necessary to get acquainted with the terminology related to logarithms.

12.6.1 Characteristic and Mantissa of Logarithm

Definition: If the logarithm of any number is partly integral (i.e., it is an integer) and partly fractional, the integral portion of the logarithm is called its characteristic and the decimal portion is called its mantissa.

For example, log 795 = 2.9004. Here, the number 2 is the *characteristic* and 0.9004 is the *mantissa.*

12.6.2 Method of Expressing Negative Logarithm

The characteristic of a logarithm may be any real number (positive, zero, or negative), but the mantissa "x" is always *expressed as a non-negative number* $(0 \leq x < 1)$. The method of expressing a *negative logarithm, with positive mantissa* is made clear from the following example.

From the table of logarithms, it will be found that, log 2 = 0.3010[9]

Then, we have $\log\left(\frac{1}{2}\right) = \log 1 - \log 2 = 0 - 0.3010 = -0.3010$ (since $\log 1 = 0$) *which is a negative number*.

This is a case of a negative logarithm, wherein the characteristic is zero, and hence the mantissa is a negative number.

To express the mantissa as a positive number, we write

$$\begin{aligned} -0.3010 &= -1 + 1 - 0.3010 \\ &= -1 + 0.6990. \end{aligned}$$

For *shortness*, we write this latter expression as $\bar{1}.6990$. The horizontal line over the number 1 denotes that *the integral part (i.e., characteristic)* is a negative number; *the decimal part (i.e., mantissa),* however, is positive. Thus, $\bar{2}.3276$, stands for $-2 + 0.3276$.

There is an advantage in expressing the mantissa as a positive number *with reference to the base 10.* This is explained in point (b) given in Section 12.7.

12.7 ADVANTAGES OF COMMON LOGARITHMS

(a) The characteristic of the logarithm of any number can always be determined by inspection.

Case (I): *Let the number be greater than unity*

Since $10^0 = 1$, therefore log 1 = 0,
since $10^1 = 10$, therefore log 10 = 1,
since $10^2 = 100$, therefore log 100 = 2, and so on.

[9] Here the characteristic (i.e., integral part of logarithm) is zero.

Hence, *the logarithm of any number between 1 and 10* must lie between 0 and 1. From the log tables, it may be seen that,

$$\log 3 = 0.4771;\ \log 7 = 0.8451;$$
$$\log 8.3 = 0.9191;\ \log 9.9 = 0.9958.$$

Similarly, *the logarithm of any number between 10 and 100 must lie between 1 and 2*, the logarithm of any number between 100 and 1000 must lie between 2 and 3, and so on. Thus, the logarithm of any number between 10^n and 10^{n+1} must lie between n and $n+1$. From the log tables, we have

$$\log 27.6 = 1.4409;\ \log 153.2 = 2.1853;$$
$$\log 1623 = 3.2067;\ \log 7295 = 3.8576.$$

Case (II): Let the number be less than unity.

Since $10^0 = 1$, therefore $\log 1 = 0$,
since $10^{-1} = 1/10 = 0.1$, therefore $\log 0.1 = -1$,
since $10^{-2} = 1/10^2 = 0.01$, therefore $\log 0.01 = -2$,
since $10^{-3} = 1/10^3 = 0.001$, therefore $\log 0.001 = -3$ and so on.

Thus, the logarithm of any number between 0.1 and 1, lies between -1 and 0, and so it is equal to "-1" + some number in decimal (i.e., its characteristic is $\bar{1}$). Similarly, the logarithm of any number between 0.01 and 1 lies between -2 and -1 and hence it is equal to "-2" + some number in decimal (i.e., its characteristic is $\bar{2}$). From the log tables, we get,

$$\log 0.35 = \bar{1}.5441;\ \log 0.057 = \bar{2}.7559;$$
$$\log 0.0091 = \bar{3}.9590;\ \log 0.0006 = \bar{4}.7782.$$

(b) A very important property of logarithms to the base 10, is that the mantissa (i.e., the decimal portion) of the logarithms of all numbers, consisting of the same significant digits, are the same. The following example makes this point clear.

Suppose we are given that $\log 66818 = 4.8249$. Then, consider the numbers 66818, 668.18, 0.66818, and 0.00066818, which consist of the *same significant figures*, but differ only in the position of the decimal point. Let us find the logarithms of these numbers, to the base 10. From the log tables, we have $\log_{10} 66818 = 4.8249$.
Now, consider

$$\log 668.18 = \log\frac{66818}{100} = \log 66818 - \log 100$$
$$= 4.8249 - 2$$
$$= 2.8249$$
$$\log 0.66818 = \log\frac{66818}{100000} = \log 66818 - \log 100000$$
$$= 4.8249 - 5$$
$$= \bar{1}.8249$$
$$\log 0.00066818 = \log\frac{66818}{10^8}$$
$$= 4.8249 - 8$$
$$= \bar{4}.8249$$

Observe that, the logarithms of the above numbers have the same decimal portion (i.e., the same mantissa), and they differ only in the characteristic.

Remark: In view of the above, we say that the mantissa of a logarithm is by convention positive. However, the above property (possessed by logarithms to the base 10) is not possessed by logarithms to the base "e" (or other bases such as 2, 3, 5, 7, . . . etc.).

12.8 CHANGE OF BASE

We will now show that, if we are given the logarithm of a number, to any base, then we can *easily* compute the logarithm of that number to any other base. The following relation states the rule.

$$\log_a x = \frac{\log_b x}{\log_b a}$$

$$\text{or } \log_a x \cdot \log_b a = \log_b x \qquad (5)^{(10)}$$

Let us prove this relation.

Proof: Let

$$\log_b x = y \text{ and } \log_b a = c \qquad (6)$$

$$b^y = x \text{ and } b^c = a \qquad (7)$$

We must *eliminate* "b". For this purpose, we obtain from (7)

$$b = x^{1/y} \text{ and } b = a^{1/c}$$

$$\therefore x^{1/y} = a^{1/c} \ \therefore x = a^{y/c}$$

$$\therefore y/c = \log_a x$$

$$\therefore \log_a x = y/c = \frac{\log_b x}{\log_b a}, \quad \text{[using (6)]}$$

$$\therefore \log_a x = \frac{\log_b x}{\log_b a} \quad \text{(Proved)}$$

In the same manner, it can be proved that

$$\log_a x = \log_b x \cdot \log_a b \qquad (8)$$

Thus, if we know the logarithm of any number to a base "b" then we can easily find its logarithm to any other (desired) base "a".

(10) Look at the following algebraic identity. $(x/a)\cdot(a/b) = (x/b)$. It is useful, in writing the rule for change of base, for logarithms.

12.8.1 Corollary

An important property of logarithms

$$\log_b a \cdot \log_a b = 1 \text{ or } \log_b a = \frac{1}{\log_a b}$$

Proof: Let

$$\log_b a = c \quad \therefore b^c = a \tag{9}$$

and

$$\log_a b = d \quad \therefore a^d = b \tag{10}$$

From (10), we obtain

$$b^c = a^{cd} \tag{11}$$

Now, from (9) and (11) we have (by eliminating "b"),

$$a = a^{cd}, \text{ which means } c \cdot d = 1$$

$$\therefore \log_b a \cdot \log_a b = 1$$

or

$$\log_b a = \frac{1}{\log_a b} \tag{12}$$

Remark: Using (12), result (5) can be written in the form (6).

12.8.1.1 To Express Any Positive Number in the Exponential Form Now, it is easy to show that

$$a^{(\log_a b)} = b,\ c^{(\log_c b)} = b,\ a^{(\log_a x)} = x, \text{ and so on.}$$

Now, we will show that, $a^{(\log_a x)} = x$.

Let, $\log_a x = t$

$$\therefore x = a^t \text{ (by definition of logarithm)}$$
$$= a^{(\log_a x)},\ (\because t = \log_a x)$$

Remark: The above result tells us that *any positive number* "x" can be expressed in the *exponential form* by choosing an *arbitrary positive base* "a" ($a \neq 1$) and raising it to *the power* $\log_a x$.

12.8.2 Antilogarithm: Definition

If $\log_a c = b$, then c is called *the antilogarithm* of b, to the base a. We read, $c = \text{antilog}\, b$, to the base a. Thus, the process of finding the antilogarithm is *just the reverse of the procedure for finding the logarithm of a given number*.

Now, we shall work a few numerical examples to show the application of logarithms and antilogarithms for calculations. Here, it is assumed that the reader is familiar with how to use the tables for logarithms and antilogarithms. (Later in the text, we have discussed the method of using these tables.)

12.8.3 Application of Logarithms

(a) *Multiplication of Numbers*

To find the product (0.035681)(2763.5)

Let $x = (0.035681)(2763.5)$

$$\begin{aligned}\log x &= \log 0.035681 + \log 2763.5\\ &= \bar{2}.5524 + 3.4254\\ &= 1.9778\\ \therefore x &= \text{antilog}\,(1.9778)\\ &= 95.01.\end{aligned}$$

(b) *Powers and Roots*

(i) To find $(5.978)^4$

Let $x = (5.978)^4$

$$\begin{aligned}\therefore \quad \log x &= 4 \log 5.978\\ &= 4 \times 0.7766\\ &= 3.1064\end{aligned}$$

$$\begin{aligned}\therefore x &= \text{antilog}(3.1064)\\ &= 1277\end{aligned}$$

Check $6^4 = 1296$.

(ii) To find cube root of 79507

Let $x = [79507]^{1/3}$

$$\begin{aligned}\therefore \quad \log x &= (1/3) \log 79507\\ &= (1/3)[4.9004]\\ &= 1.6334\end{aligned}$$

$$\therefore x = \text{antilog}\, 1.6334 = 43.00$$

Thus, the cube root of 79507 is 43.

From the above, we note that the simpler process of addition has replaced the process of multiplication, and the simpler process of division has replaced the difficult process of extracting the cube root.

12.9 WHY WERE LOGARITHMS INVENTED?

To speed up and simplify calculations, naturally. Indeed, logarithms simplify and speed up calculations to a remarkable degree. They make it possible to perform operations that would otherwise be extremely difficult (e.g., extracting high-index roots).

Today, we are used to logarithms and to the extent to which they simplify the computation process so *it is hard to imagine the wonder and excitement they caused when they first appeared.*

If the logarithm of a number is an irrational number, then it cannot be *exactly* expressed in decimal form. The logarithms of such numbers are given only approximately, no matter how many decimal places are taken—*the larger the number of decimal places in the mantissa, the better the approximation.*[11]

The idea that shorter mantissas would suffice was realized recently. For most practical needs, even three place mantissas are suitable. This is because of the fact that, rarely do measurements involve more than three decimal places.

12.10 FINDING A COMMON LOGARITHM OF A (POSITIVE) NUMBER

The common logarithm of a (positive) number, consists of the sum of two parts, namely the "*characteristic*" and the "*mantissa*". Thus,

$$\log x = \text{characteristic for } x + \text{mantissa for } x.$$

The characteristic is *an integer* (positive, zero, or negative) and the mantissa is *always a non-negative number*, less than 1 in the decimal form (e.g., 0.2539, 0.0703, etc.). Characteristic is found by *inspection,* whereas mantissa is found from the *tables of (common) logarithms.*

Note (4): To find the logarithm of a number, *it is necessary to "write" the number in the decimal form*. For example, we write 635 (= 635.00), 5923/5 = 1184.6, and consider numbers of the type 2.0357, 0.8305, 0.003751, and so on.

12.10.1

To find the characteristic *by inspection*, we consider the following two cases:

(a) Characteristic of a Number Greater Than 1

Rule:
We count *the number of digits on the left of the decimal point* of a given number. Suppose, there are "n" digits, then the characteristic of that number will be $(n - 1)$.

[11] With this idea, nearly 500 types of logarithm tables have been prepared from the time logarithms were invented. They include 10-place tables (for common logarithms) by Dutch mathematician Adrian Vlacq to 260-place logarithms (to the base "e") by Adams. Interesting information about these developments is available in the book "*Mathematics can be Fun*", by Ya Perelman (p. 381), Mir Publishers, Moscow, 1979.

Example (1):

Given number	Characteristic
3073.563	3
506.335	2
93.672	1
8.359	0

(b) Characteristic of a (Positive) Number Less Than 1 (i.e., $0 < x < 1$)

Rule:
Count the number of zeros appearing *immediately after the decimal point and before the first nonzero digit*, in the decimal form of the number. Suppose, there are "z" such zeros. Then, characteristic of that number is $-(z+1)$. In this case, *the characteristic is a negative number*. However, to indicate the negative characteristic we use a *bar* over the characteristic, to emphasize the point that, only *the number below the bar is negative*.

Example (2):

Given number	Characteristic
0.23931	$\bar{1}$
0.05729	$\bar{2}$
0.00315	$\bar{3}$
0.00063	$\bar{4}$

12.10.2 Method of Finding the Mantissa (Using Logarithm Tables)

[It is useful to open the log tables(s) at the time of reading the following material.] We have to use *logarithm tables*, to find the mantissa (i.e., decimal part) of the logarithm of a number. The logarithm tables consist of rows and columns. Rows begin with the numbers 10, 11, 12, . . . upto 99, and there are *columns with headings* 0, 1, 2, 3, . . . upto 9. After these columns, there are *other columns*, with headings 1, 2, 3, . . ., 9. These are known as *mean differences*. (Reader may refer to a log table.)

The common logarithm tables are designed to find the *mantissa for four-digit numbers*. For finding the mantissa, *we ignore the decimal point, and consider only four (significant) digits of the number*. If we have numbers, which have more than four digits, after ignoring the decimal, then we proceed as follows:

- If the fifth digit is ≥ 5, then increase the fourth digit by 1 and ignore all the digits from fifth onwards.
- If the fifth digit is less than 5, then ignore all the digits from fifth onwards. For example,

Number for Finding Logarithm	Number for Finding Mantissa	Four-Digit Number for Finding Mantissa	Mantissa (Found Using the Table)
57.314	57314	5731	0.7582
57.315	57315	5732	0.7583
5.7317	57317	5732	0.7583
0.57313	57313	5731	0.7582
0.057318	57318	5732	0.7583

Remark: The *column(s) of mean differences* are prepared to ensure proper accuracy of the mantissa. Let us find the logarithm(s) of the following numbers to understand the entire procedure.

Example (3): To find the logarithm of 7452.76.

(a) Characteristic for 7452.76 is 3.

(b) Mantissa: Ignoring decimal point, we read the number as 745276 and consider the four-digit number 7453. (Why ?) Now, we look at the row starting with 74. Since the next digit (in 7453) is 5, *we find the number in this row under the column headed by 5*. Here the number at the crossing of 74 and 5 is 8722.

Now, we read the number *in the same row under the column 3 of mean differences* [3 is the next (last) digit in 7453]. The difference is 2. *We add the mean difference to 8722 and obtain the mantissa as* 0.8724.

$$\therefore \log 7452.76 = 3.8724$$

Example (4): To find logarithm of 0.035244.

(a) Characteristic for 0.035244 $= \bar{2}$

(b) Mantissa: For the four-digit number 3524

$$= 5465 + 5 \text{ (mean difference)} = 5470.$$

$$\therefore \log 0.035245 = \bar{2}.5470.$$

12.11 ANTILOGARITHM

Now, we consider the problem of finding the number m, when $\log m$ is known. We know that $\log 1000 = 3$. Therefore, we say that antilog $3 = 1000$. Our interest lies in finding, antilogarithm(s) of numbers with four-digit mantissa (i.e., numbers of the type 3.8424, 0.0134, $\bar{2}$.5470., $\bar{1}$.6133., etc.).

For finding antilog of a number, we have to refer to the table with heading *Antilogarithms*. These tables also consist of rows and columns. Here, the rows begin with numbers 0.00, 0.01, and 0.02, up to 0.99. Everything else looks similar to the log tables. The method of referring to these tables is also the same. We explain the method of finding the antilog of a number with the help of some examples. [Of course, it will be useful to refer to the antilog table(s) while solving the following examples.]

Example (5): Let us find the antilog (3.8724).

Here, we have to find *the number* whose *log* is equal to 3.8724. Suppose, the required number is y, then

$$\log y = 3.8724 = 3 + 0.8724.$$

Here, characteristic for y is 3, and mantissa is 0.8724. (Antilog tables are used with reference to the *first four digits of mantissa*. Final number is found by placing the decimal point suitably, using the characteristics.)

Step I: First we consider the mantissa part, which is 0.8724.

In the *antilog table*, the entry in the row beginning with 0.87 and under the column headed by 2 is 7447. Now, the next digit in 0.8724 is 4, so we find the number in the same row (*in the mean difference column*) headed by 4. This number is 7. We add this number to 7447 and obtain the number 7454.

Step II: Since the characteristic of the required number is 3, *the number of digits before the decimal point must be 4*. Hence, the required number is equal to 7454.0.

Note (5): In the earlier Example (3), for finding the logarithm, we have computed the logarithm of 7452.76 as 3.8724. Therefore, antilogarithm of 3.8724, should be *nearly* 7453, but here it is found to be 7454. Thus, there is a difference between the original number and the recovered number, in the fourth digit only. This indicates the accuracy achieved in obtaining the original number, by applying the antilog to the logarithm of the given number. Similarly, it will be found that, for recovering a five-digit number, *there may be a difference of (at most) two-digit number*, and so on. Let us check about this expectation.

Consider, $\log 63293 = 4.8009 + 6$ (mean difference) $= 4.8015$.

Now, let us compute antilog 4.8015.

(Recall that, for computing the antilog, we consider the *mantissa* only, and then place the decimal point suitably, depending on (the value) of characteristic.)

$$\text{Antilog } 4.8015 = 6324 + 7 \text{ (mean difference)} = 63310.0 = 6331.$$

Note that, the difference between *the original number* and *the recovered number* is $(63310 - 63293) = 17$.

Example (6): Now, consider $\log 0.07627 = \bar{2}.8824$.

It is found that, antilog $\bar{2}.8824 = 0.07628$.

(Here, the characteristic is $\bar{2}$. Hence, after finding the antilog with reference to the mantissa 0.8824, we put *one zero* to the right of decimal point) *and write the digits discovered.*

Example (7): To find antilog $\bar{1}.0352$.

(Note that, in this case *the four-digit mantissa to be considered* is 0352.)

$$\therefore \text{Antilog } \bar{1}.0352 = 0.1084.$$

Example (8): Antilog $\bar{1}.31527$

In this case, the four-digit mantissa to be considered is 3153 (Why?).

$$\therefore \text{Antilog } \bar{1}.3153 = 0.2066.$$

(Also, antilog $\bar{1}.3152 = 0.2066$.)

12.11.1 Antilogarithm of a Negative Number

In dealing with *real numbers*, our interest is restricted to the *logarithms of positive numbers only* (excluding 1). However, we may be required to compute the *antilogarithm(s)* of a (small) negative number like (−1.3256) or (−0.5913) or (−2.6512), and so on. *How to compute these antilogarithm(s)*?

[Note that, in (−1.3256) the characteristic and the mantissa both parts are negative. To express the mantissa (i.e., −0.3256) as a positive number, we add 1 and subtract 1, to keep the given number unchanged.]

$$\begin{aligned} -1.3256 &= (-1 - 0.3256) + 1 - 1 \\ &= -2 + 0.6744 \\ &= \bar{2}.6744 \end{aligned}$$

$$\begin{aligned} -0.5913 &= (-0.5913) + 1 - 1 \\ &= -1 + 0.4087 \\ &= \bar{1}.4087 \end{aligned}$$

Now, it is *simple to compute the antilogarithms* of these numbers.

12.12 METHOD OF CALCULATION USING LOGARITHM

Using tables of (common) logarithms and antilogarithms numerical calculations, involving operations of multiplication, division, raising to the power, and root extraction are easily computed, by applying the *laws of logarithms* and using the tables of common logarithm(s) and antilogarithm(s). Earlier (in Section 12.4), we have seen some examples, which explain the applications of tables of logarithms and antilogarithms. Now, having learnt the method of using log tables and antilog tables, we illustrate below, the *logarithm method* of calculation, which *is very useful in arithmetical calculations*, especially in labs.

Example (9): Using logarithm tables, let us calculate

$$\left\{\frac{(59.6)^3}{(4.7)^2 \times (7.2)}\right\}^{1/2}.$$

Solution: Let $t = \left\{\frac{(59.6)^3}{(4.7)^2 \times (7.2)}\right\}^{1/2}$

$$\begin{aligned} \therefore \log t &= \log\left\{\frac{(59.6)^3}{(4.7)^2 \times (7.2)}\right\}^{1/2} \\ &= \frac{1}{2}\log\left\{\frac{(59.6)^3}{(4.7)^2 \times (7.2)}\right\} \\ &= \frac{1}{2}\log(59.6)^3 - \frac{1}{2}\log\left[(4.7)^2 \times (7.2)\right] \\ &= \frac{1}{2}[3\log(59.6)] - \frac{1}{2}[2\log(4.7) + \log(7.2)] \end{aligned}$$

Now, 3 log 59.6 = 3 (1.7752) = 5.3256.

$$\left.\begin{aligned} 2\log 4.7 &= 2\,(0.6721) = 1.3442 \\ \log 7.2 &= 0.8573 \end{aligned}\right\} 2.2015 \text{ (Total)}$$

$$\therefore \log t = \frac{1}{2}[5.325 - 2.2015] = \frac{1}{2}[3.1241] = 1.5625$$

$$\therefore t = \text{antilog}\, 1.5625 = 36.52$$

$$\therefore \left\{\frac{(59.6)^3}{(4.7)^2 \times (7.2)}\right\}^{1/2} = 36.52 \quad \text{Ans.}$$

Example (10): To calculate $\dfrac{5.8 \times 13.6 \times 18.9}{(2.7)^3 \times 0.21}$.

Solution: Let $t = \dfrac{5.8 \times 13.6 \times 18.9}{(2.7)^3 \times 0.21}$.

$$\begin{aligned} \therefore \quad \log t &= \log\left[\frac{5.8 \times 13.6 \times 18.9}{(2.7)^3 \times 0.21}\right] \\ &= (\log 5.8 + \log 13.6 + \log 18.9) - (3\log 2.7 + \log 0.21) \end{aligned}$$

$$\text{Now,} \left.\begin{aligned} \log 5.8 &= 0.7634 \\ \log 13.6 &= 1.1335 \\ \log 18.9 &= 1.2765 \end{aligned}\right\} 3.1734 \text{ (Total)}$$

$$\left.\begin{aligned} 3\log 2.7 &= 3\,(0.4314) = 1.2942 \\ \text{and} \quad \log 0.21 &= \bar{1}.3222 \end{aligned}\right\} 0.6164 \text{ (Total)}$$

$$\therefore \quad \log t = (3.1734 - 0.6164) = 2.5570$$

$$\therefore t = \text{antilog}\, 2.5570 = 360.6$$

$$\therefore t = 360.6 \quad \text{Ans.}$$

Note (6): Besides the above applications, the computation of compound interest (on fixed deposits) or population growth, or depreciation values of houses, and so on are easily calculated by using "*log method*". The compound interest formula is

$$A = P\left[1 + \frac{r}{100}\right]^n,$$

where A, P, r, and n have their usual meaning.

Remark: The logarithm of a number *is based on representing the number in exponential form*. The mathematical operation "*raising to a power*" has *two inverse operations*. If $a^b = c$, then finding the base "*a*" is *one inverse operation* called *extraction of the root,* and finding the exponent "*b*" *is the other inverse operation called taking the logarithm.*

Note that, in the operation(s) of addition and multiplication, both the terms are *of an equal status and can be interchanged* (thus, $a+b = b+a$ and $a \cdot b = b \cdot a$). But the numbers (or terms) that take part in "*raising to a power*" are *not* of the same status, and, generally, cannot be interchanged (e.g., $3^5 \neq 5^3$). It is for this reason that, "*raising to a power*" has *two inverses.* Further, finding the *base* and finding the *exponent* are handled in *different ways.*

Let us revise the following terms, which we have frequently used in this chapter.

Power: The number of times a quantity is to be multiplied by itself. For example, $2^6 = 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 64$, is known as the *sixth power of 2.*

Exponent: A number or symbol placed as a superscript after an expression to indicate the power, to which it is raised. For example, x *is an exponent* in a^x, and in $(ay+b)^x$.

Index: A number that indicates a characteristic (or a role or a function) in a mathematical expression. For example, in y^6, the exponent 6 is also known as the index. Similarly, in $\sqrt[3]{27}$ and $\log_{10}x$, the numbers 3 and 10, respectively, are called indices (plural for index).

13a Exponential and Logarithmic Functions and Their Derivatives

13a.1 INTRODUCTION

Exponential and logarithmic functions are among the most important and most practically useful functions in calculus. The definition of logarithm of any (positive) number is based on exponents and the properties of logarithms are then proved from corresponding properties of exponents.(1)

If $a > 0$ $(a \neq 1)$, then the expression a^x makes sense for any real number x. Accordingly, for any positive base a (except $a = 1$), the expression a^x defines a sensible exponential function. In practice, $a = 2$, $a = 10$, and $a = \mathrm{e}$ are the most useful bases. Among all exponential functions, the one with base e (i.e., the function e^x) turns out to be especially useful and convenient. For day-to-day calculations such as those in the field of engineering, the base 10 is found to be very useful. Logarithms to the base 10 are called *common logarithms*. Important in calculus are logarithms to the base e called *natural logarithms*.(2)

In many books, systematic and excellent information about the number e is available. It possesses certain unique properties valuable in many branches of mathematics, particularly *calculus*. For a student of mathematics, the knowledge of this unique number (and the related functions: e^x, $\log_e x$, a^x, $\log_a x$) is very essential. We give here a brief account of the number e. Its approximate value is given by

$$\mathrm{e} = 2.718281828459045235 36\ldots$$

For this number, the symbol e was first adopted by the great Swiss mathematician Leonard Euler.

What must you know to learn calculus? 13a-The number "*e*," its origin, value, and properties Exponential and logarithmic functions ($\mathbf{e}^x$, $\log_e x$, a^x, $\log_a x$), their derivatives and the applications of e^x (exponential growth and decay)

(1) Recall that, if three numbers a, b, and c are related such that, $a^b = c$ $(a > 0, a \neq 1)$, then the exponent b is called the logarithm of c to the base a. Observe that, for $a > 0$, c is always a positive number.

(2) In mathematics, the two numbers, namely, π and e are very important. They arise in a natural way in *geometry* and *calculus*, respectively. Both π and e are special types of irrational numbers, known as transcendental numbers. They arise not as the result of a simple algebraic relationship, but as a basic property of mathematics. (Transcendental numbers are defined as numbers that are not the roots of any algebraic equation with rational coefficients.) In this chapter, we shall discuss why e is important in mathematics.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition. Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

13a.2 ORIGIN OF e

The idea of the number e comes from the practice of money lending. Consider a quantity growing in such a way that the increment of its growth, during a given time, shall always be proportional to its own magnitude. This situation resembles the process of computing interest on money lent at some fixed rate, since the bigger the capital, the bigger the amount of interest. Here, we must distinguish clearly between two ways for calculating interest on the capital: (a) at simple interest, and (b) at compound interest.

13a.2.1 At Simple Interest

We know that in this case the *capital remains fixed*, so the interest is always calculated on the fixed capital for a given time. Thus, if the initial capital is Rs. 100 and the rate of interest is 10% per annum, then the owner will earn Rs. 10 every year. If this earning continues for a period of 10 years, then the owner must have received 10 increments of Rs. 10 each (so the total interest earned is Rs. 100) and thus, his initial capital will be doubled in 10 years. (In this case, the value of the yearly interest is 1/10 of the capital.)

If the rate of interest is 5% (i.e., 5/100 = 1/20), then the initial capital will be doubled in 20 years, and if the rate of interest is 1% (i.e., 1/100), then it will take 100 years for the initial capital to be doubled. It is easy to see that if the value of the yearly interest is $(1/n)^{\text{th}}$ of the initial capital, the owner must go on hoarding for n years in order to double his capital. In other words, if p is the initial capital and the yearly interest is p/n, then at the end of n years his final amount will be

$$p + \frac{p}{n} \cdot n = 2p$$

13a.2.2 At Compound Interest

In this case, the interest is added to the capital every year (or every half or every quarter of the year, and so on, as the terms may be); so the capital increases by successive additions of the interest part to it at the end of every term.

As before, let the owner begin with an initial capital of Rs. 100, earning an interest at the rate of 10% per annum. Then, at the end of first year, the capital will grow to Rs. 110 and in the second year this new capital will earn (assuming the interest rate is still 10%) Rs. 11 as interest. Accordingly, he will start the third year with Rs. 121 as capital and the interest on this amount will be Rs. 12.10. He will, thus, start the fourth year with Rs. 133.10 as capital, and so on.

If p is the initial capital that grows by compound interest at the rate of 10% per annum, then at the end of 10 years the capital will grow to the amount A given by

$$\begin{aligned} \text{A} &= \text{Rs.}\, p\left(1 + \frac{10}{100}\right)^{10} \\ &= \text{Rs.}\, p\left(1 + \frac{1}{10}\right)^{10} \\ &\approx \text{Rs.}\, p \times 2.594 \end{aligned}$$

However, this mode of calculating compound interest, once a year, is not quite fair because it is possible to earn more by computing the interest at the end of every half-year. This demands that instead of computing the interest at the rate of 10% per year, we should compute it at the rate of

5% per half year. Thus, during the period of 10 years, there will be 20 operations involved, at the end of which the initial capital is multiplied by 21/20.

Now, since $(1+1/20)^{20} = (21/20)^{20} = 2.654$, the original capital (of Rs. 100) will be multiplied by the factor 2.654, showing that the capital must grow to Rs. 265.40.

But even so, the process is still not quite fair since by further reducing the period of each term, it is possible to earn more and more. Suppose we divide the year into 10 parts and reckon a 1% interest for each tenth of the year. In this case, we will have 100 operations lasting over the period of 10 years. Thus, at the end of 10 years, the capital will be multiplied by factor 101/100, thereby obtaining the amount A given by

$$A = \text{Rs. } 100\left(1+\frac{1}{100}\right)^{100}$$

which works out to approximately Rs. 270.40.

Even this is not final. Let 10 years be divided into 1000 periods (each of 1/100 of a year), the interest being 1/10% for each such period. Then,

$$A = \text{Rs. } 100\left(1+\frac{1}{1000}\right)^{1000}$$

which works out to approximately Rs. 271.71.

Let $1/n$ be the fraction added on at each of the n operations, then the value of the capital p at the end of n operations is given by $p(1+(1/n))^n$.

Now, it must be clear that what we are trying to find is in reality the ultimate value of the expression $(1 + 1/n)^n$ as $n \to \infty$. As we take n larger and larger, the number $(1 + 1/n)^n$ grows closer and closer to a particular limiting value. However large we make n, the value of this expression grows nearer and nearer to the figure 2.718281828459..., a number never to be forgotten. To this number, the mathematicians have assigned the English letter "e".

13a.2.3 Compound Interest and True Compound Interest

In the process of computing compound interest, the capital p has its interest added to it at regular periods of time and thus increases by jumps at the end of each period. If we calculate the interest at shorter and shorter intervals, then in the *limiting case* it will signify in a sense that the interest is compounded continuously at each instant. When the interest is compounded in this way, we say that *true compound interest* is calculated.

13a.2.4 What Is e

Suppose we are to let 1 grow at simple interest till it becomes 2, and if at the same nominal rate of interest and for the same period of time we were to let 1 grow at true compound interest instead of simple, then it would grow to the value e.

Further Explanation for e:

Let us take 100% as the unit of rate and any fixed period as the unit of time. Then, the result of letting 1 grow *arithmetically* (i.e., by simple interest) at the unit rate for the unit time will be 2, while the result of letting 1 grow by true compound interest at the unit rate for the unit time will be 2.71828..., which is the number e.

Accordingly, we write

$$\lim_{n \to \infty} (1 + (1/n))^n = \mathrm{e}$$

13a.3 DISTINCTION BETWEEN EXPONENTIAL AND POWER FUNCTIONS

The expression 2^x can be carelessly mistaken for the expression x^2 as typographically they are similar; however, the resemblance ends here. They in fact define entirely different functions. The function x^2 is an algebraic *power function* in which the *base* is a *variable* and the *exponent* is a *constant*. On the other hand, the function 2^x is an *exponential function* in which the *base* is a *constant* and the *exponent* is a *variable*. The difference in their pattern of behavior is illustrated in Table 13a.1.

TABLE 13a.1 Comparative Values of the Function x^2 and 2^x

x	x^2	2^x
0	0	1
1	1	2
2	4	4
3	9	8
4	16	16
5	25	32
6	36	64
7	49	128

As can be seen from Table 13a.1, the exponential function $y = 2^x$ increases more slowly for small values of x and is actually less than the power function $y = x^2$ between $x = 2$ and $x = 4$. However, $y = 2^x$ increases more and more rapidly as compared to $y = x^2$. This is because the exponent in the exponential function increases with x (which means that the base is multiplied to itself more number of times), whereas for the power function the exponent remains constant and only the base increases with x.[3]

Another important difference between the two functions is as follows: Corresponding to the fact that $2^x \to 0$ as $x \to -\infty$, the graph of $y = 2^x$ has the line $y = 0$ (i.e., the x-axis) as a *horizontal asymptote*. In fact, every exponential function $y = a^x$ ($a > 0$, $a \neq 1$) has the line $y = 0$ as a horizontal asymptote. By contrast, no power function x^α (where α is a real number) has a horizontal asymptote.

13a.4 THE VALUE OF e

We know that $\lim_{n \to \infty} (1 + 1/n)^n = \mathrm{e}$. A good number of values obtained for this expression, taking $n = 2$, $n = 5$, $n = 10$, and so on up to $n = 10{,}000$, are given below.

[3] Each of the expressions 2^x, e^x, 4^x, $(1/2)^x$, and so on defines an exponential function. Note that, the name, exponential, function is chosen, since the value of the function depends on the exponent x. The most general exponential function is of the form $[f(x)]^{g(x)}$, where both $f(x)$ and $g(x)$ are variables.

$$\left(1+\frac{1}{2}\right)^2 = 2.25$$

$$\left(1+\frac{1}{5}\right)^5 = 2.489$$

$$\left(1+\frac{1}{10}\right)^{10} = 2.594$$

$$\left(1+\frac{1}{20}\right)^{20} = 2.653$$

$$\left(1+\frac{1}{100}\right)^{100} = 2.704$$

$$\left(1+\frac{1}{1000}\right)^{1000} = 2.7171$$

$$\left(1+\frac{1}{10,000}\right)^{10000} = 2.7182$$

For practical purposes, we can obtain the above values with the help of a pocket calculator. Besides, the value of e can be computed to any prescribed degree of accuracy using Taylor's Theorem (introduced later in Chapter 22).

It is, however, worthwhile to find another way of calculating this immensely important figure. First, observe that since n is infinitely large, the number $1/n$ is very small and hence <1.

Therefore, by using the binomial theorem, we can expand the expression $(1+(1/n))^n$ and have

$$\left(1+\frac{1}{n}\right)^n = 1 + n\cdot\frac{1}{n}\cdot\frac{1}{1!} + \frac{n(n-1)}{2!}\cdot\frac{1}{n^2} + \frac{n(n-1)(n-2)}{3!}\cdot\frac{1}{n^3} + \cdots$$

$$= 1 + 1 + \frac{n\cdot n\left(1-\frac{1}{n}\right)}{2!}\cdot\frac{1}{n^2} + \frac{n\cdot n\left(1-\frac{1}{n}\right)\cdot n\left(1-\frac{2}{n}\right)}{3!}\cdot\frac{1}{n^3} + \cdots$$

$$\left(1+\frac{1}{n}\right)^n = 1 + 1 + \frac{\left(1-\frac{1}{n}\right)}{2!} + \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)}{3!} + \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\left(1-\frac{3}{n}\right)}{4!} + \cdots$$

Now, when $n\to\infty$, $1/n$, $2/n$, $3/n$, and so on all tend to 0. This permits us to write

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots$$

$$\text{or}\quad e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots$$

We know that $n!$ grows very fast, therefore $(1/n!)$ goes on reducing rapidly with increasing n. We can work out the sum to any prescribed degree of accuracy by considering the necessary number of terms and ignoring the rest.

Here is the working for 10 terms:

1st term	= 1.000000
2nd term = (dividing 1st term by 1)	= 1.000000
3rd term = (dividing 2nd term by 2)	= 0.500000
4th term = (dividing 3rd term by 3)	= 0.166667
5th term = (dividing 4th term by 4)	= 0.041667
6th term = (dividing 5th term by 5)	= 0.008333
7th term = (dividing 6th term by 6)	= 0.001389
8th term = (dividing 7th term by 7)	= 0.000198
9th term = (dividing 8th term by 8)	= 0.000025
10th term = (dividing 9th term by 9)	= 0.000002
Total	2.718281

Remark: It might seem that the unbounded increase in the exponent would imply an unbounded increase in the function $(1 + 1/n)^n$. But the growth in the exponent is compensated by the fact that the base $(1 + 1/n)$ tends to 1 as $n \to \infty$. The integral function $(1 + 1/n)^n$ increases as $n \to \infty$, but remains bounded. The bounded character of $(1 + 1/n)^n$ can be easily proved. It can be shown that e lies between 2 and 3.

13a.5 THE EXPONENTIAL SERIES

Now, we will show that,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

For this purpose, consider the expression $(1 + 1/n)^{nx}$. Note that, for $n > 1$, $(1/n) < 1$. Therefore, by making use of the binomial theorem, we can expand this expression and get

$$\left(1+\frac{1}{n}\right)^{nx} = 1 + nx \cdot \frac{1}{n} + \frac{nx(nx-1)}{2!} \cdot \frac{1}{n^2} + \frac{nx(nx-1)(nx-2)}{3!} \cdot \frac{1}{n^3} + \cdots$$

$$= 1 + x + \frac{n^2x(x-(1/n))}{2!} \cdot \frac{1}{n^2} + \frac{n^3x(x-(1/n))(x-(2/n))}{3!} \cdot \frac{1}{n^3} + \cdots$$

$$= 1 + x + \frac{x(x-(1/n))}{2!} + \frac{x(x-(1/n))(x-(2/n))}{3!} + \cdots$$

But, as $n \to \infty$, the terms $1/n$, $2/n$, and so on approach 0. Therefore, the right-hand side simplifies to the following:

$$\text{R.H.S.} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

Moreover, the number of terms (being $n + 1$) becomes infinitely large as $n \to \infty$, whatever x may be. Hence, the series continues to infinity.

$$\text{Also, L.H.S.} = \left(1+\frac{1}{n}\right)^{nx} = \left[\left(1+\frac{1}{n}\right)^{n}\right]^{x}$$

$$\therefore \lim_{n\to\infty}\left[\left(1+\frac{1}{n}\right)^{n}\right]^{x} = \left[\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{n}\right]^{x} = e^{x}$$

$$\therefore \text{We get, } e^{x} = 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\cdots$$

This series is called the *exponential series.* It can be shown that this infinite power series is a rapidly convergent series for all real values of x.

13a.6 PROPERTIES OF e AND THOSE OF RELATED FUNCTIONS

The greatest reason why e is regarded important is that the function e^x possesses a property that is not possessed by any other function of x, that is, when e^x is differentiated, the result is the same (i.e., e^x).[(4)]

This can be easily seen by differentiating e^x with respect to x. We have

$$y = e^{x} = 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\cdots, \text{we get}$$

$$\frac{dy}{dx} = \frac{d}{dx}(e^{x})$$

$$= 0+1+\frac{2x}{2!}+\frac{3x^2}{3!}+\frac{4x^3}{4!}+\frac{5x^4}{5!}+\cdots$$

$$= 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\cdots$$

which is exactly the same as the original function e^x.

13a.6.1 Another Way of Obtaining the Exponential Series

Let us try to find a function of x such that its derivative is the same as the function itself. We may also ask: Is there any expression involving only powers of x that is unchanged by differentiation?

We will show that such a function is

$$y = 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\frac{x^5}{5!}+\cdots$$

(4) This is equivalent to saying that $(e^x)' = e^x, \forall x \in R$, that is, $(d/dx)(e^x) = e^x$.

As a general expression (involving only powers of x), let the required function be

$$y = A + Bx + Cx^2 + Dx^3 + Ex^4 + \ldots \tag{1}$$

where the coefficients A, B, C, and so on are to be determined.

By differentiating (1) we get,

$$\frac{dy}{dx} = B + 2Cx + 3Dx^2 + 4Ex^3 + \cdots \tag{2}$$

Now, if this new expression at (2) is to be the same as that at (1), from which it was derived, then by comparing coefficients it is clear that

$$\begin{aligned}
B &= A \\
2C &= B, \quad \therefore C = \frac{B}{2} = \frac{A}{1\cdot 2} = \frac{A}{2!} \\
3D &= C, \quad \therefore D = \frac{C}{3} = \frac{A}{1\cdot 2\cdot 3} = \frac{A}{3!} \\
4E &= D, \quad \therefore E = \frac{D}{4} = \frac{A}{1\cdot 2\cdot 3\cdot 4} = \frac{A}{4!}
\end{aligned}$$

Using these values of B, C, D, E, and so on in equation (1), we get the general expression of the desired function to be

$$y = A\left[1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots\right] \tag{3}$$

If we compare this expression with that of e^x, we observe that both the expressions will be the same if we choose $A = 1$. Note that, in the general expression in equation (3), we can assume $A = 1$, without any loss of generality. Therefore, for the sake of simplicity, we take $A = 1$ and get

$$y = 1 + x + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots \tag{4}$$

which is the function of x (involving only powers of x) having the *desired property*. Moreover, it represents e^x.

Thus,

$$y = e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots \tag{5}$$

is the only function (in powers of x) that has the property that differentiating it any number of times will always give the same function. The function $f(x) = e^x$ (with base e) is often called the exponential function, or sometimes the natural exponential function.

In Chapter 2, we have shown that the exponential function $f(x) = e^x$, $x \in R$ has one-to-one mapping from $(-\infty, \infty)$ onto $(0, \infty)$. Hence, its inverse function exists. The inverse of the exponential function is called the *logarithmic function*.

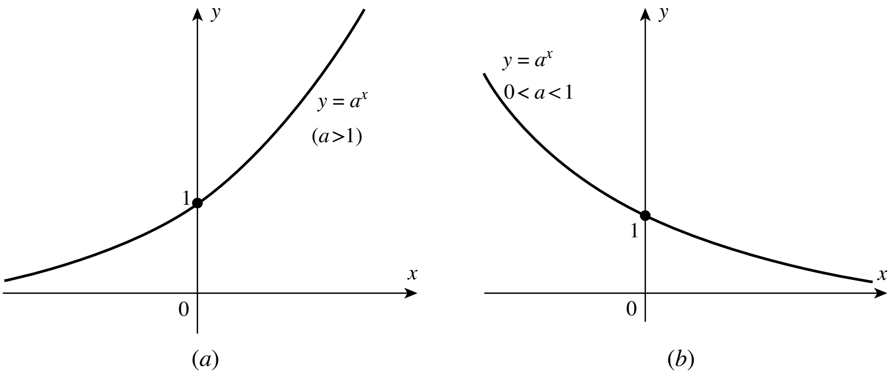


FIGURE 13a.1 Two graphs of the function, $y = a^x$ for different positive values of the base 'a'.

Note (1): Exponential and logarithmic functions come in *pairs*. An exponential function with base e corresponds to a logarithmic function with the same base. What makes base e special for both exponential and logarithmic functions will become clearer when we study the derivatives of these functions. Instead of taking e as the base, we can choose any other positive number a as the base. Then this function is called an exponential function to the base a. We now define the function a^x and $\log_a x$.

Definition: The exponential function, $y = a^x$ ($a > 0$, $a \neq 1$) is defined at every point on the number line R and its range is the set of positive numbers. This function monotonically increases, if the base is $a > 1$ and monotonically decreases if $0 < a < 1$ (see Figures 13a.1a and 13a.1b).

To define the logarithm function, we use the exponential function.

Definition: The logarithm function with positive base a is denoted by

$$f(x) = \log_a x \ (x > 0)$$

and defined by the condition

$$y = \log_a x \Leftrightarrow a^y = x$$

Note (2): For any positive base a ($a \neq 1$), the value of a^y, $y \in R$, is always positive. Let $a^y = x$. This equation also stands for the statement $\log_a x = y$. It follows that the logarithm function is defined only for positive numbers and that the logarithm of a positive number will be a real number (positive, zero, or negative).

The logarithm function, $y = \log_a x$ is defined for all positive x, and its range is the interval $(-\infty, \infty)$. This function monotonically increases if $a > 1$, and monotonically decreases if $0 < a < 1$ (see Figures 13a.2a and 13a.2b).

The logarithmic function, $y = \log_a x$ is the inverse of the exponential function $y = a^x$ and vice versa. [The logarithmic function to the base e is called the natural logarithm (or Naperian logarithm) and is usually denoted by $\ln x$ (or $\log_e x$).] The logarithmic function to the base 10 is called the common logarithm and sometimes denoted by $\log x$.

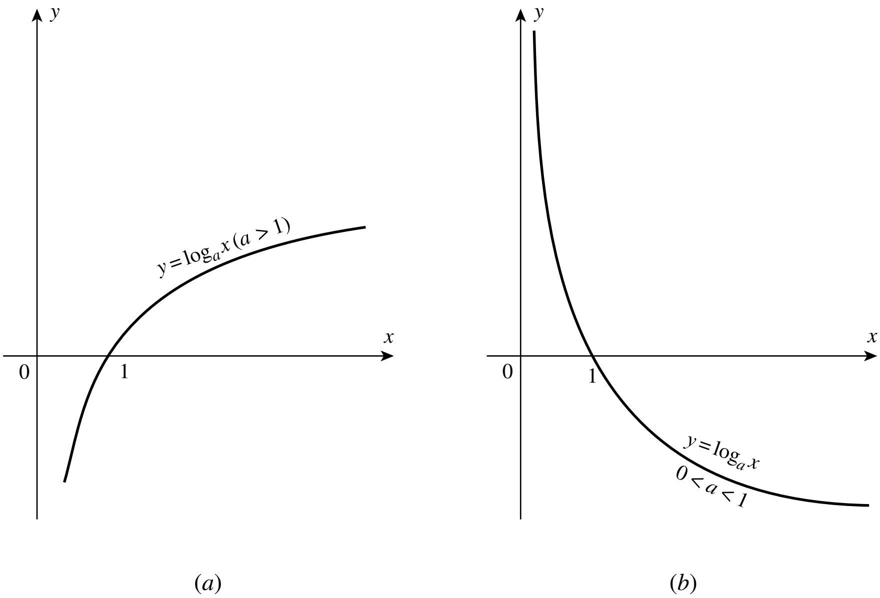


FIGURE 13a.2 Two graphs of the function, $\log_a x$, for positive values of 'a'.

Thus, $\log_e x = \ln x$ and $\log_{10} x = \log x$. We observe the following:

(i) If the base $a > 1$, then for $x \geq 1$, $\log_a x \geq 0$ and for $0 < x < 1$, $\log_a x < 0$ (see Figure 13a.2a).

(ii) If the base a is such that $0 < a < 1$, then for $x \geq 1$, $\log_a x \leq 0$ and for $0 < x < 1$, $\log_a x > 0$ (see Figure 13a.2b).

To get the proper feel of our observation at (ii), let us find out what happens if the base a lies between 0 and 1. For convenience, let us consider the base $a = 1/2$.

We have

$$(1/2)^3 = 1/8 \qquad \therefore \quad \log_{1/2}(1/8) = 3$$

$$(1/2)^1 = 1/2 \qquad \therefore \quad \log_{1/2}(1/2) = 1$$

$$(1/2)^0 = 1 \qquad \therefore \quad \log_{1/2}(1) = 0$$

$$(1/2)^{-5} = 1/(1/2)^5 = 32 \qquad \therefore \quad \log_{1/2}(32) = -5$$

Thus, if the base a is such that $0 < a < 1$, then for $x > 1$, $\log_a(x) < 0$, and for $0 < x < 1$, $\log_a(x) > 0$.

In Table 13a.2, we give values of matched pair of exponential and logarithm functions with base $a = 2$. The values illustrate the correspondence between the functions 2^x and $\log_2 x$.

TABLE 13a.2 Exponential and Logarithmic Function Values for the Base $a = 2$

x	-2	-1	0	1	2	3	10	13.28771
2^x	1/4	1/2	1	2	4	8	1024	10,000
x	1/4	1/2	1	2	4	8	1024	10,000
$\log_2 x$	-2	-1	0	1	2	3	10	13.28771

TABLE 13a.3 Comparative Values of Logarithms (of some numbers) to the bases 10 and e.

$\log_{10}50 = 1.6990$	$\log_{10}500 = 2.6990$	$\log_{10}5000 = 3.6990$
$\log_e 50 = 3.9120$	$\log_e 500 = 6.2146$	$\log_{10}5000 = 8.5172$

Note (3): The rules of common logarithms hold good for natural logarithms also.
Thus,

(i) $\log_e ab = \log_e a + \log_e b$
(ii) $\log_e(a/b) = log_e a - \log_e b$
(iii) $\log_e a^n = n \log_e a$

But, as 10 is no longer the base, one cannot write down the logarithm of $100x$ or $1000x$ by merely adding 2 or 3 to the index. What does this mean?

Table 13a.3 clarifies the point.

13a.7 COMPARISON OF PROPERTIES OF LOGARITHM(S) TO THE BASES 10 AND e

(a) *Common logarithms* (i.e., log to the base 10) are usually studied in elementary mathematics. They are the most convenient to use for most arithmetical calculations, because their base coincides with the decimal base of our number system.

(b) *Natural logarithms* (i.e., log to the base e) are useful in calculus. The logarithmic base e is "natural" only in the sense that it is "naturally convenient" in making the standard process of differentiation work out simply for a logarithmic function.[(5)]

For all practical purposes, we can always convert back and forth between natural and common logarithms of the same number by the following relations:

$$\log_{10}x = \log_{10}\mathrm{e} \times \log_e x \tag{6}$$

$$\log_e x = \log_e 10 \times \log_{10}x \tag{7}$$

It is simple to establish these relationships between common and natural logarithms of the same number x.

Let

$$y = \log_{10}x \tag{8}$$

(5) Later we will see that if the base is e, then the result of differentiating logarithmic and exponential functions assumes simpler forms.

or

$$x = 10^y \tag{9}$$

Consider equation (9) and take logarithms of both sides to the base e. We get,

$$\log_e x = y \log_e 10$$

$$\log_e x = \log_{10} x \times \log_e 10$$

$$\log_{10} x = \frac{1}{\log_e 10} \times \log_e x$$

$$\log_{10} x = \log_{10} \mathrm{e} \times \log_e x$$

To remember the above relationships, it is useful to remember the algebraic identity,

$$\frac{a}{b} = \frac{a}{c} \times \frac{c}{b}$$

Now, the identity $(x/10) = (x/\mathrm{e}) \cdot (\mathrm{e}/10)$ may be looked upon as suggesting

$$\log_{10} x = \log_e x \times \log_{10} \mathrm{e}$$

and similarly we can remember the other one.

But, $\log_{10}\mathrm{e} = \log_{10} 2.718 = 0.4343$ and $\log_e 10 = 2.3026$. Therefore,

$$\log_{10} x = 0.4343(\log_e x)$$

$$\log_e x = 2.3026(\log_{10} x)$$

(c) The characteristic of common logarithm of any (positive) number N changes from 0 to 1 at a convenient point, where $N = 10$, and from 1 to 2 at a convenient point, where $N = 10^2 = 100$, and so on. But the corresponding part of the natural logarithm of N changes from 0 to 1 at the (inconvenient) point, where $N = \mathrm{e} = 2.718\ldots$, and from 1 to 2 at the (inconvenient) point, where $N = \mathrm{e}^2 = 7.389\ldots$, and so on. Thus, a drawback of natural logarithms is that their integral parts (i.e., characteristic or digits to the left of decimal point) are not obvious, as in the case of common logarithms.

Thus, the *naturalness* of natural logarithms has nothing whatsoever to do with the mathematical nature of our decimal number system. For this reason, tables of natural logarithms must include digits to the left of the decimal point as well as to the right (corresponding to both the *characteristic* and *mantissas* of common logarithms) (Table 13a.4).

Besides, additional tables are computed for certain functions involving e. The value of e^x, e^{-x}, and $(1 - \mathrm{e}^{-x})$ are frequently required in different branches of physics. Some of the values of these functions are tabulated here for convenience.

TABLE 13a.4 A Useful Table of Naperian Logarithms (Also Called Natural Logarithms)

No.	$\log_e$	No.	$\log_e$	No.	$\log_e$
1	0.0000	3.0	1.0986	20	2.9957
1.1	0.0953	3.5	1.2528	50	3.9120
1.2	0.1823	4.0	1.3863	100	4.6052
1.5	0.4055	4.5	1.5041	200	5.2983
1.7	0.5306	5.0	1.6094	400	6.2146
2.0	0.6931	6	1.7918	1000	6.9078
2.2	0.7885	7	1.9459	2000	7.6010
2.5	0.9163	8	2.0794	5000	8.5172
2.7	0.9933	9	2.1972	10000	9.2104
2.8	1.0296	10	2.3026	20000	9.9035

TABLE 13a.5

x	e^x	e^{-x}	$1 - e^{-x}$
0.00	1.0000	1.0000	0.0000
0.10	1.1052	0.9048	0.0952
0.20	1.2214	0.8187	0.1813
0.50	1.6487	0.6065	0.3935
0.75	2.1170	0.4724	0.5276
0.90	2.4596	0.4066	0.5934
1.00	2.7183	0.3679	0.6321
1.10	3.0042	0.3329	0.6671
1.20	3.3201	0.3012	0.6988
1.25	3.4903	0.2865	0.7135
1.50	4.4817	0.2231	0.7769
1.75	5.754	0.1738	0.8262
2.00	7.389	0.1353	0.8647
2.50	12.183	0.0821	0.9179
3.00	20.085	0.0498	0.9502
3.50	33.115	0.0302	0.9698
4.00	54.598	0.0183	0.9817
4.50	90.017	0.0111	0.9889
5.00	148.41	0.0067	0.9933
5.50	244.69	0.0041	0.9959
6.00	403.43	0.00248	0.99752
7.50	1808.04	0.00053	0.99947
10.00	22026.5	0.000045	0.999955

13a.8 A LITTLE MORE ABOUT e

The number $e \approx 2.718281\ldots$ plays a vital role in higher mathematics, physics, astronomy, and other sciences. It often appears in a situation where it is least expected. For example, let us have a look at the following problems:

(i) It is required to partition a given positive number a so that the product of all its parts is a maximum. How to do this? Of course, each part must be greater than 1. It is known that the largest product for a constant sum can be obtained when the numbers are all equal. Clearly, then the number a must be partitioned into equal parts. But into how many equal parts? Two, three, five, or what?

Techniques in higher mathematics enable us to establish that the largest product is obtained when the parts are as close as possible to e.

For example, if we want to partition 10 into a number of equal parts such that they are as close as possible to 2.718..., then we have to find the quotient $(10/2.718) = 3.678\ldots$ that is approximately 4.

Then, we get the product

$$(2.5)^4 = 39.0625$$

which is the largest product that can be obtained from multiplying together (four) equal parts of the number 10.

Observe that by dividing 10 into three or five equal parts, we get smaller products:

$$\left(\frac{10}{3}\right)^3 = 37, \quad \left(\frac{10}{5}\right)^5 = 32$$

Again, in order to obtain the largest product of the parts of 20, the number has to be partitioned into seven equal parts, because $20 \div 2.718 = 7.36 \approx 7$.

Similarly, the number 50 has to be partitioned into 18 parts and the number 100 into 37 parts as

$$50 \div 2.718 = 18.4$$
$$100 \div 2.718 = 36.8$$

(ii) *Stirling's Formula:* To compute the product $n!$ that stands for the product of all natural numbers, from 1 to a certain number n is a tedious exercise. It may be verified that

$$10! = 362800$$
$$25! = 15511210043330985984000000$$

In the eighteenth century, the Scottish mathematician James Stirling elaborated a formula that could *calculate factorials approximately*:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

where $\pi = 3.141\ldots$ and $e = 2.718\ldots$. Both these numbers play an important role in various mathematical problems. Applying Stirling's formula and using the tables of logarithms, it is easy to obtain $25! \approx 1.55 \times 10^{25}$.

There are many other questions, considered mathematically, that involve e.

(**Note (4):** Both the problems given above are taken from the book *Mathematics Can Be Fun* by Yakov Perelman, Mir Publishers, Moscow).

13a.9 GRAPHS OF EXPONENTIAL FUNCTION(S)

The general exponential function (to the base a) is expressed by the formula

$$y = a^x, (a > 0, a \neq 1)$$

It is defined for all values of x.

Note (5): The restriction $a \neq 1$ merely excludes from our consideration the rather trivial constant function $y = f(x) = 1^x = 1$.

Note (6): Since the exponent x (in a^x) can be any real number, the question comes up as to how shall we define something like $a^{\sqrt{2}}$? Stated simply, we use an approximation method as follows: First, $a^{\sqrt{2}} \approx a^{1.4} = a^{7/5}$, which is defined. Better approximations are $a^{1.41} = a^{141/100} = \sqrt[100]{a^{141}}$ and $a^{1.414}$, and so on. In this way, a meaning of $a^{\sqrt{2}}$ becomes clear. Thus, we can say that a^x is defined for all real x.

It is simple to calculate the points for drawing the graphs of $y = 2^x$, $y = 3^x$, and $y = (1/2)^x = 2^{-x}$. The following table gives the value(s) of the function(s) corresponding to some values of independent variable x. We consider the following three cases, concerning different positive values of the base a in Table 13a.6.

The ordered pairs are now plotted in a two-dimensional Cartesian frame of reference. Since the domain is the set of all real numbers, we join these points by a smooth continuous curve, as shown in Figure 13a.3.

TABLE 13a.6

	x	-2	-1	0	1	2	3
Case (I)	$y = 2^x$	1/4	1/2	1	2	4	8
$a > 1$	$y = 3^x$	1/9	1/3	1	3	9	27
Case (II) $a = 1$	$y = 1^x$	1	1	1	1	1	1
Case (III) $0 < a < 1$	$y = (1/2)^x = 2^{-x}$	4	2	1	1/2	1/4	1/8

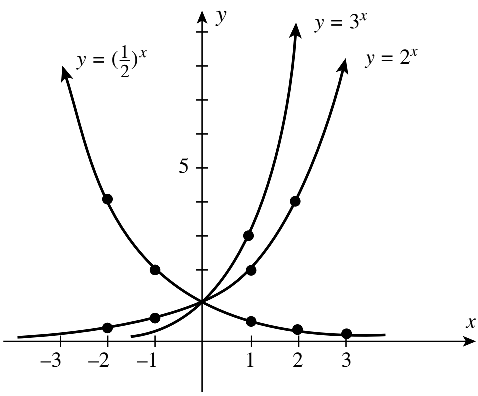


FIGURE 13a.3 Three graphs of the exponential function, $y = a^x$, showing their behavior.

Case (I): $a > 1$: Let $a = 2$ and $a = 3$. The curve $y = 3^x$ draws a comparison with $y = 2^x$ and shows how the graph changes as the base a (arbitrary constant) changes from 2 to 3.

Case (II): $a = 1$: The graph of $y = 1^x = 1$ is a line parallel to the x-axis, passing through the point (0, 1). (We have excluded this case from our consideration and hence from the graph.)

Case (III): $0 < a < 1$: Let $a = 1/2$. The graph of $y = (1/2)^x$ is drawn to see how the curve changes as a changes to less than 1, while remaining positive.

Observations: We make the following observations regarding the nature of the graph(s) of $y = a^x$ for different (positive) values of a and real values of x.

1. In each case, the curve always passes through the point (0, 1), whatever be the value of a.
2. In each case, whatever be the value of x (+ve, zero, or −ve), the value of $y = a^x$ is always positive. Hence, the graph of every exponential function must completely be above x-axis (i.e., in the first and second quadrants only). No part of the graph(s) will lie in the third or fourth quadrant.
3. When $a > 1$ (see the graphs for 2^x and 3^x), the following are the observation:
   (i) **For** $x > 0$: As x increases (or decreases), the value of the function $y = a^x$ increases (or decreases) at a faster rate.[6]
   (ii) **For** $x < 0$: As x decreases by having negative values, y also goes on decreasing and the curve comes nearer and nearer to the x-axis on the negative side. However, since $y = a^x$ can never actually become equal to zero, whatever x may be, the graph will not touch x-axis. The x-axis is said to be an *asymptote* to the curve on the negative side of the axis.
   For negative values of x, the function $y = a^x$ $(a > 1)$ increases (or decreases) with x, but at a slower rate. Thus, the change of y depends, directly on the change of x. The function $y = a^x$ $(a > 1)$ is an increasing function.
4. When $0 < a < 1$ [see the graph of $y = (1/2)^x$], the observation is just the reverse.
   (i) **For** $x < 0$: As x decreases (by having negative values), the value of the function $y = a^x$ increases at a faster rate.
   (ii) **For** $x > 0$: As x increases, y can take on values closer and closer to 0, that is, the curve comes nearer and nearer to the x-axis on the positive side. However, since $y = a^x$ can never actually become equal to zero, whatever x may be, the graph will never touch x-axis. The x-axis is then said to be an asymptote to the curve on the positive side of the axis.

   For positive values of x, the value of the function $y = a^x$ $(0 < a < 1)$ decreases at a slower rate as x increases. Thus, the function $y = a^x$ $(0 < a < 1)$ is a decreasing function.
5. The graphs of the functions $y = a^x$ and $y = (1/a)^x = a^{-x}$, are symmetric with respect to the y-axis (see the graphs of 2^x and $(1/2)^x$).

Note that, the function $y = (1/a)^x$ can be written as $y = a^{-x}$. It then follows that the values assumed by a^{-x} for positive x's are the same as those assumed by a^x for negative x's, having the same absolute values and vice versa. This means that the graphs of the functions $y = a^x$ and $y = (1/a)^x = a^{-x}$ are symmetric, relative to the axis of ordinates.

[6] It means that when x increases (or decreases) by one unit, the corresponding increase (or decrease) in the value of $y = a^x$ is more than one unit.

13a.9.1 Notation

The exponential function to the base a is denoted by $\exp_a$. It relates x to a^x. We write

$$\exp_a: x \to a^x$$

Similarly, the exponential function to the base e is denoted by $\exp_e$. It relates x to e^x. We write,

$$\exp_e: x \to e^x$$

The logarithmic function to the base a ($a > 0, a \neq 1$) is denoted by $\log_a$ (read as log to the base a). It relates x ($x > 0$) to $\log_a x$. We write

$$\log_a: x \to \log_a x \text{ (for each positive } x)$$

13a.10 GENERAL LOGARITHMIC FUNCTION

Definition: The general logarithmic function of $x > 0$ (to the base a) is denoted by $\log_a x$ ($a > 0, a \neq 1$) and defined by $y = \log_a x$, if and only if, $a^y = x$.

(It is important to remember that to define a logarithmic function, we use an exponential function.)

We know that if f and f^{-1} are the functions that are inverses of one another, then their composite in either order is the identity function. In other words,

$$f(f^{-1}(x)) = x$$

$$f^{-1}(f(x)) = x$$

Since the functions $\exp_a$ and $\log_a$ are inverses of one another, we obtain the equations:

$$\log_a(\exp_a(x)) = x, \text{ or } \log_a a^x = x \text{ (for all } x) \tag{10}$$

$$\exp_a(\log_a(x)) = x, \text{ or } a^{\log_a x} = x \text{ (for each positive } x) \tag{11}$$

Note: Since $\log_a$ is the inverse of $\exp_a$, the domain of $\log_a$ is the range of $\exp_a$, which is the set of all positive numbers. Hence, we say that equation (11) is defined for each positive x.

Remark: Equation (11) says that we can represent any positive number x in an exponential form. For this purpose, we must choose any positive number a (except $a = 1$) as base and raise it to the power of the logarithm of x to the base a. Thus, we can write,

$$a^{\log_a x} = x \, (a > 0, \, a \neq 1, \, x > 0)$$

$$e^{\log_e x} = x \, (x > 0)^{7}$$

(7) This equation, together with the expansion of e^x, permits us to expand $e^{\log_e x}$ as an infinite power series. We may write $x = e^{\log_e x} = 1 + \log_e x + \frac{(\log_e x)^2}{2!} + \frac{(\log_e x)^3}{3!} + \cdots$ This expansion will be found useful in evaluating certain limits, including derivatives of a^x ($a > 0$, $a \neq 1$).

In other words, the logarithm to the positive base a ($a \neq 1$) is the exponent to which we raise a to get x.

Besides, equation (11) can be used to represent any power function x^α ($x > 0$), with an arbitrary exponent $\propto$ in the form of a function of a function composed of the logarithmic and exponential functions.

$$y = x^\alpha = (a^{\log_a x})^\alpha = a^{\alpha \log_a x}$$ [8]

13a.10.1 Graphs of Logarithmic Functions $y = \log_a x$

The graph of a logarithmic function can be obtained just by interchanging the domain and range of the equivalent exponential function. Thus, to plot points for a logarithmic function $\log_2 x$, we use the equivalent exponential form: $2^y = x$.

If $y = 0$, then $x = 1$, giving the point (1, 0). Similarly, we find other points as follows:

y	-2	-1	0	1	2	3
x	1/4	1/2	1	2	4	8
(x, y)	(1/4, -2)	(1/2, -1)	(1, 0)	(2, 1)	(4, 2)	(8, 3)

[Remember that the domain of $(\log_2 x)$ is the set of all positive numbers and the range is the set of all real numbers.]

Similarly, we may plot points for $y = \log_{1/2} x$, using the equivalent exponential function $(1/2)^y = x$ or $2^{-y} = x$.

y	-2	-1	0	1	2	3
x	4	2	1	1/2	1/4	1/8
(x, y)	(4, -2)	(2, -1)	(1, 0)	(1/2, 1)	(1/4, 2)	(1/8, 3)

We expect the graph of the $\log_a$ function to be the curve that behaves with respect to the x-axis as the $\exp_a$ curve does with respect to the y-axis and vice versa (or, more informally, we may say that if the axes of x and y are interchanged, the $\exp_a$ curve becomes the $\log_a$ curve and conversely).

13a.10.2 Observations from the Graphs of Logarithmic Functions

The observations we make are mostly similar to those we have made for the graphs of exponential functions.

1. All logarithmic curves pass through the point (1, 0).
2. The graphs lie in the first and fourth quadrants only. That is, the graphs lie entirely on the right of the y-axis. From this observation, we get that logarithmic functions are not defined for negative values of x. (What about $x = 0$?).[9]
3. For $a > 1$, the function is an increasing function. That is, as x increases, so does y and conversely (this is clear from the equation $2^y = x$, which stands for $\log_2 x = y$).

[8] *Mathematical Analysis* (English Translation) by A.F. Bermant and I.G. Aramanovich (p. 59), Mir Publishers, Moscow, 1975.

[9] Remember that for $a > 0$, the equation $\log_a x = y$ means $a^y = x$. Therefore, for $x = 0$, we get $a^y = 0$, which is not possible (why?). Hence, logarithmic function is not defined for $x = 0$.

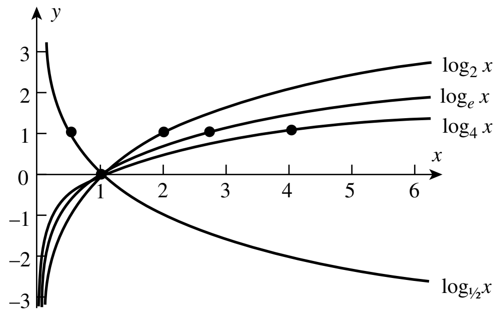


FIGURE 13a.4 Logarithm functions with various bases.

For $0 < a < 1$, it is a decreasing function, that is, as x increases, y decreases and conversely (this is clear from the equation $(1/2)^y = x$, which stands for $\log_{1/2}x = y$).

4. The curve never meets the y-axis, since a^y cannot be zero for any value of y.
5. For $a > 1$, the curve $y = \log_a x$ approaches the y-axis on its negative side but never crosses it. The y-axis is said to be an asymptote to the curve on its negative side (see the graphs of $y = \log_2 x$, $y = \log_e x$, and $y = \log_4 x$).
   For $a < 1$, the same observation is made for the positive side of the y-axis. In this case, the y-axis is an asymptote to the curve on its positive side (see the graphs for $y = \log_{1/2} x$, Figure 13a.4.

The graphs of the logarithmic functions to the bases a and $1/a$ are symmetric with respect to the x-axis (see graphs of $\log_2 x$ and $\log_{1/2} x$ in Figure 13a.4).

13a.10.3 Geometrical Relationship Between the Graphs of Mutually Inverse Functions

We know that the following two functions are mutually inverse functions:

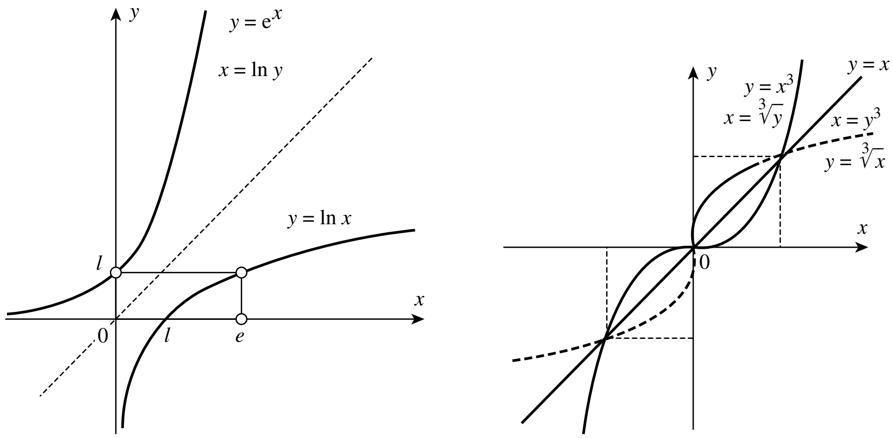
$$\left.\begin{array}{ll}\text{(i)} & y = e^x \\ \text{(ii)} & x = \log_e y\end{array}\right\} \quad (x \in R,\ y > 0)$$

The relation (i) is in the *index form* and the relation (ii) expresses the same thing in the *log form*. It is important to note that these two equations really mean the same thing and that they describe one and the same curve in the xy-plane. For the function e^x, the axis of argument is the x-axis; while for the function $\log_e y$, this role is played by the y-axis (Figure 13a.5a). Similarly, if $y = x^3$, then $x = \sqrt[3]{y}$. The graph of these relationships is a *cubical parabola* (Figure 13a.5b).

It is seen that although the calculations yield different points for plotting, the result is identical, that is, the two curves are the same.

Important Note (8): Generally, we construct the graphs of two mutually inverse functions in such a way that the x-axis is the axis of argument for both of them. For this purpose, we express the two mutually inverse functions in the forms $y = f(x)$ and $y = \phi(x)$. The ordered pairs (for plotting) are then calculated and graphed. All such graphs are symmetric about the line $y = x$ (for more details, see Chapter 2).

Note that, the following graphs of functions, are symmetric about the line $y = x$.



(*a*) (*b*)

FIGURE 13a.5

13a.11 DERIVATIVES OF EXPONENTIAL AND LOGARITHMIC FUNCTIONS

We know that

(a) $\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = \mathrm{e} = 1+\frac{1}{1!}+\frac{1}{2!}+\frac{1}{3!}+\frac{1}{4!}+\cdots$

(b) $\mathrm{e}^x = 1+\frac{x}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\cdots$

(c) $\frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{e}^x) = \mathrm{e}^x$ [This result is obtained by differentiating both sides of the result in equation (b).]

Now, to find the derivative of the logarithmic function $y = \log_e x$, we use the following two results:

(i) The logarithmic function $y = \log_e x$ and the exponential function $x = \mathrm{e}^y$ are mutually inverse functions, of which we know the derivative of the exponential function e^y.

(ii) For any pair of mutually inverse functions $y = f(x)$ and $x = f^{-1}(y)$, their derivatives are related by the condition

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{(\mathrm{d}x/\mathrm{d}y)}, \qquad \text{provided } \frac{\mathrm{d}x}{\mathrm{d}y} \neq 0^{(10)}$$

(10) The rule for the derivative of inverse function states as follows: If $x = f(y)$ is a differentiable function of y such that the inverse function $y = f^{-1}(x)$ exists, then $(\mathrm{d}y/\mathrm{d}x) = 1/(\mathrm{d}x/\mathrm{d}y)$ provided $(\mathrm{d}x/\mathrm{d}y) \neq 0$ (see Chapter 10).

13a.11.1 Finding the Derivative of the Logarithmic Function

To find the derivative of the logarithmic function $y = \log_e x$, consider the equation

$$y = \log_e x \tag{12}$$

We transform equation (12) into its equivalent exponential form.

We have,

$$x = e^y \tag{13}$$

[Here, x stands for the function (i.e., the dependent variable) and y for the independent variable.]

Hence, differentiating both sides of (13) with respect to y, we get,

$$\frac{dx}{dy} = \frac{d}{dy}(e^y) = e^y \tag{14}$$

The derivative of e^y with respect to y is the original function unchanged.

Now, to compute the derivative of $y = \log_e x$, we use the formula

$$\frac{dy}{dx} = \frac{1}{(dx/dy)}, \text{ provided } \frac{dx}{dy} \neq 0^{(11)}$$

$$= \frac{1}{e^y} \quad \left[\therefore \frac{dx}{dy} = e^y, \text{ by equation } (14)\right]$$

$$= \frac{1}{x} \quad [\therefore e^y = x, \text{ by equation } (13)]$$

Therefore, for the function $y = \log_e x$, we get

$$\frac{dy}{dx} = \frac{d}{dx}(\log_e x) = \frac{1}{x} = x^{-1} \tag{15}$$

This is a very curious result. Note that, x^{-1} is a result that we obtained by differentiating the function $\log_e x$ with respect to x, and that we could never have got it by differentiating the power functions, as can be seen from the following results.

$$\frac{d}{dx}\left(\frac{x^3}{3}\right) = x^2, \ \frac{d}{dx}\left(\frac{x^2}{2}\right) = x^1$$

$$\frac{d}{dx}(x) = x^0 = 1$$

$$\frac{d}{dx}(????) = x^{-1} = \frac{1}{x}$$

(11) Note that, $dx/dy = e^y \neq 0$, for any value of y.

$$\frac{d}{dx}\left(-x^{-1}\right) = x^{-2}, \frac{d}{dx}\left(\frac{x^{-2}}{2}\right) = x^{-3}$$

From the above list of derivatives, we note that by differentiating any power function, we can never get the result x^{-1}. Thus, we can say that if there exists any function whose derivative is $(1/x)$, then such a function must be a new function other than a power function. We ask the question:

Is there any function whose derivative is $(1/x)$?[(12)]

Note that, we have obtained the function $\log_e x$ whose derivative is $(1/x)$. Thus, $\log_e x$ is the *desired* (*new*) *function* that fills up the gap noticed above. We call it the *natural logarithm function*.

Recall that the definition of logarithmic function was encountered in algebra and it was based on exponents. The properties of logarithms were then proved from the corresponding properties of exponents.[(13)]

Definition: The *natural logarithmic function* denoted by ln (or $\log_e$) is defined by

$$\ln x = \log_e x = \int_1^x \frac{1}{t}\,dt, \quad x > 0$$

The properties of logarithms can be proved by means of this definition. However, to understand this definition, we have to study the properties of definite integrals and the first fundamental theorem of calculus. These topics are discussed in Part II of this book.

Now, let us try to differentiate $y = \log_e(x+a)$.

$$\text{Consider, } y = \log_e(x+a)$$
$$\therefore x+a = e^y$$

Differentiating both the sides with respect to y, we get

$$\therefore \frac{d}{dy}(x+a) = e^y \quad \left[\because \frac{d}{dy}(e^y) = e^y\right]$$

This gives

$$\frac{dx}{dy} = e^y = x+a$$

[(12)] This is equivalent to asking the question: Is there any function that is antiderivative of $1/x$? [Here, antiderivative is a new term that stands for a (new) function such that its derivative must be equal to the given function].

In other words, we can say that if there is any antiderivative of $1/x$ denoted by $f(x)$, then we must have $(d/dx)[f(x)] = 1/x$. Later in Part II of the book, it is shown that the antiderivative of $1/x$ is $\log_e x$.

[(13)] In particular, to define the logarithm function $y = \log_e x$ $(x > 0)$, we used the exponential function $x = e^y$ $(y \in R)$. The condition, $y = \log_e x \Leftrightarrow x = e^y$ means that a logarithm function and an exponential function with the same base are inverse of each other.

Now, for reverting to the original function, we use the formula

$$\lim_{x\to 0} \frac{1}{x} \cdot \log_a(1+x)$$

We get

$$\frac{dy}{dx} = \frac{1}{x+a} = \left[\text{note that } \frac{dx}{dy} = x+a = e^y \neq 0\right]$$

Thus, for $y = \log_e (x+a)$, we have

$$\frac{dy}{dx} = \frac{1}{x+a}$$

Next, let us try to differentiate $y = \log_a x$. First, we must change $\log_a x$ to natural logarithms (why?). We get

$$y = \log_e x \cdot \log_a e \quad (\text{here, } \log_a e \text{ is a constant})$$

$$= \log_e x \cdot \frac{1}{\log_e a}^{(14)}$$

$$\therefore \frac{dy}{dx} = \frac{1}{x} \cdot \frac{1}{\log_e a} \; (\text{where } \log_e a \text{ is constant})$$

In particular, for the function $y = \log_{10} x$, we have

$$\frac{dy}{dx} = \frac{d}{dx}(\log_{10} x) = \frac{1}{x} \cdot \frac{1}{\log_e 10} = \frac{0.4343}{x} \left[\because \frac{1}{\log_e 10} = \log_{10} e = 0.4343\right]$$

13a.11.2 Finding the Derivative of the Exponential Function

To find the derivative of the exponential function $y = a^x$ $(a > 0, a \neq 1)$ is not very simple.

Taking the natural logarithm of both the sides, we get

$$\log_e y = x \log_e a$$

$$\therefore x = \log_e y \cdot \frac{1}{\log_e a} \tag{16}$$

Note that, here the independent variable is y; hence, differentiating both sides of equation (16) with respect to y, we get

$$\frac{dx}{dy} = \frac{1}{y} \cdot \frac{1}{\log_e a} = \frac{1}{a^x} \cdot \frac{1}{\log_e a}$$

(14) Any expression of the form $\log_a e$ must be expressed in the form $\log_e a$, since tables are available for log to the base e.

Now, reverting to the original function, we get

$$\frac{dy}{dx} = \frac{1}{(dx/dy)} = a^x \cdot \log_e a$$

Thus, for the function $y = a^x$ ($a > 0$, $a \neq 1$), we have

$$\frac{dy}{dx} = \frac{d}{dx}(a^x) = a^x \cdot \log_e a$$

Remark: We have obtained the following results:

$$\frac{d}{dx}(e^x) = e^x \tag{17}$$

$$\frac{d}{dx}(a^x) = a^x \cdot \log_e a \tag{18}$$

$$\frac{d}{dx}(\log_e x) = \frac{1}{x} \tag{19}$$

$$\frac{d}{dx}(\log_a x) = \frac{1}{x} \cdot \log_e a \tag{20}$$

From the above results (17)–(20), observe that the derivatives of exponential and logarithmic functions assume simplest forms, if the number e is chosen as the base.

Also, note that for any constant

$$\frac{d}{dx}(e^{kx}) = ke^{kx}$$

$$\begin{aligned}\frac{d}{dx}(a^{kx}) &= \frac{d}{dx}(b^x) \text{ (where } b = a^k\text{)} \\ &= b^x \log_e b \text{ [using (ii)]} \\ &= a^{kx} \cdot \log_e a^k = ka^{kx} \cdot \log_e a\end{aligned}$$

Note (9): We have obtained the derivatives of exponential functions (e^x and a^x) and those of logarithmic functions ($\log_e x$ and $\log_a x$) using the special property that $(d/dx)(e^x) = e^x$ and (the relationship) that the functions a^t and $\log_a t$ (with the same base) are mutually inverse. In fact, this is an indirect approach by which we could obtain their derivatives. Subsequently, we shall obtain the derivatives of these functions by applying the *definition of derivative*, that is, by the *first principle*.

13a.12 EXPONENTIAL RATE OF GROWTH

The process of growing proportionately at every instant to the magnitude at that instant is called the exponential rate of growth.

Definition: Unit Exponential Rate of Growth: It is the rate of growth that in unit time will cause 1 to grow to the value 2.718281... (i.e., the value of e).

13a.13 HIGHER EXPONENTIAL RATES OF GROWTH

Let us find out what should be the meaning of twice (or thrice) the logarithmic rate of growth?

Observe that a unit logarithmic rate causes 1 to grow (in unit time) to e. Now, if the rate of growth is doubled then 1 will grow to e in half the time, and during the remaining half the time (at the same rate), the quantity e must again grow e times, resulting in total growth to e^2 times. Similarly, if the rate of growth is thrice the logarithmic rate, then 1 must grow to e^3, in unit time, and so on. In general, if a quantity grows at the exponential rate, x units per unit time, then it causes 1 to grow to e^x in unit time.

Remark: Note that in the statement $y = e^x$, the exponent x is the logarithm of the (positive) number y to the base e. Since the value of the function e^x changes with the exponent x, some people say that the (exponential) function e^x grows at the logarithmic rate x. But since the rate of change of the function $y = e^x$ is also given by the value of e^x for every $x \in R$ [since, $(d/dx)(e^x) = e^x$], it is logical to say that the exponential function e^x grows at the exponential rate e^x (see Figure 13a.6). Thus, both the statements about the rate of growth of e^x given above mean the same thing.

Note (10): In the case of any other exponential function whose base is a (positive) number a other than e ($a \neq 1$), then its rate of change is different from the rate of change of e^x for obvious

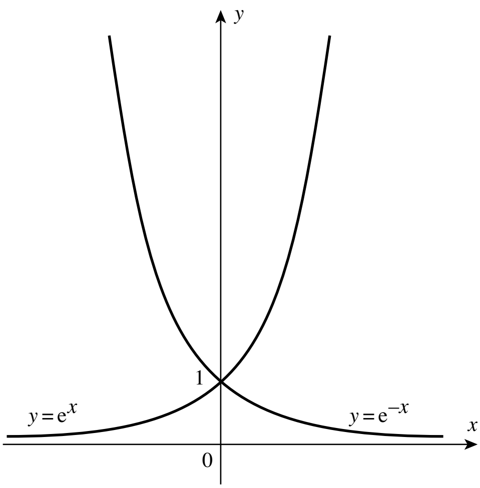
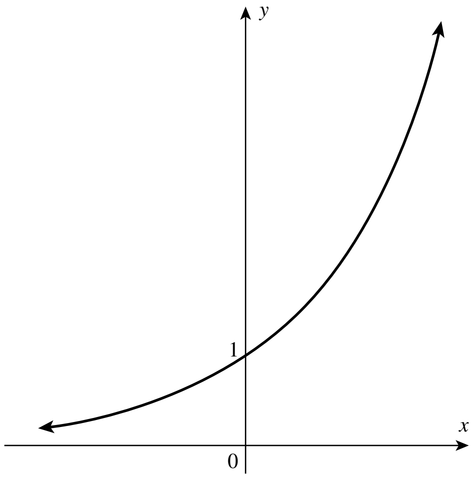


FIGURE 13a.6 Graphs indicating growth (and decay) of the function, $y = e^x$, $x \in R$.



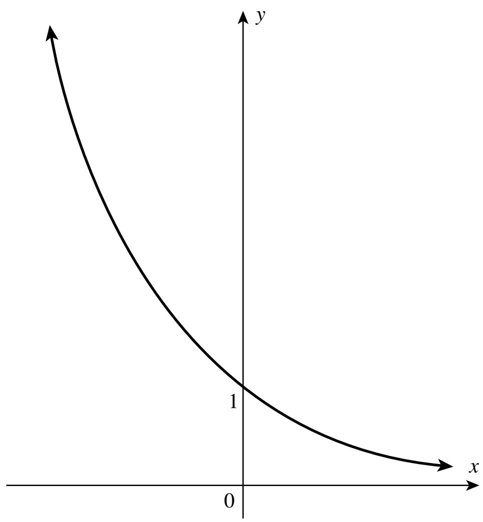
$y = a^x,\ a > 1$

FIGURE 13a.7 Exponential growth, $y = a^x$, $a>1$ $x \in R$.

reasons. Of course, all such exponential functions are said to grow (or decline) exponentially (see Figures 13a.7 and 13a.8).

We know that

$$\lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = \mathrm{e} \quad (n \in N) \tag{21}$$



$y = a^x, 0 < a < 1$

FIGURE 13a.8 Exponential decay, $y = a^x$, $0 < a < 1$ $x \in R$.

It can also be shown that

$$\lim_{x \to \infty} \left(1 + \frac{1}{x}\right)^x = \mathrm{e} \quad (x \in R) \tag{22}$$

$$\lim_{x \to 0} \ (1 + x)^{1/x} = \mathrm{e} \quad (x \in R) \tag{23}$$

Furthermore, if $f(x) \to 0$ as $x \to 0$, then

$$\lim_{x \to 0} (1 + kf(x))^{1/(kf(x))} = \mathrm{e} \quad (k \neq 0) \tag{24}$$

In the process of defining the number e, we accepted the result (21). By applying (21), we can prove the results at (22) and (23). Here, we shall accept the results without proof. Furthermore, the result at (24) can be proved if we put $kf(x) = t$ (where $k \neq 0$), since it can be then expressed in the form (23).[15]

The limits at (21)–(23) will be used as *standard results*. They are used for computing the derivatives of logarithmic functions. Besides, they are used for establishing the following *standard limit* that is needed for computing the derivative(s) of exponential functions.

13a.14 AN IMPORTANT STANDARD LIMIT

To prove $\lim_{x \to 0} \dfrac{a^x - 1}{x} = \log_e a$.

To prove the above result, we shall prove the following prerequisite results:

$$\lim_{x \to 0} \frac{\log_a(1 + x)}{x} = \frac{1}{\log_e a}$$

Proof: Consider

$$\begin{aligned}
\lim_{x \to 0} \frac{\log_a(1 + x)}{x} &= \lim_{x \to 0} \frac{1}{x} \cdot \log_a(1 + x) \\
&= \lim_{x \to 0} \log_a(1 + x)^{1/x} \\
&= \log_a \left[\lim_{x \to 0} (1 + x)^{1/x}\right] \\
&= \log_a e \quad \text{[using equation (23)]} \\
&= \frac{1}{\log_e a} \quad \text{(by change of base)}
\end{aligned}$$

[15] *Differential and Integral Calculus* (Second Edition, Vol. I, pp. 49–50; revised from 1972 Russian edition) by N. Piskunor, Mir Publishers, Moscow, 1974.

Corollary:

$$\lim_{x \to 0} \frac{\log_e(1+x)}{x} = \log_e \mathrm{e} = 1$$

Now, it is easy to prove the result

$$\lim_{x \to 0} \frac{a^x - 1}{x} = \log_e a$$

Solution: Put $a^x - 1 = \mathrm{y}$

$$\therefore a^x = 1 + y$$

$$\therefore x = \log_a(1+y) \qquad \text{(by definition of logarithm)}$$

Also, as $x \to 0$, $y \to 0$ ($\because$ as $x \to 0$, $\mathrm{a}^x - 1 = y \to 0$).

$$\lim_{x \to 0} \frac{a^x - 1}{x} = \lim_{y \to 0} \frac{y}{\log_a(1+y)} = \frac{1}{\lim\limits_{y \to 0} \dfrac{\log_a(1+y)}{y}}$$

$$= \frac{1}{\log_a e} = \log_e a$$

$$\therefore \lim_{x \to 0} \frac{a^x - 1}{x} = \log_e a \qquad (25)$$

Corollary:

$$\lim_{x \to 0} \frac{\mathrm{e}^x - 1}{x} = \log_e \mathrm{e} = 1$$

In particular,

$$\lim_{x \to 0} \frac{7^x - 1}{x} = \log_e 7, \ \lim_{x \to 0} \frac{\mathrm{e}^x - 1}{x} = \log_e \mathrm{e} = 1.$$

Remark: Note that if $f(x) \to 0$, as $x \to 0$ and $k \neq 0$, then $t = k{\cdot}f(x) \to 0$ as $x \to 0$.

$$\therefore \lim_{x \to 0} \frac{a^{kf(x)} - 1}{kf(x)} = \lim_{t \to 0} \frac{a^t - 1}{t} = \log_e a$$

13a.14.1 Derivative of Exponential Function a^x (by the First Principle)

To prove $(\mathrm{d}/\mathrm{d}x)(a^x) = a^x \cdot \log_e a \ (a > 0, \ a \neq 1)$.

Proof: Let $f(x) = a^x \ (a > 0,\ a \neq 1)$

$\therefore f(x+h) = a^{x+h}$

$$\begin{aligned}
\text{Now, } \frac{\mathrm{d}}{\mathrm{d}x}(a^x) &= \frac{\mathrm{d}}{\mathrm{d}x} f(x). \\
&= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \quad \text{(definition of derivative)} \\
&= \lim_{h \to 0} \frac{a^{x+h} - a^x}{h} \\
&= \lim_{h \to 0} \frac{a^x(a^h - 1)}{h} \\
&= a^x \lim_{h \to 0} \frac{(a^h - 1)}{h} \\
&= a^x \cdot \log_e a \qquad \text{[using equation (25)]}
\end{aligned}$$

$$\therefore \frac{\mathrm{d}}{\mathrm{d}x}(a^x) = a^x \cdot \log_e a \tag{26}$$

In particular,

$$\frac{\mathrm{d}}{\mathrm{d}x}(5^x) = 5^x \cdot \log_e 5$$

and

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{e}^x) &= \mathrm{e}^x \cdot \log_e \mathrm{e} \\
&= \mathrm{e}^x (\because \log_e \mathrm{e} = 1)
\end{aligned}$$

Recall that earlier we had proved this result by differentiating both sides of the result.

$$\mathrm{e}^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

The result (26) can also be obtained as follows:

We have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}(a^x) &= \lim_{h \to 0} \frac{a^{x+h} - a^x}{h}, \quad \text{(provided the limit exists)} \\
&= a^x \lim_{h \to 0} \frac{(a^h - 1)}{h}
\end{aligned}$$

Now, we can put $a^h = e^{h \cdot \log_e a}$(16)

$$\therefore \frac{d}{dx}(a^x) = a^x \lim_{h \to 0} \frac{(e^{h \log_e a} - 1)}{h}$$

$$= a^x \lim_{h \to 0} \frac{1}{h}\left[\left(1 + h \log_e a + \frac{(h \log_e a)^2}{2!} + \frac{(h \log_e a)^3}{3!} + \cdots\right) - 1\right]$$

$$= a^x[\log_e a + 0 + 0 + 0 + \cdots] \text{ (taking } \lim_{h \to 0})$$

$$= a^x \cdot \log_e a$$

Note (11): By expressing a^h in the form $e^{h \log_e a}$, we can expand it by using the *exponential series.*

Corollary: Derivative of the Logarithmic Function $\log_a x$ (by the First Principle):

We have

$$\frac{d}{dx}(\log_a x) = \lim_{h \to 0} \frac{\log_a(x+h) - \log_a x}{h} \text{(provided the limit exists)}$$

$$= \lim_{h \to 0}\left[\frac{1}{h}\log_a\left(\frac{(x+h)}{x}\right)\right]$$

$$\frac{d}{dx}(\log_a x) = \lim_{h \to 0}\left[\frac{1}{h}\,\frac{x}{x}\log_a\left(\frac{x+h}{x}\right)\right]$$

$$= \lim_{h \to 0}\left[\frac{1}{h}\log_a\left(1 + \frac{h}{x}\right)\right] = \lim_{h \to 0}\left[\frac{1}{x}\log_a\left(1 + \frac{h}{x}\right)^{x/h}\right]$$

Put $(h/x) = t$. Therefore, as $h \to 0$, $t \to 0$. We get

$$\frac{d}{dx}(\log_a x) = \frac{1}{x} \lim_{t \to 0}\left[\log_a(1+t)^{1/t}\right] = \frac{1}{x}\log_a e \quad \left[\because \lim_{t \to 0}(1+t)^{1/t} = e\right]$$

$$= \frac{1}{x} \cdot \frac{1}{\log_e a} \text{(by change of base)}$$

$$\frac{d}{dx}(\log_a^x) = \frac{1}{x} \cdot \frac{1}{\log_e a}$$

(16) We know that any positive number x can be expressed in the exponential form as the following: $x = a^{\log_a x} = b^{\log_b x} = e^{\log_e x}$, ... where a, b, e, and so on are *positive* numbers other than 1. Now $a = e^{\log_e a}$. Therefore, $a^h = e^{h \log_e a}$.

In particular,

$$\frac{\mathrm{d}}{\mathrm{d}x}(\log_e x) = \frac{1}{x} \cdot \frac{1}{\log_e \mathrm{e}} = \frac{1}{x} \quad (\because \log_e \mathrm{e} = 1)$$

Corollary: If $y = \log_a[f(x)]$ $(a > 0,\ a \neq 1)$
Then,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\log_a[f(x)] = \frac{1}{f(x)\log_e a} \cdot f'(x) \text{ (by chain rule)}$$

In particular, for $y = \log_e[f(x)]$,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{f(x)\log_e \mathrm{e}} \cdot f'(x) = \frac{f'(x)}{f(x)} \quad (\because \log_e \mathrm{e} = 1)$$

13a.14.2 Derivatives of Different Exponential Functions: Graphical View

We know that, if $f(x) = \mathrm{e}^x$, then $f'(x) = \mathrm{e}^x$.

We ask the question: Does something similar hold for exponential functions, with other bases?

Consider the exponential functions, $y = 4^x$, $y = \mathrm{e}^x$, and $y = 2^x$ whose graphs are given in Figure 13a.9. We have the following:

$$\text{For, } y = \mathrm{e}^x, \frac{\mathrm{d}y}{\mathrm{d}x} = \mathrm{e}^x$$

$$\text{For, } y = 2^x = \mathrm{e}^{x\log_e 2}, \therefore \frac{\mathrm{d}y}{\mathrm{d}x} = (\log_e 2) \cdot \mathrm{e}^x$$

$$\text{For, } y = 4^x = \mathrm{e}^{x\log_e 4}, \therefore \frac{\mathrm{d}y}{\mathrm{d}x} = (\log_e 4) \cdot \mathrm{e}^x$$

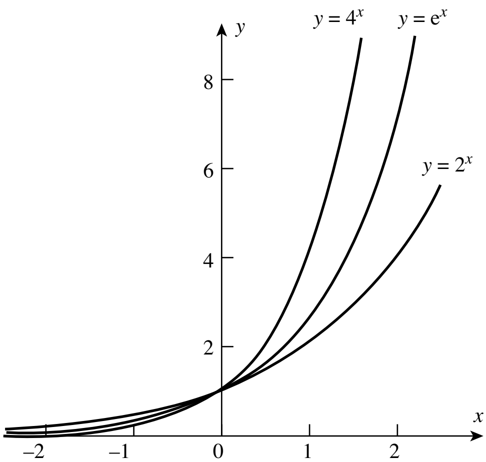


FIGURE 13a.9 Graphs of three exponentials.

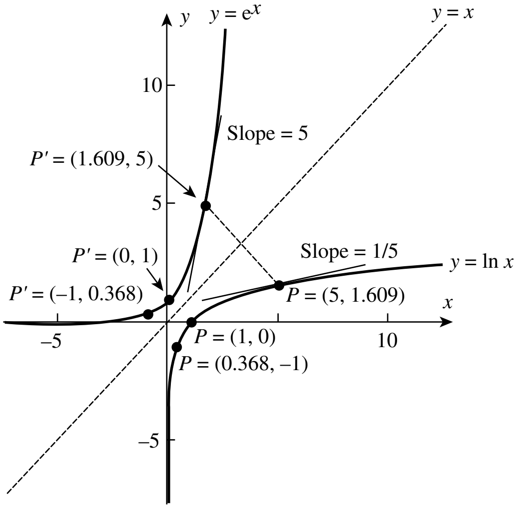


FIGURE 13a.10 Inverse functions and their derivatives.

Note that, at $x = 0$, the e^x graph has the slope 1, the graph of $y = 2^x$ has the slope $\log_e 2$, and the graph of $y = 4^x$ has the slope $\log_e 4$. The three graphs cross at the point (0, 1), but have different slopes there.

For any other value of x, the slope of a^x depends on a and it is $\log_e a$ times the slope of e^x at that value of x.

13a.14.3 Derivative of the Logarithmic Function: A Graphical View

We know that for any positive base, $a \neq 1$, the functions a^x and $\log_a x$ are inverses of one another. Geometrically, this relationship means that either graph can be obtained from the other by reflection around the line $y = x$. Figure 13a.10 shows the graph for the case $a = \mathrm{e}$ and illustrates a crucial point. It is interesting to observe how such a reflection affects the tangent lines.

We notice the following features of the graph:

(i) *Symmetric Points*: A point $p(x, y)$ lies on one graph if and only if the point $p'(y, x)$ lies on the other. Several such points are shown.

(ii) *Symmetric Tangent Lines*: Like the graphs themselves, the tangent lines at the symmetric points are symmetric to the line $y = x$. It follows that the slopes of the tangent lines at P and P' are reciprocals. (This is a key fact!) At points (0, 1) and (1, 0), the graphs of $y = e^x$ and $y = \ln x$ are parallel to each other and to the line $y = x$ (why?).

13a.15 APPLICATIONS OF THE FUNCTION e^x: EXPONENTIAL GROWTH AND DECAY

Earlier in this chapter, we have discussed exponential and logarithmic functions to the base e and respectively expressed them by the following equations:

$$y = e^x \ (x \in R) \tag{27}$$

$$x = \log y \, (x \in R, \ y > 0) \tag{28}$$

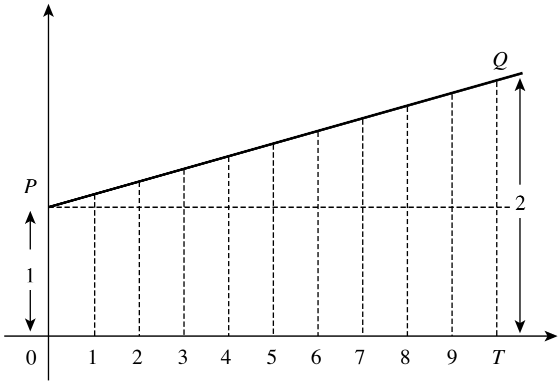


FIGURE 13a.11

We know that the relation (27) is in the *index form*, and the relation (28) expresses the same thing in the *log form*. We emphasize that these two equations mean the same thing and that they describe the same curve in the xy-plane.

Now, our interest is to discuss the applications of the exponential function e^x. For this purpose, we shall review (through *geometrical illustrations*) the growth of a capital by both simple interest and compound interest, so that the capital is doubled in a given interval of time.

In Figure 13a.11, *OP* stands for the original value of a function representing the capital (or the quantity). *OT* is the *whole time* during which the value is increasing (we treat this time interval as the unit time). Let the time interval be divided into 10 (equal) periods, each of which have an equal step-up; meaning there is an equal increase in the capital in each interval. We say that the value *OP* increases at a constant rate. This is also clear from the straight line *PQ* sloping up by equal steps. Here, dy/dx is a constant.

Note (12): To learn the subject of calculus, it is important to understand clearly the meaning of the symbol dy/dx. We know that if y is a function of x given by $y = f(x)$, then dy/dx stands for the (instantaneous) rate of change of y $[=f(x)]$ with respect to x and it is generally different from the "average rate of change" of y with respect to x, which we denote by $(\Delta y/\Delta x)$. It is only in the case of functions of the form $y = ax + b$ that both these rates are equal. This is so because dy/dx also stands for the "slope of the curve" at a point that varies from point to point, but the slope of the straight line is same at each point. Here, each step-up is 1/10 of the original *OP*; so with 10 such steps, the height is doubled. If we had taken 20 steps, each being half the height shown, at the end (of 20 steps) the height would still have just doubled. Obviously, n such steps, each being $1/n$ of the original height *OP*, would suffice to double the height. This is the case of simple interest.

Figure 13a.12 illustrates the corresponding geometrical progression. Each of the successive ordinates is to be $1 + (1/n)$ [i.e., $(n+1)/n$] times as high as its predecessor.[17]

[17] Recall that in a geometric progression (or a geometric sequence), the ratio of each term to the one after it is a constant. The general term of a geometric progression is expressed by ar^n, where r represents the common ratio and a stands for the first term (i.e., when $n = 0$). If $r > 1$, each term of the sequence increases with n, but if r is a proper fraction (i.e., $0 < r < 1$), then the terms keep on decreasing as n increases. Obviously, the general term of a geometric sequence represents the $(n + 1)^{\text{th}}$ term.

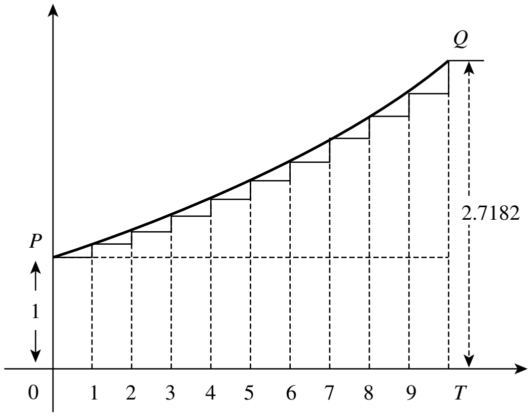


FIGURE 13a.12

The step-ups are not equal because each step-up is now 1/*n* of the ordinate at that part of the curve. If we literally had 10 steps, with (1 + (1/10)) for the multiplying factor, the final total would be $(1+(1/10))^{10}$ or 2.593 times the original 1. But, if we take *n* sufficiently large (and the corresponding (1/*n*) sufficiently small), then the final value $(1+(1/n))^n$ to which 1 will grow will be 2.7182818, and so on. Mathematicians have assigned this (mysterious) number the English letter e. It is an *irrational number* (and a transcendental number). It is known that e is even more important than π.

The process of growing proportionality at every instant to the magnitude at that instant is called the *exponential rate of growth*. Some people call it a *logarithmic rate of growth*, since the quantity y $(=e^x)$ grows with the exponent x, which is the logarithm of y to the base e. It might also be called the *organic rate of growing*, because it is characteristic of organic growth that (in certain circumstances) the increment of the organism in a given time is proportional to the magnitude of the organism itself. Some such examples are as follows:

- Some models of population growth assume that the rate of change of the population at any time t is proportional to the number y of individuals present at that time.
- In biology, under certain circumstances, the rate of growth of a culture of bacteria is proportional to the amount of bacteria present at any specific time.
- We may recall an application of these phenomena in business. When interest grows continuously, the number e originates in the process, as we have already seen.
- It is known from experiments that the rate of decay of radium is proportional to the amount of radium present at the given moment.

Besides, there are many physical processes in which something is gradually dying away. Mathematical models of these processes involve differential equations whose solutions contain powers of e. Some such processes will be discussed subsequently.

Now, we are in a position to discuss the phenomenon of growth and decay of a number (or a quantity or an amount) whose rate of increase (or decrease) is proportional to the number present at any given time.

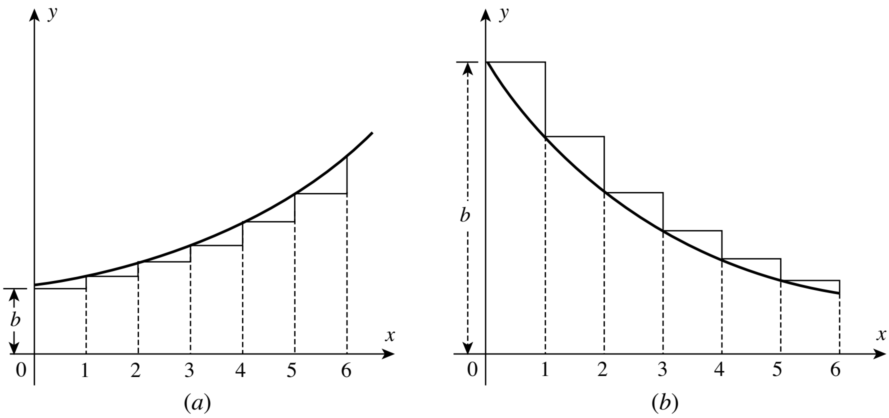


FIGURE 13a.13 (a) The curve showing exponential growth $f(x) = bp^{kx}$, $k > 0$. (b) The curve showing exponential decay $f(x) = bp^{kx}$, $k < 0$.

Suppose $y = f(t)$ represents the number at time t and (dy/dt) or $f'(t)$ represents the (instantaneous) rate of increase of the number at time t, then there is a positive constant k such that

$$\frac{dy}{dt} = k \cdot y[\text{i.e.,} f'(t) = kf(t)]^{(18)} \tag{29}$$

Also, we may consider a *different physical phenomenon* in which the number (or the quantity) is decreasing at a rate proportional to the number present at that time. (This happens in the case of a radioactive substance.) Then, the function $f(t)$ satisfies equation (29) for an appropriate negative constant k. Thus, equation (29) describes two quite different physical phenomena:

(a) The situation(s) in which something representing a number (or a quantity) is increasing at every instant proportional to the number at that instant. In such cases, the number k is a positive constant (see Figure 13a.13a).
(b) The situation(s) in which something representing a number (or a quantity) is decreasing at every instant proportional to the number at that instant. In such cases, the number k is a negative constant (Figure 13a.13b).

Remark: Note that both the situations (defining the phenomenon of growth or decay) are described only by a *geometrical progression* that is a function of the type $f(x) = bp^{kx}$, where k is a positive constant in the case of growth and a negative constant in the case of decay.

It would therefore be of interest to determine all such functions which satisfy equation (29). [In other words, we have to solve the differential equation (29).]

(18) Equation (29) involving the derivative f'(t) [i.e., dy/dt] of the function $y = f(t)$ is called a *differential equation*. We shall discuss the formation of differential equations and their solutions in Chapter 9a (in Part II of this book).

First, observe that the constant function 0 satisfies equation (29). Also, for $f(t) = e^{kt}$, we have $f'(t) = ke^{kt} = k \cdot f(t)$ [$\because$ $e^{kt} = f(t)$]. It follows that functions such as e^{kt} also satisfy equation (1) [recall that the exponential function e^t is the only function such that $(d/dt)(e^t) = e^t$].

Before we proceed to solve the differential equation (29), we make the following assumption that is in fact an important *idealization* of the function $f(t)$.

We know that if y [$=f(t)$] is the population of a certain community, then by definition y is a positive integer. However, to apply calculus to the phenomenon, we assume that y can be any positive real number, such that $y = f(t)$ represents a *continuous function* of t. The same logic is applicable when $f(t)$ represents an amount or a quantity.

13a.15.1 Solving the Differential Equation (29)

Consider a mathematical model given by the equation, $y = f(t)$, involving the law of *natural growth or decay* and the *initial condition* that $y = y_0$, when $t = 0$. Then, the differential equation formed at equation (29) is

$$\frac{dy}{dt} = ky[\text{or} f'(t) = kf(t)]$$

where k is a *constant* and $y > 0$ for all $t \geq 0$, with the initial conditions $y = y_0$ when $t = 0$ [i.e., $f(t) = f(0)$ when $t = 0$]. [In the above equation, the time is represented by t units (from $t = 0$ onward) and y represents the number of y units present (in the process) at any time t.]

Separating the variables, we obtain,

$$\frac{dy}{y} = k\,dt$$

On integrating, we get

$$\int \frac{dy}{y} = k \int dt$$

$$\therefore \log_e|y| = kt + c, \text{where } c \text{ is an arbitary constant}$$

$$\text{or } |y| = e^{kt+c} = e^c \cdot e^{kt}$$

Letting $e^c = b$, we get $|y| = be^{kt}$, and because y is positive, we can omit the absolute value bars and write,

$$y = be^{kt}, t \geq 0$$

Also, since $y = y_0$ when $t = 0$, we obtain from the above equation $b = y_0$ when $t = 0$. Thus, we get

$$y = y_0 \cdot e^{kt} \qquad (30)^{(19)}$$

(19) Note that, the function $y = y_0 \cdot e^{kt}$ represents a *geometrical progression* in which the base $e > 1$ and k is a constant for the given phenomenon. Also, it is clear that $y > 0$ for all $t \geq 0$ and that for $t = 0$, $y = y_0$.

Equation (30) gives us the form of the functions satisfying equation (29). We call it the solution of the differential equation (29).

If $k > 0$, then equation (29) is the law of natural growth and equation (30) defines a function that has *exponential growth.*

If $k < 0$, then equation (29) is the law of *natural decay* and equation (30) defines a function that has *exponential decay*.

It must also be noted that in the case of exponential growth, $f(t)$ increases without bound; whereas, in the case of exponential decay, $f(t)$ approaches 0 through positive values.

In fact, e^{-kt} serves as a *die away factor* for all those phenomena in which the rate of decrease (in our usual symbols dy/dt) is proportional at every moment to the value that is decreasing at that moment.

13a.15.2

We give below some processes in which the solution (of the above differential equation) given at equation (30) is applicable

- The cooling of a hot body is represented (in Newton's celebrated "Law of Cooling") by the equation $\theta_t = \theta_0 e^{-kt}$, where θ_0 is the *original excess of temperature* of a hot body over that of its surroundings, θ_t is the excess of temperature at the end of time t, and k is a constant, namely, the *constant of decrement* (here k depends on the amount of surface exposed by the body and on its coefficients of conductivity and emissivity, and other parameters).
- The formula

$$Q_t = Q_0 e^{-kt}$$

is used to express the *charge of an electrified body*, originally having a charge Q_0 that is *leaking away* with a constant of decrement k (here k depends on the capacity of the body and on the resistance of the leakage path).
- The oscillations given to a flexible spring die out after a time; and the dying out of the magnitude of the motion may be expressed in a similar way.

13a.15.2.1 The Time Constant In the expression for the *die away factor* e^{-kt}, let us replace k by another quantity $(1/T)$. Then, the die away factor will be written as $e^{(-t/T)}$ (note that the quantity k is represented by the reciprocal of another quantity T, which we call the time constant). Now, we may explain the meaning of T as follows:

In the die away factor $e^{(-t/T)}$, if we put $t = T$, the meaning of T [or of $(1/k)$] becomes clear. It means that T is the length of time that the (die away) process takes for the original quantity (θ_0 or Q_0, in the proceeding instances) *to die away to* $(1/e)^{\text{th}}$ part (i.e., to 0.3678) of its original value.

Example (1): Consider a hot body that is cooling. Suppose at the beginning of the experiment (i.e., when $t = 0$), it is 72° hotter than its surrounding objects. Let the time constant of this cooling be 20 min [which means that it takes 20 min for its excess of temperature to fall to $(1/e)^{\text{th}}$ part of 72°]. We can then calculate its temperature at any given time. For instance, suppose we wish to find the temperature of the hot body after 60 min. Here, $t = 60$ min and $T = 20$ min. Therefore, $(t/T) = 60 \div 20 = 3$. Now we can find the value of e^{-3} (from Table 13a.5] and then multiply the *original difference of temperature*, that is, 72°, by this number. The table shows that e^{-3} is 0.0498. Hence, at the end of 60 min, the excess of temperature will have fallen to $72° \times 0.0498 = 3.586°$.[20]

Similarly, we can compute the temperature of the hot body after 30 min. Here, $t = 30$ min and T (i.e., time constant) $= 20$ min. $\therefore \frac{t}{T} = \frac{30}{20} = 1.5$

From Table 13a.5, we get that $e^{-1.5} = 0.2231$. Hence, the temperature of the hot body after 30 min will be around $72° \times 0.2231 = 16.063°$.

Now, we proceed to discuss some real-life problems.

Example (2): The rate of increase of the population of a certain city is proportional to its population. In 1950, the population was 50,000 and in 1980 it was 75,000. (a) If y is the population after t years since 1950, express y as a function of t. (b) Estimate analytically what the population will be in 2010.

Solution: Let $y = f(t)$ denote the population after t years for $t \geq 0$. The differential equation is

$$\frac{dy}{dt} = ky \, [\text{or } f'(t) = k \cdot f(t)] \tag{31}$$

We know that the solution of the differential equation (31) is given by

$$y = y_0 e^{kt} \, [\text{or } f(t) = f(0)e^{kt}] \tag{32}$$

where k is a constant and $y_0 = 50{,}000$, when $t = 0$ [$f(0) = 50{,}000$, when $t = 0$] (this is the situation in the year 1950).

The following table indicates the boundary conditions:

Units of time t (in the year)	$t = 0$(in 1950)	$t = 30$ (in 1980)	$t = 60$ (in 2010)
Units of population y (in numbers)	$y_0 = 50{,}000 = f(0)$	$y_{30} = f(30) = 75{,}000$	$y_{60} = f(60) = ?$

Now, in view of the *solution* at equation (32) and the *boundary conditions*, we have

$$f(t) = 50{,}000e^{kt} \tag{33}$$

and our interest is to find $f(60)$.

In order to find $f(60)$, we first determine the value of k and then apply the formula (32) to obtain the value of $f(60)$ (or y_{60}). It is given that in 1980 (i.e., after 30 from 1950) the population has grown to 75,000. (Thus, $y(30) = 75{,}000$).

Using this information in equation (33), we get

$$75{,}000 = 50{,}000e^{30k}$$

$$\therefore e^{30k} = \frac{75{,}000}{50{,}000} = \frac{3}{2} \tag{34}$$

[20] *Calculus Made Easy* by S.P. Thomson (a fellow of Royal society), published in 1948.

Now, it is easy to compute $f(60)$. We can utilize the information of equation (34) in equation (33) as follows:

We have

$$f(t) = 50,000e^{kt}$$

or

$$f(t) = 50,000\,(e^{30k})^{t/30}\left[\because kt = (30k)\frac{t}{30}\right]$$

Now, we can put $t = 60$ in the above equation so that we get

$$\begin{aligned} f(60) &= 50,000\left(\frac{3}{2}\right)^2 = 50,000 \cdot \frac{9}{4} \\ &= 12,500 \times 9 \\ &= 1,12,500 \end{aligned}$$

Thus, the population in the year 2010 will be 1,12,500. Ans.

Example (3): The rate of decay of radium is proportional to the amount present at any time. The half-life of radium is 1690 years and 20 mg of radium is present now.

(a) If y mg of radium will be present t years from now, express y as a function of t.
(b) Estimate how much radium will be present 1000 years from now.

Solution: The boundary conditions are recorded in the table given below, where it is indicated by Y_{1000}, the units of (the material) radium in milligrams that will be present after 1000 years from now.

Units of time t (in years)	0	1690 (half-life of radium)	1000 years
Units of radium (in mg)	20	10 [$=y$]	Y_{1000} (remaining quantity)?

The differential equation is

$$\frac{dy}{dt} = ky \tag{35}$$

where k is a constant and $y = 20$ when $t = 0$ [we say $y_0 = 20 = f(0)$ when $t = 0$]. The solution of the differential equation (35) is known to be

$$y = y_0 e^{kt} = 20e^{kt} = f(t) \tag{36}$$

Our interest is to find the value of y_{1000} or $f(1000)$.

It is given that $y = 10$ when $t = 1690$.

Therefore, from equation (36), we get

$$10 = 20e^{1690k}$$

or

$$e^{1690k} = \frac{10}{20} = \frac{1}{2} \tag{37}$$

Now, we can compute the quantity Y_{1000} [or $f(1000)$] using equation (36), along with the information available at equation (37). From equation (36), we have

$$\begin{aligned} f(t) &= 20e^{kt} \\ \therefore f(t) &= 20 \cdot (e^{1690k})^{t/1960} \text{ (where } t = 1000) \\ \therefore f(1000) &= 20 \cdot \left(\frac{1}{2}\right)^{t/1960} \\ &= 20 \cdot \left(\frac{1}{2}\right)^{1000/1960} \quad (\because t = 1000) \\ &= 20 \cdot \left(\frac{1}{2}\right)^{0.5917} \\ &= 13.27 \end{aligned}$$

Thus, 1000 years from now, 13.27 mg of radium will be present out of 20 mg. Ans.

Example (4): In a certain culture, the rate of growth of bacteria is proportional to the amount present. Initially, 1000 bacteria are present and the amount doubles in 12 min.

(a) If y bacteria are present at t min, express y as a function of t.
(b) Estimate to the nearest minute how long will it take for 10,000 bacteria to be present.

Solution: The following table gives the boundary conditions where y bacteria are present at t min. Suppose that it will take t min for 10,000 bacteria to be present.

Units of time t (in min)	0	12	t?
Units of bacteria (y in numbers)	1000	2000	10,000

The differential equation is

$$\frac{dy}{dt} = ky \tag{38}$$

where k is a constant and $y = 1000$ when $t = 0$ [we say $y_0 = 1000 = f(0)$ when $t = 0$].

The solution of differential equation (38) is known to be

$$y = y_0 e^{kt} \text{ or } f(t) = 1000e^{kt} \tag{39}$$

Our interest is to find the value of t when $f(t) = 10{,}000$. But, it is given that $y = 2000$ when $t = 12$. From this information, we obtain from equation (39)
$f(12) = 1000e^{12k}$ or $2000 = 1000e^{12k}$ [$\therefore f(12) = 2000$].

$$\therefore e^{12k} = \frac{2000}{1000} = 2 \tag{40}$$

We use equation (40) in equation (39) to obtain the value of t for which $f(t) = 10000$. Equation (39) tells us that $f(t) = 1000e^{12kt}$ or $f(t) = 1000(e^{12k})^{t/12}$. Now, $10{,}000 = 1000(2)^{t/12}$ [$\therefore kt = 12k(t/12)$].

$$\therefore\ 10{,}000 = 1000(2)^{t/12}$$

$$\therefore\ (2)^{t/12} = 10$$

$$\therefore\ \frac{t}{12}\log_e 2 = \log_e 10$$

$$\therefore\ t = \frac{12\log_e 10}{\log_e 2}$$

$$= \frac{12[2.3026]}{0.6931} \qquad \text{(using Table 13a.4)}$$

$$= \frac{27.6312}{0.6931} = 39.86 \qquad \text{(using calculator)}$$

Thus, in 40 min, 10,000 bacteria will be present. Ans.

Definition: The time that a population takes to double is called its *doubling time*.
If the population grows exponentially with doubling time d; then, that time t is given by

$$F(t) = f(0)2^{t/d}$$

Definition: The half-life of a radioactive substance is the length of time it takes for half of a given amount of the substance to disintegrate through radiation.

Note: The half-life of C^{14} by international agreement is 5568 years. However, recent measurements indicate that the half-life of C^{14} is actually closer to 5730 years.

If the half-life of a substance having exponential decay is h years and $f(0)$ units of the substance are present now, then $f(t)$ units will be present in t years, where $f(t) = f(0)\left(\frac{t}{12}\right)^{t/h}$

13b Methods for Computing Limits of Exponential and Logarithmic Functions

13b.1 INTRODUCTION

In Chapter 13a, we have studied *exponential* and *logarithmic functions* and plotted their graphs. In this Chapter, our interest lies in learning the methods that help in computing limits of functions, which are in the exponential form, and those that involve exponential or logarithmic functions.[(1)]

For this purpose, it is useful to review in brief the topic of logarithms and then see how to evaluate limits of these functions. For convenience, we will also list some *basic limits* and some *standard limits*, which were proved (or accepted) earlier.

13b.2 REVIEW OF LOGARITHMS

If three numbers a, b, and c are related such that,

$$a^b = c(a > 0,\ a \neq 1) \tag{I}$$

then, the exponent b is called the logarithm of c to the base a.

We write,

$$\log_a c = b \tag{II}$$

Definition: Let a be a positive real number ($a \neq 1$) and y be any given real number. If there is a number x such that

$$a^x = y$$

then x is called the logarithm of y to the base a and we write $\log_a y = x$.

13b-Methods for computing limits of exponential and logarithmic functions

[(1)] Here are some examples of the type of limit(s) that we will learn to evaluate:

$\lim_{x\to 0}\left(\frac{5+x}{5-x}\right)^{1/x}$; $\lim_{x\to 0}\frac{a^x - b^x}{x}$; $\lim_{x\to 4}(x-3)^{1/(x-4)}$; $\lim_{x\to 2}\left(\frac{\log x - \log 2}{x-2}\right)$; $\lim_{x\to e}\left(\frac{\log x - 1}{x-e}\right)$; $\lim_{x\to 0}\left(\frac{8^{\sin x} - 1}{\sin x}\right)$; $\lim_{x\to 0}\left(\frac{12^x - 4^x - 3^x + 1}{x\sin x}\right)$; and so on

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and A. K. Ghosh.

Note (1): From the above definition, we have

(i) If $a^x = y$ (1)

then $\log_a y = x$ (2)

Substituting for y [from equation (1)] in equation (2), we get

$$\log_a a^x = x \quad (3)^{(2)}$$

Again, substituting for x [from equation (2)] in equation (1), we get

$$a^{\log_a y} = y \quad (4)^{(3)}$$

(ii) $a^1 = a, \quad \therefore \quad \log_a a = 1$ (5)

(iii) $a^0 = 1, \quad \log_1 a = 0$ (6)

13b.2.1 Laws of Logarithms

(i) $\log_a xy = \log_a x + \log_a y$
(ii) $\log_a \frac{x}{y} = \log_a x - \log_a y$
(iii) $\log_a x^m = m \log_a x$
(iv) Change of base:

$$\log_a x = \frac{\log_b x}{\log_a b} \quad (7)$$

If we write $x = b$ in the above equation, we get

$$\log_a b = \frac{\log_b b}{\log_b a} = \frac{1}{\log_b a} \quad (\because \quad \log_b b = 1) \quad (8)$$

The relation (7) tells that we can express $\log_a x$ in terms of $\log_b x$, wherein the base a is changed to a new base b, $\log_b a$ being a constant.

Next, the relation (8) tells us that $\log_b a \cdot \log_a b = 1$.

We know that there are two important bases: 10 and e. In the system of logarithms, which we use in our day-to-day calculations (such as in the field of engineering), the base 10 is found to be the most useful. Logarithms to the base 10 are called *common logarithms*. Logarithms to the base e are called *natural logarithms* and they are useful in calculus.[4]

(Recall that if the base is e, then the result of differentiating the functions $\log_e x$ and e^x assume simpler forms.) Besides, for all practical purposes, we can always convert back and forth between natural and common logarithms. Therefore, throughout this course, we are going

[2] Equation (3) tells that any real number x can be expressed in log form.
[3] Equation (4) tells that any real number y can be expressed in *exponential form*.
[4] The logarithmic base e is "natural" only in the sense that it is "naturally convenient" in order to make the standard process of differentiation work out simply for a logarithmic function (for details, see Chapter 13a).

to use natural logarithms only. We may or may not write the base. Thus, even when we write $\log x$, we shall mean $\log_e x$.

Note (2):

(i) $\log \mathrm{e} = \log_e \mathrm{e} = 1$ [see equation (5)]

(ii) $\log \mathrm{e}^x = \log_e \mathrm{e}^x = x$ [see equation (3)]

(iii) $\mathrm{e}^{\log x} = \mathrm{e}^{\log_e y} = y$ [see equation (4)]

13b.3 SOME BASIC LIMITS

The following *basic limits* are used for evaluating the limits of exponential and logarithmic functions.

1. $\lim\limits_{x \to p} a^x = a^p$, whenever a^p is defined.
2. (i) If $0 < a < 1$, then $\lim\limits_{x \to \infty} a^x = 0$ $(x \in R)$.

   (ii) If $a > 1$, then $\lim\limits_{x \to \infty} a^x = \infty$ $(x \in R)$.
3. If $p > 1$, then $\lim\limits_{x \to p} \log_a x = \log_a p$.
4. (i) If $a > 1$, then $\lim\limits_{x \to \infty} \log_a x = \infty$.

   (ii) If $0 < a < 1$, then $\lim\limits_{x \to \infty} \log_a x = -\infty$.

In Chapter 13a, we have seen that

- $$\lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = \mathrm{e} \quad (n \in N) \qquad (9)$$

  It can also be shown (by *substitution*) that

- $$\lim_{x \to \infty} \left(1 + \frac{1}{x}\right)^x = \mathrm{e} \quad (x \in R) \qquad (10)$$
- $$\lim_{x \to 0} (1 + x)^{1/x} = \mathrm{e} \quad (x \in R) \qquad (11)$$
- $$\lim_{x \to 0} (1 + kf(x))^{1/(kf(x))} = \mathrm{e} \quad (k \neq 0) \qquad (12)$$

Note: Limits at (9)–(11) are considered as standard limits. The limit at (12) can be expressed in the form (11) by substitution, and then we can use the standard result (11).

The following *important limits* have already been proved in Chapter 13a.

1. $$\lim_{x \to 0} \frac{\log_a(1 + x)}{x} = \log_a \mathrm{e} = \frac{1}{\log_e a} \qquad (13)$$
2. $$\lim_{x \to 0} \frac{a^x - 1}{x} = \log_e a \qquad (14)$$

Corollary: $\lim\limits_{x \to 0} \dfrac{\mathrm{e}^x - 1}{x} = \log_e \mathrm{e} = 1.$

Note (4): The limit at equation (13) is not considered as a standard limit. However, it is a very important limit, since it is used in proving the limit at equation (14), which is treated as a standard limit.

First, we will evaluate certain limits that can be evaluated by applying the results given in equations (9)–(12). Later on, we will recall the proof of the result (13) and then evaluate certain limits wherein the result (14) is applicable.

Example (1): $\lim\limits_{x \to 0} \left(\dfrac{3+2x}{3-2x}\right)^{1/x}$

Solution: $\lim\limits_{x \to 0} \left(\dfrac{3+2x}{3-2x}\right)^{1/x} = \lim\limits_{x \to 0} \left(\dfrac{1+(2x/3)}{1-(2x/3)}\right)^{1/x} = L\,(\text{say})$

First consider $\lim\limits_{x \to 0} \left(1+\dfrac{2x}{3}\right)^{1/x}$

(If we put $2x/3 = t$, then $3/(2x) = 1/t$. Furthermore, note that as $x \to 0$, $t \to 0$ and $1/t \to \infty$.)

$$= \left[\lim_{x \to 0}\left(1+\frac{2x}{3}\right)^{3/2x}\right]^{2/3} = \left[\lim_{t \to 0}(1+t)^{1/t}\right]^{2/3}$$

$$= e^{2/3} \text{ since } \left[\lim_{t \to 0}(1+t)^{1/t} = e\right]$$

Next, consider,

$$\frac{1}{\lim\limits_{x \to 0}(1-(2x/3))^{1/x}}$$

$$= \lim_{x \to 0}\left(1-\frac{2x}{3}\right)^{-(1/x)} = \left[\lim_{x \to 0}\left(1-\frac{2x}{3}\right)^{-(3/2x)}\right]^{2/3}$$

$$= \left[\lim_{x \to 0}\left(1-\frac{2x}{3}\right)^{(3/2x)}\right]^{-(2/3)} = e^{-(2/3)}$$

$$\therefore \quad L = \frac{e^{(2/3)}}{e^{-(2/3)}} = e^{(4/3)} \qquad \text{Ans.}$$

Example (2): $\lim\limits_{x \to \infty} \left(\dfrac{2x+3}{2x-1}\right)^{x+1}$

Solution: $L = \lim_{x\to\infty}\left(\frac{2x+3}{2x-1}\right)^x \lim_{x\to\infty}\left(\frac{2x+3}{2x-1}\right)$

$$\lim_{x\to\infty}\left(\frac{1+(3/2x)}{1-(1/2x)}\right)^x \cdot \left(\frac{1+(3/2x)}{1-(1/2x)}\right)$$

$$\lim_{x\to\infty}\frac{(1+(3/2x))^x}{(1-(1/2x))^x}\cdot\frac{(1+(3/2x))}{(1-(1/2x))}$$

Now, consider,

$$\left(1+\frac{3}{2x}\right)^x$$

$$= \left[\lim_{x\to\infty}\left(1+\frac{3}{2x}\right)^{(2x/3)}\right]^{3/2} \quad \left[\text{As } x\to\infty, \quad \frac{3}{2x}\to 0, \quad \text{and} \quad \frac{2x}{3}\to\infty\right]$$

$$= e^{3/2}, \text{ since } \lim_{x\to 0}(1+t)^{1/t} = e, \text{ where } t = 3/(2x)$$

Next, consider,

$$\lim_{x\to\infty}\left(1-\frac{1}{2x}\right)^x$$

$$= \left[\lim_{x\to\infty}\left(1-\frac{1}{2x}\right)^{-2x}\right]^{-(1/2)} = e^{-(1/2)}$$

$$\therefore \quad L = \frac{e^{(3/2)}}{e^{-(1/2)}}\cdot\frac{(1+0)}{(1+0)} = e^{(3/2)+(1/2)} = e^2 \qquad \text{Ans.}$$

Note that,

$$\lim_{x\to\infty}\left(1-\frac{1}{2x}\right)^{-2x} = \lim_{x\to\infty}\left(1+\frac{1}{-2x}\right)^{(-2x)} = \frac{1}{\lim_{x\to\infty}\left[1+\left(\frac{1}{-2x}\right)\right]^{2x}}$$

Furthermore, note that as $x\to\infty$, $(1/-2x)\to 0$ and $2x\to\infty$. Accordingly, the limit in question (in the denominator) is $e^{-1/2}$

Example (3): Evaluate; $\lim_{x\to 1} x^{(1/(x-1))}$

Solution: Let $\lim_{x\to 1} x^{(1/(x-1))} = L$

Put $x-1=t$. Therefore, $x=1+t$. Note that, as $x\to 1,\ t\to 0$,

$\therefore \quad L = \lim_{t\to 0}(1+t)^{1/t}$

But $\lim_{x\to 1}(1+t)^{1/t} = \text{e}$.

$\therefore$ the limit in question = e. Ans.

Example (4): Evaluate; $\lim_{x\to 4}(x-3)^{(1/(x-4))}$

Solution: Let $\lim_{x\to 4}(x-3)^{(1/(x-4))} = L$.

Put $x-4=t$. Therefore, $x=4+t$. Note that, as $x\to 4,\ t\to 0$,

$$\therefore \quad L = \lim_{t\to 0}(4+t-3)^{1/t} = \lim_{t\to 0}(1+t)^{1/t}$$

(Note that, as $t\to 0$, $(1/t)\to\infty$.)

$$\therefore \quad L = \lim_{t\to 0}(1+t)^{1/t} = \text{e}. \qquad \text{Ans.}$$

Exercise (1): Evaluate the following limits:

(i) $\lim_{x\to 0}\left(\dfrac{1-x}{1+x}\right)^{(1/x)}$

(ii) $\lim_{x\to 0}\left(1-\dfrac{x}{2+3x}\right)^{(1/x)}$

(iii) $\lim_{x\to 0}\left(\dfrac{3+2x}{3-x}\right)^{(1/x)}$

(iv) $\lim_{x\to 0}\left(\dfrac{5+2x}{5-x}\right)^{(1/x)}$

(v) $\lim_{x\to 0}\left(\dfrac{3+2x}{3-2x}\right)^{(1/x)}$

(vi) $\lim_{x\to\infty}\left(1+\dfrac{2}{x}\right)^{x}$

(vii) $\lim_{x\to\infty}\left(\dfrac{3x+1}{3x-1}\right)^{x}$

(viii) $\lim_{x\to 2}(x-3)^{(1/(x-4))}$

Answer:

(i) e^{-2}

(ii) $\text{e}^{-1/2}$

(iii) e

(iv) $\text{e}^{3/5}$

(v) $\text{e}^{4/3}$

(vi) e^2
(vii) 1/e
(viii) e

(3) Now, we recall for convenience, the proof of the result (15) that was proved in Chapter 13a. This is an important limit.

To show that

$$\lim_{x \to 0} \frac{\log_a(1+x)}{x} = \frac{1}{\log_e a} \tag{15}$$

Solution:

Consider

$$\begin{aligned}
\lim_{x \to 0} \frac{\log_a(1+x)}{x} &= \lim_{x \to 0} \frac{1}{x} \cdot \log_a(1+x) \\
&= \lim_{x \to 0} \log_a(1+x)^{(1/x)} \\
&= \log_a\left[\lim_{x \to 0} (1+x)^{(1/x)}\right] \\
&= \log_a \mathrm{e} \text{ (using C)} \\
&= \frac{1}{\log_e a} \text{ (by change of base)} \quad \text{(Proved)}
\end{aligned}$$

Corollary:

$$\lim_{x \to 0} \frac{\log_e(1+x)}{x} = 1$$

Solution:

$$\begin{aligned}
&= \lim_{x \to 0} \frac{\log_e(1+x)}{x} = \lim_{x \to 0} \frac{1}{x} \log_e(1+x) = \lim_{x \to 0} \log_e(1+x)^{(1/x)} \\
&= \log_e \mathrm{e} = 1 \quad \text{(Proved)}
\end{aligned}$$

Important Note (5): We know that the limit (15) is *not considered as a standard limit*. Hence, we cannot use this result directly in evaluating other limits. Accordingly, to evaluate any limit that involves *logarithm of a function*, we must express the *inner function* in a suitable standard form whose limit can be evaluated using standard limits. We proceed as follows:

(i) Use suitable substitutions and the properties of logarithms to simplify the given limit.

(ii) Express the given expression in the form of a *logarithm of an expression* and then modify the inner function suitably so that its limit can be evaluated using standard limit(s).

(iii) Evaluate the given limit by expressing the limit $\lim_{x \to a} \log_e[f(x)]$ in the form $\log_e\left[\lim_{x \to a} f(x)\right]$.

The following solved examples indicate the various steps involved.

Example (5): Evaluate

$$\lim_{x \to 3} \frac{\log x - \log 3}{x - 3}$$

Solution: Let

$$\lim_{x \to 3} \frac{\log x - \log 3}{x - 3} = L^{(5)}$$

Put $x - 3 = t$. Therefore, $x = 3 + t$. Note that as $x \to 3$, $t \to 0$.

Thus,

$$L = \lim_{t \to 0} \frac{\log(3+t) - \log 3}{t} = \lim_{t \to 0} \frac{1}{t}\left(\log\left(\frac{3+t}{3}\right)\right)$$

$$= \lim_{t \to 0} \log\left(1 + \frac{t}{3}\right)^{(1/t)} = \log\left[\lim_{t \to 0}\left(1 + \frac{t}{3}\right)^{(3/t)}\right]^{(1/3)}$$

[Note that, as $t \to 0$, $(t/3) \to 0$.]

$$\therefore \quad L = \log \mathrm{e}^{1/3} = (1/3)\log_e \mathrm{e} = (1/3) \qquad \text{Ans.}$$

Note (6): When we are considering logarithms to the base e, it is conventional not to write the base e. Thus, $\log x$ means $\log_e x$.

Example (6): Evaluate

$$\lim_{x \to \mathrm{e}} \frac{\log x - 1}{x - \mathrm{e}}$$

Solution: Let

$$\lim_{x \to \mathrm{e}} \frac{\log x - 1}{x - \mathrm{e}} = L$$

(5) Note that the method of substitution is very important in converting the given expression to the desired form.

Put $x - \mathrm{e} = t$. Therefore, $x = \mathrm{e} + t$. Also, note that as $x \to \mathrm{e}$, $t \to 0$.

$$L = \lim_{t \to 0}\left(\frac{\log(\mathrm{e}+t) - 1}{t}\right)$$

$$= \lim_{t \to 0}\left(\frac{\log(\mathrm{e}+t) - \log \mathrm{e}}{t}\right) \quad [\because \quad \log_e \mathrm{e} = 1]$$

$$= \lim_{t \to 0}\left(\frac{1}{t}\log\frac{\mathrm{e}+t}{\mathrm{e}}\right) = \lim_{t \to 0} \log\left(1+\frac{t}{\mathrm{e}}\right)^{(1/t)}$$

$$= \log\left[\lim_{t \to 0}\left(1+\frac{t}{\mathrm{e}}\right)^{(\mathrm{e}/t)}\right]^{(1/\mathrm{e})}$$

[Note that, as $t \to 0$, $(t/\mathrm{e}) \to 0$.]

$$\therefore \quad L = \log \mathrm{e}^{1/\mathrm{e}} = \frac{1}{\mathrm{e}}\log_e \mathrm{e} = \frac{1}{\mathrm{e}} \qquad \text{Ans.}$$

Example (7): Evaluate

$$\lim_{x \to 0} \frac{\log 10 + \log(x + 0.1)}{x}$$

Solution: Let

$$\lim_{x \to 0} \frac{\log 10 + \log(x + 0.1)}{x} = L$$

Consider, $\log 10 + \log(x + 0.1)$

$$= \log 10 + \log\left(\frac{10x+1}{10}\right)$$

$$= \log 10 + \log(10x + 1) - \log 10 = \log(1 + 10x)$$

Therefore, the given limit can be expressed in the form

$$L = \lim_{x \to 0}\frac{\log(1+10x)}{x} = \lim_{x \to 0}\frac{1}{x}\log(1+10x) = \lim_{x \to 0}\log(1+10x)^{(1/x)}$$

$$= \lim_{x \to 0}\log\left[(1+10x)^{(1/10x)}\right]^{10} = \log\left[\lim_{x \to 0}(1+10x)^{(1/10x)}\right]^{10}$$

$$= \log \mathrm{e}^{10} = 10\log_e \mathrm{e} = 10 \qquad \text{Ans.}$$

Exercise (2): Evaluate the following limits:

(i) $\lim_{x \to 2} \dfrac{\log x + \log 2}{x - 2}$ [Ans: (1/2)]

(ii) $\lim_{x \to 0} \dfrac{\log(1 + (8x/3))}{x}$ [Ans: (8/3)]

(iii) $\lim_{x \to 0} \dfrac{\log(5 + x) - \log(5 - x)}{x}$ [Ans: (2/5)]

(iv) $\lim_{x \to 2} \dfrac{1}{x}[\log(3 + x) - \log(3 - x)]$ [Ans: (2/3)]

13b.4 EVALUATION OF LIMITS BASED ON THE STANDARD LIMIT

$\lim_{x \to 0}((a^x - 1)/x) = \log_e a$, where $a > 0$ (we have proved this limit in Chapter 13a).

Corollary: Replacing a by e, we get

$$\lim_{x \to 0} \frac{e^x - 1}{x} = \log_e e = 1$$

Furthermore, if $f(x) \to 0$ as $x \to 0$ and k is a *nonzero number*, then

$$t = k \cdot f(x) \to 0 \text{ as } x \to 0$$

$$\therefore \quad \lim_{x \to 0} \frac{a^{k \cdot f(x)} - 1}{k \cdot f(x)}$$

$$= \lim_{t \to 0} \frac{a^t - 1}{t} = \log_e a$$

Example (8): Evaluate

$$\lim_{x \to 0} \frac{a^x - b^x}{x}$$

Solution: $\lim_{x \to 0} \dfrac{a^x - b^x}{x} = \lim_{x \to 0} \dfrac{(a^x - 1) - (b^x - 1)}{x}$

$$= \lim_{x \to 0} \frac{(a^x - 1)}{x} - \frac{(b^x - 1)}{x}$$

$$= \log_e a - \log_e b$$

$$= \log_e \frac{a}{b} \qquad \text{Ans.}$$

Example (9): Evaluate

$$\lim_{x \to 0} \frac{3^{8x} - 1}{x}$$

Solution: Let

$$\lim_{x \to 0} \left(\frac{3^{8x} - 1}{x} \right) = L$$

$$L = \lim_{x \to 0} \frac{3^{8x} - 1}{x} = \left[\lim_{x \to 0} \frac{3^{8x} - 1}{8x} \right] \cdot 8$$

Put $8x = t$. Then, as $x \to 0$, $t \to 0$.

$$L = \lim_{t \to 0} \left(\frac{3^t - 1}{t} \right) \cdot 8$$

$$= (\log_e 3) \cdot 8$$

$$= 8 \cdot \log_e 3 \qquad \text{Ans.}$$

Example (10): Evaluate

$$\lim_{x \to 0} \frac{e^x - e^{-x}}{\sin x}$$

Solution:

$$\lim_{x \to 0} \frac{e^x - e^{-x}}{\sin x} = \lim_{x \to 0} \frac{e^{2x} - 1}{e^x \sin x} = L, \text{ say.}$$

$$\therefore \quad L = \lim_{x \to 0} \left[(e^{2x} - 1) \cdot \frac{1}{\sin x} \cdot \frac{1}{e^x} \right] = \lim_{x \to 0} \left[\frac{e^{2x} - 1}{2x} \cdot \frac{2x}{\sin x} \cdot \frac{1}{e^x} \right]$$

$$\text{or } \; L = \lim_{x \to 0} \left[\frac{(e^{2x} - 1)}{2x} \cdot \frac{2x}{\sin x} \cdot \frac{1}{e^x} \right]$$

$$= \lim_{x \to 0} \left(\frac{e^{2x} - 1}{2x} \right) \cdot 2 \left(\lim_{x \to 0} \frac{x}{\sin x} \right) \cdot \lim_{x \to 0} \left(\frac{1}{e^x} \right)$$

$$\therefore \quad L = \lim_{x \to 0} \left(\frac{e^{2x} - 1}{2x} \right) \cdot 2 \left(\lim_{x \to 0} \frac{1}{\sin x / x} \right) \cdot \lim_{x \to 0} \left(\frac{1}{e^x} \right)$$

$$= \lim_{x \to 0} \left(\frac{e^{2x} - 1}{2x} \right) \cdot 2 \left(\frac{1}{\lim_{x \to 0} (\sin x / x)} \right) \cdot \lim_{x \to 0} \left(\frac{1}{e^x} \right)$$

$$= (1) \cdot (2) \cdot (1) \cdot (1) \left[\because \quad \lim_{t \to 0} \frac{e^t - 1}{t} = \log_e e = 1, \quad \lim_{x \to 0} \frac{\sin x}{x} = 1, \text{ and } \lim_{x \to 0} e^x = 1 \right]$$

$$= 2 \qquad \text{Ans.}$$

Example (11): Evaluate

$$\lim_{x\to 0} \frac{(ab)^x - a^x - b^x + 1}{x^2}$$

Solution: Consider, $(ab)^x - a^x - b^x + 1$

$$= a^x b^x - a^x - b^x + 1$$
$$= a^x(b^x - 1) - 1(b^x - 1)$$
$$= (b^x - 1)\cdot(a^x - 1)$$

$$\therefore \quad \text{The required limit} = \lim_{x\to 0} \frac{(b^x - 1)(a^x - 1)}{x^2}$$
$$= \lim_{x\to 0} \frac{(a^x - 1)}{x} \cdot \frac{(b^x - 1)}{x}$$
$$= \log_e a \cdot \log_e b \qquad \text{Ans.}$$

Example (12): Evaluate

$$\lim_{x\to 0} \frac{a^x + a^{-x} - 2}{x^2}$$

Solution: Consider $a^x + a^{-}x - 2$

$$= a^x + \frac{1}{a^x} - 2 = \frac{a^{2x} - 2a^x + 1}{x^2} = \frac{(a^x - 1)^2}{a^x} \text{ (This is the simplified numerator.)}$$

$$\therefore \quad \text{The required limit} = \lim_{x\to 0} \frac{(a^x - 1)^2}{x^2 \cdot a^x}$$

$$= \lim_{x\to 0} \left(\frac{a^x - 1}{x}\right)^2 \lim_{x\to 0} \frac{1}{a^x} = \left[\lim_{x\to 0} \frac{a^x - 1}{x}\right]^2 \lim_{x\to 0} \frac{1}{a^x}$$

$$= (\log_e a)^2 \frac{1}{a^0} = (\log_e a)^2 \cdot 1 = (\log_e a)^2 \qquad \text{Ans.}$$

The following example explains clearly the approach for evaluating limits involving exponential functions:

Example (13): Evaluate

$$\lim_{x\to 0} \frac{3^{5x} - 1}{\tan 3x}$$

Solution: Let

$$\lim_{x \to 0} \frac{3^{5x} - 1}{\tan 3x} = L$$

First, consider only the numerator:

$$3^{5x} - 1 = \left(\frac{3^{5x} - 1}{5x}\right) \cdot 5x$$

Now, consider the denominator:

$$\frac{1}{\tan 3x} = \frac{1}{(\sin 3x)(1/\cos 3x)} = \frac{1}{(\sin 3x/3x) \cdot 3x \cdot (1/\cos 3x)}$$

$$\therefore \quad L = \lim_{x \to 0} \frac{[((3^{5x} - 1)/5x) \cdot 5x]}{[((\sin 3x)/3x) \cdot 3x \cdot (1/\cos 3x)]}$$

$$\therefore \quad L = \lim_{x \to 0} \frac{((3^{5x} - 1)/5x) \cdot 5x}{(\sin 3x/3x)} (5x/3x)(\cos 3x)$$

$$= (\log 3) \cdot \frac{5}{3} \cdot 1 = \frac{5}{3}(\log 3) \qquad \text{Ans.}$$

Note carefully the important points in evaluating limits in the following two examples:

Example (14): Evaluate

$$\lim_{x \to 0} \frac{12^x + 4^x - 3^x - 1}{x}$$

Solution: Let

$$\lim_{x \to 0} \frac{12^x + 4^x - 3^x - 1}{x} = L$$

$$\therefore \quad L = \lim_{x \to 0} \frac{(12^x - 1) + (4^x - 1) - (3^x - 1)}{x}$$

(Note that the last two bracketed terms in the numerator keep the numerator unchanged.

$$= \lim_{x \to 0} \frac{(12^x - 1)}{x} + \frac{(4^x - 1)}{x} - \frac{(3^x - 1)}{x}$$

$$= \log_e 12 + \log_e 4 - \log_e 3$$

$$= \log_e\left(\frac{12 \times 4}{3}\right)$$

$$= \log_e 16 \qquad \text{Ans.}$$

Example (15): Evaluate

$$\lim_{x \to 0} \frac{12^x - 4^x - 3^x + 1}{x \sin x}$$

Solution: Let

$$\lim_{x \to 0} \frac{12^x - 4^x - 3^x + 1}{x \sin x} = L$$

[The points of difference between Examples (6) and (7) are as follows: In this example $12^x = (-4)^x(-3)^x$. This suggests that numerator can be factorized. Note that this was not the case in Example (6). Furthermore, the denominator has a product of two functions that may be suitably adjusted to apply the standard result(s).]

$$\therefore L = \lim_{x \to 0} \frac{4^x \cdot 3^x - 4^x - 3^x + 1}{x \sin x}$$

$$= \lim_{x \to 0} \frac{4^x(3^x - 1) - 1(3^x - 1)}{x \sin x}$$

This form of expression suggests that both the denominator and the numerator must be multiplied by x. Thus, we get

$$= \lim_{x \to 0} \frac{(3^x - 1)(4^x - 1)}{x \sin x} \cdot \frac{x}{x}$$

$$= \lim_{x \to 0} \frac{(3^x - 1)}{x} \cdot \frac{(4^x - 1)}{x} \cdot \left[\frac{1}{\left(\sin \frac{x}{x} \right)} \right]$$

$$L = \log_e 3 \cdot \log_e 4 \cdot 1 = (\log_e 3)(\log_e 4) \qquad \text{Ans.}$$

An important point here is that both problems look alike at a glance, but the distinction between the two must be carefully noted. This should help in solving similar problems.

Exercise (3): Evaluate the following limits:

(i) $\lim_{x \to 0} \frac{a^x + b^x - 2^{x+1}}{x}$

(ii) $\lim_{x \to 0} \frac{a^{3x} - a^{2x} - a^x + 1}{x \sin x}$

(iii) $\lim_{x \to 0} \frac{5^x - 5^{-x} - 2}{x^2}$

(iv) $\lim_{x \to 0} \frac{15^x - 5^x - 3^x + 1}{\sqrt{2 - \cos 2x} - 1}$

(v) $\lim_{x \to 0} \dfrac{e^x - e^{-x} - 2}{\cos 3x - \cos 7x}$

(vi) $\lim_{x \to 0} \dfrac{5^{\sin x} - 1}{\sin x}$

(vii) $\lim_{x \to 0} \dfrac{6^x - 3^x - 2^x + 1}{x^2}$

(viii) $\lim_{x \to 0} \dfrac{5^{2x} - 1}{\tan x}$

Answers

(i) $\log\left(\dfrac{ab}{4}\right)$

(ii) $\dfrac{1}{4}\log\dfrac{3}{2}$

(iii) $(\log 5)^2$

(iv) 1

(v) -2

(vi) log5

(vii) (log2) (log3)

(viii) 2 log 5

14 Inverse Trigonometric Functions and Their Derivatives

14.1 INTRODUCTION

We introduced *the concept of the inverse of a function in Chapter 2*. It is useful to review this concept before we discuss *inverse trigonometric functions*. Functions that always give different outputs for different inputs are called one-to-one. Since each output of a one-to-one function comes from just one input, any one-to-one function can be reversed to turn the outputs back into the inputs from which they came. Thus, *a function has an inverse if and only if* it is one-to-one. The function defined by reversing a one-to-one function f [which means that each ordered pair (a, b) belonging to f, is replaced by a corresponding ordered pair (b, a) in the new function] is called the inverse of f and denoted by f^{-1}.[(1)]

Example (1): Consider the function $y = x^3$. It gives different output(s) for different input(s). Hence, it is a one-to-one function. On the other hand, the function $y = x^2$ can give the same outputs for different inputs. (Check for the inputs 1 and -1, $\sqrt{2}$ and $-\sqrt{2}$, -3 and 3, etc.) Hence this function is *not* one-to-one. However, if we restrict the domain of this function to *non-negative numbers* then the same expression (with restricted domain), that is, $y = x^2$, $x > 0$, defines a one-to-one function. This example tells us that by restricting the domain of a function suitably, it is possible that a given formula (expression) defines a one-to-one function. This fact is specially used when we consider inverse trigonometric functions.

Now, consider the graph of $y = f(x) = \sqrt{x}$ shown in Figure 14.1.

The function $y = \sqrt{x}$ is defined for all $x \geq 0$ and its range is $y \geq 0$. For each input x_0, the function f gives a single output $y = \sqrt{x_0}$. Since every non-negative y is the image of just one x under this function, we can reverse the construction. That is, we can start with $y \geq 0$ and then go over to the curve and down to $x = y^2$, on the x-axis. [This is indicated by the arrows starting from y_0 (on y-axis) and reaching (on to the x-axis) the point $x = y_0^2$]. *This construction in reverse* defines the function $g(y) = y^2$, the inverse of $f(x) = \sqrt{x}$. Thus, the inverse of $y = f(x) = \sqrt{x}$ is given by $x = g(y) = y^2$ [or $x = f^{-1}(y) = y^2$].

14-Inverse trigonometric functions and their derivatives

(1) The term one-to-one function stands for a function which is one-one and onto (Chapter 2).

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

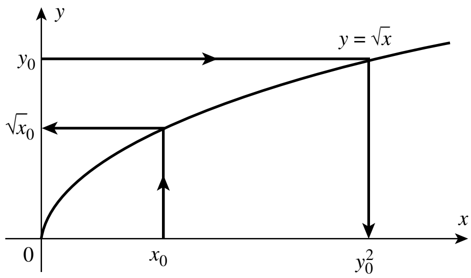


FIGURE 14.1

Note (1): Each pair of inverse functions (here, f and g) behave opposite to each other in the sense that one function undoes (i.e., reverses) what the other does. The algebraic description of what we see in Figure 14.1 is that

$$\left.\begin{aligned} g(f(x)) &= (\sqrt{x})^2 = x \\ f(g(y)) &= (\sqrt{y})^2 = y \end{aligned}\right\}$$

Observe that, in the above equations f is the inverse of g. It must be noted that an inverse function associates the same pair of elements, as in the original function, but with the object and the image interchanged. In the inverse notation,

$$g = f^{-1}$$

Note (2): Not every function has an inverse, as in the case of $y = x^2$ $(x \in R)$.

Whenever a function

$$y = f(x) \tag{1}$$

has an inverse, we can write it as

$$x = f^{-1}(y) \tag{2}$$

provided (1) can be solved for x uniquely.

Both the functions at (1) and (2), if they are defined, *describe one and the same curve in the xy-plane*.

14.1.1

The independent variable for the function f is x, while for the function f^{-1} the independent variable is y. *If we wish to denote the argument in formula (ii) by* x [i.e., if we wish to write $x = f^{-1}(y)$ in the form $y = f^{-1}(x)$] *in a single coordinate system, we get two different graphs which are symmetric about the line* $y = x$. *They represent two mutually inverse functions.*

The graphs of the *two mutually inverse functions* are given in Figure 14.2. The graph of a function and its inverse are symmetric with respect to the line $y = x$.

14.1.2 Distinguishing Geometrical Properties of One-to-One Functions

We know that a *vertical line* can intersect the graph of a function *at one point* only. For a one-to-one function, it is also true that a horizontal line can intersect a graph in at most one point. This is

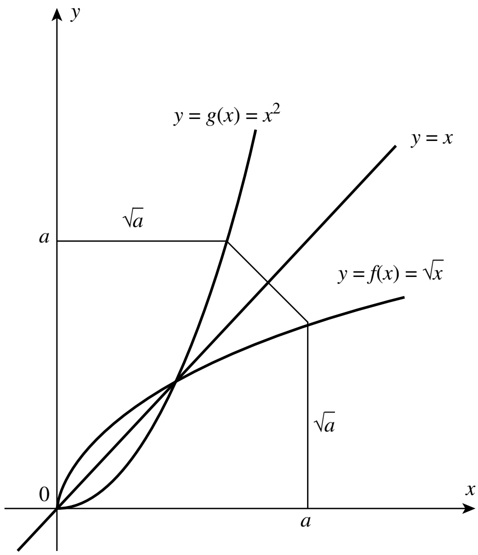


FIGURE 14.2

the situation for the one-to-one function defined by $y = x^3$ whose graph appears in Figure 14.3. On the other hand, observe in Figure 14.4 that for the function defined by $y = x^2$, which is *not one-to-one*, any horizontal line above the x-axis intersects the graph in two points. We have, therefore, the following *geometric test for determining if a function is one-to-one*.

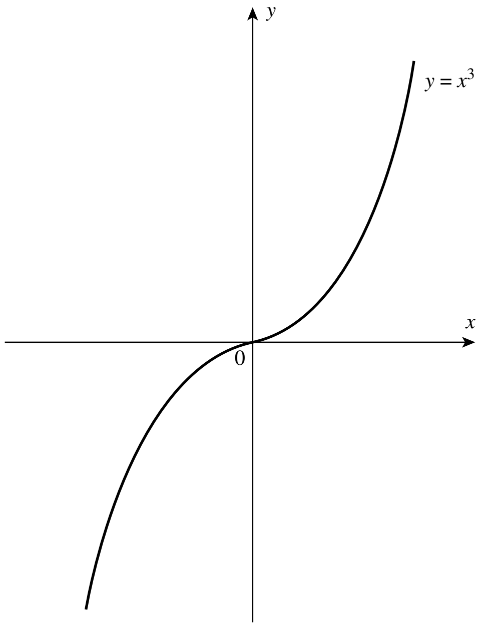


FIGURE 14.3

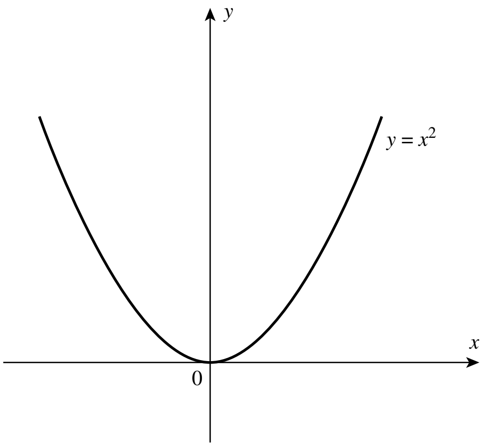


FIGURE 14.4 Graph of a function which is not one-to-one.

14.1.3 Horizontal-Line Test

A function is one-to-one if and only if every horizontal line intersects the graph of a function in at most one point.

Note (3): We use the terminology "*inverse functions*" only when referring to a function and its inverse.

Note (4): The criterion that a function be one-to-one, in order to have an inverse may be very hard to apply in a given situation, since it demands that we have complete knowledge of the graph. *A more practical criterion is that a function be strictly monotonic* (i.e., either strictly increasing or strictly decreasing). This is a practical result, *because we have an easy way of deciding if a function f is strictly monotonic*. We simply examine the sign of $f'(x)$. If $f'(x) > 0$ the function f is strictly increasing on its domain but if $f'(x) < 0, f$ is strictly decreasing. These results are proved in Chapter 19a. Later on, in Chapter 20, it is proved that a strictly monotonic function is one-to-one, showing that all such functions have inverses.

The *six basic trigonometric functions* (sin x, cos x, tan x, cot x, sec x, and cosec x of the real variable x) are defined in Chapter 5. Since all these functions are *periodic (and hence not one-to-one)*, none of them has an inverse. We can however, *restrict the domains of these functions in a way to allow for an inverse*.

14.2 TRIGONOMETRIC FUNCTIONS (WITH RESTRICTED DOMAINS) AND THEIR INVERSES

We begin with the sine function, $y = \sin x$, whose graph appears in Figure 14.5. Observe from the figure that *the sine function is strictly increasing on the interval* $[-(1/2)\pi$ and $(1/2)\pi]$.

Consequently, from the horizontal-line test (see Section 14.1.3), the function f_1, for which

$$f_1(x) = \sin x; \quad x \in \left[-\frac{1}{2}\pi,\ \frac{1}{2}\pi\right] \tag{3}$$ [2]

[2] Later on in Chapter 19a, we will show that the function $f_1(x)$ is strictly increasing on $[-(1/2)\pi,\ (1/2)\pi]$.

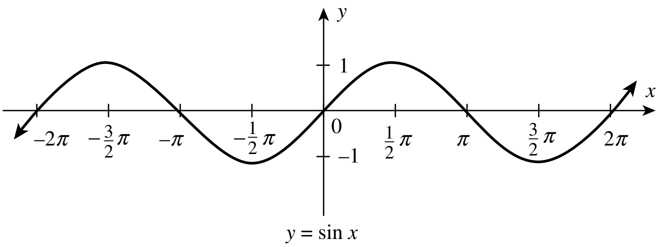


FIGURE 14.5

is one-to-one, and hence it does have an inverse in this interval. The graph of $f_1(x)$ is sketched in Figure 14.6. Its domain is $[-(1/2)\pi, (1/2)\pi]$ and its range is $[-1, 1]$. *The inverse of this function is called the inverse sine function.*

14.2.1 Definition of the Inverse Sine Function

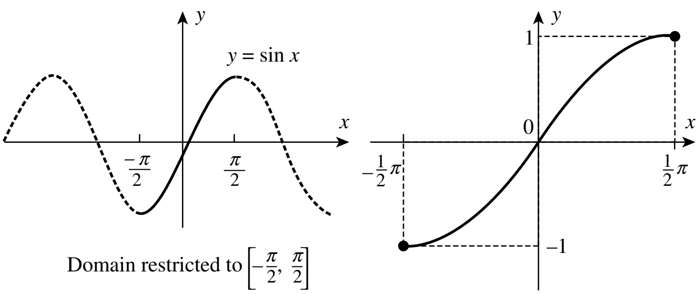
The inverse sine function, denoted by $\sin^{-1}$, is defined by

$$y = \sin^{-1}x, \text{ if and only if, } x = \sin y \text{ and } y \in \left[-\frac{1}{2}\pi, \frac{1}{2}\pi\right].$$

The *domain of* $\sin^{-1}x$ is the closed interval $[-1, 1]$ and the *range* is the closed interval $\left[-\frac{1}{2}\pi, \frac{1}{2}\pi\right]$.

Illustration:

- $\sin^{-1}(-1) = -\frac{1}{2}\pi$, because $\sin\left(-\frac{1}{2}\pi\right) = -1$.
- $\sin^{-1}(0) = 0$, because $\sin(0) = 0$.
- $\sin^{-1}\left(\frac{1}{2}\right) = \frac{1}{6}\pi$, because $\sin\left(\frac{1}{6}\pi\right) = \frac{1}{2}$.
- $\sin^{-1}\left(\frac{1}{\sqrt{2}}\right) = \frac{1}{4}\pi$, because $\sin\left(\frac{1}{4}\pi\right) = \frac{1}{\sqrt{2}}$.
- $\sin^{-1}\left(-\frac{1}{\sqrt{2}}\right) = -\frac{1}{4}\pi$, because $\sin\left(-\frac{1}{4}\pi\right) = -\frac{1}{\sqrt{2}}$.
- $\sin^{-1}(1) = \frac{1}{2}\pi$, because $\sin\left(\frac{1}{2}\pi\right) = 1$.



$f_1(x) = \sin x, \quad -\frac{1}{2}\pi \le x \le \frac{1}{2}\pi$

FIGURE 14.6

Remark: In equation (3), the domain of $f_1(x) = \sin x$ is restricted to the closed interval $[-(1/2)\pi,\ (1/2)\pi]$, so that the function is strictly monotonic and therefore has an inverse function. However, the sine function has a period of 2π and is (strictly) increasing on the other intervals as well, for instance, $[-(5/2)\pi,\ -(3/2)\pi]$ and $[(3/2)\pi,\ (5/2)\pi]$. Also, the function is strictly decreasing on certain closed intervals, in particular the intervals $[-(3/2)\pi,\ -(1/2)\pi]$ and $[(1/2)\pi,\ (3/2)\pi]$. *Any one of these intervals could just as well be chosen for the domain* of the function f_1 of equation (3). The choice of the interval $[-(1/2)\pi,\ (1/2)\pi]$, however, is customary because it is the largest interval containing the number 0, on which the function is (strictly) monotonic.

Note (5): The use of the symbol "−1" *to represent the inverse sine function* makes it necessary *to denote the reciprocal of sin x by* $(\sin x)^{-1}$, to avoid confusion.

A similar convention is applied when using any negative exponent with a trigonometric function. For instance, $1/(\tan x) = (\tan x)^{-1}$, $1/(\cos^2 x) = (\cos x)^{-2}$, and so on.

Note (6): The terminology *arc sine* is sometimes used *in place of inverse sine*, and the notation *arc sine* is then used *instead of* $\sin^{-1} x$. This notation probably comes from the fact that, *if* $t =$ *arc sin u, then* $\sin t = u$, *and t units is the length of the arc on the unit circle for which the sine is u.*

In this text, we shall be using the symbol "−1" (rather than the word arc) and thus writing $\sin^{-1}x$, $\cos^{-1}x$, and so on (instead of arc sin x, arc cos x, etc.). (This symbol is *consistent with the general notation for inverse functions.*)

We can sketch *the graph of the inverse sine function* by locating some points from values of $\sin^{-1}x$ such as those given in Table 14.1. The graph appears in Figure 14.7.

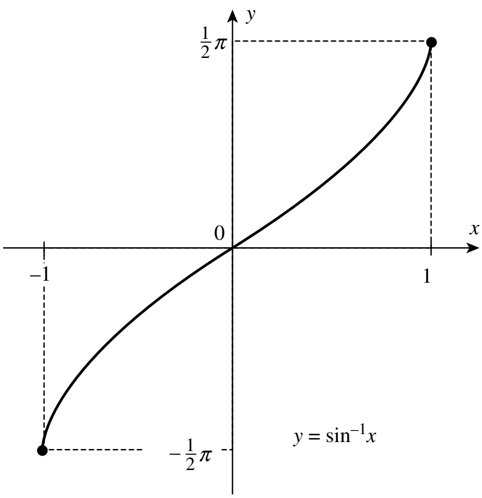


FIGURE 14.7

TABLE 14.1

x	-1	$-\frac{\sqrt{3}}{2}$	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	1
$\sin^{-1}x$	$-\frac{1}{2}\pi$	$-\frac{1}{3}\pi$	$-\frac{1}{6}\pi$	0	$\frac{1}{6}\pi$	$\frac{1}{3}\pi$	$\frac{1}{2}\pi$

From the definition of the inverse sine function (Section 14.2.1), we have

$$\sin(\sin^{-1}x) = x \quad \text{for } x \text{ in } [-1, 1]$$

$$\sin^{-1}(\sin y) = y \quad \text{for } y \text{ in } [-(1/2)\pi, (1/2)\pi]$$

Caution: Observe that, $\sin(\sin^{-1}x) = x$ is *valid for all real values of* x, it must be noted that $\sin^{-1}(\sin y) \neq y$, if y is not in the interval $[-(1/2)\pi, (1/2)\pi]$.

Example (2): Evaluate $\sin^{-1}\left(\sin\frac{5}{6}\pi\right)$

Solution: First we use the fact that

$$\sin\left(\frac{5}{6}\pi\right) = \sin\left(\pi - \frac{1}{6}\pi\right) = \sin\left(\frac{1}{6}\pi\right) = \frac{1}{2}$$

$$\sin^{-1}\left(\sin\frac{5}{6}\pi\right) = \sin^{-1}\left(\frac{1}{2}\right)$$

We know that $\sin(1/6)\pi = 1/2$, it follows that $\sin^{-1}(1/2) = \pi/6$

$$\therefore \quad \sin^{-1}\left(\sin\frac{5}{6}\pi\right) = \frac{\pi}{6}$$

Note that, $\sin^{-1}\left(\sin\frac{1}{6}\pi\right) \neq \frac{5\pi}{6}$, since $\frac{5\pi}{6} \notin \left[-\frac{1}{2}\pi, \frac{1}{2}\pi\right]$.

Similarly, $\sin^{-1}\left(\sin\frac{3}{4}\pi\right) = \frac{1}{4}\pi$ [we have $\sin^{-1}\left(\sin\frac{3}{4}\pi\right) = \sin^{-1}\left(\frac{1}{2}\right) = \frac{1}{4}\pi$] and $\sin^{-1}\left(\sin\frac{7}{4}\pi\right) = -\frac{1}{4}\pi$ [we have $\sin^{-1}\left(\sin\frac{7}{4}\pi\right) = \sin^{-1}\left(-\frac{1}{\sqrt{2}}\right) = -\frac{1}{4}\pi$].

Example (3): Find

(a) $\cos\left[\sin^{-1}\left(-\frac{1}{2}\right)\right]$
(b) $\sin^{-1}\left[\cos\frac{2}{3}\pi\right]$

Solution:
We know that *the range of the inverse sine function* is $\left[-\frac{1}{2}\pi, \frac{1}{2}\pi\right]$.

Further, $\left[\sin^{-1}\left(-\frac{1}{2}\right)\right] = -\frac{1}{6}\pi$

(a) $\cos\left[\sin^{-1}\left(-\frac{1}{2}\right)\right] = \cos\left(-\frac{1}{6}\pi\right) = \frac{\sqrt{3}}{2}$
(b) $\sin^{-1}\left[\cos\frac{2}{3}\pi\right] = \sin^{-1}\left(-\frac{1}{2}\right) = -\frac{1}{6}\pi$

14.2.2 Derivative of the Inverse Sine Function

We now obtain the formula for the derivative of the inverse sine function by applying the rule that deals with the differentiation of inverse functions. [Recall from Chapter 10, Rule 6 which states as follows: if $y = f(x)$ is a derivable function of x such that the inverse function $x = f^{-1}(y)$ is defined and dy/dx, dx/dy both exist, then derivative of the inverse function is given by $dx/dy = 1/(dy/dx)$, provided $dy/dx \neq 0$.]

Let $y = \sin^{-1}x$, *which is equivalent to*

$$x = \sin y \quad \text{and} \quad y \in \left[-\frac{1}{2}\pi,\ \frac{1}{2}\pi\right] \tag{4}$$

Differentiating both the sides of this equation with respect to y, we obtain

$$\frac{dx}{dy} = \cos y \quad \text{and} \quad y \text{ is in} \left[-\frac{1}{2}\pi,\ \frac{1}{2}\pi\right] \tag{5}$$

If y is in $\left[-\frac{1}{2}\pi,\ \frac{1}{2}\pi\right]$, $\cos y$ is non-negative.[3]

We know that,
$$\frac{dy}{dx} = \frac{1}{dx/dy} = \frac{1}{\cos y} \tag{6}$$

Here, we have to write the right-hand side in terms of x. Since, $\sin y = x$, we have

$$\cos y = \pm\sqrt{1 - \sin^2 y} = \sqrt{1 - x^2}$$

Of these two values for $\cos y$, we should take $\cos y = \sqrt{1 - x^2}$, since y lies between $-(1/2)\pi$ and $(1/2)\pi$.

$$\therefore \quad \frac{dy}{dx} = \frac{d}{dx}(\sin^{-1} x) = \frac{1}{\cos y} = \frac{1}{\sqrt{1 - x^2}}$$

$$\therefore \quad \frac{d}{dx}(\sin^{-1} x) = \frac{1}{\sqrt{1 - x^2}} \tag{7}$$

Theorem (A): If u is a differentiable function of x,

$$\frac{d}{dx}(\sin^{-1} u) = \frac{1}{\sqrt{1 - u^2}} \cdot \frac{du}{dx} \qquad \text{(by the Chain Rule)}$$

Example (4): Find $f'(x)$, if $f(x) = \sin^{-1}x^2$

Solution: From Theorem (A),

$$f'(x) = \frac{1}{\sqrt{1 - (x^2)^2}} \cdot 2x = \frac{2x}{\sqrt{1 - x^4}} \qquad \textbf{Ans.}$$

[3] Note that for $y = \pm(1/2)\pi$, $\cos y = 0$, and so $dy/dx = 1/\cos x$ is not defined. However, if y lies between $-(1/2)\pi$ and $(1/2)\pi$, then $\cos y$ is positive and so $dy/dx = 1/\cos y$ is defined. Therefore, we consider y such that it lies between $-\pi/2$ and $\pi/2$.

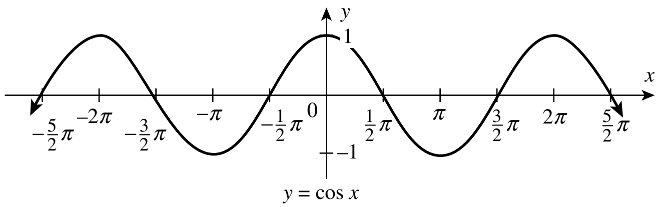


FIGURE 14.8

14.3 THE INVERSE COSINE FUNCTION

To obtain the inverse cosine function, we proceed as we did with the inverse sine function. *We restrict the cosine to an interval on which the function is (strictly) monotonic*. We choose the interval $[0, \pi]$ *on which the cosine is decreasing*, as shown by the graph of the cosine in Figure 14.8.

Let us consider the function $f_2(x)$ defined by $f_2(x) = \cos x$, $x \in [0, \pi]$.

The *domain of* $f_2(x)$ is the closed interval $[0, \pi]$ and *the range* is the closed interval $[-1, 1]$. The graph of $f_2(x)$ appears in Figure 14.9. Because $f_2(x)$ is *continuous and decreasing on its domain, it has an inverse*, which we now define.

14.3.1 Definition of the Inverse Cosine Function

The inverse cosine function, denoted by $\cos^{-1}$, is defined by $y = \cos^{-1}x$, if and only if, $x = \cos y$ and $y \in [0, \pi]$. The domain of $\cos^{-1}$ is the closed interval $[-1, 1]$ and the range is the closed interval $[0, \pi]$[4].

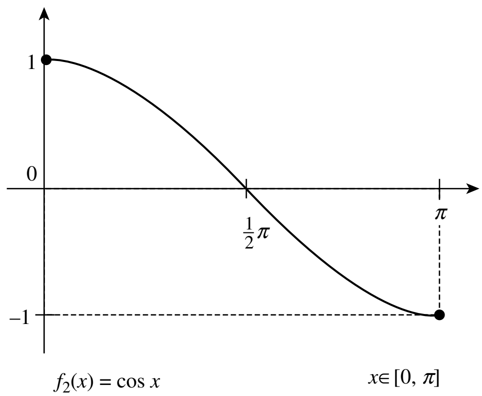
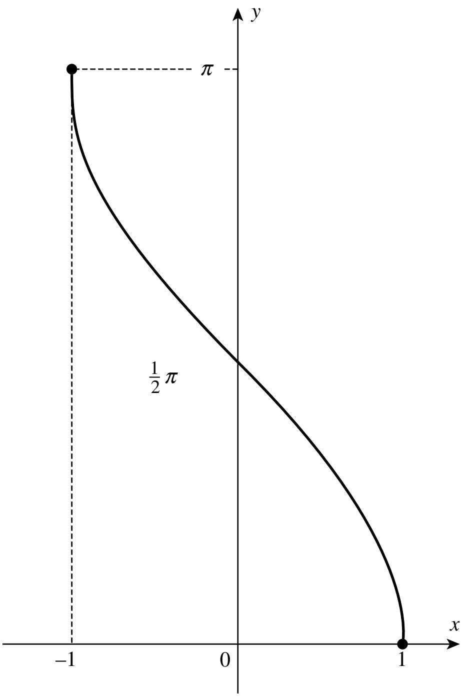


FIGURE 14.9

[4] Note that, the domain of $y = \cos^{-1}x$ is the set of numbers x such that $x = \cos y$. But, the value $\cos y$ lies in the interval $[-1, 1]$. Hence, domain of $\cos^{-1}$ is the interval $[-1, 1]$ and range is $[0, \pi]$.



$y = \cos^{-1} x$

FIGURE 14.10

The graph of the inverse cosine function appears in Figure 14.10. From the definition of the inverse cosine function (Section 14.3.1), we have

$$\cos(\cos^{-1} x) = x, \quad \text{for } x \text{ in } [-1, 1]$$

$$\cos^{-1}(\cos y) = y, \quad \text{for } y \text{ in } [0, \pi]$$

Note (7): Observe that *there is again a restriction on y in order to have the equality.*

$$\cos^{-1}(\cos y) = y,$$

For example, because $(3/4)\pi$ is in $[0, \pi]$.

$$\cos^{-1}\left(\cos \frac{3}{4}\pi\right) = \frac{3}{4}\pi$$

However, $\cos^{-1}\left(\cos \frac{5}{4}\pi\right) = \cos^{-1}\left(-\frac{1}{\sqrt{2}}\right) = \frac{3}{4}\pi$, and

$$\cos^{-1}\left(\cos \frac{7}{4}\pi\right) = \cos^{-1}\left(\frac{1}{\sqrt{2}}\right) = \frac{1}{4}\pi.$$

14.3.2 Formula for the Derivative of the Inverse Cosine Function

Let $y = \cos^{-1}x$, which is equivalent to

$$x = \cos y \quad \text{and} \quad y \in [0, \pi] \tag{8}$$

Differentiating both sides with respect to y, we have

$$\frac{dx}{dy} = -\sin y, \quad \text{and} \quad y \in [0, \pi] \tag{9}$$

If y is in $[0, \pi]$, $\sin y$ is non-negative, making the above term on the RHS negative

But,
$$\frac{dy}{dx} = \frac{1}{dx/dy} = \frac{-1}{\sin y} \qquad (10)^{(5)}$$

Here, we have to express the right-hand side in terms of x. Since $\cos y = x$, we have

$$\sin y = \pm\sqrt{1-\cos^2 x} = \pm\sqrt{1-x^2}$$

Of these two values for $\sin y$, we should take $\sin y = \sqrt{1-x^2}$, since y lies between 0 and π.

$$\therefore \quad \frac{dy}{dx} = \frac{d}{dx}(\cos^{-1} x) = \frac{-1}{\sin y} = \frac{-1}{\sqrt{1-x^2}}$$

$$\therefore \quad \frac{d}{dx}(\cos^{-1}x) = \frac{-1}{\sqrt{1-x^2}} \tag{11}$$

14.3.3 Important Identities Involving Inverse Trigonometric Functions

The following identities involving inverse trigonometric functions are very important.

(i) $\sin^{-1} x + \cos^{-1} x = \frac{\pi}{2}$
(ii) $\tan^{-1} x + \cot^{-1} x = \frac{\pi}{2}$
(iii) $\sec^{-1} x + \text{cosec}^{-1} x = \frac{\pi}{2}$

Let us prove (i)

Let
$$\sin^{-1}x = t \tag{12}$$

$$\therefore \quad x = \sin t = \cos\left(\frac{\pi}{2} - t\right)$$

$$\therefore \quad \frac{\pi}{2} - t = \cos^{-1} x \tag{13}$$

Adding (12) and (13), we get

$$\sin^{-1} x + \cos^{-1} x = \frac{\pi}{2}$$

Similarly (ii) and (iii) can be proved.

(5) Refer to Chapter 10, Corollary to Rule 6, Page 303.

Note (8): Now, using the identity at (i) above and the result,

$$\frac{d}{dx}(\sin^{-1} x) = \frac{1}{\sqrt{1-x^2}}$$

We will now show that,

$$\frac{d}{dx}(\cos^{-1} x) = \frac{-1}{\sqrt{1-x^2}}$$

Consider the identity $\sin^{-1} x + \cos^{-1} x = \frac{\pi}{2}$

Differentiating both sides with respect to x, we get,

$$\frac{d}{dx}(\sin^{-1} x) + \frac{d}{dx}(\cos^{-1} x) = 0$$

$$\therefore \quad \frac{d}{dx}(\cos^{-1} x) = -\frac{d}{dx}(\sin^{-1} x)$$

$$= \frac{-1}{\sqrt{1-x^2}} \quad \left[\text{Since, } \frac{d}{dx}(\sin^{-1} x) = \frac{1}{\sqrt{1-x^2}}\right] \quad \text{(Proved)}$$

Theorem (B): If u is a differentiable function of x, then

$$\frac{d}{dx}(\cos^{-1} u) = \frac{-1}{\sqrt{1-u^2}} \cdot \frac{du}{dx}$$

Example (5): Find $\frac{dy}{dx}$ if $y = \cos^{-1} e^{2x}$

Solution: Given $y = \cos^{-1} e^{2x}$,
We get, from theorem (B),

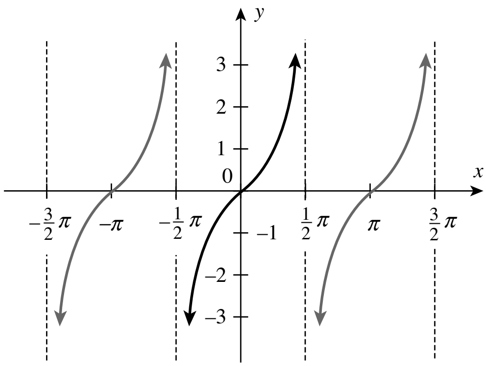
$$\frac{dy}{dx} = \frac{-1}{\sqrt{1-(e^{2x})^2}} \frac{d}{dx}(e^{2x})$$

$$= \frac{-2e^{2x}}{\sqrt{1-(e^{2x})^2}} = \frac{-2e^{2x}}{\sqrt{1-e^{4x}}} \qquad \textbf{Ans.}$$

14.4 THE INVERSE TANGENT FUNCTION

To develop the *inverse tangent function*, observe from the graph in Figure 14.11, that the tangent function is *continuous* and *(strictly) increasing on the open interval* $(-(1/2)\pi, (1/2)\pi)$. We restrict the tangent function to this interval, denote it by f_3 and define it by $f_3(x) = \tan x$ and $-(1/2)\pi < x < (1/2)\pi$.

The *domain* of $f_3(x)$ is the *open interval* $(-(1/2)\pi, (1/2)\pi)$ and the range is the set R of real numbers. The graph of $f_3(x)$ is given in Figure 14.12. *This function has an inverse called the inverse tangent function.*

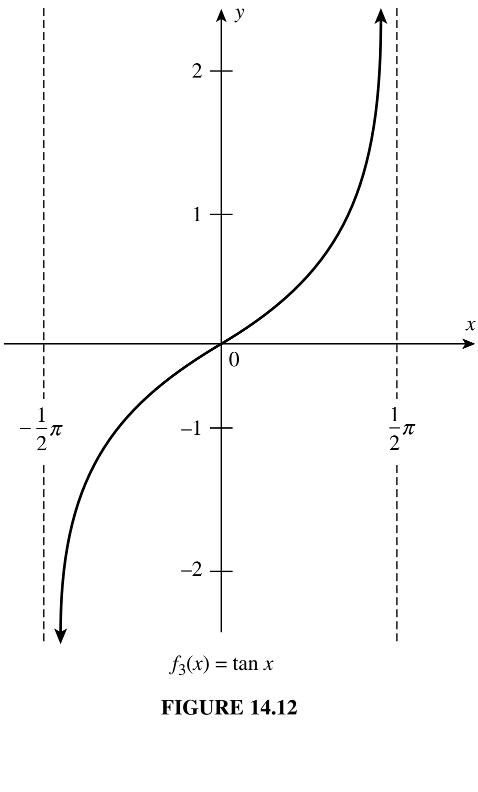


$y = \tan x$

FIGURE 14.11

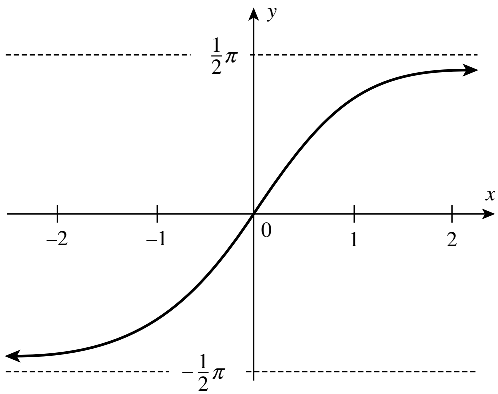
14.4.1 Definition of the Inverse Tangent Function

The *inverse tangent function*, denoted by $\tan^{-1}$, is defined by $y = \tan^{-1}x$, if and only if, $x = \tan y$ and $-(1/2)\pi < y < (1/2)\pi$. The domain of $\tan^{-1}$ is the set R of real numbers and the range is the open interval $(-(1/2)\pi, (1/2)\pi)$. The graph of the inverse tangent function is shown in Figure 14.13.



$f_3(x) = \tan x$

FIGURE 14.12



$y = \tan^{-1}x, \qquad (-\frac{1}{2}\pi < x < \frac{1}{2}\pi)$

FIGURE 14.13

From the definition of *inverse tangent function* (Section 14.4.1), we have

$$\tan(\tan^{-1} x) = x, \quad \text{for } x \text{ in } (-\infty, \ +\infty)$$

$$\tan^{-1}(\tan y) = y, \quad \text{for } y \text{ in } \left(\frac{-1}{2}\pi, \ \frac{1}{2}\pi\right)$$

The restrictions on y are discussed through the following examples.

Example (6): $\tan^{-1}\left(\tan\frac{1}{4}\pi\right) = \frac{1}{4}\pi$ and $\tan^{-1}\left[\tan\left(-\frac{1}{4}\pi\right)\right] = -\frac{1}{4}\pi$

However, $\tan^{-1}\left(\tan\frac{3}{4}\pi\right) = \tan^{-1}(-1) = -\frac{1}{4}\pi$ and $\tan^{-1}\left(\tan\frac{5}{4}\pi\right) = \tan^{-1}(1) = -\frac{1}{4}\pi$.

14.4.2 Formula for the Derivative of the Inverse Tangent Function

Let $y = \tan^{-1}x$. Then,

$$x = \tan y \text{ and } y \text{ is in} \left(-\frac{1}{2}\pi, \ \frac{1}{2}\pi\right) \tag{14}$$

Differentiating both the sides of this equation, with respect to y, we obtain

$$\frac{dx}{dy} = \sec^2 y \quad \text{and} \quad y \text{ is in} \left(-\frac{1}{2}\pi, \ \frac{1}{2}\pi\right) \tag{15}$$

From the identity $\sec^2 y = 1 + \tan^2 y$, and replacing $\tan y$ by x, we have

$$\sec^2 y = 1 + x^2$$

$$\text{But, } \frac{dy}{dx} = \frac{1}{(dx/dy)}$$

$$\therefore \quad \frac{dy}{dx} = \frac{1}{1+x^2}$$

$$\text{Thus, } \frac{d}{dx}(\tan^{-1}x) = \frac{1}{1+x^2} \tag{16}$$

The domain of the derivative of the inverse tangent function is the set R of real numbers.

Theorem (C): If u is a differentiable function of x,

$$\frac{d}{dx}(\tan^{-1}u) = \frac{1}{1+u^2} \cdot \frac{du}{dx}$$

Example (7): Find $f'(x)$, if $f(x) = \tan^{-1}\frac{1}{x+1}$

Solution: From Theorem (C),

$$f'(x) = \frac{1}{1+\left(1/(x+1)^2\right)} \cdot \frac{d}{dx}\left(\frac{1}{1+x}\right)$$

or

$$f'(x) = \frac{1}{1+\left(1/(x+1)^2\right)} \cdot \frac{-1}{(1+x)^2}$$

$$= \frac{-1}{(x+1)^2+1} = \frac{-1}{x^2+2x+2} \qquad \textbf{Ans.}$$

Example (8): Differentiate $\tan^{-1}\log x$

Solution: $$\frac{d}{dx}\left[\tan^{-1}(\log x)\right] = \frac{1}{1+(\log x)^2} \cdot \frac{d}{dx}(\log x)$$

$$= \frac{1}{1+(\log x)^2} \cdot \frac{1}{x}$$

$$= \frac{1}{x[1+(\log x)^2]} \qquad \textbf{Ans.}$$

14.5 DEFINITION OF THE INVERSE COTANGENT FUNCTION

To define the *inverse cotangent function*, we use the *identity* $\tan^{-1}x + \cot^{-1}x = \pi/2$, (see Section 14.3.3) where x is any real number.

Definition: The inverse cotangent function, denoted by $\cot^{-1}$, is defined by

$$y = \cot^{-1}x = \frac{1}{2}\pi - \tan^{-1}x \text{ where } x \text{ is any real number} \tag{17}$$

The domain and the range of $\cot^{-1}$.[(6)]

[(6)] *The Calculus 7 of a Single Variable* (Sixth Edition) by Louis Leithold (pp. 501–502), HarperCollins College Publishers.

By definition, the domain of $\cot^{-1}$ is the set R of real numbers. To obtain the range, we write the equation in the definition as

$$\tan^{-1} x = \frac{1}{2}\pi - \cot^{-1} x \tag{18}$$

We know that,
$$-\frac{1}{2}\pi < \tan^{-1} x < \frac{1}{2}\pi \tag{19}$$

Using (18) in (19), we get

$$-\frac{1}{2}\pi < \frac{1}{2}\pi - \cot^{-1} x < \frac{1}{2}\pi$$

Subtracting $(1/2)\pi$ from each member, we get

$$-\pi < -\cot^{-1} x < 0$$

Now, multiplying each member by -1, we get

$$\pi > \cot^{-1} x > 0$$

Reversing the direction of inequality signs, we obtain

$$0 < \cot^{-1} x < \pi$$

The range of the inverse cotangent function is therefore the open interval $(0, \pi)$. Its graph is sketched in Figure 14.14.

Illustration:

(a) $\tan^{-1}(1) = \frac{1}{4}\pi$
(b) $\tan^{-1}(-1) = -\frac{1}{4}\pi$
(c) $\cot^{-1}(1) = \frac{1}{2}\pi - \tan^{-1}(1) = \frac{1}{2}\pi - \frac{1}{4}\pi = \frac{1}{4}\pi$
(d) $\cot^{-1}(-1) = \frac{1}{2}\pi - \tan^{-1}(-1) = \frac{1}{2}\pi - \left(\frac{-1}{4}\pi\right) = \frac{3}{4}\pi$

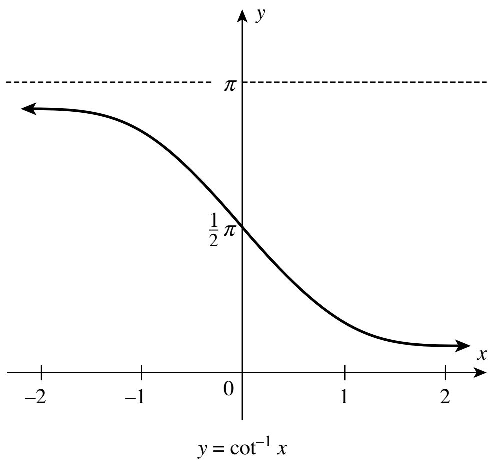


FIGURE 14.14

14.5.1 Formula for the Derivative of $\cot^{-1}x$

From the definition of inverse cotangent function, we have

$$\cot^{-1}x = \frac{1}{2}\pi - \tan^{-1}x,$$

Differentiating both sides with respect to x, we get

$$\frac{d}{dx}\cot^{-1}x = \frac{d}{dx}\left(\frac{1}{2}\pi - \tan^{-1}x\right)$$

$$\therefore \quad \frac{d}{dx}\left(\cot^{-1}x\right) = -\frac{1}{1+x^2} \tag{20}$$

Theorem (D): If u is a differentiable function of x,

$$\frac{d}{dx}\left(\cot^{-1}u\right) = -\frac{1}{1+u^2}\cdot\frac{du}{dx}$$

Before we define the inverse secant and the inverse cosecant functions, let us again look at the graphs of basic trigonometric functions and the inverse trigonometric functions.

The graphs of six trigonometric functions are shown in Figure 14.15a–f. None of these functions has an inverse, since *a horizontal line $y = c$ may cross each graph at more points.*

Now consider the six functions (f_1–f_6), which as graphs have heavily marked portions of the six trigonometric functions in the same graph (Figure 14.15). (In fact, these portions of the graph define the respective trigonometric functions with restricted domain.) Each of these graphs represents a new function, which has the same range as the corresponding trigonometric function, and each new function has an inverse. We call them the *principal branches of the basic trigonometric functions.*

By abuse of terminology, the inverses of $f_1, f_2, \ldots, f_6$ are called the inverse trigonometric functions, so that f_1^{-1} is the inverse sine, denoted by $x = \sin^{-1}y$, f_2^{-1} is the inverse cosine, denoted by $x = \cos^{-1}y$, and so on. Similar notations are used for the remaining four inverse trigonometric functions. The graphs of the inverse trigonometric functions as functions of the independent variable x are shown in Figure 14.16.

Note (9): As can be seen from the graph of sec x and cosec x (Figure 14.15), it is impossible to choose "branches" of these functions so that the inverse functions become continuous. The branches of $\sec^{-1}x$ and $\text{cosec}^{-1}x$ (Figure 14.16) are chosen to make the formulas for the derivatives of these functions come out nicely, without ambiguity to sign. Now, the derivatives of $\sec^{-1}x$ and $\text{cosec}^{-1}x$ can easily be found just as we found the derivatives in other cases.[(7)]

14.6 FORMULA FOR THE DERIVATIVE OF INVERSE SECANT FUNCTION

Let $$y = \sec^{-1}x$$

$$\therefore \quad x = \sec y$$

[(7)] For more details, refer to *Calculus with Analytic Geometry* by John B. Fraleigh (p. 261), Addison-Wesley.

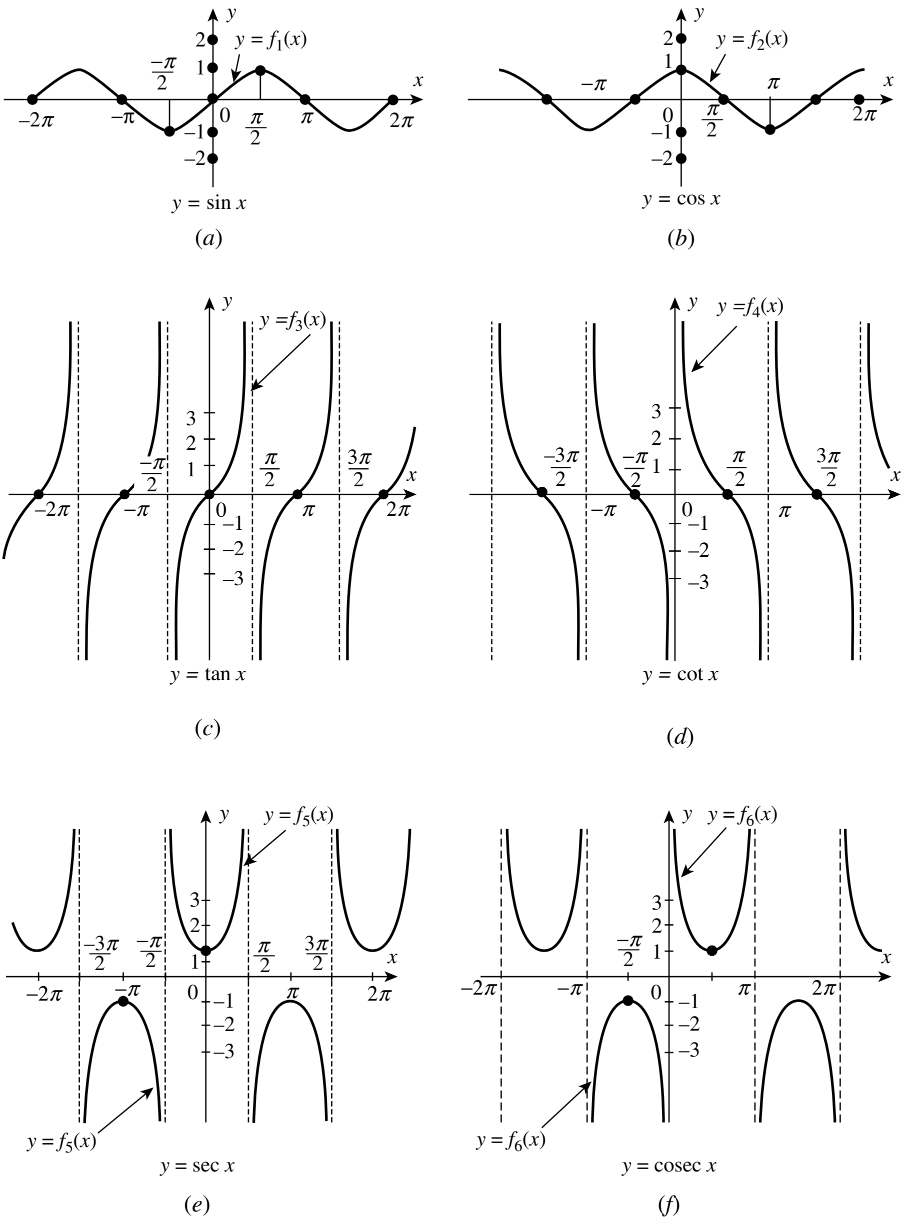

y = f1(x)
y = sin x
(a)
y = f2(x)
y = cos x
(b)
y = f3(x)
y = tan x
(c)
y = f4(x)
y = cot x
(d)
y = f5(x)
y = sec x
(e)
y = f6(x)
y = cosec x
(f)

FIGURE 14.15

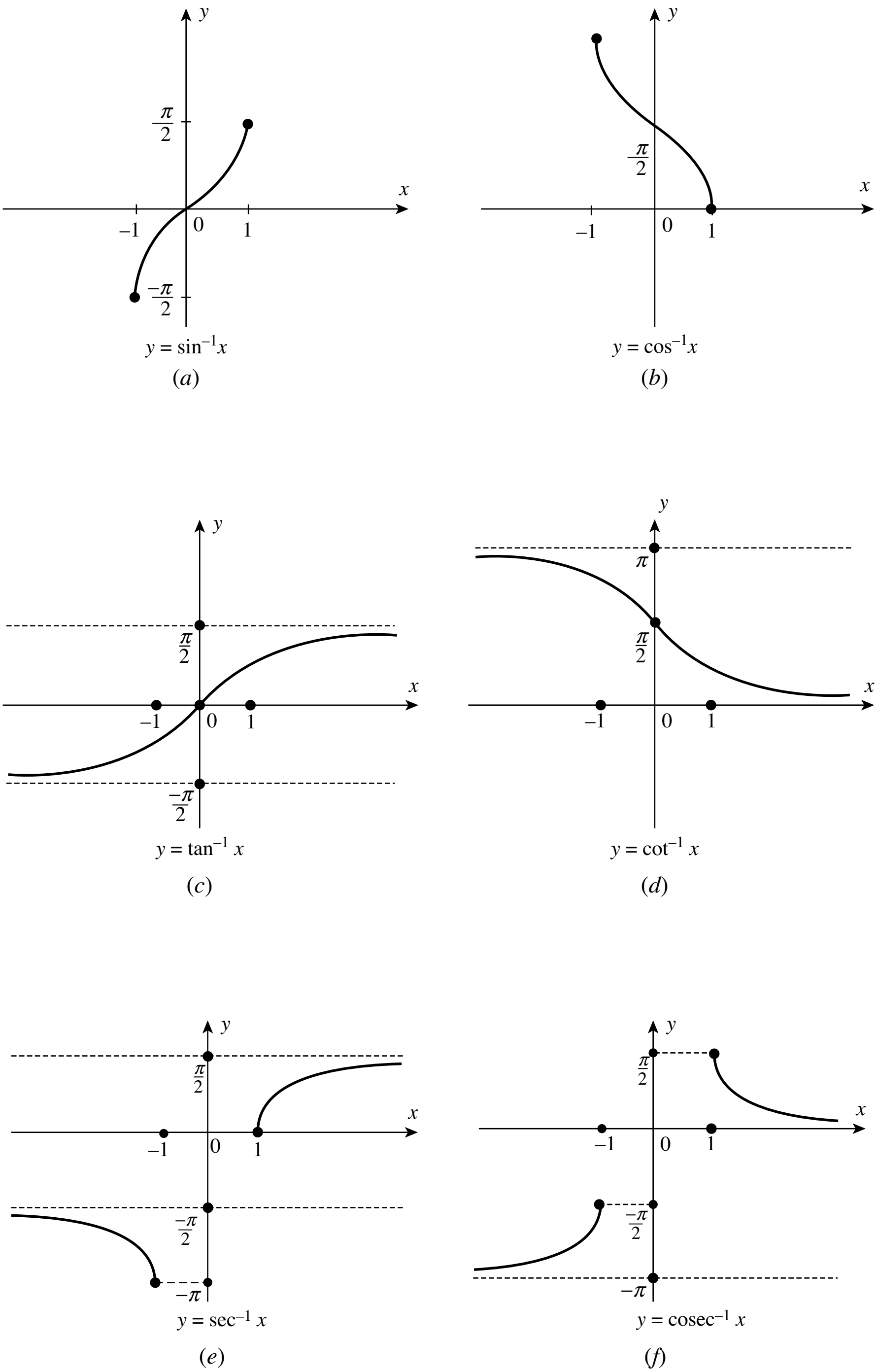

y
x
$\frac{\pi}{2}$
–1
0
1
$\frac{-\pi}{2}$
$y = \sin^{-1}x$
(a)
y
x
$\frac{\pi}{2}$
–1
0
1
$y = \cos^{-1}x$
(b)
y
x
$\frac{\pi}{2}$
–1
0
1
$\frac{-\pi}{2}$
$y = \tan^{-1} x$
(c)
y
x
π
$\frac{\pi}{2}$
–1
0
1
$y = \cot^{-1} x$
(d)
y
x
$\frac{\pi}{2}$
–1
0
1
$\frac{-\pi}{2}$
$-\pi$
$y = \sec^{-1} x$
(e)
y
x
$\frac{\pi}{2}$
–1
0
1
$\frac{-\pi}{2}$
$-\pi$
$y = \text{cosec}^{-1} x$
(f)

FIGURE 14.16

Differentiating with respect to y, we get

$$\frac{dx}{dy} = \sec y \cdot \tan y = \sec y\sqrt{\tan^2 y}$$

$$= \sec y\sqrt{\sec^2 y - 1}$$

$$= x\sqrt{x^2 - 1}$$

$$\frac{dy}{dx} = \frac{1}{(dy/dx)} = \frac{1}{x\sqrt{x^2 - 1}}$$

$$\therefore \quad \frac{d(\sec^{-1}x)}{dx} = \frac{1}{x\sqrt{x^2 - 1}}, \quad |x| > 1 \qquad (21)$$

Theorem (E): If y is a differentiable function of x,

$$\frac{d}{dx}(\sec^{-1}u) = \frac{1}{u\sqrt{u^2 - 1}} \cdot \frac{du}{dx}, \quad |u| > 1$$

14.7 FORMULA FOR THE DERIVATIVE OF INVERSE COSECANT FUNCTION

Let $\quad y = \text{cosec}^{-1}x$

Then, $\quad x = \text{cosec}\, y$

Differentiating with respect to y, we get,

$$\frac{dx}{dy} = -\text{cosec}\, y \cdot \cot y = -\text{cosec}\, y\sqrt{\text{cosec}^2 y - 1}$$

$$= -\text{cosec}\, y\sqrt{\text{cosec}^2 y - 1}$$

$$= -x\sqrt{x^2 - 1}$$

Now, $$\frac{dy}{dx} = \frac{1}{(dy/dx)} = \frac{-1}{x\sqrt{x^2 - 1}}, \quad |x| > 1$$

$$\therefore \quad \frac{d(\text{cosec}^{-1}x)}{dx} = \frac{-1}{x\sqrt{x^2 - 1}}, \quad |x| > 1 \qquad (22)$$

Theorem (F): If u is a differentiable function of x,

$$\frac{d}{dx}(\text{cosec}^{-1}u) = \frac{-1}{u\sqrt{u^2 - 1}} \cdot \frac{du}{dx}, \quad |u| > 1$$

Table 14.2 summarizes the data that we should remember regarding inverse trigonometric functions.

TABLE 14.2

Function	Domain	Range	Derivative
$\sin^{-1}x$	$[-1, 1]$	$\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$	$\frac{1}{\sqrt{1-x^2}}$
$\cos^{-1}x$	$[-1, 1]$	$[0, \pi]$	$\frac{-1}{\sqrt{1-x^2}}$
$\tan^{-1}x$	All x	$-\frac{\pi}{2} < y < \frac{\pi}{2}$	$\frac{1}{1+x^2}$
$\cot^{-1}x$	All x	$0 < y < \pi$	$\frac{-1}{1+x^2}$
$\sec^{-1}x$	$x \leq -1$ or $x \geq 1$	$-\pi \leq y < -\frac{\pi}{2}$ or $0 \leq y < \frac{\pi}{2}$	$\frac{1}{x\sqrt{x^2-1}}$
$\text{cosec}^{-1}x$	$x \leq -1$ or $x \geq 1$	$-\pi < y \leq -\frac{\pi}{2}$ or $0 < y \leq \frac{\pi}{2}$	$\frac{-1}{x\sqrt{x^2-1}}$

Source: *Calculus with Analytic Geometry* by John B. Fraleigh (p. 263), Addison-Wesley.

From the theorems stated at (A)–(F) above, we know that if u is a function of independent variable x, then we may write the formulas for derivatives of inverse trigonometric functions of u, using the chain rule.

$\frac{d}{dx}\sin^{-1}u = \frac{1}{1-u^2} \cdot \frac{du}{dx}$, and so on. These results may also be written as

$$\frac{d}{dx}\sin^{-1}[f(x)] = \frac{f'(x)}{\sqrt{1-[f(x)]^2}}; \quad \frac{d}{dx}\cos^{-1}[f(x)] = \frac{-f'(x)}{\sqrt{1-[f(x)]^2}}$$

$$\frac{d}{dx}\tan^{-1}[f(x)] = \frac{f'(x)}{1+[f(x)]^2}; \quad \text{and so on.}$$

These formulas are primarily important for evaluation of certain definite integrals. In fact, this is the main reason for studying the calculus of inverse trigonometric functions.

14.8 IMPORTANT SETS OF RESULTS AND THEIR APPLICATIONS

The following sets of results [set (1) to set (5)] connecting trigonometric (circular) functions and inverse trigonometric functions are useful in simplifying certain inverse trigonometric functions for computing their derivatives.

In the above results (or formulas) *it is assumed that we are dealing with the principal branch (es) of the functions and their appropriate domain(s)*. Their applications are given below:

Set (1)	**Set (2)**
$\sin^{-1}(\sin x) = x$	$\sin(\sin^{-1}x) = x$
$\cos^{-1}(\cos x) = x$	$\cos(\cos^{-1}x) = x$
$\tan^{-1}(\tan x) = x$	$\tan(\tan^{-1}x) = x$
and so on.	and so on.

Applications of set (1) and (2) (differentiate with respect to x).

$$y = \sin^{-1}(\sin 5x)$$

$$\text{Put } 5x = t \quad \therefore y = \sin^{-1}(\sin t) = t$$

$$\text{or} \quad y = 5x \quad \therefore \frac{dy}{dx} = 5.$$

Set (3): We know that,

$$\sin^{-1}(\cos x) = \sin^{-1}\left[\sin\left(\frac{\pi}{2} - x\right)\right] = \frac{\pi}{2} - x$$

$$\cos^{-1}(\sin x) = \cos^{-1}\left[\cos\left(\frac{\pi}{2} - x\right)\right] = \frac{\pi}{2} - x$$

$$\tan^{-1}(\cot x) = \tan^{-1}\left[\tan\left(\frac{\pi}{2} - x\right)\right] = \frac{\pi}{2} - x$$

and so on.

Application of set (3)

$$\text{Let } y = \sin^{-1}[\cos 3x]$$

$$\therefore \quad y = \sin^{-1}\left[\sin\left(\frac{\pi}{2} - 3x\right)\right] = \frac{\pi}{2} - 3x$$

$$\therefore \quad \frac{dy}{dx} = 0 - 3 = -3 \quad \textbf{Ans.}$$

Set (4)

$$\left[\begin{array}{l} \tan^{-1}x + \tan^{-1}y = \tan^{-1}\left(\dfrac{x+y}{1-xy}\right) \\[2ex] \tan^{-1}x - \tan^{-1}y = \tan^{-1}\left(\dfrac{x-y}{1+xy}\right) \end{array}\right]$$

These results are very useful as can be seen from the solved examples (it is proposed to prove these results at the end of this chapter).

Note that the expression $((x+y)/(1-xy))$ can be converted to the form $\tan(p+q)$ by proper substitution and similarly $((x-y)/(1+xy))$ can be converted to the form $\tan(p-q)$. Thus, in any expression of the type $\tan^{-1}[f(x)]$, if it is possible to break up $f(x)$ in any of the two above forms, then the given function $\tan^{-1}[f(x)]$ can be simplified for the purpose of the differentiation as will be clear from the following solved examples.

Application of set (4)

(a) Let $y = \tan^{-1}\left(\dfrac{5x}{1-6x^2}\right)$,

$$\therefore \quad y = \tan^{-1}\left(\frac{3x+2x}{1-(3x)\cdot(2x)}\right)$$

$$= \tan^{-1}(3x) + \tan^{-1}(2x)$$

$$\therefore \quad \frac{dy}{dx} = \frac{1}{1+(3x)^2}\cdot\frac{d}{dx}(3x) + \frac{1}{1+(2x)^2}\cdot\frac{d}{dx}(2x)$$

$$= \frac{3}{1+9x^2} + \frac{2}{1+4x^2} \quad \textbf{Ans.}$$

(b) Let $y = \tan^{-1}\left[\dfrac{\sin 7x - \cos 7x}{\sin 7x + \cos 7x}\right]$

Dividing numerator and denominator by $\cos 7x$.

$$\therefore \quad y = \tan^{-1}\left[\frac{\tan 7x - 1}{\tan 7x + 1}\right] = \tan^{-1}\left[\frac{\tan 7x - 1}{1 + \tan 7x}\right]$$

$$= \tan^{-1}\left[\frac{\tan 7x - \tan(\pi/4)}{1 + \tan 7x\cdot\tan(\pi/4)}\right]^{(8)}$$

$$= \tan^{-1}\left[\tan\left(7x - \frac{\pi}{4}\right)\right] = 7x - \frac{\pi}{4}$$

$$\therefore \quad \frac{dy}{dx} = 7 \quad \textbf{Ans.}$$

Set (5)

$$\left[\begin{aligned} \sin^{-1}x &= \operatorname{cosec}^{-1}\left(\frac{1}{x}\right) \\ \cos^{-1}x &= \sec^{-1}\left(\frac{1}{x}\right) \\ \tan^{-1}x &= \cot^{-1}\left(\frac{1}{x}\right) \\ \cot^{-1}x &= \tan^{-1}\left(\frac{1}{x}\right) \\ \sec^{-1}x &= \cos^{-1}\left(\frac{1}{x}\right) \\ \operatorname{cosec}^{-1}x &= \sin^{-1}\left(\frac{1}{x}\right) \end{aligned}\right]$$

Application of set (5) (differentiate the following with respect to x).

(8) Here the expression inside the bracket can be simplified (using trigonometric identities) to the form $\tan^{-1}(7x - (\pi/4))$, or else we may use the formula of set (4) to write the right-hand side as $\tan^{-1}[\tan(7x)] - \tan^{-1}[\tan(\pi/4)] = 7x - (\pi/4)$, which can be easily differentiated.

Example (9): Let $y = \sin\left[\operatorname{cosec}^{-1}\left(\frac{1}{x}\right)\right]$

$$\therefore \quad y = \sin(\sin^{-1}x) \quad \text{[Using set (5)]}$$
$$= x$$
$$\therefore \quad \frac{dy}{dx} = 1$$

Example (10): Let $y = \sec\left[\cos^{-1}\left(\frac{2}{5x}\right)\right]$

$$\therefore \quad y = \sec\left(\sec^{-1}\frac{5x}{2}\right) = \frac{5x}{2}$$
$$\therefore \quad \frac{dy}{dx} = \frac{5}{2} \qquad \textbf{Ans.}$$

Example (11): Let $y = \cot^{-1}\left[\frac{3 - 2\tan x}{2 + 3\tan x}\right]$

Note that, using the formula $\cot^{-1}x = \tan^{-1}\left(\frac{1}{x}\right)$,
we can write,

$$y = \tan^{-1}\left[\frac{2 + 3\tan x}{3 - 2\tan x}\right]$$

Observe that the expression on the right-hand side can be simplified *if the denominator is expressed in the form* $(1 - k \tan x)$. This can be done by dividing the numerator and denominator by 3. We then get,

$$y = \tan^{-1}\left[\frac{(2/3) + \tan x}{1 - (2/3)\tan x}\right]$$
$$= \tan^{-1}\left(\frac{2}{3}\right) + \tan^{-1}(\tan x) \quad \left[\because \quad \tan^{-1}\left(\frac{a+b}{1 - a \cdot b}\right) = \tan^{-1}a + \tan^{-1}b\right]$$
$$= \tan^{-1}\left(\frac{2}{3}\right) + x$$
$$\therefore \quad \frac{dy}{dx} = 0 + 1 = 1 \quad \textbf{Ans.}$$

Note (10): It is normally preferred to express $\cot^{-1}x$, $\sec^{-1}x$, and $\operatorname{cosec}^{-1}x$ in the forms $\tan^{-1}t$, $\cos^{-1}t$, and $\sin^{-1}t$, respectively, where t stands for $(1/x)$.

Example (12): Let $y = \cot^{-1}\left(\frac{5 + 4x}{5x - 4}\right)$

$$= \tan^{-1}\left(\frac{5x - 4}{5 + 4x}\right) \quad \left[\because \cot^{-1}t = \tan^{-1}\frac{1}{t}\right]$$

Dividing numerator and denominator by 5, we get

$$y = \tan^{-1}\left(\frac{x - (4/5)}{1 + (4x/5)}\right) = \tan^{-1}\left(\frac{x - (4/5)}{1 + x \cdot (4/5)}\right)$$

$$= \tan^{-1}x - \tan^{-1}\left(\frac{4}{5}\right)$$

$$\therefore \quad \frac{dy}{dx} = \frac{1}{1+x^2} - 0 = \frac{1}{1+x^2} \quad \textbf{Ans.}$$

14.9 APPLICATION OF TRIGONOMETRIC IDENTITIES IN SIMPLIFICATION OF FUNCTIONS AND EVALUATION OF DERIVATIVES OF FUNCTIONS INVOLVING INVERSE TRIGONOMETRIC FUNCTIONS

Sometimes a simplification of the function makes the differentiation easier. It is useful to learn the methods of manipulation on certain trigonometric expressions so that they can be expressed in the desired form(s), which can in turn be simplified, using the relations given above [i.e., sets (1)–(5)]. Such simplifications are possible only in certain functions.

Example (13): Differentiate $\tan^{-1}(\sec x + \tan x)$

Let $$y = \tan^{-1}(\sec x + \tan x) \tag{23}$$

Consider, $\sec x + \tan x = \dfrac{1}{\cos x} + \dfrac{\sin x}{\cos x}$

$$= \frac{1 + \sin x}{\cos x} = \frac{\cos^2(x/2) + \sin^2(x/2) + 2\sin(x/2)\cos(x/2)}{\cos^2(x/2) - \sin^2(x/2)}$$

$$= \frac{\cos(x/2) + \sin(x/2)}{\cos(x/2) - \sin(x/2)} = \frac{1 + \tan(x/2)}{1 - \tan(x/2)} = \frac{\tan(\pi/4) + \tan(x/2)}{1 - \tan(\pi/4) \cdot \tan(x/2)} = \tan\left(\frac{\pi}{4} + \frac{x}{2}\right)$$

$$\therefore \quad y = \tan^{-1}\left[\tan\left(\frac{\pi}{4} + \frac{x}{2}\right)\right]$$

$$= \frac{\pi}{4} + \frac{x}{2} \quad [\because \quad \tan^{-1}(\tan x) = x]$$

$$\frac{dy}{dx} = \frac{1}{2} \qquad \textbf{Ans.}$$

Given below are some trigonometric functions with necessary simplifications, to help understand the approach.

1. $$\frac{\sin x}{1+\cos x} = \frac{2\sin(x/2)\cdot\cos(x/2)}{1+2\cos^2(x/2)-1} = \tan\frac{x}{2}$$

$$\therefore \quad \tan^{-1}\left[\frac{\sin x}{1+\cos x}\right] = \tan^{-1}\left(\tan\frac{x}{2}\right) = \frac{x}{2}$$

$$\frac{\cos x}{1+\sin x} = \frac{\sin((\pi/2)-x)}{1+\cos((\pi/2)-x)}$$

2. $$\tan\frac{1}{2}\left(\frac{\pi}{2}-x\right) = \tan\left(\frac{\pi}{4}-\frac{x}{2}\right)$$

$$\therefore \quad \tan^{-1}\left[\frac{\sin x}{1+\cos x}\right] = \tan^{-1}\left(\tan\left(\frac{\pi}{4}-\frac{x}{2}\right)\right) = \frac{\pi}{4}-\frac{x}{2}$$

3. $$\frac{1+\sin x}{1-\sin x} = \frac{\sin^2(x/2)+\cos^2(x/2)+2\cdot\sin(x/2)\cdot\cos(x/2)}{\sin^2(x/2)+\cos^2(x/2)-2\cdot\sin(x/2)\cdot\cos(x/2)}$$

$$= \left[\frac{\cos(x/2)+\sin(x/2)}{\cos(x/2)-\sin(x/2)}\right]^2$$

$$\therefore \quad \frac{1+\sin x}{1-\sin x} = \left[\frac{1+\tan(x/2)}{1-\tan(x/2)}\right]^2$$

$$= \left[\tan\left(\frac{\pi}{4}+\frac{x}{2}\right)\right]$$

$$\therefore \quad \sqrt{\frac{1+\sin x}{1-\sin x}} = \tan\left(\frac{\pi}{4}+\frac{x}{2}\right)$$

$$\therefore \quad \tan^{-1}\sqrt{\frac{1+\sin x}{1-\sin x}} = \left(\frac{\pi}{4}+\frac{x}{2}\right)$$

4. $$\therefore \quad \tan^{-1}\sqrt{\frac{1+\cos x}{1-\cos x}} = \tan^{-1}\left[\tan^{-1}\frac{x}{2}\right] = \frac{x}{2}$$

5. $$\therefore \quad \tan^{-1}\left(\frac{1}{\sec x+\tan x}\right) = \tan^{-1}\left[\cot\left(\frac{\pi}{4}+\frac{x}{2}\right)\right]$$

$$= \tan^{-1}\left[\tan\left(\frac{\pi}{4}-\frac{x}{2}\right)\right] = \frac{\pi}{4}-\frac{x}{2}$$

6. $$\tan^{-1}(\sec x-\tan x) = \frac{1-\sin x}{\cos x} = \frac{1-\tan(x/2)}{1+\tan(x/2)} = \tan\left(\frac{\pi}{4}-\frac{x}{2}\right)$$

7. $\therefore \quad \tan^{-1}\left(\dfrac{1}{\sec x - \tan x}\right) = \tan^{-1}\left[\cot\left(\dfrac{\pi}{4} - \dfrac{x}{2}\right)\right]$

$$= \tan^{-1}\left[\tan\left(\frac{\pi}{4} + \frac{x}{2}\right)\right] = \frac{\pi}{4} + \frac{x}{2}$$

8. $\operatorname{cosec}x + \cot x = \dfrac{1 - \cos x}{\sin x} = \cot\dfrac{x}{2}$

$\therefore \quad \cot^{-1}(\operatorname{cosec}x + \cot x) = \cot^{-1}\left(\cot\dfrac{x}{2}\right) = \dfrac{x}{2}$

And, $\tan^{-1}(\operatorname{cosec}x + \cot x) = \tan^{-1}\left(\cot\dfrac{x}{2}\right)$

$$= \tan^{-1}\left(\tan\left(\frac{\pi}{2} - \frac{x}{2}\right)\right)$$

$$= \frac{\pi}{4} - \frac{x}{2} \qquad \textbf{Ans.}$$

9. $\dfrac{1}{\operatorname{cosec}x + \cot x} = \dfrac{1}{\cot(x/2)} = \tan\dfrac{x}{2}$

10. $\operatorname{cosec}x - \cot x = \dfrac{1 - \cos x}{\sin x} = \tan\dfrac{x}{2}$

11. $\dfrac{1}{\operatorname{cosec}x - \cot x} = \dfrac{1}{\tan(x/2)} = \cot\dfrac{x}{2} = \tan\left(\dfrac{\pi}{2} - \dfrac{x}{2}\right)$

14.9.1 Evaluation of Derivatives of Inverse Trigonometric Functions by Making Substitutions (Usually Trigonometric Substitutions)

Sometimes appropriate trigonometric substitutions can be made to simplify inverse trigonometric functions in order to compute their derivatives. The following trigonometric formulas give us a clue regarding suitable substitutions. The expressions in question can be simplified using the trigonometric formulas, and the sets of results [set (1)–(5)] given above.

Trigonometric Formulas	Examples of Inverse Trignometric Functions for Simplification
[A] $\sin 2x = \dfrac{2\tan x}{1 + \tan^2 x}$	$\sin^{-1}\left(\dfrac{4x}{1 + 4x^2}\right)$
[B] $\cos 2x = \dfrac{1 - \tan^2 x}{1 + \tan^2 x}$	$\cos^{-1}\left(\dfrac{1 - 4x^2}{1 + 4x^2}\right)$
[C] $\tan 2x = \dfrac{2\tan x}{1 - \tan^2 x}$	$\cot^{-1}\left(\dfrac{1 - x^2}{2x}\right)$
[D] $\tan(x \pm y) = \dfrac{\tan x \pm \tan y}{1 \pm \tan x.\tan y}$	$\tan^{-1}\left(\dfrac{2e^x}{1 - e^{2x}}\right)$

(*Continued*)

Trigonometric Formulas	Examples of Inverse Trignometric Functions for Simplification
[d(i)] $\tan\left(\frac{\pi}{4}+x\right) = \frac{1+\tan x}{1-\tan x}$	$\tan^{-1}\left(\frac{1+x}{1-x}\right)$
[d(ii)] $\tan\left(\frac{\pi}{4}-x\right) = \frac{1-\tan x}{1+\tan x}$	$\tan^{-1}\left(\frac{1-x}{1+x}\right)$
[E]	
[e(i)] $\cos\left(\frac{\pi}{2}-x\right) = \sin x,$	$\sin^{-1}x(\cos x)$
[e(ii)] $\sin\left(\frac{\pi}{2}-x\right) = \cos x,$	$= \sin^{-1}\left[\sin\left(\frac{\pi}{2}-x\right)\right]$
and so on.	$= \frac{\pi}{2}-x$
[F]	
[f(i)] $\cos 2x = \cos^2 x - \sin^2 x$	$\sin^{-1}(1-2x^2)$
[f(ii)] $\cos 2x = 1 - 2\sin^2 x$	$\sec^{-1}\left(\frac{1}{2x^2-1}\right)$
[f(iii)] $\cos 2x = 2\cos^2 x - 1$	$\operatorname{cosec}^{-1}\left(\frac{1}{2x\sqrt{1-x^2}}\right)$
[f(iv)] $\sin 2x = 2\sin x\cos x$	
[G] $\sin^2 x + \cos^2 x = 1$	
[g(i)] $\sin^2 x = 1 - \cos^2 x$	$\tan^{-1}\frac{\sqrt{1-x^2}}{x}$
[g(ii)] $\cos^2 x = 1 - \sin^2 x$	$\cos^{-1}\frac{1}{\sqrt{1+x^2}}$
[g(iii)] $\sec^2 x = 1 + \tan^2 x$	$\sec^{-1}\sqrt{1+x^2}$
[g(iv)] $\operatorname{cosec}^2 x = 1 + \cot^2 x$	$\operatorname{cosec}^{-1}\left(\frac{\sqrt{1+x^2}}{x}\right)$
[h(i)] $\sin 3x = 3\sin x - 4\sin^3 x$	$\operatorname{cosec}^{-1}\left(\frac{1}{3x-4x^3}\right)$
[h(ii)] $\cos 3x = 4\cos^3 x - 3\cos x$	$\cos^{-1}(4x^3-3x)$
[h(iii)] $\tan 3x = \frac{3\tan x - \tan^3 x}{1-3\tan^2 x}$	$\tan^{-1}\left(\frac{3x-x^3}{1-3x^2}\right)$

Given below are some solved examples, which indicate the usefulness of such substitutions.

Example (14): If $y = \tan^{-1}\left(\dfrac{1+x}{1-x}\right)$, find $\dfrac{dy}{dx}$

Solution: Put $x = \tan t$ and $1 = \tan\dfrac{\pi}{4}$

$$\therefore \quad y = \tan^{-1}\left[\frac{\tan(\pi/4) + \tan t}{1 - \tan(\pi/4)\cdot\tan t}\right] = \tan^{-1}\left[\tan\left(\frac{\pi}{4} + t\right)\right]$$

$$= \frac{\pi}{4} + t$$

$$y = \frac{\pi}{4} + \tan^{-1}x. \quad (\because \quad t = \tan^{-1}x)$$

$$\therefore \quad \frac{dy}{dx} = \frac{1}{1+x^2}$$ **Ans.**

Example (15): If $y = \cos^{-1}\left(\dfrac{1-e^{2x}}{1+e^{2x}}\right)$, find $\dfrac{dy}{dx}$

Solution: We have, $y = \cos^{-1}\left[\dfrac{1-(e^x)^2}{1+(e^x)^2}\right]$

Put $e^x = \tan t \therefore t = \tan^{-1}e^x$

$$\therefore \quad y = \cos^{-1}\left[\frac{1-\tan^2 t}{1+\tan^2 t}\right]$$

$$= \cos^{-1}(\cos 2t) = 2t = 2\tan^{-1}e^x$$

$$\therefore \quad \frac{dy}{dx} = \frac{2}{1+(e^x)^2}\cdot\frac{d}{dx}e^x = \frac{2e^x}{1+e^{2x}}$$ **Ans.**

Example (16):

$$y = \tan^{-1}\left(\frac{\sqrt{1+x^2}+\sqrt{1-x^2}}{\sqrt{1+x^2}-\sqrt{1-x^2}}\right), \quad \text{find } \frac{dy}{dx}$$

Solution: We know that,

$$\cos 2t = 2\cos^2 t - 1$$
$$= 1 - 2\sin^2 t$$

$$\left.\begin{aligned} &\text{(i)} \quad \therefore \quad \cos 2t + 1 = 2\cos^2 t \\ &\text{(ii)} \quad \text{or } 1 - \cos 2t = 2\sin^2 t \end{aligned}\right\}^{(9)}$$

[9] It is useful to remember these relations, and use them whenever similar expressions appear.

In view of these relations, we put

$$x^2 = \cos 2t, \quad \text{so that} \quad 2t = \cos^{-1}x^2 \quad \text{or} \quad t = \frac{1}{2}\cos^{-1}x^2$$

$$\therefore \quad 1 + x^2 = 1 + \cos 2t = 2\cos^2 t$$

$$\text{and} \quad 1 - x^2 = 1 - \cos 2t = 2\sin^2 t$$

$$\therefore \quad y = \tan^{-1}\left[\frac{\cos t + \sin t}{\cos t - \sin t}\right]$$

$$= \tan^{-1}\left[\frac{1 + \tan t}{1 - \tan t}\right]$$

$$= \tan^{-1}\left[\tan\left(\frac{\pi}{4} + t\right)\right] = \frac{\pi}{4} + t$$

$$\therefore \quad y = \frac{\pi}{4} + \frac{1}{2}\cos^{-1}x^2 \qquad \left[\because \quad t = \frac{1}{2}\cos^{-1}x^2\right]$$

$$\therefore \quad \frac{dy}{dx} = \frac{1}{2}\left[\frac{-1}{\sqrt{1 - x^4}}\right]\cdot 2x = \frac{-x}{\sqrt{1 - x^4}} \qquad \textbf{Ans.}$$

Example (17): Differentiate $y = \tan^{-1}\left(\frac{\sqrt{1+x^2}-1}{x}\right)$ with respect to x.

Solution: We have, $y = \tan^{-1}\left(\frac{\sqrt{1+x^2}-1}{x}\right)$

Consider the expression, $\frac{\sqrt{1+x^2}-1}{x} = E$ (say)

Put,

$$x = \tan t$$

$$\therefore \quad t = \tan^{-1}x$$

$$\therefore \quad E = \frac{\sqrt{1 + \tan^2 t} - 1}{\tan t} = \frac{\sqrt{\sec^2 t} - 1}{\tan t}$$

$$= \frac{\sec t - 1}{\tan t} = \frac{1 - \cos t}{\sin t}$$

$$= \frac{2\sin^2(t/2)}{2\sin(t/2)\cdot\cos(t/2)} = \tan\frac{t}{2}$$

$$y = \tan^{-1}\left(\tan\frac{t}{2}\right) = \frac{t}{2}$$

$$= \frac{1}{2}\tan^{-1}x \quad \left[\because \quad t = \tan^{-1}x\right]$$

$$\therefore \quad \frac{dy}{dx} = \frac{1}{2}\cdot\frac{d}{dx}\tan^{-1}x$$

$$= \frac{1}{2(1 + x^2)} \qquad \textbf{Ans.}$$

Example (18): Differentiate $y = \sin^{-1}\left(x\sqrt{1-x} - \sqrt{x}\sqrt{1-x^2}\right)$

or
$$y = \sin^{-1}\left(x\sqrt{1-\left(\sqrt{x}\right)^2} - \sqrt{x}\sqrt{1-x^2}\right)$$

Put $x = \sin A$ and $\sqrt{x} = \sin B$

$$\therefore \quad y = \sin^{-1}\left(\sin A\sqrt{1-\sin^2 B} - \sin B\sqrt{1-\sin^2 A}\right)$$
$$= \sin^{-1}(\sin A \cdot \cos B - \cos A \cdot \sin B)$$
$$= \sin^{-1}[\sin(A-B)]$$
$$= A - B$$
$$= \sin^{-1}x - \sin^{-1}\sqrt{x} \quad [\because \quad A = \sin^{-1}x] \quad \text{and} \quad B = \sin^{-1}\sqrt{x}$$

$$\therefore \quad \frac{dy}{dx} = \frac{1}{\sqrt{1-x^2}} - \frac{1}{\sqrt{1-(\sqrt{x})^2}}\frac{d(\sqrt{x})}{dx}$$
$$= \frac{1}{\sqrt{1-x^2}} - \frac{1}{\sqrt{1-x^2}}\cdot\left(\frac{1}{2\sqrt{x}}\right)$$
$$= \frac{1}{\sqrt{1-x^2}} - \frac{1}{2\sqrt{x-x^2}} \qquad \textbf{Ans.}$$

Example (19): If $y = \sin\left[2\tan^{-1}\sqrt{\frac{1-x}{1+x}}\right]$, find $\frac{dy}{dx}$

Solution: We have, $y = \sin\left[2\tan^{-1}\sqrt{\frac{1-x}{1+x}}\right]$

If we put $x = \cos t$, then it can be shown that $\sqrt{\frac{1-x}{1+x}} = \tan\frac{t}{2}$

$$\therefore \quad y = \sin\left[2\tan^{-1}\left(\tan\frac{t}{2}\right)\right]$$
$$= \sin\left[2\cdot\frac{t}{2}\right] = \sin t^{(10)}$$
$$= \sqrt{1-\cos^2 t}$$
$$= \sqrt{1-x^2} \qquad [\because \quad \cos t = x]$$
$$\therefore \quad \frac{dy}{dx} = \frac{d}{dx}\left(1-x^2\right)^{1/2}$$
$$= \frac{1}{2}\cdot\left(1-x^2\right)^{1/2}(-2x)$$
$$= -\frac{x}{\sqrt{1-x^2}} \qquad \textbf{Ans.}$$

(10) At this stage, if we use the relation $x = \cos t$, we get $y = \sin(\cos^{-1}x)$.

$\frac{dy}{dx} = \cos(\cos^{-1}x).\ \frac{d}{dx}\cos^{-1}x = x.\frac{-1}{\sqrt{1-x^2}} = \frac{-x}{\sqrt{1-x^2}}$.

Example (20):

$$y = \sec^{-1}\left(\frac{1+4^x}{1-4^x}\right)$$

$$\therefore \quad = \cos^{-1}\left(\frac{1-4^x}{1+4^x}\right) \qquad \left[\because \sec^{-1}x = \cos^{-1}\left(\frac{1}{x}\right)\right]$$

$$= \cos^{-1}\left(\frac{1-(2^x)^2}{1+(2^x)^2}\right)$$

$$\text{Put } 2^x = \tan t \quad \therefore t = \tan^{-1}2^x \qquad (1)$$

$$\therefore \quad y = \cos^{-1}\left[\frac{1-\tan^2 t}{1+\tan^2 t}\right]$$

$$= \cos^{-1}(\cos 2t) = 2t$$

$$\therefore \quad y = 2\tan^{-1}2^x \qquad \text{[by using (1)]}$$

$$\frac{dy}{dx} = \frac{2}{1+(2^x)^2}\cdot\frac{d}{dx}(2^x)$$

$$= \frac{2}{1+4^x}\cdot 2^x\cdot\log_e 2$$

$$\frac{dy}{dx} = \frac{2^{x+1}}{1+4^x}\log_e 2 \qquad \textbf{Ans.}$$

Exercise (1)

Differentiate the following with respect to x.

(1) $\tan^{-1}\left(\frac{a+x}{1-ax}\right)$

(2) $\tan^{-1}\left(\frac{6x}{1-8x^2}\right)$

(3) $\tan^{-1}\left(\frac{\sin x}{1+\cos x}\right)$

(4) $\tan^{-1}\left(\frac{\cos x+\sin x}{\cos x-\sin x}\right)$

(5) $\tan^{-1}\left(\frac{\cos x}{1+\sin x}\right)$

(6) $\tan^{-1}\left(\frac{1-\cos x}{\sin x}\right)$

(7) $\cos^{-1}\left(\sqrt{\frac{1+\cos x}{2}}\right)$

(8) $\cot^{-1}\left(\frac{1-x}{1+x}\right)$

(9) $\tan^{-1}\left(\frac{1+\sin x}{\cos x}\right)$

(10) $\tan^{-1}\left(\sqrt{\frac{1+\sin x}{1-\sin x}}\right)$

(11) $\tan^{-1}\left(\sqrt{\frac{1+\cos x}{1-\cos x}}\right)$

Answers:

(1) $\frac{1}{1+x^2}$ **(2)** $\frac{2}{1+4x^2}+\frac{4}{1+16x^2}$ **(3)** $\frac{1}{2}$

(4) 1 **(5)** $-\frac{1}{2}$ **(6)** $\frac{1}{2}$

(7) $\frac{1}{2}$ **(8)** $\frac{1}{1+x^2}$ **(9)** $\frac{1}{2}$

(10) $\frac{1}{2}$ **(11)** $-\frac{1}{2}$

Exercise (2)

Differentiate the following with respect to x.

(1) $\sin^{-1}\left(\frac{x}{\sqrt{1+x^2}}\right)$ **(2)** $\sin^{-1}\left(\frac{2\sec x}{1+\sec^2 x}\right)$

(3) $\sin^{-1}\left(\frac{x+\sqrt{1+x^2}}{\sqrt{2}}\right)$ **(4)** $\sin^{-1}\left(\frac{2^{x+1}}{1+4^x}\right)$

(5) $\cos^{-1}\left(\frac{1}{\sqrt{1+x^2}}\right)$ **(6)** $\cos^{-1}\left(\frac{1-x^2}{1+x^2}\right)$

(7) $\tan^{-1}\left(\frac{\sqrt{1+x^2}}{x}\right)$ **(8)** $\tan^{-1}\left(\frac{3x-x^3}{1+3x^2}\right)$

(9) $\sec^{-1}\left(\sqrt{1+x^2}\right)$ **(10)** $\sec^{-1}\left(\frac{1+x^2}{1-x^2}\right)$

(11) $\operatorname{cosec}^{-1}\left(\frac{1}{3x-4x^3}\right)$ **(12)** $\cot^{-1}\left(\frac{\sqrt{1+x^2}}{x}\right)$

Answers:

(1) $\frac{1}{1+x^2}$ **(2)** $\frac{2\sin x}{1+\cos^2 x}$

(3) $\frac{1}{\sqrt{1-x^2}}$ **(4)** $\frac{2^{x+1}\log 2}{1+4^x}$

(5) $\frac{1}{1+x^2}$ **(6)** $\frac{2}{1+x^2}$

(7) $\dfrac{-1}{\sqrt{1-x^2}}$ **(8)** $\dfrac{3}{1+x^2}$

(9) $\dfrac{1}{1+x^2}$ **(10)** $\dfrac{2}{1+x^2}$

(11) $-\dfrac{3}{\sqrt{1-x^2}}$ **(12)** $\dfrac{1}{\sqrt{1-x^2}}$

Note (11): *The inverse trigonometric functions discussed above are of a special type* and as we have seen, their derivatives can be computed using *special methods involving substitution* and/ or simplification. On the other hand, there can be any number of functions involving inverse trigonometric functions whose derivatives are computed simply by applying the rules of differentiation. Of course, substitution may also be useful as an intermediate step. Consider the following examples.

(a) Let $y = x^2 \cdot \cos^{-1}x$

Here, we have a product of two functions, and therefore we must use the *product rule* for derivatives.

$$\frac{dy}{dx} = x^2 \cdot \frac{d}{dx}\cos^{-1}x + \cos^{-1}x \cdot \frac{d}{dx}x^2$$

$$= x^2 \cdot \frac{-1}{\sqrt{1-x^2}} + \cos^{-1}x \cdot (2x)$$

$$\therefore \quad \frac{dy}{dx} = 2x \cdot \cos^{-1}x - \frac{x^2}{\sqrt{1-x^2}} \qquad \textbf{Ans.}$$

(b) Let $y = \dfrac{\sin^{-1}x}{x^2+1}$

Here we must use the *quotient rule* for derivatives.

(c) Let $y = \sin(\tan^{-1}x)$

Here we have to use the *chain rule.*

$$\frac{dy}{dx} = \cos(\tan^{-1}x) \cdot \frac{d}{dx}(\tan^{-1}x)$$

$$\therefore \quad \frac{dy}{dx} = \frac{\cos(\tan^{-1}x)}{1+x^2} \qquad \textbf{Ans.}$$

Exercise (3)

Differentiate the following with respect to x.

(1) $y = \sin^{-1}\sqrt{x}$ **(2)** $y = \sin^{-1}ax$

(3) $y = \sin^{-1}(2x)$ **(4)** $y = \cos^{-1}(\sqrt{\cos x})$

(5) $y = \sin^{-1}\left(\dfrac{x}{a}\right)$ **(6)** $\dfrac{\tan^{-1}x}{1+x^2}$

(7) $\log(\tan^{-1}x)$ **(8)** $\sin^{-1}(3x+2)$

(9) $\cos^{-1}x^2$

Answers:

(1) $\dfrac{1}{2\sqrt{x}\sqrt{1-x}}$

(2) $\dfrac{a}{\sqrt{1-a^2x^2}}$

(3) $\dfrac{2}{\sqrt{1-4x^2}}$

(4) $\dfrac{\sin x}{2\sqrt{1-\cos x}\sqrt{\cos x}}$

(5) $\dfrac{1}{\sqrt{a^2-x^2}}$

(6) $\dfrac{1-2x\tan^{-1}x}{(1+x^2)^2}$

(7) $\dfrac{1}{(1+x^2)\tan^{-1}x}$

(8) $\dfrac{3}{\sqrt{1+(3x+2)^2}}$

(9) $\dfrac{-2x}{\sqrt{1-x^2}}$

15a Implicit Functions and Their Differentiation

15a.1 INTRODUCTION

First, let us distinguish between *explicit* and *implicit functions*. Functions of the form, $y = f(x)$, in which y (alone) is directly expressed in terms of the function(s) of x, are called *explicit functions*.

Example (1):

$$\begin{aligned} &y = x^2 + 3x - 2; \qquad && y = \sin x + 2e^x \\ &y = (x+3)/(1+x^2); \qquad && y = \cos x + \log_e(1+x^2), \quad \text{and so on} \end{aligned} \tag{1}$$

Not all functions, however, can be defined by equations of this type. For example, we cannot solve the following equations for y (alone) in terms of the functions of x.

Examples (2):

$$\begin{aligned} &x^3 + y^3 = 2xy; \qquad && y^5 + 3y^2 - 2x^2 = -4; \qquad && x^2 + y^2 = 36; \\ &\sin y = x \sin(a+y); \qquad && y^3 + 7y = x^3, \text{ and so on} \end{aligned} \tag{2}$$

Such relations connecting x and y are called *implicit relations. An implicit relation (in x and y) may represent jointly two or more functions of x.*

As an example, the relation $x^2 + y^2 = 36$ jointly represents two functions:

$$y = \sqrt{36 - x^2} \quad \text{and} \quad y = -\sqrt{36 - x^2}.$$

Remark: *Every explicit function $y = f(x)$ can also be expressed as an implicit function.* For example, we may write the above equation in the form $y - f(x) = 0$ and call it an implicit function of x. Thus, the term *explicit function and implicit function do not characterize the nature of a function* but merely the way a function is defined.[1]

(Implicit functions may be expressed in the form $f = \{(x, y) | y = f(x)\}$.)

15a-Differentiation technique for implicit functions and the method of logarithmic differentiation (For general exponential functions and other expressions involving products, quotients and powers of functions)

[1] *Differential and Integral Calculus* by N. Piskunov (vol. I, p. 86), Mir Publishers Moscow, 1974.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

Note (1): In the case of an implicit function in the form, $y - f(x) = 0$, it is quite simple to compute the derivative dy/dx since it is as good as if we are handling an explicit function. Hence, here onward we shall consider the implicit functions such as those given in (II) above.[(2)]

Note (2): *It is assumed that an implicit relation defines y as at least one differentiable function of x.* With this *assumption,* the derivate of y with respect to x can be found *without transforming it into the explicit form.*

(*This assumption is important* since certain relations in x and y may not represent any function. For example, the relation $x^2 + y^2 = -36$ does not represent any function.)

Note (3): *The technique of implicit differentiation is based on the chain rule.*

For example, consider the equation

$$y^3 + 7y = x^3 \qquad (3)^{(3)}$$

Differentiating both the sides with respect to x, *treating y as a function of x*, we get (via the *rule for differentiating a composite function*)

$$3y^2 \frac{dy}{dx} + 7\frac{dy}{dx} = 3x^2 \qquad (4)$$

Now solving (4) for $\frac{dy}{dx}$, we get

$$\frac{dy}{dx}(3y^2 + 7) = 3x^2 \quad \therefore \quad \frac{dy}{dx} = \frac{3x^2}{3y^2 + 7}$$

Note that, the above expression for dy/dx *involves both x and y*. If it is required *to find the value of the derivative of an implicit function for a given value of x, then we have to first find the corresponding value of y, using the given relation* (such as in 3). This will help in computing the value of dy/dx(or the *slope of the curve) at those points that lie on the graph of the given equation.*

For example, the point (2, 1) *satisfies equation* (3); hence, it must be on its graph. At (2, 1), we have

$$\frac{dy}{dx} = \frac{3(2)^2}{3(1)^2 + 7} = \frac{12}{10} = \frac{6}{5}$$

Thus, the slope of the curve at (2, 1) is 6/5.

On the other hand, if we have to find the gradient at the point (1, 1) of the curve $x^2 + y^2 - 3x + 4y - 3 = 0$, then it is a simpler situation. It can be seen that $dy/dx = (-2x + 3)/(2y + 4) = 1/6$ at (1, 1).

[(2)] From this point of view, a relation like $x^y \cdot y^x = a^b$ may also be looked upon as an implicit function; however, to compute the derivative dy/dx in such cases, there is only one method available, namely, the logarithmic differentiation (to be discussed later in this chapter). No other method is helpful.

[(3)] Note that, in this equation though all the terms involving y are on LHS, the value of y (alone) is not expressed in terms of the functions of x, and hence it is an implicit function of our interest.

Now, we ask the question: *Is the method of implicit differentiation legitimate? Does it give the right answer?* We can give *evidence for the correctness of the method through examples*, which can be solved in two ways.

Let us find $\mathrm{d}y/\mathrm{d}x$, if $4x^2y - 3y = x^3 - 1$.

Method (1): Here, we have, $y = (x^2 - 1)/(4x^2 - 3)$, which defines y explicitly.

We get,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{4x^4 - 9x^2 + 8x}{(4x^2 - 3)^2} \qquad \text{(by quotient rule)} \tag{5}$$

Method (2): (Implicit Differentiation)

Now, after using the product rule in the first term, we obtain

$$\begin{aligned} 4x^2\frac{\mathrm{d}y}{\mathrm{d}x} + y\cdot 8x - 3\frac{\mathrm{d}y}{\mathrm{d}x} &= 3x^2 \\ \therefore\ \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{3x^2 - 8xy}{4x^2 - 3} \end{aligned} \tag{6}$$

This answer looks different from the one obtained at (5). However, if we substitute $y = (x^3 - 1)/(4x^2 - 3)$ in (6), we get the same expression for $\mathrm{d}y/\mathrm{d}x$, as in (5).

Thus, we observe that, *if* an equation in x and y determines a function $y = f(x)$ and if *this function is differentiable*, *then the method of implicit differentiation* will yield a correct expression for $\mathrm{d}y/\mathrm{d}x$.

(Note the "*two ifs*" in this statement.)

15a.2 CLOSER LOOK AT THE DIFFICULTIES INVOLVED

The equation $x^2 + y^2 = -1$ has no solution and, therefore, does not determine a function.

On the other hand,

$$x^2 + y^2 = 25 \tag{7}$$

represents a circle with center at the origin and radius 5 units (Figure 15a.1). It does not represent any function of x.

For each x in the open interval $(-5, 5)$, there are two corresponding values of y, namely,

$$y = \sqrt{25 - x^2} \quad \text{and} \quad y = -\sqrt{25 - x^2}$$

They represent two functions, in the interval $(-5, 5)$, given by

$$y = f(x) = \sqrt{25 - x^2} \tag{8a}$$

and

$$y = g(x) = -\sqrt{25 - x^2} \tag{8b}$$

Their graphs are the upper and lower semicircles, respectively, as shown below in Figures 15a.2a and 15a.2b.

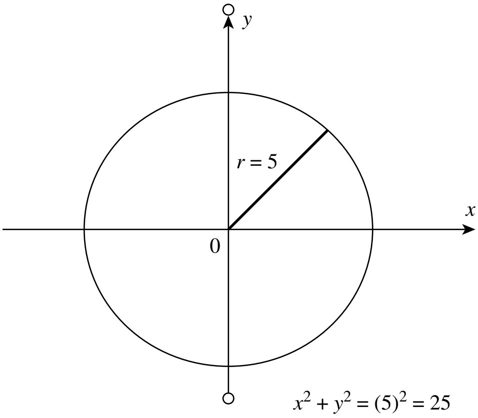


FIGURE 15a.1

It may be noted that *both functions are differentiable in the open interval* (−5, 5), but not at $x = \pm 5$ (since their graphs have vertical tangents at those (end) points. Let us find their derivatives.

First, consider $f(x) = \sqrt{25 - x^2}$. It satisfies $x^2 + [f(x)]^2 = 25$, where

$$f(x) = y.$$

When we differentiate $f(x)$ implicitly and solve for $f'(x)$, we obtain

$$2x + 2f(x)f'(x) = 0$$

$$\therefore\ f'(x) = \frac{-2x}{2f(x)} = -\frac{x}{\sqrt{25 - x^2}}$$

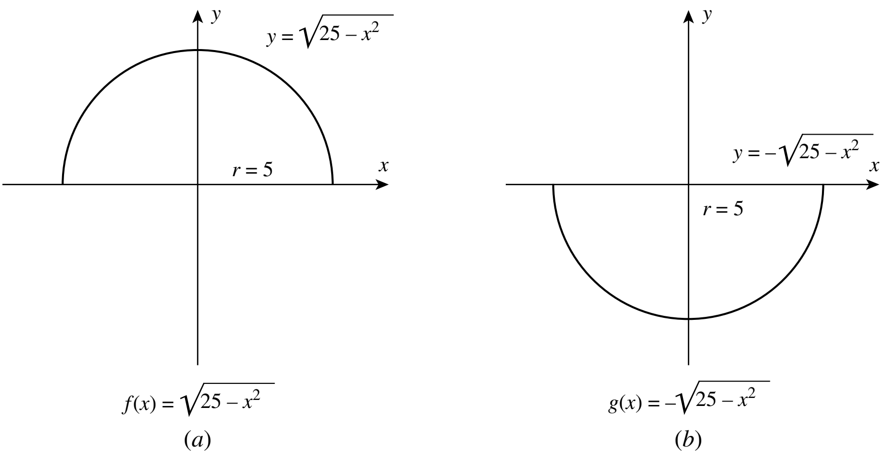


FIGURE 15a.2

A completely similar treatment of $g(x)$ yields

$$g'(x) = \frac{-x}{g(x)} = \frac{-x}{-\sqrt{25-x^2}} = \frac{x}{\sqrt{25-x^2}}$$

For practical purposes, *we can obtain both these results simultaneously by the implicit differentiation* of $x^2 + y^2 = 25$. We get

$$2x + 2y\frac{dy}{dx} = 0$$

$$\therefore \quad \frac{dy}{dx} = \frac{-x}{y} = \begin{cases} -x/\sqrt{25-x^2}, & \text{if } y = f(x) \\ -x/-\sqrt{25-x^2}, & \text{if } y = g(x) \end{cases}$$

It is enough to know that $dy/dx = -x/y$. Suppose, we want to know the slope of the tangent line to the circle $x^2 + y^2 = 25$, when $x = 3$. *The corresponding y-values are 4 and −4.* The slope at (3, 4) is $-3/4$, and that at (3, −4) is $3/4$.

15a.2.1

When an equation of the form $\phi(x, y) = 0$ is differentiated implicitly, we get dy/dx in the form of a *quotient.* At certain points (x, y) on the curve, *the denominator of this quotient, representing* dy/dx, *may become zero.* In fact, these are the points where the tangent line is vertical and hence the slope of the curve (i.e., dy/dx) is not defined.(4)

Example (3): Let us find dy/dx, if $y^5 + 3y^2 - 2x^2 = -4$.

Differentiating both sides of the given equation "with respect to x" (using the chain rule), we obtain

$$5y^4\frac{dy}{dx} + 6y\frac{dy}{dx} - 4x = 0$$

We now solve for dy/dx, obtaining

$$\frac{dy}{dx} = \frac{4x}{5y^4 + 6y}$$

This formula gives dy/dx *at any point* (x, y) on the curve *where the denominator* $5y^4 + 6y$ *is nonzero.*

For example, it is easily seen that the point (2, 1) satisfies $y^5 + 3y^2 - 2x^2 = -4$, and therefore it lies on the curve. Then

$$\left.\frac{dy}{dx}\right|_{(2,1)} = \left.\frac{4x}{5y^4 + 6y}\right|_{(2,1)} = \frac{8}{11}$$

(4) The subject of implicit functions leads to some other difficult technical questions, which are dealt with in *Advanced Calculus*. The problems we study here have straightforward solutions. (For details, refer to *Calculus with Analytic Geometry* (Fifth Edition) by Edwin J. Purcell and Dale Varberg (p. 135), Prentice-Hall Inc., New Jersey.)

Two intersecting curves are said to be *orthogonal* to each other if the tangent lines at the point of their intersection are perpendicular.

Example (4): Let us show that the curve $y - x^2 = 0$ is orthogonal to the curve $x^2 + 2y^2 = 3$, at the point (1, 1) of intersection.

Solution: The given curve is $y = x^2$. The slope of the tangent line to this curve is given by

$$\frac{dy}{dx} = 2x \quad \therefore \left.\frac{dy}{dx}\right|_{(1,\,1)} = 2 = m_1 \text{ (say)}$$

The other curve is $x^2 + 2y^2 = 3$.

Differentiating *implicitly* w.r.t. x, we get

$$2x + 4y\frac{dy}{dx} = 0$$

$$\therefore \frac{dy}{dx} = \frac{-2x}{4y} = \frac{-x}{2y} \quad \therefore \left.\frac{dy}{dx}\right|_{(1,\,1)} = \frac{-1}{2} = -\frac{1}{2} = m_2 \quad \text{(say)}$$

Since $m_1 \cdot m_2 = -1$, the curve $y = x^2$ is orthogonal to the curve $x^2 + 2y^2 = 3$, at the point (1, 1), of their intersection.

Remark: Whenever it is required to find the value of dy/dx at a particular point on given curve, we can *easily check that the point in question lies on the curve*.

Use implicit differentiation to find the derivative of y with respect to x, at the given point.[5]

(a) $x^2 - y^2 = 1$; $(\sqrt{3}, \sqrt{2})$
(b) $x^4 + xy^3 = 0$; $(-1, 1)$
(c) $(2x + y)^5 = 31 - 1/x$; $(-1, 4)$

Note (4): Implicit differentiation is useful in computing *related rates*.

This topic is discussed in Chapter 18.

Example (5): If $x^3 + y^3 = 3axy$, find dy/dx.

Solution: We have $x^3 + y^3 = 3axy$.

Differentiating implicitly both the sides, w.r.t. x, we get

$$3x^2 + 3y^2\frac{dy}{dx} = 3a\left(x\frac{dy}{dx} + y \cdot 1\right) = 3a\,x\,\frac{dy}{dx} + 3a\,y$$

$$(3y^2 - 3ax)\,\frac{dy}{dx} = 3ay - 3x^2$$

$$\therefore \frac{dy}{dx} = \frac{3(ay - x^2)}{3(y^2 - ax)} = \frac{ay - x^2}{y^2 - ax} \quad \textbf{Ans.}$$

[5] *Calculus with Analytic Geometry* (Alternate Edition) by Robert Ellis and Denny Gulick (pp. 151–155).

Example (6): If $x^y = e^{x-y}$, show that $dy/dx = (\log_e x)/(1+\log_e x)^2$.

Solution: We have $x^y = e^{x-y}$.

$$\therefore\ x - y = \log_e x^y = y\log_e x \quad \text{(by definition of logarithm)}$$

$$\therefore\ x = y + y\log_e x = y(1+\log_e x)$$

$$\therefore\ y = \frac{x}{(1+\log_e x)}$$

Differentiating both the sides w.r.t. x, we get

$$\frac{dy}{dx} = \frac{(1+\log_e x)(1) - x(0+1/x)}{(1+\log_e x)^2}$$

$$= \frac{1+\log_e x - 1}{(1+\log_e x)^2} = \frac{\log_e x}{(1+\log_e x)^2} \qquad \textbf{Ans.}$$

Exercise (1)

Q1. If $x^a y^b = (x+y)^{a+b}$, and $ay \neq bx$, prove that $dy/dx = y/x$.

Q2. If $\sin y = x \sin(a+y)$, show that $dy/dx = \sin^2(a+y)/\sin a$.

Q3. If $y = \sin(x+y)$, find dy/dx.

Ans. $\dfrac{\cos(x+y)}{1-\cos(x+y)}$ or $\dfrac{\sqrt{1-y^2}}{1-\sqrt{1-y^2}}$

Q4. If $y = x\,e^y$, show that $dy/dx = y/x(1-y)$.

Q5. Find the equation of the tangent line to the curve

$y^3 - xy^2 + \cos xy = 2$ at the point (0, 1).

Ans. $\dfrac{1}{3}x + 1$

Q6. If $x \sin 2y = 2\cos 2x$, find dy/dx.

Ans. $\dfrac{2y\sin 2x + \sin 2y}{\cos 2x - 2x\cos 2y}$

Q7. If tan $(x+y)+\tan(x-y)=1$, find dy/dx.

$$\textbf{Ans.}\ \frac{\sec^2(x+y)+\sec^2(x-y)}{\sec^2(x-y)-\sec^2(x+y)}$$

Q8. If sin $y=x\cos(a+y)$, then prove that $dy/dx=\cos^2(a+y)/\cos a$.

Q9. If $\sqrt{1-x^2}+\sqrt{1-y^2}=a(x-y)$, then prove that

$$\frac{dy}{dx}=\sqrt{\frac{1-y^2}{1-x^2}}$$

Q10. If $x=y\log(xy)$, then prove that $dy/dx=(y(x-y))/(x(x+y))$.

We give below the solutions of the first five problems.

Q1. If $x^a y^b=(x+y)^{a+b}$ and $ay\neq bx$, prove that $dy/dx=y/x$.

Solution: From the given relation, on taking logarithms, we have

$$a\log_e x+b\log_e y=(a+b)\log_e(x+y)$$

Differentiating w.r.t. x, we get

$$a\frac{1}{x}+b\frac{1}{y}\frac{dy}{dx}=(a+b)\frac{1}{x+y}\left(1+\frac{dy}{dx}\right)$$

$$\text{or}\quad \frac{a}{x}+\frac{b}{y}\frac{dy}{dx}=\frac{a+b}{x+y}+\frac{(a+b)}{(x+y)}\frac{dy}{dx}$$

$$\text{or}\quad \left[\frac{b}{y}-\frac{a+b}{x+y}\right]\frac{dy}{dx}=\frac{a+b}{x+y}-\frac{a}{x}$$

$$\text{that is}\quad \frac{bx+by-ay-by}{y(x+y)}\frac{dy}{dx}=\frac{ax+bx-ax-ay}{x(x+y)}$$

$$\text{or}\quad \frac{bx-ay}{y(x+y)}\frac{dy}{dx}=\frac{bx-ay}{x(x+y)} \tag{9}$$

But it is given that $ay\neq bx$ (i.e., $bx-ay\neq 0$).
$\therefore$ From (9), we get

$$\frac{1}{y}\frac{dy}{dx}=\frac{1}{x}$$

$$\text{or}\quad \frac{dy}{dx}=\frac{y}{x}\qquad \textbf{Ans.}$$

Q2. If sin $y=x\sin(a+y)$, show that $dy/dx=(\sin^2(a+y))/\sin a$.

Solution: We have sin $y=x\sin(a+y)$ (10)

Differentiating implicitly both sides w.r.t. x, we get

$$\cos y\frac{\mathrm{d}y}{\mathrm{d}x} = x\cos(a+y)\frac{\mathrm{d}y}{\mathrm{d}x} + \sin(a+y)$$

$$\therefore\ [\cos y - x\cos(a+y)]\frac{\mathrm{d}y}{\mathrm{d}x} = \sin(a+y)$$

$$\therefore\ \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\sin(a+y)}{\cos y - x\cos(a+y)} \tag{11}$$

Now, observe that the above result contains x whereas the desired result does not. Hence, we try to remove x from the above result using (10) and get

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\sin(a+y)}{\cos y - \dfrac{(\sin y)}{\sin(a+y)}\cos(a+y)}$$

$$= \frac{\sin^2(a+y)}{\sin(a+y)\cos y - \cos(a+y)\sin y} = \frac{\sin^2(a+y)}{\sin(a+y-y)}$$

$$= \frac{\sin^2(a+y)}{\sin a}$$

Method II: From (10), we get $x = \sin y/\sin(a+y)$.

Now, differentiating both sides w.r.t. x (by applying quotient rule to the RHS), we can easily prove the desired result, as follows:

$$\text{We get, } 1 = \frac{\sin(a+y)\cos y\dfrac{\mathrm{d}y}{\mathrm{d}x} - \sin y\cos(a+y)\dfrac{\mathrm{d}y}{\mathrm{d}x}}{[\sin(a+y)]^2}$$

$$= \frac{\dfrac{\mathrm{d}y}{\mathrm{d}x}[\sin(a+y-y)]}{[\sin(a+y)]^2}$$

$$= \frac{\dfrac{\mathrm{d}y}{\mathrm{d}x}\sin a}{\sin^2(a+y)}$$

$$\therefore\ \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\sin^2(a+y)}{\sin a}$$

Q3. If $y = \sin(x+y)$, find $\frac{\mathrm{d}y}{\mathrm{d}x}$.

Solution: Given $y = \sin(x+y)$

Differentiating both sides w.r.t. x, we get

$$\frac{dy}{dx} = \cos(x+y)\frac{d}{dx}(x+y)$$

$$= \cos(x+y)\left(1+\frac{dy}{dx}\right)$$

$$\therefore \frac{dy}{dx} - \cos(x+y)\frac{dy}{dx} = \cos(x+y)$$

$$\text{or} \quad \frac{dy}{dx}[1-\cos(x+y)] = \cos(x+y)$$

$$\therefore \frac{dy}{dx} = \frac{\cos(x+y)}{1-\cos(x+y)} \tag{12}$$

Method II: Given, $y = \sin(x+y)$

$$\therefore x + y = \sin^{-1} y$$

Differentiating both sides w.r.t. x, we get

$$1 + \frac{dy}{dx} = \frac{1}{\sqrt{1-y^2}}\frac{dy}{dx}^{(6)}$$

$$1 = \left(\frac{1}{\sqrt{1-y^2}} - 1\right)\frac{dy}{dx} = \left(\frac{1-\sqrt{1-y^2}}{\sqrt{1-y^2}}\right)\frac{dy}{dx}$$

$$\therefore \frac{dy}{dx} = \frac{\sqrt{1-y^2}}{1-\sqrt{1-y^2}} \tag{13}$$

Check that (12) and (13) are the same.

Q4. If $y = x\,e^y$, show that $dy/dx = y/x(1-y)$.

Solution: Given, $y = x\,e^y$ (14)

Differentiating both sides w.r.t. x, we get

$$\frac{dy}{dx} = x\frac{d}{dx}(e^y) + e^y\frac{d}{dx}(x)$$

$$= x\,e^y\frac{dy}{dx} + e^y$$

$$\therefore (1 - x\,e^y)\frac{dy}{dx} = e^y$$

$$\therefore \frac{dy}{dx} = \frac{e^y}{1 - x\,e^y}$$

[6] Derivatives of inverse trigonometric functions are discussed in Chapter 14.

Now, observe that the term e^y does not appear in the desired result. Hence, we eliminate it using (14) and get

$$\frac{dy}{dx} = \frac{y}{x}\frac{1}{1-y} = \frac{y}{x(1-y)} \quad \textbf{Ans.}$$

Q5. Find the equation of the tangent line to the curve

$$y^3 - xy^2 + \cos xy = 2 \text{ at the point } (0,\ 1)$$

Solution: Given $y^3 - xy^2 + \cos\ xy = 2$

Differentiating both sides implicitly w.r.t. x, we get

$$3y^2\frac{dy}{dx} - \left[x\cdot 2y\frac{dy}{dx} + y^2\right] - \sin xy\left(x\frac{dy}{dx} + y\right) = 0$$

$$3y^2\frac{dy}{dx} - 2xy\frac{dy}{dx} - y^2 - x\sin xy\frac{dy}{dx} - y\sin xy = 0$$

$$\frac{dy}{dx}\left[3y^2 - 2xy - x\sin xy\right] = y^2 + y\sin xy$$

$$\therefore\ \frac{dy}{dx} = \frac{y^2 + y\sin xy}{3y^2 - 2xy - x\sin xy}$$

$$\therefore\ \left.\frac{dy}{dx}\right|_{(0,1)} = \frac{1^2 + 0}{3(1)^2 - 0 - 0} = \frac{1}{3}$$

Thus, the equation of the tangent line at (0, 1) is

$$y - 1 = \frac{1}{3}(x - 0)$$

$$\text{or} \quad y = \frac{1}{3}x + 1 \quad \textbf{Ans.}$$

15a.3 THE METHOD OF LOGARITHMIC DIFFERENTIATION

(For (complicated) functions such as general exponential functions and other expressions involving products, quotients, and powers of functions.)

Recall that to find the derivative $d(x^n)/dx$, we use the *power rule*:

$$\frac{d}{dx}(x^n) = nx^{n-1}$$

Also, we get

$$\frac{d}{dx}[f(x)^n] = n[f(x)]^{n-1}f'(x)$$

using power rule and the chain rule.

But, we cannot use the power rule to find $d(e^x)/dx$. Thus, $d(e^x)/dx \neq x\cdot e^{x-1}$

Recall that, $\mathrm{d}(a^x)/\mathrm{d}x = a^x \log_e a$, which is the differentiation formula for the exponential function.

Thus, we get,

$$\frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{e}^x) = \mathrm{e}^x \log_e \mathrm{e} = \mathrm{e}^x \; [\because \log_e \mathrm{e} = 1]$$

$$\text{and } \frac{\mathrm{d}}{\mathrm{d}x}\left[a^{f(x)}\right] = a^{f(x)} \log_e a \cdot f'(x)$$

using differentiation formula for exponential function and the *chain rule*.

15a.3.1

Now, we ask the question; *what can we write for* $\mathrm{d}(x^x)/\mathrm{d}x$?

Of course, it would be sheer nonsense to write $\mathrm{d}(x^x)/\mathrm{d}x = x \cdot x^{x-1}$.

It is for these types of functions, and *more generally for functions of the type* $y = [f(x)]^{g(x)}$, where both $f(x)$ and $g(x)$ are *differentiable functions* of x, that we can use *the technique of logarithmic differentiation* for computing their derivatives.

This technique is also used *to simplify differentiation of many (complicated) functions* involving *products, quotients, and powers of different functions.*

We list below the *right technique for differentiating each of the following forms of functions:*

$$[f(x)]^n \rightarrow \text{Power rule}$$

$$y = a^{f(x)} \rightarrow \text{Differentiation formula for exponential functions}$$

$$[f(x)]^{g(x)} \rightarrow \text{Logarithmic differentiation}$$

Remark: *The technique of logarithmic differentiation is so powerful that it can be used for each of these forms.*

15a.4 PROCEDURE OF LOGARITHMIC DIFFERENTIATION

The procedure of logarithmic differentiation involves taking *natural logarithm* of each side of the given equation. After simplifying (by using properties of logarithms), we differentiate both sides w.r.t. x. *The usefulness of the process is due to the fact that the differentiation of the product of functions is reduced to that of a sum; of their quotients to that of a difference; and of the general exponential to that of the product of simpler functions.*

The following solved examples will illustrate the process of *logarithmic differentiation*. First, we start with the differentiation of certain (complicated) function involving *products, quotients, and powers of functions*.

Example (7): If $y = \mathrm{e}^{5x} \sin 2x \cos x$, find $\mathrm{d}y/\mathrm{d}x$.

We have, $y = \mathrm{e}^{5x} \sin 2x \cos x$

Taking the *natural logarithm* of both sides, we get

$$\log_e y = \log_e \mathrm{e}^{5x} + \log_e \sin 2x + \log_e \cos x$$

Differentiating w.r.t. x, we get

$$
\begin{aligned}
\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{1}{\mathrm{e}^{5x}}\frac{\mathrm{d}}{\mathrm{d}x}\left(\mathrm{e}^{5x}\right) + \frac{1}{\sin 2x}\frac{\mathrm{d}}{\mathrm{d}x}(\sin 2x) + \frac{1}{\cos x}\frac{\mathrm{d}}{\mathrm{d}x}(\cos x) \\
&= \frac{1}{\mathrm{e}^{5x}}\mathrm{e}^{5x}\cdot 5 + \frac{1}{\sin 2x}\cos 2x \cdot 2 + \frac{1}{\cos x}(-\sin x) \\
&= 5 + 2\cot 2x - \tan x \\
&= \therefore \ \frac{\mathrm{d}y}{\mathrm{d}x} = y[5 + 2\cot 2x - \tan x] \\
&= \mathrm{e}^{5x}\sin 2x \cos x[5 + 2\cot 2x - \tan x] \quad \textbf{Ans.}
\end{aligned}
$$

Example (8): If $y = \mathrm{e}^{4x}\sin^2 x\tan^3 x$, find $\mathrm{d}y/\mathrm{d}x$.

We have $y = \mathrm{e}^{4x}\sin^2 x\tan^3 x$

Taking the *natural logarithms* of both sides, we get

$$
\begin{aligned}
\log_e y &= \log_e \mathrm{e}^{4x} + \log_e \sin^2 x + \log_e \tan^3 x \\
&= 4x + 2\log_e \sin x + 3\log_e \tan x
\end{aligned}
$$

Differentiating w.r.t. x, we get

$$
\begin{aligned}
\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} &= 4 + 2\frac{1}{\sin x}\cos x + 3\frac{1}{\tan x}\sec^2 x \\
&= 4 + 2\cot x + \frac{3}{\sin x \cdot \cos x} \\
\therefore \ \frac{\mathrm{d}y}{\mathrm{d}x} &= y\left[4 + 2\cot x + \frac{3}{\sin x \cdot \cos x}\right] \\
&= \mathrm{e}^{4x}\sin^2 x\tan^3 x\left[4 + 2\cot x + \frac{3}{\sin x \cdot \cos x}\right] \quad \textbf{Ans.}
\end{aligned}
$$

Example (9): If $y = \sqrt{\frac{(1+x)(2+x)}{(1-x)(2-x)}}$, find $\frac{\mathrm{d}y}{\mathrm{d}x}$.

Taking *natural logarithm* of both sides, we get

$$
\begin{aligned}
\log_e y &= \frac{1}{2}[\log_e(1+x)(2+x) - \log_e(1-x)(2-x)] \\
&= \frac{1}{2}[\log_e(1+x) + \log_e(2+x) - \log_e(1-x) - \log_e(2-x)]
\end{aligned}
$$

Differentiating w.r.t. x, we get

$$\frac{1}{y}\frac{dy}{dx} = \frac{1}{2}\left[\frac{1}{1+x} + \frac{1}{2+x} - \frac{1}{1-x}(-1) - \frac{1}{2-x}(-1)\right]$$

$$\therefore\ \frac{dy}{dx} = \frac{y}{2}\left[\frac{1}{1+x} + \frac{1}{2+x} + \frac{1}{1-x} + \frac{1}{2-x}\right]$$

$$= \frac{y}{2}\left[\frac{(1-x)+(1+x)}{(1+x)(1-x)} + \frac{(2-x)+(2+x)}{(2+x)(2-x)}\right] \quad \text{(Imp.)}$$

$$= \frac{y}{2}\left[\frac{2}{1-x^2} + \frac{4}{4-x^2}\right]$$

$$= y\left[\frac{1}{1-x^2} + \frac{2}{4-x^2}\right]$$

$$= y\left[\frac{4-x^2+2-2x^2}{(1-x^2)(4-x^2)}\right]$$

$$= \sqrt{\frac{(1+x)(2+x)}{(1-x)(2-x)}}\left[\frac{6-3x^2}{(1-x^2)(4-x^2)}\right] \quad \textbf{Ans.}$$

Now, we consider functions of the type $[f(x)]^{g(x)}$. Here, it may be mentioned that such functions do not occur naturally. However, to demonstrate the power of technique of the logarithmic differentiation, we solve the following examples.

Example (10): If $y = 5^{\tan x}$, find dy/dx.

We have $y = 5^{\tan x}$

Taking natural logarithm of each side, we get

$$\log_e y = \tan x \cdot \log_e 5$$

Differentiating w.r.t. x, we get

$$\frac{1}{y}\frac{dy}{dx} = \sec^2 x \cdot \log_e 5$$

$$\therefore\ \frac{dy}{dx} = y[\sec^2 x \cdot \log_e 5]$$

$$= 5^{\tan x}[\sec^2 x \cdot \log_e 5] \quad \textbf{Ans.}$$

Example (11): If $y = x^x$, find dy/dx.

We have $y = x^x$ [7]

Taking the natural logarithm of each side, we obtain

$$\log_e y = x \log_e x$$

Differentiating both sides w.r.t. x, we have

$$\frac{1}{y}\frac{dy}{dx} = x\left(\frac{1}{x}\right) + (\log_e x)(1) = 1 + \log_e x$$

$$\therefore \frac{dy}{dx} = y(1 + \log_e x) = x^x(1 + \log_e x) \quad \textbf{Ans.}$$

Example (12): If $y = x^{x^x}$, find dy/dx.

We have $y = (x)^{x^x}$

Taking the *natural logarithm* of each side, we get

$$\log_e y = x^x \log_e x$$

Differentiating both sides w.r.t. x, we get

$$\frac{1}{y}\frac{dy}{dx} = x^x \frac{d}{dx}\log_e x + \log_e x \frac{d}{dx}(x^x)$$

$$= x^x \frac{1}{x} + \log_e x \frac{d}{dx}(x^x)$$

$$= x^{x-1} + \log_e x[x^x(1 + \log_e x)] \quad \left[\because \frac{d}{dx}(x^x) = x^x(1 + \log_e x), \quad \text{from Example (1).}\right]$$

$$= x^{x-1} + x^x \log_e x(1 + \log_e x)$$

$$\therefore \frac{dy}{dx} = y[x^{x-1} + x^x \log_e x(1 + \log_e x)] = x^{x^x}[x^{x-1} + x^x \log_e x(1 + \log_e x)]$$

Method II: If $y = x^{x^x}$, find dy/dx.
We have

$$y = (x)^{x^x} \tag{15}$$

Taking the natural logarithm of both sides, we get

$$\log_e y = x^x \log_e x \tag{16}$$

[7] Recall that $a^b = e^{\log_e a^b}$, $(a > 0)$.
$= e^{b \log_e a}$

Taking logarithms again, we get

$$\log_e(\log_e y) = x\log_e x + \log_e(\log_e x)$$

Differentiating both sides w.r.t. x, we get

$$\frac{1}{\log_e y}\frac{1}{y}\frac{dy}{dx} = \left(x\frac{1}{x} + \log_e x(1)\right) + \frac{1}{\log_e x}\frac{1}{x}$$

$$\therefore \frac{dy}{dx} = y\log_e y\left[1 + \log_e x + \frac{1}{x\log_e x}\right]$$

$$= x^{x^x}\cdot x^x\log_e x\left[1 + \log_e x + \frac{1}{x\log_e x}\right]$$

[using (2)] **Ans**.

Example (13): If $y = (x^x)^x$, then find dy/dx.

We have, $y = (x^x)^x = x^{x\cdot x} = x^{x^2}$

Taking *natural logarithm* of both sides, we get

$$\log_e y = x^2\log_e x$$

Differentiating w.r.t. x, we get

$$\frac{1}{y}\frac{dy}{dx} = x^2\frac{1}{x} + (\log_e x)(2x)$$

$$= x + 2x\log_e x$$

$$\therefore \frac{dy}{dx} = y[x + 2x\log_e x]$$

$$= x^{x^2}\cdot x[1 + 2\log_e x]$$

$$= x^{x^{2+1}}[1 + 2\log_e x]$$ **Ans**.

Example (14): If $y = (\log_e x)^x$, find dy/dx.

We have $y = (\log_e x)^x$

Taking *natural logarithm* of both the sides, we get

$$\log_e y = x\log_e(\log_e x)$$

Differentiating both sides w.r.t. x, we get

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} = x\frac{\mathrm{d}}{\mathrm{d}x}[\log_e(\log_e x)] + \log_e(\log_e x)\frac{\mathrm{d}}{\mathrm{d}x}(x)$$

$$= x\frac{1}{\log_e x}\frac{1}{x} + \log_e(\log_e x)\cdot 1$$

$$= \frac{1}{\log_e x} + \log_e(\log_e x)$$

$$\therefore\ \frac{\mathrm{d}y}{\mathrm{d}x} = y\left[\frac{1}{\log_e x} + \log_e(\log_e x)\right]$$

$$= (\log_e x)^x\left[\frac{1}{\log_e x} + \log_e(\log_e x)\right] \qquad \textbf{Ans.}$$

Example (15): If $y=(\cos x)^{\sin x}$, find $\mathrm{d}y/\mathrm{d}x$.

We have $y=(\cos x)^{\sin x}$

Taking *natural logarithm* of both sides, we get

$$\log_e y = \sin x\cdot\log_e \cos x$$

Differentiating both sides w.r.t. x, we get

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} = \sin x\left[\frac{1}{\cos x}(-\sin x)\right] + (\log_e \cos x)(\cos x)$$

$$= -\frac{\sin^2 x}{\cos x} + \cos x\cdot\log_e \cos x$$

$$\therefore\ \frac{\mathrm{d}y}{\mathrm{d}x} = y\left[\cos x\cdot\log_e \cos x - \frac{\sin^2 x}{\cos x}\right]$$

$$= (\cos x)^{\sin x}\left[\cos x\cdot\log_e \cos x - \frac{\sin^2 x}{\cos x}\right] \qquad \textbf{Ans.}$$

Example (16): If $y = (\tan x)^{\log_e x}$, find $\mathrm{d}y/\mathrm{d}x$.

We have, $y = (\tan x)^{\log_e x}$

Taking *natural logarithm* of each side, we get

$$\log_e y = \log_e x\cdot\log_e(\tan x)$$

Differentiating both sides w.r.t. x, we get

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} = \log_e x \frac{1}{\tan x}\sec^2 x + \log_e(\tan x)\frac{1}{x}$$

$$= \log_e x \frac{\cos x}{\sin x}\frac{1}{\cos^2 x} + \frac{\log_e(\tan x)}{x}$$

$$\therefore \frac{\mathrm{d}y}{\mathrm{d}x} = y\left[\frac{\log_e x}{\sin x \cos x} + \frac{\log_e(\tan x)}{x}\right]$$

$$= (\tan x)^{\log_e x}\left[\frac{\log_e x}{\sin x \cos x} + \frac{\log_e(\tan x)}{x}\right] \quad \textbf{Ans.}$$

Example (17): If $y = (\sin x)^{\tan x}$, find $\mathrm{d}y/\mathrm{d}x$.

We have, $y = (\sin x)^{\tan x}$

Taking the *natural logarithm* of each side, we get

$$\log_e y = \tan x \cdot \log_e \sin x$$

Differentiating both sides w.r.t. x, we have

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} = \tan x \frac{1}{\sin x}\cos x + \log_e \sin x \sec^2 x$$

$$= 1 + \sec^2 x \log_e \sin x$$

$$\therefore \frac{\mathrm{d}y}{\mathrm{d}x} = y[1 + \sec^2 x \log_e \sin x] \quad \textbf{Ans.}$$

Example (18): If $y = (\cos x)^{\log_e x}$, find $\mathrm{d}y/\mathrm{d}x$.

We have $y = (\cos x)^{\log_e x}$

Taking the *natural logarithm* of each side, we get

$$\log_e y = \log_e x \cdot \log_e(\cos x)$$

Differentiating both sides w.r.t. x, we get

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} = \log_e x \frac{1}{\cos x}(-\sin x) + \log_e(\cos x)\left(\frac{1}{x}\right)$$

$$= \log_e x(-\tan x) + \frac{1}{x}\log_e(\cos x)$$

$$= \left[\frac{1}{x}\log_e(\cos x) - \log_e x \tan x\right]$$

$$\therefore \frac{\mathrm{d}y}{\mathrm{d}x} = y\left[\frac{1}{x}\log_e(\cos x) - \log_e x \tan x\right]$$

$$= (\cos x)^{\log_e x}\left[\frac{1}{x}\log_e(\cos x) - \log_e x \tan x\right] \quad \textbf{Ans.}$$

Example (19): $x^y \cdot y^x = 1$, then prove that $dy/dx = \frac{-y(y+x\log_e y)}{x(x+y\log_e x)}$.

Solution: Given

$$x^y \cdot y^x = 1$$

Taking *natural logarithm* of both sides, we get

$$\log_e x^y + \log_e y^x = \log 1$$

or $\log_e x^y + \log_e y^x = 0$ [$\because$ $\log 1 = 0$]

$$\therefore\ y \log_e x + x \log y = 0$$

Differentiating w.r.t. x, we get

$$y\frac{1}{x} + \log_e x \frac{dy}{dx} + x\frac{1}{y}\frac{dy}{dx} + \log_e y \cdot 1 = 0$$

$$\Rightarrow \left(\log_e x + \frac{x}{y}\right)\frac{dy}{dx} = -\left(\log_e y + \frac{y}{x}\right)$$

$$\Rightarrow \left(\frac{y\log_e x + x}{y}\right)\frac{dy}{dx} = -\left(\frac{y + x\log_e y}{x}\right)$$

$$\Rightarrow \frac{dy}{dx} = -\frac{(y + x\log_e y)/x}{(x + y\log_e x)/y}$$

$$= \frac{y(y\log_e x + x)}{x(x\log_e y + y)} \quad \textbf{Ans}.$$

Example (20): $x^y + y^x = a^b$, find dy/dx.

Solution: Given

$$x^y + y^x = a^b$$

Putting $u = x^y$ and $v = y^x$, we get

$$u + v = a^b$$

$$\therefore\ \frac{du}{dx} + \frac{dv}{dx} = 0 \qquad (17)$$

Now, consider $u = x^y$

Taking *natural logarithm* of both sides, we get

$$\log_e u = y \log_e x$$

Differentiating both sides w.r.t x, we get

$$\frac{1}{u}\frac{\mathrm{d}u}{\mathrm{d}x} = y\frac{\mathrm{d}}{\mathrm{d}x}(\log_e x) + \log_e x\frac{\mathrm{d}(y)}{\mathrm{d}x}$$

$$\frac{\mathrm{d}u}{\mathrm{d}x} = u\left(\frac{y}{x} + \log_e x\frac{\mathrm{d}y}{\mathrm{d}x}\right) = x^y\left(\frac{y}{x} + \log_e x\frac{\mathrm{d}y}{\mathrm{d}x}\right) \tag{18}$$

Now, consider $u = y^x$

Taking *natural logarithm* of both sides, we get

$$\log_e v = x\log_e y$$

Differentiating both sides w.r.t x, we get

$$\frac{1}{v}\frac{\mathrm{d}v}{\mathrm{d}x} = x\frac{\mathrm{d}}{\mathrm{d}x}(\log_e y) + \log_e y\frac{\mathrm{d}(x)}{\mathrm{d}x}$$

$$\frac{\mathrm{d}v}{\mathrm{d}x} = v\left(\frac{x}{y}\frac{\mathrm{d}y}{\mathrm{d}x} + \log_e y\right) = y^x\left(\frac{x}{y}\frac{\mathrm{d}y}{\mathrm{d}x} + \log_e y\right) \tag{19}$$

Using (18) and (19) in (17), we get

$$x^y\left(\frac{y}{x} + \log_e x\frac{\mathrm{d}y}{\mathrm{d}x}\right) + y^x\left(\frac{x}{y}\frac{\mathrm{d}y}{\mathrm{d}x} + \log_e y\right) = 0$$

$$\therefore\ \frac{\mathrm{d}y}{\mathrm{d}x}\left[x^y\log_e x + y^x\frac{x}{y}\right] = -\left[x^y\frac{y}{x} + y^x\log_e y\right]$$

$$\therefore\ \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{x^y(y/x) + y^x\log_e y}{x^y\log_e x + y^x(x/y)}\quad \textbf{Ans}.$$

15b Parametric Functions and Their Differentiation

15b.1 INTRODUCTION

Let a body be moving in the x, y-plane, perhaps in the direction of the arrows on the curve shown in Figure 15b.1. Suppose the Cartesian coordinates (x, y) of its position at any time t are given by the pair of equations

$$x = f(t) \text{ and } y = g(t) \tag{1}$$

Then, *for every number t in the domain common to f and g*, the body is at a point $(f(t), g(t))$ and these points trace a *plane curve c* traveled by the body. Equation (1) is called a *parametric equation* of c and the variable t is called a *parameter*.[1]

The curve c is also called the graph of the parametric equation (1).

15b.1.1 Definition

If a functional relationship between *two variables* is specified so that each variable is determined separately *as a function of one and the same auxiliary variable*, we say that this functional relationship is represented parametrically and call the auxiliary variable a parameter.

It may be noted that *the curve c represented by parametric equations need not be the graph of a function.* If the parameter t is eliminated from the pair of Equations (1), we obtain one equation of the curve in x and y, of the form,

$$\phi(x, y) = 0 \tag{2}$$

called a *Cartesian equation* of the curve c.

If a plane curve is defined by *an equation of the form* $y = f(x)$, where f is continuous, then its *parametric equations may be obtained by letting* $x = t$ *and* $y = f(t)$, where t is in the *domain* of f. Other substitutions for x may also give parametric equations of the curve provided x *assumes every value in the domain of f.*

15b-Derivatives of functions in parametric forms. Derivative of one function w.r.t. another function and the method of substitution

[1] The equation $x = f(t)$ describes the motion on the x-axis of the x-projection, which always stays right under the main body. Similarly, $y = g(t)$ gives the motion on the y-axis of the y-projection, which stays opposite the main body.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

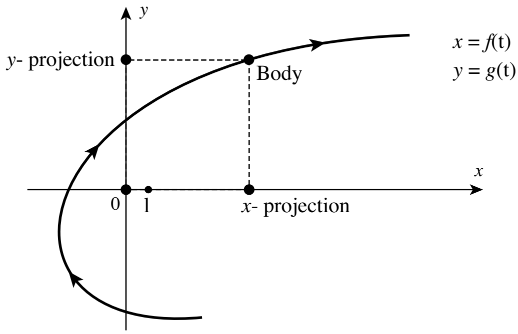


FIGURE 15b.1 A curve indicating motion (of a body in a plane) represented by $x = f(t)$ and $y = g(t)$.

Example (1): A parabola having the equation

$$y = x^2 \tag{3}$$

is also defined by parametric equations

$$x = t \text{ and } y = t^2 \tag{4a}$$

as well as by the parametric equations

$$x = t^3 \text{ and } y = t^6 \tag{4b}$$

Note (1): The above observation suggests that we may *write any number of parametric equations* for the parabola at (3) above. However, the parametric equations

$$x = t^2 \text{ and } y = t^4 \tag{5}$$

define only *the right-hand side of the parabola* where $x \geq 0$.[(2)]

Let us try to find the parametric equations for the parabola $y^2 = 4ax$. From this equation, we get $y = 2\sqrt{ax}$. Now to get the parametric equations in a simple form, we try to get rid of the "square root" on the right-hand side. We put $x = at^2$, which gives $y = 2at$. Thus, we get the parametric equations as $x = at^2$ and $y = 2at$.

Note (2): Every relation in x and y cannot be expressed in the form of parametric equations.

15b.1.2 What is a Parameter?

The *term parameter* is one that is widely used in mathematics and in engineering and it is *not easy to give a definition that covers all its applications.* If a moving point is tracing a curve, time "t" can be taken as a parameter. In writing the equations of certain planar curves, parameter t represents

[(2)] Note that, by eliminating parameter t from equations (5), we get the Cartesian equation $y = x^2$ that is defined for all real values of x and hence its graph consists of the parabola represented by (3), whereas the parametric equations at (5) define only the right-hand side of the parabola. Also, note that this situation does not occur with the parametric equations at (4a) and (4). (Why?)

the *radian measure* of the angle measured from the positive side of the x-axis to the line segment from the origin to the point (x, y) on the graph. Another useful parameter is arc length.

However, *a parameter need not have any physical significance*. Any quantity that is *algebraically convenient* can be used as a parameter. In such cases, the purpose of using parametric representation is *usually to simplify the algebra.* We might think of parameter t as an *independent variable that controls the values of x and y.*

A parameter can be described as a quantity (appearing in a formula) that can take different values, and these values indicate different individual members of a family or different states of a physical system. For example, a family of curves can be represented by the equation $F(x, y, \alpha) = 0$, where α *is a parameter defining the different curves of the family*. In fact, there are curves that can be *conveniently represented only by parametric equations*. Thus, *parametric functions are unavoidable in coordinate geometry and in calculus.* (see the equation of a cycloid in Example (5) below.) Let us now consider some examples.

Example (2): Circle

A circle with center at the coordinate origin and with radius "a" is defined by the relation:

$$x^2 + y^2 = a^2 \tag{6a}$$

This relation is satisfied by every x and y, given by,

$$\left.\begin{aligned} x &= a\cos t \\ y &= a\sin t \end{aligned}\right\} (0 \le t \le 2\pi) \tag{6b}$$

for any value of t. The pair of equations (6b) are called *parametric equations of the curve* (6a) and t is called a parameter.

Example (3): Find a Cartesian equation of the graph given by the parametric equations, $x = 2\cos\, t$ and $y = 2\sin\, t$, $(0 \le t \le 2\pi)$ and sketch the graph.

Solution: To eliminate t from the two parametric equations, we square both sides of each equation and add, which gives

$$x^2 + y^2 = 4\cos^2 t + 4\sin^2 t = 4$$

The graph of the equation $x^2 + y^2 = 4$ is a circle with center at the origin and radius 2. By letting t take on all numbers in the closed interval $[0, 2\pi]$, we obtain the entire circle *starting at the point* (2, 0) and moving (along the circle) in the counterclockwise direction, as indicated in Figure 15b.2. Note that, in this example, the parameter t *represents the radian measure of the angle* measured from the positive side of the x-axis to the line segment from the origin to the point $P(x, y)$ on the circle, as indicated in Figure 15b.3.

Example (4): Ellipse

The equation of an ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{7a}$$

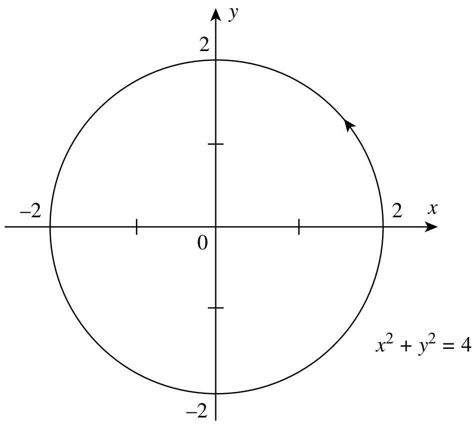


FIGURE 15b.2

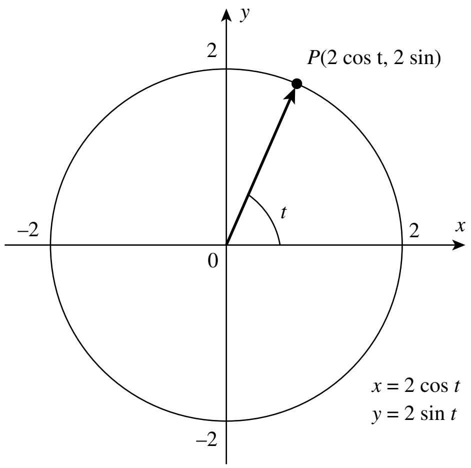


FIGURE 15b.3

is represented by the parametric equations,

$$\left.\begin{array}{l} x = a\cos t \\ y = b\sin t \end{array}\right\} \quad (0 \leq t \leq 2\pi) \tag{7b}$$

The equation of the curve (7a) is obtained by eliminating parameter *t* from the pair of equations in (7b).

Example (5): Cycloid

The cycloid is a curve described by a point *M* lying on the circumference of a circle if the circle rolls upon a straight line without sliding. The equations

$$\left.\begin{aligned} x &= a(t - \sin t) \\ y &= a(1 - \cos t) \end{aligned}\right\} \quad 0 \le t \le 2\pi \tag{8a}$$

are the parametric equations of the cycloid, where "a" is the radius of the circle and t is a parameter. As t varies between 0 and 2π, *the point M (on the circle) describes one arc of the cycloid.* By eliminating parameter t from equation 6(A), we get x as a function of y directly, given by

$$x = 2\pi\, a - \left[a \cos^{-1}\left[\frac{a-y}{a}\right] - \sqrt{(ay - y^2)}\right], \quad \pi a \le x \le 2\pi a \tag{8b}$$

It will be noted that *this is the simplest form in which the relation between x and y can be expressed* and that *y cannot be expressed in terms of elementary functions of x.*

Remark: Equation (8b) of the cycloid clearly shows that, *in certain cases, it is more convenient to use parametric equations for studying functions and curves rather than the direct relationship of y and x.*[3]

15b.2 THE DERIVATIVE OF A FUNCTION REPRESENTED PARAMETRICALLY

We now prove the theorem that helps us find the *derivatives of functions represented parametrically.*

Theorem: If $x = f(t)$ and $y = g(t)$ are differentiable functions of t, then[4]

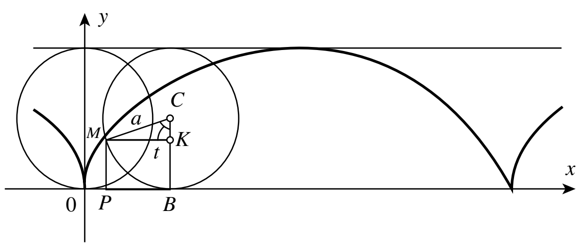
$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}, \quad \text{if} \quad \frac{dx}{dt} \neq 0^{(4)}$$

Proof: As t changes to $t + \delta t$, let x change to $x + \delta x$ and y change to $y + \delta y$.

$$\therefore \text{As } \delta t \to 0,\ \delta x \to 0, \text{ and } \delta y \to 0 \tag{9}$$

(It means that at any stage $\delta t \neq 0$, $\delta x \neq 0$, and $\delta y \neq 0$.)

[3] *Differential and Integral Calculus* (Second Edition) by N. Piskunov (vol. I, pp. 100–101), Mir Publishers, 1974.



Cycloid

[4] It is given that $x = f(t)$ and $y = g(t)$ are differentiable functions of t, and we assume that $x = f(t)$ has an inverse, $t = h(x)$, which also has a derivative. Accordingly, $y = g(t) = g(h(x))$ is also a differentiable function because it is a composite of differentiable functions. Now, the derivatives dy/dx, dy/dt, and dx/dt, all exist and, with $dx/dt \neq 0$, we may solve the equation $dy/dt = (dy/dx)(dx/dt)$ for dy/dx, and obtain $dy/dx = (dy/dt)/(dx/dt)$.

Now, consider the algebraic identity

$$\frac{\delta y}{\delta x} = \frac{\delta y/\delta t}{\delta x/\delta t} \quad (\because \delta t \neq 0 \text{ and } \delta x \neq 0)$$

$$\therefore \lim_{\delta t \to 0} \frac{\delta y}{\delta x} = \lim_{\delta t \to 0} \frac{\delta y/\delta t}{\delta x/\delta t} = \frac{\lim\limits_{\delta t \to 0} (\delta y/\delta t)}{\lim\limits_{\delta t \to 0} (\delta x/\delta t)} \tag{10}$$

Now, since $x = f(t)$ and $y = g(t)$ are *differentiable functions of* t,

$$\therefore \lim_{\delta t \to 0} \frac{\delta y}{\delta t} = \frac{dy}{dt} \quad \text{and} \quad \lim_{\delta t \to 0} \frac{\delta x}{\delta t} = \frac{dx}{dt} \tag{11}$$

Now, if $dx/dt \neq 0$, limit on RHS of (10) exists.

$\therefore$ limit on LHS. of (10) also exists (i.e., $\lim\limits_{\delta t \to} \delta y/\delta x$ exists)

But as $\delta t \to 0$, $\delta y \to 0$ and $\delta x \to 0$ (see (A) above)

$$\therefore \lim_{\delta t \to 0} \frac{\delta y}{\delta x} = \lim_{\delta t \to 0} \frac{\delta y}{\delta x} = \frac{dy}{dx} \tag{12}$$

Using (11) and (12) in (10), we get

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}, \quad \text{provided} \left(\frac{dx}{dt} \neq 0 \right) \tag{13}$$

Remark: The derived formula, $dy/dx = (dy/dt)/(dx/dt)$, $dx/dt \neq 0$, permits us to calculate the derivative dy/dx *as a function of t*, from the derivatives, dy/dt and dx/dt.

An important fact is that if a function is defined parametrically, as x = f(t), y = g(t), then we can find dy/fx *without having to find y as a function of x.*

Now, let us consider some examples.

Example (6): If $x = 2t + 3$, $y = t^2 - 1$, find the value of dy/dx at $t = 6$.

Also, find dy/dx as a function of x.

Solution: The result (13) gives dy/dx as a function of t:

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{2t}{2} = t$$

When $t = 6$, $dy/dx = 6$.

Note (3): It may be noted that dy/dx is expressed in terms of the parameter only without directly involving the main variables x and y.

Example (7): If $x = at^2$, $y = 2at$, find dy/dx.

We have,

$$\frac{dx}{dt} = 2at, \quad \frac{dy}{dt} = 2a$$

$$\therefore \frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{2a}{2at} = \frac{1}{t} \qquad \textbf{Ans.}$$

Example (8): Find dy/dx, if $x = \sin(\log_e t)$, $y = \cos(\log_e t)$

We have

$$\begin{aligned}\frac{dx}{dt} &= \cos(\log_e t)\frac{d}{dt}(\log_e t)\\ &= \cos(\log_e t)\frac{1}{t}\end{aligned}$$

and

$$\begin{aligned}\frac{dy}{dt} &= -\sin(\log_e t)\frac{d}{dt}(\log_e t)\\ &= -\sin(\log_e t)\frac{1}{t}\end{aligned}$$

Now,

$$\begin{aligned}\frac{dy}{dx} = \frac{dy/dt}{dx/dt} &= \frac{-\sin(\log_e t)(1/t)}{\cos(\log_e t)(1/t)}\\ &= -\tan(\log_e t) \qquad \textbf{Ans}.\end{aligned}$$

Example (9): Find the slope of the tangent to the following cycloid at an arbitrary point ($0 \le t \le 2\pi$):

$$\left.\begin{aligned}x &= a(t - \sin t)\\ y &= a(1 - \cos t)\end{aligned}\right\} \tag{14}$$

Solution: *The slope of the tangent at any point of the curve (1) is equal to the value of the derivative* dy/dx *for the value of parameter t at that point.*

(See Example (5), on computation of speed.)

We know that,

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}$$

Now,

$$\frac{dx}{dt} = a(1 - \cos t),\ \frac{dy}{dt} = a[0 - (-\sin t)] = a \sin t$$

$$\therefore\ \frac{dy}{dx} = \frac{a \sin t}{a(1 - \cos t)} = \frac{\sin t}{1 - \cos t} = \frac{2\sin(t/2)\cos(t/2)}{2\sin^2(t/2)} = \cot\frac{t}{2} = \tan\left(\frac{\pi}{2} - \frac{t}{2}\right)$$

Hence, the slope of the tangent to a cycloid *at every point is equal to* $\tan((\pi/2) - (t/2))$, *where t is the value of the parameter corresponding to that point.*

15b.3 LINE OF APPROACH FOR COMPUTING THE SPEED OF A MOVING PARTICLE

To compute the speed of a moving particle whose x, y-coordinates at any instant t are given by the parametric equations:

$$x = f(t), \; y = g(t)$$

where t is the *time parameter.*

Suppose, a body is *moving in the x–y plane*. Let the x-coordinate of the body's position at time t be some function $x = f(t)$, and the y-coordinate of its position be given by the function $y = g(t)$. At each instant t, the body is moving in the direction tangent to the curve of its motion.

Suppose, it were possible to command the motion of the body, with the instruction "*stop curving and keep going in the same direction* (as at the particular instant "t") and at the same speed (as at that instant)"; *then the body would go off along a tangent line to the curve* at that point (or time instant). Now, if the body were to go off on the tangent line *and keep the same speed as it had at the instant t, then in one unit of time it would travel a distance* $|dx/dt|$ in the *x-direction* and a distance $|dy/dt|$ in the *y-direction*. Hence, the actual distance traveled along the hypotenuse, in one unit of time, would then be $\sqrt{[(dx/dt)^2 + (dy/dt)^2]}$. This represents the *magnitude of the speed* along the *tangent line to the curve at the instant* "t" *under consideration* (see Figure 15b.4).

To illustrate the above, consider the discussion in the following example.

Example (10): Let the position of a body in the plane at time t be given by $x = t^3 - 3t$, $y = 2t^2 + 7t$. Then, compute its speed and the slope of the curve traced by the body at $t = 1$, $t = 2$.

Solution: At any instant t,

$$x\text{-component of velocity} = \frac{dx}{dt} = 3t^2 - 3$$

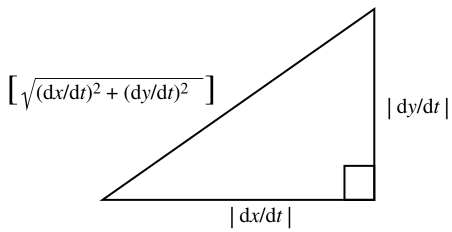


FIGURE 15b.4

and

$$y\text{-component of velocity} = \frac{dy}{dt} = 4t + 7$$

$$\therefore\ t = 1,\ dx/dt = 3(1)^2 - 3 = 0,\ \text{and } dy/dt = 4(1) + 7 = 11.$$

$$\text{At } t = 2,\ dx/dt = 3(2)^2 - 3 = 9,\ \text{and } dy/dt = 4(2) + 7 = 15.$$

We have,

$$\text{speed} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}$$

$$\therefore\ \text{speed (at } t = 1) = \sqrt{(0)^2 + (11)^2} = 11$$

and

$$\text{speed (at } t = 2) = \sqrt{(9)^2 + (15)^2} = \sqrt{306}$$

$$\left.\begin{array}{l}\text{Slope of curve}\\ (\text{at } t = 1)\end{array}\right\} = \frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{11}{0}^{(5)}$$

$$\left.\begin{array}{l}\text{Slope of curve}\\ (\text{at } t = 2)\end{array}\right\} = \frac{dy}{dx} = \frac{15}{9} = \frac{5}{3}$$

$$\text{Now, for } t = 1\left\{\begin{array}{l} x = (1)^3 - 3(1) = -2\\ y = 2(1)^2 + 7(1) = 9\end{array}\right.$$

$$\text{and for } t = 2\left\{\begin{array}{l} x = (2)^3 - 3(2) = 2\\ y = 2(2)^2 + 7(2) = 22\end{array}\right.$$

Therefore, at the point (−2, 9) on the curve, *the tangent line is vertical, whereas at the point (2, 22) the tangent line makes an angle of* $\tan^{-1}(5/3)$ *with the positive x-axis.*

15b.4 MEANING OF dy/dx WITH REFERENCE TO THE CARTESIAN FORM $y = f(x)$ AND PARAMETRIC FORMS $x = f(t)$, $y = g(t)$ OF THE FUNCTION

In the case of a function $y = f(x)$, the derivative dy/dx represents the instantaneous rate of change of y with respect to x. It also represents the slope of the curve $[y = f(x)]$ at an arbitrary point (x, y) on the curve.

However, *when a function is expressed by parametric equations* (such as $x = f(t)$, $y = g(t)$), then *the notation* dy/dx, *though it represents the slope of the curve, for each value of t, it does not represent the speed of the particle in any direction.* This is so because in the parametric representation of a function, the variations in x and y are controlled by an *independent variable (parameter) t.*

(5) It means that the tangent line is vertical to x-axis, for $t = 1$. Also, note that slope of the curve (represented by parametric equations) for any value of t need not represent the velocity (speed, in this case) of the moving object.

The curve represented by parametric equations indicates the actual path of motion of the particle, *which need not be a straight line*. Accordingly, dy/dx gives the rate of change of y w.r.t. x (which represents the slope of the curve for different values of t, *but it does not represent the speed of the particle, in any direction*).

In order to find the speed of the particle when equations are given in parametric form, we must use the following formulas:[6]

$$\text{Speed} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}$$

$$\text{Acceleration} = \sqrt{\left(\frac{d^2x}{dt^2}\right)^2 + \left(\frac{d^2y}{dt^2}\right)^2}$$

Note (4): Whenever the equations of a curve are given parametrically, *the Cartesian coordinates of a point (on the curve) can be obtained, corresponding to each value of the parameter "t", as needed in the following example.*

Example (11): Let us find the equation of the line normal to the curve given parametrically by

$$\left.\begin{aligned} x &= t^2 + 1 \\ y &= 2t^3 - 6t \end{aligned}\right\} \tag{15}$$

at the point where $t = 2$.

Solution: To find the equation of the line, *we need to know a point on the line and the slope of the line.*

Point: When $t = 2$, we get, $x = 5$ and $y = 4$, so the point on the curve is (5, 4).

Slope: The tangent line has slope,

$$\left.\frac{dy}{dx}\right|_{t=2} = \left.\frac{dy/dt}{dx/dt}\right|_{t=2} = \left.\frac{6t^2 - 6}{2t}\right|_{t=2} = \frac{18}{4} = \frac{9}{2}$$

Therefore, the normal line has slope $= -2/9$.

$\therefore$ Equation of the line, normal to the curve (15) at (5, 4) is given by

$$y - 4 = -\frac{2}{9}(x - 5) \text{ or } 9y + 2x = 46 \qquad \textbf{Ans.}$$

It is important to realize that an innocent looking problem (as in Example (11)) requires proper logical thinking for its solution. We will now show how the *derivative of one function with respect to another function can be obtained by treating the functions as parametric equations.*

[6] *Calculus with Analytic Geometry* by John B. Fraleigh (p. 86).

15b.5 DERIVATIVE OF ONE FUNCTION WITH RESPECT TO THE OTHER

Let $u = f(x)$ and $v = g(x)$ be two *differentiable functions* of x. Then, we can easily compute the derivatives $\mathrm{d}u/\mathrm{d}v$, and treating x as a parameter.

Example (12): Differentiate $\log_e(1+x^2)$ w.r.t. $\tan^{-1}x$.

Solution: Let $u = \log_e(1+x^2)$ and $v = \tan^{-1}x$

$\therefore$ We have to find $\mathrm{d}u/\mathrm{d}v$.
Now,

$$\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\left[\log_e(1+x^2)\right] = \frac{1}{1+x^2}2x = \frac{2x}{1+x^2} \quad \text{and} \quad \frac{\mathrm{d}v}{\mathrm{d}x} = \frac{1}{1+x^2}$$

$$\therefore \frac{\mathrm{d}u}{\mathrm{d}v} = \frac{(\mathrm{d}u/\mathrm{d}x)}{(\mathrm{d}v/\mathrm{d}x)} = \left(\frac{2x/(1+x^2)}{1/(1+x^2)}\right) = \frac{2x}{1+x^2}\frac{1+x^2}{1}$$
$$= 2x \qquad \textbf{Ans.}$$

Example (13): Differentiate $x \cdot \mathrm{e}^x$ w.r.t. $x \cdot \log x$.

Solution: Let $u = x \cdot \mathrm{e}^x$ and $v = x \cdot \log_e x$.

We have to find $\mathrm{d}u/\mathrm{d}v$.
Now,

$$\frac{\mathrm{d}u}{\mathrm{d}x} = x(e^x) + e^x(1) = e^x(x+1)$$

$$\text{and} \quad \frac{\mathrm{d}v}{\mathrm{d}x} = x\left(\frac{1}{x}\right) + \log_e x \cdot (1) = 1 + \log_e x$$

$$\therefore \frac{\mathrm{d}u}{\mathrm{d}v} = \frac{(\mathrm{d}u/dx)}{(\mathrm{d}v/\mathrm{d}x)} = \frac{\mathrm{e}^x(x+1)}{1+\log_e x} \qquad \textbf{Ans.}$$

Example (14): Differentiate $\mathrm{e}^x \cdot \cos x$ w.r.t. $\mathrm{e}^{-x} \cdot \sin x$.

Solution: Let $u = \mathrm{e}^x \cdot \cos x$ and $v = \mathrm{e}^{-x} \cdot \sin x$.

Then, we have to find $\mathrm{d}u/\mathrm{d}v$.
Now,

$$\frac{\mathrm{d}u}{\mathrm{d}x} = \mathrm{e}^x(-\sin x) + \cos x(\mathrm{e}^x) = \mathrm{e}^x(\cos x - \sin x)$$

and

$$\frac{\mathrm{d}v}{\mathrm{d}x} = \mathrm{e}^{-x}(\cos x) + \sin x \cdot (-\mathrm{e}^{-x})$$
$$= \mathrm{e}^{-x}(\cos x - \sin x)$$

$$\therefore \frac{\mathrm{d}u}{\mathrm{d}v} = \frac{(\mathrm{d}u/\mathrm{d}x)}{(\mathrm{d}v/\mathrm{d}x)} = \frac{\mathrm{e}^x(\cos x - \sin x)}{\mathrm{e}^{-x}(\cos x - \sin x)} = \mathrm{e}^{2x} \qquad \textbf{Ans.}$$

Example (15): Differentiate 7^x w.r.t. $\log_x 7$.

Solution: Let $u = 7^x$ and $v = \log_x 7$.
Then, we have to find $\mathrm{d}u/\mathrm{d}v$.
Now,

$$\frac{\mathrm{d}u}{\mathrm{d}x} = 7^x \log_e 7 \left[\text{If } t = a^x, \text{ then } \frac{\mathrm{d}t}{\mathrm{d}x} = a^x \log_e a\right]$$

Furthermore, $v = \log_x 7$ means $x^v = 7$ (by definition of logarithm).(7)

$$\therefore\ v \log_e x = \log_e 7$$

Differentiating both sides w.r.t. x, we get

$$v\frac{1}{x} + \log_e x \frac{\mathrm{d}v}{\mathrm{d}x} = 0$$

$$\therefore \frac{\mathrm{d}v}{\mathrm{d}x} = -\frac{v}{x}\frac{1}{\log_e x} = \frac{-\log_x 7}{x \log_e x}$$

$$= \frac{-1}{x \log_e x \log_7 x} \quad \left[\because \log_b a = \frac{1}{\log_a b}\right]$$

Now,

$$\frac{\mathrm{d}u}{\mathrm{d}v} = \frac{(\mathrm{d}u/\mathrm{d}x)}{(\mathrm{d}v/\mathrm{d}x)} = 7^x \log_e 7 \cdot (-x \log_e x \log_7 x)$$

$$= -x \cdot 7x \log_e 7 \log_e x \log_7 \mathrm{e} \log_e x \quad [\because \log_7 x = \log_e x \log_7 \mathrm{e}]$$

$$= -x7^x (\log_e x)^2 (\text{Since, } \log_e 7 \log_7 \mathrm{e} = 1) \qquad \textbf{Ans.}$$

15b.5.1 Method of Substitution (Usefulness of Trigonometric Identities)

In the process of differentiating one function with respect to another function, it is at times more convenient to use the *method of substitution* employing *trigonometric identities*, wherever applicable. This results in a change of parameter.

Example (16): Differentiate e^{x^2} w.r.t. x^2.

Let $u = \mathrm{e}^{x^2}$ and $v = x^2$.
We have to find $\mathrm{d}u/\mathrm{d}v$.

Solution: In u and v, we substitute $x^2 = t$, thus getting $u = \mathrm{e}^t$ and $v = t$.

(7) In calculus, we always express logarithm to the base "e". The reason for this choice is the simplicity of the relation $\mathrm{d}(\log_{\mathrm{e}} x)/\mathrm{d}x = 1/x$.

(Here, the parameter x changes to parameter t.) Now,

$$\frac{du}{dt} = e^t \quad \text{and} \quad \frac{dv}{dt} = 1 \quad \therefore \frac{du}{dv} = \frac{(du/dt)}{(dv/dt)} = \frac{e^t}{1} = e^t = e^{x^2} \qquad \textbf{Ans.}$$

Example (17): Differentiate $\log_e(1+x^2)$ w.r.t. $\sqrt{1+x^2}$.

Let $u = \log_e(1+x^2)$ and $v = \sqrt{1+x^2}$.(8)

We have to find du/dv.

Solution: If we substitute $1+x^2 = t$, we get

$$u = \log_e t \quad \text{and } v = \sqrt{t} = (t)^{1/2}$$

(Here, parameter x changes to parameter t.)
Now,

$$\frac{du}{dt} = \frac{1}{t} \quad \text{and} \quad \frac{dv}{dt} = \frac{1}{2}(t)^{-1/2} = \frac{1}{2\sqrt{t}}$$

$$\therefore \quad \frac{du}{dv} = \frac{(du/dt)}{(dv/dt)} = \frac{1}{t} \div \frac{1}{2\sqrt{t}} = \frac{1}{t} \times \frac{2\sqrt{t}}{1}$$

$$= \frac{2}{\sqrt{t}} = \frac{2}{\sqrt{1+x^2}} \qquad \textbf{Ans.}$$

Example (18): Differentiate $\tan^{-1}((3x-x^3)/(1-3x))$ w.r.t. $\tan^{-1}x$.

Let $u = \tan^{-1}((3x-x^3)/(1-3x))$ and $v = \tan^{-1}x$.(9)
We have to find du/dv.

Solution:
Put $x = \tan t$.

$$\therefore u = \tan^{-1}\left[\frac{3\tan t - \tan^3 t}{1 - 3\tan t}\right] = \tan^{-1}(\tan 3t) = 3t$$

$$\text{and } v = \tan^{-1}(\tan t) = t$$

$$\therefore \frac{du}{dv} = 3 \quad \text{and} \quad \frac{dv}{dt} = 1$$

$$\therefore \frac{du}{dv} = \frac{(du/dt)}{(dv/dt)} = \frac{3}{1} = 3 \qquad \textbf{Ans.}$$

(9) The expression $(3x-x^3)/(1-3x)$ suggests to recall the trigonometric identity $\tan 3t = ((3\tan t - \tan^3 t)/(1-3\tan t))$ and the corresponding substitution (i.e., $x = \tan t$). If we do this, parameter x will change to parameter t. Even otherwise, from the equation $v = \tan^{-1}x$, we get $x = \tan v$. By substituting $x = \tan v$, we get $u = \tan^{-1}(\tan 3v) = 3v$. From this relation, we get $du/dv = 3$, which is the desired answer.

Example (19):
If $u = \cos^{-1}((1 - x^2)/(1 + x^2))$ and $v = \tan^{-1}(2x/(1 - x^2))$, then find du/dv.

Solution: Put $x = \tan t$.

$$\therefore u = \cos^{-1}\left(\frac{1 - \tan^2 t}{1 + \tan^2 t}\right) = \cos^{-1}(\cos 2t) = 2t$$

and

$$v = \tan^{-1}\left(\frac{2\tan t}{1 - \tan^2 t}\right) = \tan^{-1}(\tan 2t) = 2t$$

$$\therefore \frac{du}{dt} = 2 \quad \text{and} \quad \frac{dv}{dt} = 2$$

Now,

$$\frac{du}{dv} = \frac{(du/dt)}{(dv/dt)} = \frac{2}{2} = 1$$

$$\therefore \frac{du}{dv} = 1 \qquad \textbf{Ans.}$$

Exercise

Q1. Differentiate e^x w.r.t. $\sqrt{x}$.

Ans. $2\sqrt{x}\,e^x$

Q2. Differentiate $\sin^{-1}((1 - x)/(1 + x))$ w.r.t. $\sqrt{x}$.

Ans. $\dfrac{-2}{1 + x}$

Q3. Differentiate $\tan^{-1}\left(\left(\sqrt{1 - x^2} - 1\right)/x\right)$ w.r.t. $\tan^{-1}x$.

Ans. $\dfrac{1}{2}$

Q4. Differentiate $\sin^{-1}x$ w.r.t. $\cos^{-1}\sqrt{1 - x^2}$.

Ans. 1

Q5. Differentiate $\tan^{-1}(2x/(1 - x^2))$ w.r.t. $\sin^{-1}(2x/(1 + x^2))$.

Ans. 1

16 Differentials "dy" and "dx": Meanings and Applications

16.1 INTRODUCTION

We now introduce the concept of the *differential*, which enables us to approximate *changes in function values*, where the function is differentiable. Even though, the application of differentials for approximating the function values is not very important in the age of technology (since better tools are available), *differentials are important as a convenient notational device for the computation of antiderivatives*, as we will learn later in Part II of this book.

For a differentiable function $y = f(x)$, we have been using Leibnitz notation dy/dx to mean the derivative of y with respect to x. Although this notation has the appearance of a quotient, it is treated as a *single entity, since it is a symbol for the limit*

$$\lim_{\delta x \to 0} \frac{f(x+\delta x) - f(x)}{\delta x} = \lim_{\delta x \to 0} \frac{\delta y}{\delta x} = \frac{\mathrm{d}y}{\mathrm{d}x} = f'(x)$$

if this limit exists. It is now proposed to give separate meanings to the symbols dy and dx.[1]

The concept of the differential of a function is closely related to the derivative of the function. To understand this, refer to Figure 16.1. In this figure, an equation of a curve is $y = f(x)$. The line PT is tangent to the curve at $P(x, f(x))$, Q is the point $(x + \Delta x, f(x + \Delta x))$, and the directed distance $\overline{MQ}$ is

$$\Delta y = f(x + \Delta x) - f(x)$$

which represents *the actual change in the value of f*, when x is changed to $(x + \Delta x)$.

In the following figure, Δx and Δy are both positive; however they could be negative. For a *small value of* Δx, the slope of the secant line PQ and the slope of the tangent line at P are *approximately equal*; so that we can write,

$$\frac{\Delta y}{\Delta x} \approx f'(x)$$

$$\text{or } \Delta y \approx f'(x)\Delta x \tag{1}$$

The right-hand side of equation (1) is defined to be the *differential* of y.

We give the following definition.

16-The differentials dy, dx, and the derivative dy/dx as a ratio of differentials.

[1] The symbols dy and dx should be understood as individual symbols and not as product(s) of d and y or d and x, respectively.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

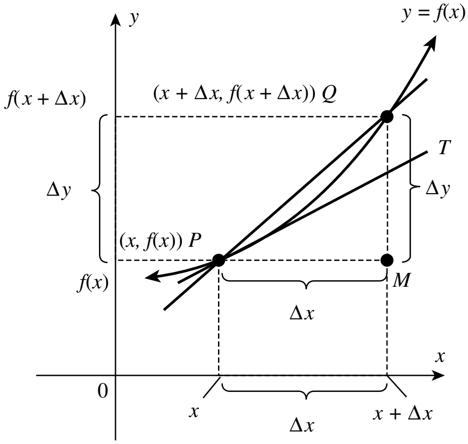


FIGURE 16.1

16.1.1 Definition: Differential of the Dependent Variable *y*

Let the function f be defined by the equation $y = f(x)$, then the *differential of y* is denoted by "dy" [or d$f(x)$], and is given by

$$\mathrm{d}y = f'(x)\Delta x \tag{2}$$

where x *is in the domain of* f' *and* Δx *is an arbitrary increment to* x.

Refer now to Figure 16.2, which is the same as Figure 16.1, except that the vertical distance segment MR is shown, *where the directed line* $\overline{MR} = \mathrm{d}y$. Observe that d$y$ represents the change in y along the tangent line to the graph of the equation $y = f(x)$ at the point $P(x, f(x))$, when x is changed by Δx.(2)

Note that, $\mathrm{d}y \neq \Delta y$, but for small values of Δx, dy *is very close to* Δy. Also, note that [with reference to equation (2)] since variable x can be any number in the domain of f' and Δx can be any number whatsoever, the differential dy or [d$f(x)$] is a function of two variables x and Δx.

We now wish to define the differential of the independent variable or "dx". To arrive at a suitable definition consistent with the definition of dy, we consider the identity function denoted by $f(x) = x$. For this function, $f'(x) = 1$ and $y = x$. Thus, from (2), we get $\mathrm{d}y = 1 \cdot \Delta x$, that is, if $y = x$, then, $\mathrm{d}y = \Delta x$. For the identity function, we would want that dx be equal to dy. This permits us to write $\Delta x = \mathrm{d}x$. This reasoning leads us to the following definition.

16.1.2 Definition: Differential of the Independent Variable

If the function f is defined by the equation $y = f(x)$, then the differential of x, denoted by dx is given by

$$\mathrm{d}x = \Delta x$$

where x is any number in the domain of f' and Δx is an arbitrary increment of x. The relation (2) can now be written as

(2) Note that, the definition of differential does not involve the notion of the derivative, though the derivative $f'(x)$ appears in the expression for the differential dy.

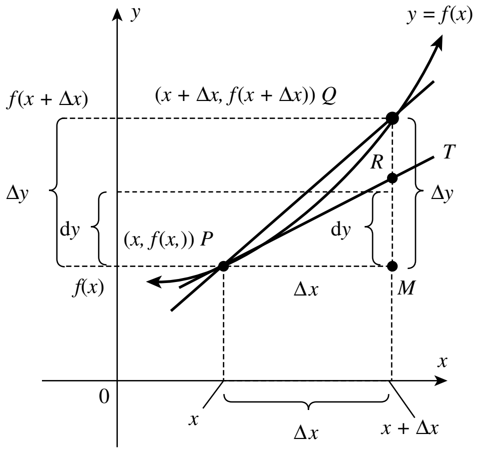


FIGURE 16.2

$$dy = f'(x)dx \tag{3}$$[3]

We treat (3) as *the definition of the differential of* y (i.e., the dependent variable). It tells us that, knowing the derivative of a function $y = f(x)$, we can readily find its differential. Further, by dividing both sides of (3) by dx, *we can, if we wish, interpret the derivative as a quotient of two differentials.*

$$\frac{dy}{dx} = f'(x), \text{ if } dx \neq 0 \tag{4}$$

This representation of the derivative, as the ratio of two differentials, is extremely important for mathematical analysis.

Remark (1): By defining the differential of a function, we have attached meanings to the symbols "dy" and "dx", and given a new meaning to dy/dx (the derivative of y with respect to x) as a ratio of dy to dx, retaining the meaning of the symbol dy/dx in question.

Remark (2): In equation (3), dx being arbitrary, can have any (finite) magnitude big or small. Also, since the magnitude of dy depends on two variables x and Δx, it can have any (finite) magnitude. Thus, in equation (3), dy and dx need not be small.

However, if we think of dx and dy as being small, then the equation (3) proves to be very useful since it gives the approximate changes in function values, where the function is differentiable.

Note (1): When we introduced the notation dy/dx (for the derivative of y with respect to x), we emphasized that dy and dx had not been given independent meaning. But, now we can also treat the symbol dy/dx as a ratio of two differentials. It is only when we think of differentials, that we can write $dy = f'(x)dx$. This permits us to write $dy/dx = 3x^2$ in the form $dy = 3x^2\,dx$ and similarly $dy/dx = \cos x$ in the form $dy = \cos x\,dx$, and so on.

[3] In this expression for dy, the derivative $f'(x)$ appears as the coefficient of dx, which is the differential of independent variable. Hence, the derivative of a function is called the differential coefficient.

16.1.3 Geometrical Interpretation of the Differential dy

Let $y = f(x)$ be a differentiable function of x and consider a fixed value of x, say x_0. Then, the differential of f at x_0 is given by

$$\mathrm{d}y = f'(x_0)\mathrm{d}x \tag{5}$$

Note that, in this case, dy is a *linear function* of *the single variable* dx, $f'(x_0)$ being a constant.[4]

Also, if an increment dx is given to x_0, the corresponding increment Δy in y, is given by

$$f(x_0 + \mathrm{d}x) - f(x_0) = \Delta y$$

or

$$f(x_0 + \mathrm{d}x) = f(x_0) + \Delta y$$

But, we know that, *for small values of* dx, Δy *is very close to* dy. *Hence, replacing* Δy *by* dy, we can write,

$$\begin{aligned} f(x_0 + \mathrm{d}x) &\approx f(x_0) + \mathrm{d}y \\ &\approx f(x_0) + f'(x_0)\mathrm{d}x \end{aligned}$$

Since, f is differentiable we may *drop the subscript* "0", and write the above equation as

$$f(x + \mathrm{d}x) \approx f(x) + f'(x)\mathrm{d}x \qquad (6)^{(5)}$$

The relation (6) gives us an *approximate* value of $f(x + \mathrm{d}x)$ in terms of fully known quantities [i.e., $f(x)$, $f'(x)$ and dx] where x is a number at which f is differentiable. *We shall make use of this equation to estimate values of functions that are difficult or impossible to obtain exactly.* The *approximation* given by this equation is *most useful* when $f(x)$ and $f'(x)$ are easy to compute. This will be clear from the solved examples which follow shortly.

Note that, *when we approximate* $f(x + \mathrm{d}x)$ by $f(x) + \mathrm{d}y$, *we are approximating the ordinate of the point Q on the curve by the ordinate of the point R* on the tangent line (see Figure 16.2).[6]

Note (2): One should not think that the increment Δy is always greater than dy. The situation becomes clear from the Figure 16.3a and 16.3b. It may be noted from Figure 16.3b that $\Delta y < \mathrm{d}y$.

Note (3): It should also be noted that, if $f'(x) = 0$ at a point x, *the differential is equal to zero*: $\mathrm{d}y = 0$. In this case, dy *is not compared with the increment* Δy *of the function*. Now, let us compute the differentials of some functions:

(4) We know that the differential $\mathrm{d}y = f'(x)\mathrm{d}x$ is a function of two variables, x and dx, which are independent of each other, since, in general, $f'(x)$ varies with x and the increment "dx" can be chosen arbitrarily.

(5) In a sufficiently small neighborhood of the point x, this replacement leads to small errors. A demerit of this formula is that although we know that the relative error ($\Delta y/y$) tends to zero as $\mathrm{d}x \to 0$, it does not provide any estimation of the error for a numerical value of dx. This is natural, because the error depends on the nature of f. Of course, we can measure the error for a given function and the given value of dx. Here, it may be mentioned that for all practical purposes the error is generally found to be negligible, as will be seen in the solved examples.

(6) We know that tangent lines and derivatives are closely related. Since a straight line is simpler than curves, and since the tangent line to a differentiable curve runs close to the curve near the point of tangency, the tangent line can provide a useful approximation to the function values near the point of tangency. Equation (6) tells us that to approximate the value $f(x + \mathrm{d}x)$, we add the tangent line increment $f'(x)\mathrm{d}x$ to the value of $f(x)$. Thus, an approximation of $f(x + \mathrm{d}x)$ given by (6) is called a linear approximation. Of course, we can measure the error, but for all practical purposes it is found to be negligible, as will be seen in the solved examples. However, we shall discuss the error and its estimation later in Chapter 22.

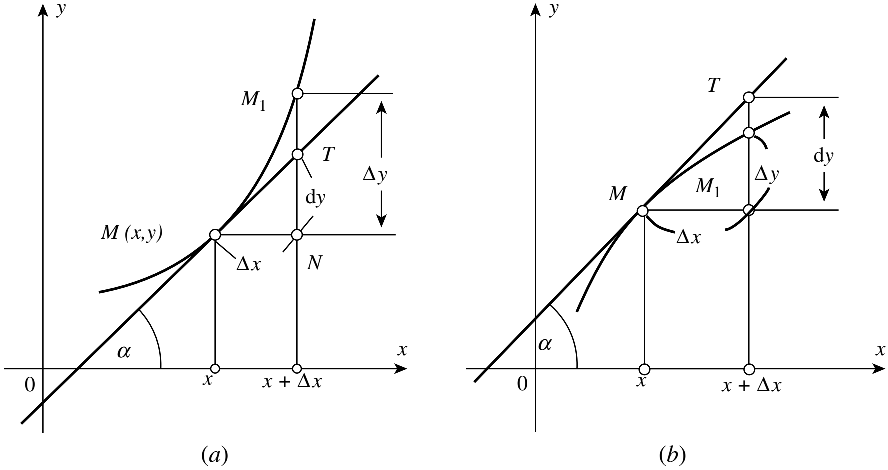


FIGURE 16.3 (a) $\Delta y > dy$, (b) $\Delta y < dy$.

Example (1):

(i) $y = x^3 + 5x^2 - 1$

$$\therefore \text{d}y = D_x[x^3 + 5x^2 - 1]\Delta x$$

(where D_x is the "derivative with respect to x").

$$\therefore \text{d}y = [3x^2 + 10x]\Delta x$$

Now, if $x = 1$ and $\Delta x = 0.02$, then $\text{d}y = [3(1)^2 + 10(1)(0.02)] = 0.26$.

(ii) $y = \sin x$

$$\begin{aligned} \therefore \text{d}y &= d(\sin x) \\ &= D_x[\sin x]\Delta x \\ &= \cos x \cdot \Delta x \end{aligned}$$

(iii) $y = \text{e}^{3x}$

$$\therefore \text{d}y = d(\text{e}^{3x}) = D_x[\text{e}^{3x}] \cdot \Delta x = 3\text{e}^{3x} \cdot \Delta x$$

(iv) Let $y = f(x) = \log_e(x^2 + 1)$

$$\begin{aligned} \therefore \text{d}y &= \text{d}[f(x)] \\ &= \text{d}[\log_e(x^2 + 1)] \\ &= D_x[\log_e(x^2 + 1)] \cdot \Delta x \\ &= \frac{1}{x^2 + 1} \cdot 2x \cdot \Delta x = \frac{2x \cdot \Delta x}{x^2 + 1} \end{aligned}$$

Example (2): Given $y = 4x^2 - 3x + 1$, find Δy, dy and $\Delta y - dy$ for

(a) any x and Δx
(b) $x = 2$, $\Delta x = 0.1$
(c) $x = 2$, $\Delta x = 0.01$
(d) $x = 2$, $\Delta x = 0.001$

Solution:

(a) We are given $y = f(x) = 4x^2 - 3x + 1$

$$\begin{aligned} \therefore \Delta y &= f(x + \Delta x) - f(x) \\ &= 4(x + \Delta x)^2 - 3(x + \Delta x) + 1 - (4x^2 - 3x + 1) \\ &= 4x^2 + 8x \cdot \Delta x + 4(\Delta x)^2 - 3x - 3 \cdot \Delta x + 1 - 4x^2 + 3x - 1 \\ &= (8x - 3) \cdot \Delta x + 4 \cdot (\Delta x)^2 \end{aligned}$$

From the definition of the differential in equation (3) above, we have

$$dy = f'(x)dx = (8x - 3)\Delta x \qquad (\text{since } dx = \Delta x)$$

$$\text{Thus, } \Delta y - dy = 4(\Delta x)^2$$

The results of the parts (b), (c), and (d) are given in the table below:

	x	Δx	Δy	dy	$\Delta y - dy$
(b)	2	0.1	1.34	1.3	0.04
(c)	2	0.01	0.1304	0.13	0.0004
(d)	2	0.001	0.013004	0.013	0.000004

From the above table, we note that *the closer* Δx *is to zero, the smaller is the difference between* Δy *and* dy. Furthermore, observe that for each value of Δx, the corresponding value of $\Delta y - dy$ is smaller than the value of Δx. More generally, dy *is an approximation of* Δy *when* Δx *is small, and the approximation is of better accuracy than the size of* Δx.

16.2 APPLYING DIFFERENTIALS TO APPROXIMATE CALCULATIONS

The application of the differential to approximate calculations is based on the replacement of the increment.

$$\therefore \Delta y = f(x_0 + dx) - f(x_0)$$

by the differential, dy [$= f'(x_0)dx$], since for small values of dx we have $\Delta y \approx dy$. Therefore, we write,

$$f(x_0 + dx) - f(x_0) \approx dy = f'(x_0)dx \tag{7}$$

Note that, even though the increment Δy may depend on dx in a complicated manner, the differential dy can be easily obtained by differentiation.

This *approximate equality* can be immediately used to solve the following problem.

Given the values of $f(x_0)$, $f'(x_0)$, *and* $\mathrm{d}x$, *it is required to compute an approximation to the value* $f(x_0 + \mathrm{d}x)$ *of the function.*

Relation (7) directly gives us the desired formula:

$$f(x_0 + \mathrm{d}x) \approx f(x_0) + f'(x_0)\mathrm{d}x$$

Let us consider some illustrative examples. (*For brevity, we shall write* x *in place of* x_0 *and denote* $\mathrm{d}x$ *by* h.)

(I) Consider the function $y = \sqrt{x}$

Its differential is $\mathrm{d}y = \frac{1}{2\sqrt{x}}\mathrm{d}x$

Hence, $\sqrt{x+h} \approx \sqrt{x} + \frac{h}{2\sqrt{x}}$

- In particular, for $x = 1$, we obtain,

$$\sqrt{1+h} \approx 1 + \frac{h}{2}$$

- In the general case, for $x = a^2$ $(a > 0)$, we have,

$$\sqrt{a^2+h} \approx a + \frac{h}{2a}$$

These approximate formulas are extremely simple and make it possible to compute square roots with a sufficient accuracy when $|h|$ is small compared to a^2.

For instance, the application of these results yields

$$\sqrt{1.21} = \sqrt{1+0.21} \approx 1 + \frac{0.21}{(2)(1)} = 1.105$$

The exact value of the root is equal to 1.1.

To compute the root $\sqrt{408}$ we represent it in the form $\sqrt{408} = \sqrt{20^2+8}$, and thus obtain $\sqrt{408} \approx 20 + (8/(2)(20)) = 20.2$.

Now, let us take $\sqrt{390}$. Here, it is convenient to put $h = -10$, then

$$\sqrt{390} = \sqrt{20^2-10} \approx 20 - \frac{10}{(2)(20)} = 19.75$$

- If $y = \sqrt[n]{x}$, then, $\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}(x^n) = \frac{1}{n}(x^{(1/n)-1}) = \frac{1}{n}x^{1/n} \cdot x^{-1} = \frac{1}{n}\frac{\sqrt[n]{x}}{x}$

$$\therefore\ \mathrm{d}y = \frac{1}{n}\frac{\sqrt[n]{x}}{x} \cdot \mathrm{d}x = \frac{1}{n}\frac{\sqrt[n]{x}}{x}h \qquad \text{(replacing d}x\text{ by }h)$$

$$\therefore\ \sqrt[n]{x+h} = \sqrt[n]{x} + \frac{1}{n}\frac{\sqrt[n]{\mathrm{x}}}{\mathrm{x}} \cdot h \tag{8}$$

For $x = 1$, this yields the approximate formula

$$\sqrt[n]{1+h} \approx 1 + \frac{h}{n}$$

- A more general formula is obtained for $x = a^n$ $(a > 0)$:

$$\sqrt[n]{a^n+h} \approx \mathrm{a} + \frac{h}{n \cdot a^{n-1}}$$ [7]

[7] This formula is obtained by putting $x = a^n$ in equation (8).

Let the reader compute several roots with the aid of this formula and estimate the accuracy achieved by finding more accurate values using the table of logarithms.

(II) Let us consider the function $y = \sin x$. Its differential is $\mathrm{d}y = \cos x\,\mathrm{d}x$, and therefore

$$\sin(x+h) \approx \sin x + h\cos x$$

In particular, for $x = 0$, we derive the formula, $\sin h \approx h$.
For example, we have $\sin\frac{\pi}{180} \approx \frac{\pi}{180} \approx 0.01745$
That is, approximately, $\sin 1° = 0.1745$.[(8)]
This approximation is correct to the fifth decimal digit, that is, the error does not exceed 10^{-5}. Let us compute $\sin 31°$.

$$\begin{aligned}\sin 31° &\approx \sin 30° + \frac{\pi}{180}\cos 30° \\ &\approx 0.5 + \frac{\sqrt{3}}{2}\cdot(0.01745) \\ &\approx 0.5150\end{aligned}$$

The tabular value of $\sin 31°$ correct within 10^{-4} (i.e., the fourth place of the decimal) is 0.5150. (The reader may account for the fact that we have obtained a major approximation of $\sin 31°$.)[(9)]

(III) Now, consider the function $y = \ln x$. Here we have, $\mathrm{d}y = (1/x)\mathrm{d}x$, and

- $\ln(x+h) \approx \ln x + \frac{h}{x}$
- In particular, for $x = 1$ this yields the formula $\ln(1+h) \approx 1$

Take the known value $\ln 781 \approx 6.66058$. To compute $\ln 782$, we apply the above formulas

$$\begin{aligned}\ln 782 &\approx 6.66058 + \frac{1}{781} \\ &\approx 6.66186\end{aligned}$$

The tabular value of $\ln 782$ correct within 10^{-5} is equal to 6.66185.
(We see that the error of our approximation is small. The reader may try to find out why in this case we have also obtained a major approximation.)[(10)]

16.3 DIFFERENTIALS OF BASIC ELEMENTARY FUNCTIONS

Since the differential of a function is obtained as the product of the derivative by the differential of the independent variable, *we can readily write down the table of the differentials of all the*

[(8)] Note that $\frac{\pi}{180} \approx \frac{3.14}{180} \approx 0.01745$ On the other hand, π radians $= 180°$. Therefore $\sin\frac{\pi}{180} = \sin\left(\frac{180°}{180}\right) = \sin 1°$, and we obtain, $\sin 1° \approx 0.01745$.

[(9)] Later on, it will be found that using Taylor's theorem (to be studied in Chapter 22) these results can be obtained more easily and accurately.

[(10)] Today, in the age of technology, the application of differentials for approximating function values is not very important (since with a pocket calculator, it is easy to find very accurate values of $f(x_0)$, $f(x_0 + \mathrm{d}x)$, and $\Delta y = f(x_0 + \mathrm{d}x) - f(x_0)$). However, differentials are important as a convenient notational device for the computation of antiderivatives as we will learn later in Part II of this book.

basic elementary functions because their derivatives are known. For instance,

$$\begin{aligned} \mathrm{d}(x^n) &= nx^{n-1}\,\mathrm{d}x \\ \mathrm{d}(a^x) &= a^x \ln a\,\mathrm{d}x \\ \mathrm{d}(\ln x) &= \frac{1}{x}\mathrm{d}x \\ \mathrm{d}(\sin x) &= \cos x\,\mathrm{d}x, \text{ and so on.} \end{aligned}$$

16.3.1 Differentials of the Results of Arithmetical Operations on Functions

In accordance with the rules for finding derivatives (studied in Chapter 10), we can use the derivative formulas to write down the corresponding differentials. For example, if u and v are differentiable functions of x, then the formula

$$\frac{\mathrm{d}}{\mathrm{d}x}(u+v) = \frac{\mathrm{d}u}{\mathrm{d}x} + \frac{\mathrm{d}v}{\mathrm{d}x} \tag{9a}$$

after multiplying both the sides by $\mathrm{d}x$ becomes

$$\mathrm{d}(u+v) = \mathrm{d}u + \mathrm{d}v \tag{9b}$$

which says that the differential of the function $(u+v)$ is the differential of the function u plus the differential of the function v. *It is still assumed that u and v are differentiable functions, but the name of the independent variable no longer appears in the formula.* We do not need to mention it as long as we understand that (9b) is an abbreviation for (9a). We illustrate the major rules in the table below.

	Derivative Rule	Differential Rule
1.	$\frac{\mathrm{d}(c)}{\mathrm{d}x} = 0$	$\mathrm{d}(c) = 0$
2.	$\frac{\mathrm{d}(x^n)}{\mathrm{d}x} = nx^{n-1}$	$\mathrm{d}(x^n) = nx^{n-1}\,\mathrm{d}x$
3.	$\frac{\mathrm{d}(cu)}{\mathrm{d}x} = c\frac{\mathrm{d}u}{\mathrm{d}x}$	$\mathrm{d}(cu) = c\,\mathrm{d}u$
4.	$\frac{\mathrm{d}(u+v)}{\mathrm{d}x} = \frac{\mathrm{d}u}{\mathrm{d}x} + \frac{\mathrm{d}v}{\mathrm{d}x}$	$\mathrm{d}(u+v) = \mathrm{d}u + \mathrm{d}v$
5.	$\frac{\mathrm{d}(uv)}{\mathrm{d}x} = u\frac{\mathrm{d}v}{\mathrm{d}x} + v\frac{\mathrm{d}u}{\mathrm{d}x}$	$\mathrm{d}(uv) = u\,\mathrm{d}v + v\,\mathrm{d}u$
6.	$\frac{\mathrm{d}(u/v)}{\mathrm{d}x} = \frac{v(\mathrm{d}u/\mathrm{d}x) - u(\mathrm{d}v/\mathrm{d}x)}{v^2}$	$\mathrm{d}\left(\frac{u}{v}\right) = \frac{v\,\mathrm{d}u - u\,\mathrm{d}v}{v^2}$
7.	$\frac{\mathrm{d}(u^n)}{\mathrm{d}x} = nu^{n-1}\frac{\mathrm{d}u}{\mathrm{d}x}$	$\mathrm{d}(u^n) = nu^{n-1}\,\mathrm{d}u$

Warning: One must be careful to distinguish between derivatives and differentials. *They are not the same.* When you write D_xy or $\mathrm{d}y/\mathrm{d}x$ you are using a symbol for the derivative, but when you write "dy" you are dealing with a differential.

Note (4): It should be noted that a differential on the left-hand side of an equation (say dy), also calls for a differential usually dx on the right-hand side of the equation. *Thus, we never have* $\mathrm{d}y = 3x^2$, *but we have* $\mathrm{d}y = 3x^2\,\mathrm{d}x$.

16.3.2 Differential of a Composite Function

While the definition of dy assumes that x is an independent variable, that assumption is not important. Let $y = f(u)$ and $u = \phi(x)$ be two *functions of their arguments* possessing the derivatives $f'(u)$ and $\phi'(x)$ with respect to these arguments. If we put,

$$y = f(u) = f[\phi(x)] = F(x) \quad \text{(say)}$$

then, by differentiating both sides with respect to x, we have

$$y' = F'(x) = f'(u) \cdot \phi'(x) \tag{10}$$

On multiplying both sides of this relation by dx, we get

$$y' \cdot \mathrm{d}x = f'(u) \cdot \phi'(x)\mathrm{d}x$$

$$\text{or} \quad \mathrm{d}y = f'(u) \cdot \mathrm{d}u \quad [\text{since } y' \cdot \mathrm{d}x = \mathrm{d}y \text{ and } \phi'(x)\mathrm{d}x = \mathrm{d}u]$$

Thus, the differential has the same form *as if the magnitude u were an independent variable*. This can be stated as follows.

The differential of a function $y = f(u)$ retains the same expression irrespective of whether its argument u is an independent variable or a function of another variable.

This property is referred to as the *invariance of the form of the differential.*[(11)]

It is because of this property that *we can write down the differential in one and the same form irrespective of the nature of the argument of the function*. The equality, $\mathrm{d}y = f'(u) \cdot \mathrm{d}u$ implies

$$f'(u) = \frac{\mathrm{d}y}{\mathrm{d}u}$$

and hence in all the cases, this equation may be looked upon as follows:

The rate of change of a function relative to its argument is equal to the ratio of the differential of the function to the differential of its argument.

Relation (10) can now be written as

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}u} \cdot \frac{\mathrm{d}u}{\mathrm{d}x} \tag{11}$$

The right-hand side of equation (11) *is obtained from the left-hand side by the simultaneous multiplication and division of the former by* du (if, of course, $\mathrm{d}u \neq 0$).

Hence, the arithmetical operations on differentials can be performed as if they were ordinary numbers. *Here, lies the reason for the convenience of the representation of the derivative as the ratio of the differentials*. For instance, using this representation of derivatives we can readily write down the differentiation rule for inverse of a function.

$$D_x[y] = \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{\mathrm{d}x/\mathrm{d}y} = \frac{1}{D_y[x]}$$

(11) Note that this important property of the differential follows from the differentiation rule for a function of a function.

Remarks: Knowing the derivative of a function, we can find its differential and vice versa. Hence, *the existence of the derivative can be taken as the condition equivalent to the differentiability of the function.*[12]

From the geometrical point of view, this condition is equivalent to the existence of the tangent to the curve $y = f(x)$, not perpendicular to the x-axis.

Definition: Recall that, *a function $y = f(x)$ is said to be differentiable at a point x* (in its domain) *if it possesses a derivative at that point*. Further, if a function is differentiable at every point in its domain (i.e., the derivative exists at every point in its domain) then it is called a *differentiable function*.

Now, in view of the above discussion (about the differential of a function) we can give the following definition:

A function $y = f(x)$ is said to be differentiable at a point x, *if it has a differential at that point*.

Note (5): Any problem involving differentials, (say that of finding $\mathrm{d}y$ when y is given as a function of x), may be handled either

(a) by finding $\mathrm{d}y/\mathrm{d}x$ and multiplying by $\mathrm{d}x$ or
(b) by direct use of formulas on differentials.

Example (3): Given a function $y = \sin\sqrt{x}$. Find $\mathrm{d}y$.

Solution: Representing the given function as a composite function,

$$y = \sin u, \quad u = \sqrt{x}$$

we find, $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \cos u \cdot \dfrac{\mathrm{d}u}{\mathrm{d}x} = \cos\sqrt{x} \cdot \dfrac{1}{2\sqrt{x}}$

$$\therefore\ \mathrm{d}y = \cos\sqrt{x} \cdot \frac{1}{2\sqrt{x}}\mathrm{d}x$$

or, we write,

$$\therefore\ \mathrm{d}y = \cos u\,\mathrm{d}u, \qquad \mathrm{d}u = (\sqrt{x})'\mathrm{d}x = \frac{1}{2\sqrt{x}}\mathrm{d}x$$

$$\therefore\ \mathrm{d}y = (\cos\sqrt{x})\left(\frac{1}{2\sqrt{x}}\mathrm{d}x\right)$$

Note (6): The application of the differential of a function can also be appreciated by considering *nonuniform motion* of a particle in a straight line. Let the law of motion be expressed mathematically by

$$s = f(t) \tag{12}$$

where s is the distance traveled and t stands for the time taken. Then, the velocity of the particle at any instant t_1 is given by $f'(t_1)$. If now an additional time Δt passes, let the particle cover an

[12] It is for this reason that the operations of finding the derivative and the differential of a function are called *differentiation*.

additional distance Δs. Since the motion is nonuniform, the dependence of Δs on Δt can be complicated because the velocity of the particle varies all the time.

But if Δt is not large, the velocity will not change considerably during the period of time from t_1 to $t_1 + \Delta t$. Therefore, the motion may be regarded as "*almost uniform*" during the time interval Δt. Hence, in calculating the distance traveled, we *shall not get a serious error if we regard the motion as uniform with the constant velocity* $f'(t_1)$, *from the instant* t_1 *to* $t_1 + \Delta t$.

Thus, the (approximate) distance traveled during the interval Δt is given by $f'(t_1) \cdot \Delta t$. *This product*, as we know, *is called the differential of the distance function and is denoted by* ds. We write

$$\mathrm{d}s = f'(t_1) \cdot \Delta t \tag{13}$$

Of course, *the real distance* Δs traveled (during the interval t_1 to $t_1 + \Delta t$) differs from *the invented distance* ds given in (13) above.

It must be clear that the *accuracy of the formula* (13) *becomes greater as* Δt *is decreased and vice versa. Nevertheless, it is much easier to compute* ds *as a distance covered in uniform motion than to evaluate the real distance* Δs. *This accounts for the fact that formula* (13) *is often used even when* Δt *is not very small.*

In all such cases, *the replacement of a real change of a quantity by its differential reduces to the transition from some nonuniform processes to the uniform ones*. Such a replacement is always based upon the fact that *every process is* "*almost uniform*" *during a small interval of time*.

16.4 TWO INTERPRETATIONS OF THE NOTATION d*y*/d*x*

Leibniz used the suggestive notation dy/dx for the instantaneous rate of change of y with respect to x. This notation suggests that the instantaneous rate comes from considering an average rate (which is indeed a quotient) and computing its limit. Thus, dy/dx stands for the limit,

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \frac{\mathrm{d}y}{\mathrm{d}x}, \quad \text{provided the limit exists}$$

Here, dy and dx do not have any meaning if considered separately (since dy/dx is a single entity: a symbol for the limit, which we call the derivative).

Our investigation suggests that

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \frac{\mathrm{d}y}{\mathrm{d}x}, \quad \text{provided the limit exists}$$

It would be wrong to interpret this limiting relation in the sense that Δy tends to dy and Δx to dx, as $\Delta x \to 0$. *The correct meaning is that the ratio of the increments* $\Delta y/\Delta x$, *as* $\Delta x \to 0$, *tends to the limit denoted* by dy/dx or the $f'(x)$. On the other hand, the differential of a function $y = f(x)$ is defined by

$$\mathrm{d}y = \mathrm{d}[f(x)] = f'(x)\mathrm{d}x$$

where $f'(x)$ stands for the derivative of f at a point x and dx is an *arbitrary number* (d$x \neq 0$). Dividing both sides by dx, we get dy/d$x = f'(x)$, (d$x \neq 0$).

In this case, dy/dx stands for the ratio of two quantities namely dy and dx. In either case, we get dy/d$x = f'(x)$, d$x \neq 0$.

Now, we can also say that the ratio of increments $\Delta y/\Delta x$ *tends to the ratio of differentials* dy/dx, as $\Delta x \to 0$.

16.5 INTEGRALS IN DIFFERENTIAL NOTATION

The notation of differentials allows us to express integrals in a shorthand that often proves useful. For example, if u is a *differentiable function of* x, then the integral of du/dx with respect to x is sometimes written simply as the integral of du:

$$\int \frac{\mathrm{d}u}{\mathrm{d}x} \cdot \mathrm{d}x = \int \mathrm{d}u$$

Thus, the integral of du is required to be evaluated as

$$\int \mathrm{d}u = \int \frac{\mathrm{d}(u)}{\mathrm{d}x}\mathrm{d}x = u + c$$

or simply, $\int \mathrm{d}u = u + c$

For instance, if $u = \sin x$, then $\mathrm{d}(\sin x) = \cos x\,\mathrm{d}x$, and we can write

$$\int \mathrm{d}(\sin x) = \sin x + c$$

which is short for $\int \cos x\,\mathrm{d}x = \int \frac{\mathrm{d}}{\mathrm{d}x}(\sin x)\mathrm{d}x = \sin x + c$.

Thus, for a given integral $\int f(x)\mathrm{d}x$, we have to express the differential $f(x)\mathrm{d}x$ in the form $(\mathrm{d}/\mathrm{d}x)F(x)\mathrm{d}x$ (which also stands for $\mathrm{d}[F(x)]$). Obviously, then $(\mathrm{d}/\mathrm{d}x)F(x) = f(x)$. Whenever $f(x)\mathrm{d}x$ is expressed in the form $\mathrm{d}[F(x)]$, we say that the integrand is expressed in the standard form. Once this is done, we can immediately write down the antiderivative (or the indefinite integral), $F(x) + c$.

Solved Examples

Example (1): Find the approximate value of $(4.01)^3$ correct to two decimal places.

Solution: We have

$$f(x + \delta x) \approx f(x) + f'(x) \cdot \delta x \qquad (14)^{(13)}$$

where δx is small.

$$\therefore\ (x + \delta x)^3 \approx x^3 + 3x^2 \cdot \delta x \qquad (15)$$

We take $x = 4$ and $\delta x = 0.01$

(13) In all subsequent problems, we shall use this formula, which gives the approximate value of a differentiable function $f(x)$, at a point $(x + \delta x)$ close to x.

Substituting these values in (15), we get,

$$\begin{aligned}(4+0.01)^3 &\approx (4)^3 + 3(4)^2 \cdot (0.01)\\ &\approx 64 + (48)\cdot(0.01)\\ &\approx 64 + 0.48\\ &\approx (4.01)^3 \approx 64.48 \qquad \text{Ans.}\end{aligned}$$

Now, we must similarly compute $(3.97)^3$.

Here, we take $x = 4$ and $\delta x = 0.03$

$$\begin{aligned}\therefore (3.97)^3 &= (4-0.03)^3\\ (4-0.03)^3 &\approx (4)^3 + 3(4)^2 \cdot (-0.03)\ [\because f((x+(-\delta x)) \approx f(x) + f'(x)\cdot(-\delta x)]\\ &\approx 64 + (48)\,(-0.03)\\ &\approx 64 - 1.44\\ &\approx 62.56\end{aligned}$$

Example (5): Find an approximate value of $\sqrt[3]{8.05}$

Solution: Consider the function

$$f(x) = \sqrt[3]{x} = (\mathrm{x})^{1/3}$$

$$\therefore \mathrm{f}'(\mathrm{x}) = \frac{1}{3}\mathrm{x}^{-2/3} = \frac{1}{3\mathrm{x}^{2/3}}$$

We have,

$$f(x+\delta x) \approx f(x) + f'(x)\cdot\delta x, \quad \text{where } \delta x \text{ is small}$$

$$\therefore (x+\delta x)^{1/3} \approx x^{1/3} + \frac{1}{3x^{2/3}}\cdot\delta x \tag{16}$$

We take $x = 8$ and $\delta x = 0.05$

Substituting these rules in (16), we get

$$(8+0.05)^{1/3} \approx (8)^{1/3} + \frac{1}{3(8)^{2/3}}(0.05)$$

$$\approx 2 + \frac{0.05}{(3)(4)}$$

$$\approx 2 + 0.00417$$

$$\therefore \sqrt[3]{8.05} \approx 2.00417$$

Now, if we wish to compute $\sqrt[3]{7.95}$, we get

$$\sqrt[3]{7.95} \approx 2 - 0.00417 \approx 1.99583$$

Example (6): Estimate the value of $\sin 31^\circ$, assuming that $1^\circ = 0.0175\,\text{rad}$, and $\cos 30^\circ = 0.8660$.

Solution: Let $f(x) = \sin x \;\therefore f'(x) = \cos x$
We have, $f(x+\delta x) \approx f(x) + f'(x)\cdot\delta x$, where δx is small

$$\therefore \sin(x+\delta x) \approx \sin x + \cos x \cdot \delta x \tag{17}$$

We take $x = 30^\circ = \frac{\pi}{6}$ and $\delta x = 1^\circ = \frac{\pi}{180} = 0.0175$
Substituting in (17), we get,

$$\begin{aligned} \sin 31^\circ &= \sin(30^\circ + 1^\circ) \\ &\approx \sin 30^\circ + \cos 30^\circ \cdot (0.0175) \\ &\approx \sin(\pi/6) + \cos(\pi/6)\cdot(0.0175) \\ &\approx 0.5 + (0.8660)\cdot(0.0175) \\ &\approx 0.5 + 0.015155 \\ &\approx 0.51516 \qquad \text{Ans.} \end{aligned}$$

Note (7): Assuming that $1^\circ = 0.0175$ rad and $\sin 45^\circ = 0.7071$, we can easily estimate $\cos 46^\circ$ or $\cos 44^\circ$. (Remember that $\cos 45^\circ = \sin 45^\circ = 0.7071$.)

$$\begin{aligned} \text{Let} \quad f(x) &= \cos x \\ \therefore f'(x) &= -\sin x \end{aligned}$$

We have $f(x+\delta x) \approx f(x) + f'(x)\delta x$

$$\begin{aligned} \therefore \cos(x+\delta x) &= \cos x - \sin x \cdot \delta x \\ \therefore \cos(45^\circ + 1^\circ) &\approx \cos 45^\circ - \sin 45^\circ(0.0175) \\ &\approx 0.7071 - (0.7071)(0.0175) \\ &\approx 0.7071 - 0.01237 \\ &\approx 0.6947 \end{aligned}$$

and

$$\begin{aligned} \cos(45^\circ - 1^\circ) &\approx 0.7071 - (0.7071)(-0.0175) \\ &\approx 0.7071 + 0.01237 \\ &\approx 0.71947 \end{aligned}$$

Example (7): Approximate $\sin\frac{7\pi}{36}$

Solution: Note that, $\frac{7\pi}{36} = \frac{6\pi}{36} + \frac{\pi}{36} = \frac{\pi}{6} + \frac{\pi}{36}$. Thus, $\frac{7\pi}{36}$ is close to $\frac{\pi}{6}$.

Thus, we write $\sin\frac{7\pi}{36} = \sin\left(\frac{\pi}{6} + \frac{\pi}{36}\right)$
Let $\frac{\pi}{6} = x$ and $\frac{\pi}{36} = \delta x$

$$\therefore \sin\left(\frac{\pi}{6} + \frac{\pi}{36}\right) = \sin(x+\delta x)$$

We have, $f(x+\delta x) \approx f(x) + f'(x)\cdot\delta x$

$$\therefore \sin(x+\delta x) \approx \sin x + \cos x \cdot \delta x$$

$$\text{or}\quad \sin(\pi/6+\pi/36) \approx \sin(\pi/6)+\cos(\pi/6)\cdot(\pi/36)$$
$$\approx 0.5+(\sqrt{3}/2)\cdot(\pi/36)$$
$$\approx 0.5+0.075575 = 0.575575$$

Example (8): Find the value of $f(x)=2x^3+7x+5$, at $x=2.001$.

Solution: Let $f(x)=2x^3+7x+5$

$$\therefore f'(x) = 6x^2+7$$

We have, $f(x+\delta x) \approx f(x)+f'(x)\cdot\delta x$, where δx is small.

$$\approx (2x^3+7x+5)+(6x^2+7)\cdot\delta x$$

We take $x=2$ and $\delta x=0.001$

$$\therefore f(2.001) = [2(2)^3+7(2)+5]+[6(2)^2+7](0.001)$$
$$= (16+14+5)+(24+7)(0.001)$$
$$= 35+0.031 = 35.031 \qquad \text{Ans.}$$

Example (9): Find the approximate value of $\tan^{-1}(0.99)$.

Solution: Let $f(x) = \tan^{-1} x \quad \therefore f'(x) = \dfrac{1}{1+x^2}$

We know that,

$$\therefore f(x+\delta x) \approx f(x)+f'(x)\cdot\delta x, \text{ where } \delta x \text{ is small}$$
$$\therefore \tan^{-1}(x+\delta x) \approx \tan^{-1} x + \left(\frac{1}{1+x^2}\right)\cdot(\delta x)$$

We take $x=1$ and $\delta x=-0.01$

$$\therefore \tan^{-1}(1-0.01) \approx \tan^{-1}(1)+\left(\frac{1}{1+(1)^2}\right)\cdot(-0.01)$$
$$\approx \frac{\pi}{4}-\frac{0.01}{2}$$
$$\approx \frac{\pi}{4}-0.005$$
$$\approx 0.7854-0.005$$
$$\approx 0.780 \qquad \text{Ans.}$$

Note (8): Approximate value of $\tan^{-1}(1.001)$ is given by

$$\tan^{-1}(1+0.001) \approx \tan^{-1}(1)+\left(\frac{1}{1+1^2}\right)(0.001)$$
$$\approx \frac{\pi}{4}+\frac{1}{2}\cdot(0.001)$$
$$\approx 0.7854+0.0005$$
$$\approx 0.7855+0.0005$$
$$\approx 0.7860 \qquad \text{Ans.}$$

Example (10): Find the approximate value of $e^{1.002}$ taking $e = 2.71828$.

Solution: Let $f(x) = e^x$, then we know that $f'(x) = e^x$

We have, $f(x+\delta x) \approx f(x) + f'(x)\cdot\delta x$, where δx is small

$$\therefore e^{x+\delta x} \approx e^x + e^x \cdot \delta x$$

We take $x = 1$ and $\delta x = 0.002$

$$\begin{aligned} \therefore e^{1.002} &\approx e^1 + e^1 \cdot (0.002) \\ &\approx 2.71828 + (2.71828)(0.002) \\ &\approx 2.71828 + 0.005437 \\ &\approx 2.7237 \text{ (up to four decimal places)} \end{aligned}$$

Example (11): Taking $\log_e 10 = 2.3026$, find the approximate value of $\log_e 101$.

Solution: Let $f(x) = \log_e x, \quad \therefore f'(x) = \dfrac{1}{x}$

We have, $f(x+\delta x) \approx f(x) + f'(x)\cdot\delta x$, where δx is small

$$\therefore \log_e(x+\delta x) \approx \log_e x + \frac{1}{x}\cdot(\delta x)$$

We take $x = 100$ and $\delta x = 1$

$$\begin{aligned} \therefore \log_e(100+1) &\approx \log_e 100 + 1/100(1) \\ &\approx \log_e(10)^2 + 0.01 \\ &\approx 2(2.3026) + 0.01 \\ &\approx 4.6052 + 0.01 \\ &\approx 4.6152 \qquad \textbf{Ans}. \end{aligned}$$

16.6 TO COMPUTE (APPROXIMATE) SMALL CHANGES AND SMALL ERRORS CAUSED IN VARIOUS SITUATIONS

The measurements in physical experiments are not exact. A certain amount of error is always present. Therefore, *the measurements are in fact only approximations*. Of course, these approximate numbers [representing measurement(s) of various quantities] are very close to their exact measurement(s).

The statement, $y = f(x)$ means that *for a measured value of* x, we can calculate the corresponding value of y. If a small error δx enters in the measurement of x, then evidently, there will be an error in the calculation of the dependent variable y. These errors may be due to inaccuracies/limitation(s) of measuring instruments or due to human errors, Besides, these errors may be positive or negative in nature.[14]

[14] For example, in calculating the area of a given circle, if there is an error in measuring the radius x, then there is bound to be an error in computing the area, $y = \pi x^2$.

The resulting error in y is given by $\delta y = f(x+\delta x) - f(x)$. If δx is very small, then

$$\delta y \approx f(x) + f'(x)\cdot\delta x - f(x) \quad [\because f(x+\delta x) \approx f(x) + f'(x)\cdot\delta x]$$

or
$$\delta y \approx f'(x)\cdot\delta x \tag{18}$$

Thus, the *two errors* δx *and* δy are related by (18). This formula enables us to find approximately the *small change* δy in y corresponding to *small change* δx in x. If δx is treated as a *small error* in the measurement of x, then the formula (18) gives us the *corresponding error* δy in calculating y.

16.6.1 Definitions: Absolute Error, Relative Error, and Percentage Error

If the error δy is calculated for a given value of x (say $x = x_1$) it is called the *absolute error*. The quantity $\delta y/y$ is called the *relative error*. Sometimes, scientists are interested in the percent error in the computation of a numerical quantity. The percentage error is given by

$$\left|\frac{\delta y}{y}\cdot 100\right| = |\text{relative error} \times 100|$$ [15]

Let us see some examples:

Example (12): A spherical ball when new measures 3.00 cm in radius. What is the approximate volume of metal lost after it wears down to $r = 2.98$ cm?

Solution: Volume of the spherical ball is given by, $V = (4/3)\pi r^3$.

The *approximate change* in computing volume of the spherical ball (due to wear of 0.02 cm in its radius) is given by

$$\begin{aligned}
\delta v &\approx \frac{d}{dr}\left(\frac{4}{3}\pi r^3\right)\cdot\delta r \\
&\approx 4\pi r^2\cdot\delta r \text{ (here } r = 3 \text{ cm and } \delta r = 0.02 \text{ cm)} \\
&\approx 4\pi(3)^2\cdot(0.02) \\
&\approx 36\pi(0.02) = 0.72\pi, \text{ taking } \pi \approx 3.14 \\
&\approx 2.26\,\text{cm}^3 \qquad \text{Ans.}
\end{aligned}$$

Note (9): *If we assume that the figures* $r_1 = 3.00$ cm *and* $r_2 = 2.98$ cm *are exact, then the exact answer would be*

$$\begin{aligned}
\delta v &\approx \frac{4}{3}\pi\left[3^2 - (2.98)^2\right] \\
&\approx (0.71521066\ldots)\pi \\
&\approx 2.26\,\text{cm}^3, \text{ correct to two decimal places.}
\end{aligned}$$

[15] The percentage error has to be positive number irrespective of whether the error is positive or negative.

Example (13): A spherical ball of wood of radius 150 cm is coated by a layer of paint. If thickness of paint layer is 0.05 cm, find the volume of paint required.

Solution: The *approximate volume* of paint required is given by

$$\begin{aligned}\delta v \approx \frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{4}{3}\pi r^3\right)\cdot\delta r &= (4\pi r^2)\cdot\delta r\ \mathrm{cm}^3\\ &= 4\pi(150)^2\cdot(0.05) = 4\pi(22,500)\cdot(0.05)\ \mathrm{cm}^3\\ &= 14137.167\ \mathrm{cm}^3 = 14.14\ \mathrm{L}\ (\text{approx.}) \qquad \textbf{Ans.}\end{aligned}$$

Example (14): A hemispherical dome of a temple has radius 5 m from the inside. If the dome is to be coated by a plastic material of 0.08 cm thickness, then find the volume of material used.

Solution: The *approximate volume* of plastic material required, is given by

$$\begin{aligned}\delta v &\approx \frac{\mathrm{d}}{\mathrm{d}r}\left[\frac{1}{2}\left(\frac{4}{3}\pi r^3\right)\right]\cdot\delta r\\ &\approx 2\pi r^2\cdot\delta r \quad (\text{Here } r = 5\ \mathrm{m} = 500\ \mathrm{cm} \text{ and } \delta r = 0.08\ \mathrm{cm}.)\\ &\approx 2\pi(500)^2\cdot(0.08)\\ &\approx (0.16)(250,000)\pi\ \mathrm{cm}^3\\ &\approx 125663.7\ \mathrm{cm}^3\\ &= 125.7\ \mathrm{L}\ (\text{approx.}) \qquad \text{Ans.}\end{aligned}$$

Example (15): Find the volume of the metal of a *hollow cylindrical shell* of inner radius 2 cm and thickness 0.1 cm and length 10 cm.

Solution: Let v = inner volume of the cylinder and δv = the volume of the metal used in the hollow cylindrical shell of thickness 0.1 cm.

We have, $v = \pi r^2 h$

$$\begin{aligned}\therefore\ \delta v &\approx \frac{\mathrm{d}v}{\mathrm{d}r}\cdot\delta r\\ &\approx (2\pi rh)\cdot\delta r\\ &\approx 40\pi(0.1) \quad (\because\ r = 2,\ h = 10,\ \text{and } \delta r = 0.1)\\ &= 4\pi = 12.57\ \mathrm{cm}^3 \qquad \text{Ans.}\end{aligned}$$

Example (16): If the diameter of a sphere is measured to be 20 cm and the error in the measurement is 0.4 cm, find the error in the calculation of the surface area of the sphere.

Solution: Let surface area of the sphere be denoted by s.

Then, $s = 4\pi r^2$, where r is the radius of the sphere.
The error in measurement of the diameter of the sphere = 0.4 cm.
$\therefore$ The error in measurement of radius of the sphere = 0.2 cm.
Suppose δs is the error in the calculation of surface area of the sphere.
Then, we have,

$$\delta s \approx \frac{\mathrm{d}s}{\mathrm{d}r} \cdot \delta r = (8\pi r) \cdot \delta r \quad (\delta r = 0.2 \text{ cm})$$

For $r = 10$ cm, $\delta r = 0.2$ cm $[\because \text{dia} = 20 \text{ cm}, \ \therefore r = 10 \text{ cm}]$

$$\therefore \delta s \approx 8\pi(10) \cdot (0.2) = 16\pi \, cm^2 \qquad \text{Ans.}$$

Example (17): A right circular cone has a height of 7 cm and a base diameter of 5 cm. It is found that the diameter is not correctly measured to the extent of 0.06 cm. Find the consequent error in the calculated volume.

Solution: The volume of the right circular cone $v = (1/3)\pi r^2 h$.

Height of the cone = 7 cm (it is assumed to be *correctly measured*)
Radius of the base = 2.5 cm
(This is not correct. There is an error of 0.03 cm in its measurement).

We have to find the consequent error δv in the calculation of the volume v (i.e., to find δv).
We have,

$$\delta v \approx \frac{\mathrm{d}v}{\mathrm{d}r} \cdot \delta r = \left(\frac{2}{3}\pi r h\right) \cdot \delta r \quad \left[\begin{array}{l} \because v = \dfrac{1}{3}\pi r^2 h \\ \therefore \dfrac{\mathrm{d}v}{\mathrm{d}r} = \dfrac{2}{3}\pi r h \end{array}\right.$$

At $r = 2.5$ cm, $\delta r = 0.03$ cm. (Note *that, h is constant* = 7 cm.)

$$\begin{aligned} \therefore \delta v &\approx \frac{2}{3}\pi(2.5)(7) \cdot 0.03 \text{ cm}^3 \\ &\approx \frac{35}{100}\pi \\ &\approx 0.35\pi \text{ cm}^3 \\ &\approx 1.09 = 1.1 \text{ cm}^3 \text{ (approx)} \quad [\because \pi = 3.14] \qquad \text{Ans.} \end{aligned}$$

Example (18): Find the approximate error in computing the surface area of a cube having an edge of 3 m, if an error of 2 cm is made in measuring the edge. Also, find the percentage error in computing the surface area.

Solution: Suppose the edge of the cube is x m.

$$\therefore \text{Its surface area } A(x) = 6x^2$$

The error in measuring the edge is $\delta x = 2\,\text{cm} = 0.02\,\text{m}$. We have to find the approximate error in calculating the surface area of the cube. Suppose it is δA.
Then,

$$\begin{aligned}\delta A &\approx \frac{\mathrm{d}A}{\mathrm{d}x}\cdot\delta x\\ &\approx (12x)\cdot\delta x\\ &\approx 12(3)\cdot(0.02)\\ &\approx 0.72\,\text{m}^2\end{aligned}$$

Now, the percentage error in computing the surface area

$$\begin{aligned}&= \frac{\delta A}{A}\cdot(100) = \frac{0.72}{6(3)^2}\times 100\\ &= \frac{0.72}{54}\times 100 = \frac{72}{54} = \frac{4}{3} = 1.33\% \qquad \textbf{Ans.}\end{aligned}$$

Example (19): If the radius of a spherical balloon increases by 0.1%, find approximately the percentage error in computing the volume.

Solution: We have $v = \frac{4}{3}\pi r^3$

$$\therefore\ \delta v \approx \frac{\mathrm{d}v}{\mathrm{d}r}\cdot\delta r = 4\pi r^2\cdot\delta r$$

Now, the percentage error in computing v is given by

$$\frac{\delta v}{v}\times 100 = \frac{4\pi r^2\cdot\delta r}{(4/3)\pi r^3}\times 100 = \frac{3}{r}\times 100\cdot\delta r \tag{19}$$

But, it is given that the radius increases by 0.1% (i.e., percentage of increase in $r = 0.1$) or $\frac{\delta r}{r}\times 100 = 0.1$

$$\therefore\ \delta r = \frac{(0.1)r}{100} = \frac{r}{1000}$$

Put this value of δr in (19), we get

$$\%\ \text{error in computing}\ v = \frac{3}{r}\times 100\cdot\left(\frac{r}{1000}\right) = \frac{3}{10} = 0.3\% \qquad \textbf{Ans.}$$

Example (20): If there is an error of 0.3% in the measurement of the radius of a spherical balloon, find the *percentage error* in the calculation of its volume.

Solution: Let $x =$ radius of the sphere.

Then, its volume $v = \frac{4}{3}\pi x^3$

$$\therefore\ \delta v \approx \frac{\mathrm{d}v}{\mathrm{d}x}\cdot\delta x = (4\pi x^2)\cdot\delta x \tag{20}$$

$$\left[\therefore\ \frac{dv}{dx} = 4\pi x^2\right] \tag{21}$$

It is given that percentage error in x is 0.3.

$$\therefore\ 100\frac{\delta x}{x} = 0.3$$

$$\therefore\ \text{error in } x (\text{i.e., } \delta x) = \frac{(0.3)\cdot x}{100} \tag{22}$$

We have to compute: percentage error in calculating v [i.e., to compute the value of $((\delta v/v)\cdot 100)$].
We have $\delta v \approx \frac{dv}{dx}\cdot\delta x$ [see (20) above]
Multiplying both sides of (20) by $\frac{100}{v}$, we get,

$$\therefore\ 100\frac{\delta v}{v} \approx \frac{100}{v}\cdot\frac{dv}{dx}\cdot\delta x$$

$$\therefore\ \%\ \text{error in } v \approx \frac{100}{(4/3)\pi x^3}\cdot\frac{4\pi x^2}{1}\cdot\frac{(0.3)x}{100}$$

$$= 0.9\% \qquad \textbf{Ans.}$$

Example (21): The time T of a complete oscillation of a simple pendulum of length "l" is given by $T = 2\pi\sqrt{\frac{l}{g}}$. If there is an error of 1.2% in the measurement of l, find the percentage error in T.

Solution:

$$\text{Given } T = 2\pi\sqrt{\frac{l}{g}} = \frac{2\pi}{\sqrt{g}}\sqrt{l} \quad \text{or} \quad T = k\sqrt{l}, \quad \left[\text{where } k \text{ is a constant} = \frac{2\pi}{\sqrt{g}}\right]$$

There is an error of 1.2% in the measurement of l.

$$\therefore\ \frac{\delta l}{l}\times 100 = 1.2$$

$$\therefore\ \delta l = \frac{(1.2)l}{100} = \frac{12\,l}{1000}$$

We have to find % error in computing T $\left(\text{i.e., to compute } \frac{\delta T}{T}\times 100\right)$

$$\text{We have, } \frac{\delta T}{\delta l} \approx \frac{dT}{dl}$$

$$\therefore \delta T \approx \frac{dT}{dl} \cdot \delta l = \frac{d}{dl}\left[k\sqrt{l}\right] \cdot \delta l$$

$$= k \cdot \frac{1}{2} \cdot l^{-1/2} \cdot \delta l$$

$$= k \cdot \frac{1}{2\sqrt{l}} \cdot \delta l \quad \left[\because T = k\sqrt{l}\right]$$

$$\therefore \frac{\delta T}{T} \times 100 = \left[\frac{k}{2\sqrt{l}} \cdot \delta l\right] \cdot \left[\frac{1}{k\sqrt{l}}\right] \times 100 \quad \left[\because \frac{1}{T} = \frac{1}{k\sqrt{l}}\right]$$

$$= \frac{\delta l}{2l} \times 100 \quad \left[\because \delta l = \frac{12\,l}{1000}\right]$$

$$= \frac{12\,l}{1000} \cdot \frac{1}{2l} \cdot 100 = \frac{6}{10} = 0.6\% \qquad \text{Ans.}$$

Note (10): Exercises are not given here. The reader may refer to standard books for good exercises.

17 Derivatives and Differentials of Higher Orders

17.1 INTRODUCTION

We have studied several methods of finding derivatives of differentiable functions. If $y = f(x)$ is a *differentiable* function of x, then its derivative is denoted by

$$\frac{dy}{dx} \quad \text{or} \quad f'(x) \quad \text{or} \quad y' \quad \text{or} \quad y_1$$

The notation $f'(x)$ suggests that the derivative of $f(x)$ is also a function of x. If the function $f'(x)$ is in turn differentiable, its derivative is called the second derivative (or the derivative of the second order) of the original function $f(x)$ and is denoted by $f''(x)$. This leads us to the concept of the derivatives of higher orders.

$$f''(x) = [f'(x)]' = \lim_{\Delta x \to 0} \frac{f'(x + \Delta x) - f'(x)}{\Delta x}$$

We write,

$$\frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d^2y}{dx^2} \quad \text{or} \quad \left[\frac{d(f'(x))}{dx} = f''(x) \quad \text{or} \quad y'' \quad \text{or} \quad y_2\right]$$

Similarly, we can find the derivative of d^2y/dx^2 *provided it exists*, and is denoted by d^3y/dx^3 [or $f'''(x)$ or y''' or y_3], called the third derivative of $y = f(x)$ and so on.

17.1.1 Notations for Derivatives of $y = f(x)$

Order of Derivative	Prime Notation (′)	Leibniz Notation	y-Notation	D-Notation
1st	y' or $f'(x)$	dy/dx	y_1	Df
2nd	y'' or $f''(x)$	d^2y/dx^2	y_2	D^2f
3rd	yy''' or $f'''(x)$	d^3y/dx^3	y_3	D^3f
4th	y^{iv} or $f^{iv}(x)$	d^4y/dx^4	y_4	D^4f
⋮	⋮	⋮	⋮	⋮
nth	$y^{(n)}$ or $f^{(n)}(x)$	d^ny/dx^n	y_n	D^nf

17-Derivatives and differentials of higher order Related rates: computing unknown derivative(s) using known derivative(s).

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

Example (1): If $y = f(x) = 2x^5 - x^2 + 3$, then

$$y_1 = 10x^4 - 2x, \quad y_2 = 40x^3 - 2, \quad y_3 = 120x^2,$$

$$y_4 = 240x, \quad y_5 = 240, \quad y_6 = 0, \quad \ldots, \quad y_n = 0$$

Note that, for a polynomial function $f(x)$ of degree 5, $f^{(n)}(x) = 0$ for $n \geq 6$. More generally, the $(n+1)^{\text{th}}$ and all higher derivatives of any polynomial of degree n are equal to 0.

However, there are functions [like $\sin x$, $\cos x$, e^x, $\log_e x$, and their extended forms, [that is, $\sin(ax+b)$, $\cos(ax+b)$, e^{ax}, $\log_e(ax+b)$, or more general ones like $\sin(f(x))$, $e^{f(x)}$, and $\log_a(f(x))$] that can be differentiated any number of times and $f^{(n)}(x)$ is never 0.

The most important derivatives in physical applications are the first and the second, and these have different special meanings. For example, if x represents time and y the distance, then dy/dx represents *velocity* v. In this case, the rate of change of velocity, that is, dv/dx $(= d^2y/dx^2)$ is called the *acceleration*.

Also, the second-order derivative has other special interpretations, depending on the meaning of the related variables x and y. When the relation between x and y is graphed, then one interpretation of d^2y/dx^2 is associated with the curvature of the graph.

Note (1): The generation of successive derivatives is not merely free creation of the curious mind. A railroad engineer has to employ second derivatives to calculate the curvature of the line he constructs. He needs a precise measure of the curvature to find the exact degree of banking required to prevent trains from overturning.

An automobile designer utilizes the third derivative in order to test the ride quality of the car he designs and the structural engineer has even to go to the fourth derivative in order to measure the elasticity of the beam and the strength of the columns. Besides, we will later see that the derivatives of higher orders are needed to expand functions (to the desired degree of accuracy) in the form of polynomials.

Example (2): Let us find the nth derivatives of the following:

(i) x^n, (ii) e^x, (iii) a^x, (iv) $\sin x$, (v) $\cos x$, (vi) $1/x$, (vii) $\log_e x$

Solutions:

(i) Let $y = x^n$.

$$\therefore\ y_1 = nx^{n-1},\ y_2 = n(n-1)x^{n-2},\ y_3 = n(n-1)(n-2)x^{n-3},$$
$$y_4 = n(n-1)(n-2)(n-3)x^{n-4}, \quad \text{and so on}$$

$$\begin{aligned}\therefore\ y_n &= n(n-1)(n-2)(n-3)\ldots 2\cdot 1\cdot x^{n-n}\\ &= n(n-1)(n-2)(n-3)\ldots 2\cdot 1\\ &= n! \quad \text{Ans.}\end{aligned}$$

Remark: $y_{n+1} = 0$ (since, $y_n = n! = \text{constant}$)

(ii) Let $y = e^x$.

$$\therefore y_1 = e^x,\ y_2 = e^x,\ y_3 = e^x, \quad \text{and so on}$$

$$\therefore y_n = e^x \quad \textbf{Ans.}$$

(iii) Let $y = a^x$.

$$\therefore y_1 = a^x \log_e a = a^x \cdot k$$

where $k = \log_e a = \text{constant}$.

$$\therefore \quad y_2 = k \cdot a^x \log_e a = k^2 \cdot a^x$$
$$y_3 = k^3 \cdot a^x, \text{ and so on}$$

$$\therefore \quad y_n = k^n \cdot a^x = (\log_e a)^n \cdot a^x \quad \textbf{Ans.}$$

(iv) Let $y = \sin x$.

$$\therefore \quad y_1 = \cos x = \sin\left(\frac{\pi}{2} + x\right) \quad \left[\because \sin\left(\frac{\pi}{2} + \theta\right) = \cos\theta\right]$$

$$\therefore \quad y_2 = \cos\left(\frac{\pi}{2} + x\right) = \sin\left[\frac{\pi}{2} + \left(\frac{\pi}{2} + x\right)\right] = \sin\left(2 \cdot \frac{\pi}{2} + x\right)$$

$$y_3 = \cos\left(2 \cdot \frac{\pi}{2} + x\right) = \sin\left[\frac{\pi}{2} + \left(2 \cdot \frac{\pi}{2} + x\right)\right]$$

$$= \sin\left(3 \cdot \frac{\pi}{2} + x\right)$$

$$\therefore y_n = \sin\left(n \cdot \frac{\pi}{2} + x\right) \qquad \textbf{Ans.}$$

(v) Let $y = \cos x$.

$$\therefore \frac{dy}{dx} = y_1 = -\sin x = \cos\left(\frac{\pi}{2} + x\right) \quad \left[\because \cos\left(\frac{\pi}{2} + \theta\right) = -\sin\theta\right]$$

Now, it is easy to show that,

$$y_n = \cos\left(n \cdot \frac{\pi}{2} + x\right) \quad \textbf{Ans.}$$

(vi) Let $y = (1/x) = x^{-1}$.

$$\therefore y_1 = -1 \cdot x^{-2} = \frac{(-1)}{x^2}$$

$$y_2 = (-1)(-2)x^{-3} = \frac{(-1)^2 \cdot 1 \cdot 2}{x^3} = \frac{(-1)^2 \cdot 2!}{x^3}$$

$$y_3 = (-1)(-2)(-3) \cdot x^{-4} = \frac{(-1)^3 \cdot 1 \cdot 2 \cdot 3}{x^4} = \frac{(-1)^3 \cdot 3!}{x^4}$$

$$\vdots$$

$$\therefore y_n = \frac{(-1)^n \cdot n!}{x^{n+1}} \qquad \textbf{Ans.}$$

(vii) Let $y = \log_e x$.

$$\therefore y_1 = \frac{1}{x} = x^{-1}$$

$$\therefore y_2 = (-1)x^{-2} = \frac{(-1)}{x^2}$$

$$y_3 = (-1)(-2) \cdot x^{-3} = \frac{(-1)^2 2!}{x^3}$$

$$\vdots$$

$$\therefore y_n = \frac{(-1)(-2)(-3)\ldots(-n+1)}{x^n} = \frac{(-1)^{n-1}(n-1)!}{x^n}$$

[Compare this result with the nth derivative of $1/x$ at (vi).]

Note (2): The higher derivatives with respect to the *extended forms* of the above functions are given below at (1)–(9). The reader may easily prove these results. It is useful to remember them since they will be needed for solving problems.

1. Let $y = (ax+b)^r$, $\quad r \in R$.
   Then, $y_n = r(r-1)(r-2)\ \ldots\ (r-n+1)a^n(ax+b)^{r-n}$
   This result is true for every real value of r.
2. Let $y = (ax+b)^r$, $\quad r \in N$.
   Then,
   $$y_n = \frac{r!a^n(ax+b)^{r-n}}{(r-n)!}, \quad (n < r)$$
   where r is a positive integer.
3. Let $y = (ax+b)^n$, $\quad n \in N$.
   Then,
   $$y_n = \frac{n!a^n(ax+b)^{n-n}}{(n-n)!}, \quad (n = r)$$
   $$y_n = n!a^n \left[\begin{array}{l} \because (ax+b)^0 = 1 \\ \text{and } 0! = 1 \end{array}\right]$$

4. Let $y = \frac{1}{(ax+b)=(ax+b)^{-1}}$.
Then,

$$
\begin{aligned}
y_1 &= (-1) \cdot a \cdot (ax+b)^{-2} \\
y_2 &= (-1)(-2) \cdot a^2 \cdot (ax+b)^{-3} \\
y_3 &= (-1)(-2)(-3) \cdot a^3 \cdot (ax+b)^{-4} \\
&\vdots \\
y_n &= (-1)^n n! a^n (ax+b)^{-(n+1)} \\
y_n &= \frac{(-1)^n n! a^n}{(ax+b)^{n+1}}
\end{aligned}
$$

5. Let $y = \log\,(ax+b)$.
Then,

$$y_n = \frac{(-1)^{n-1}(n-1)!a^n}{(ax+b)^n}$$

6. Let $y = \mathrm{e}^{ax}$.
Then,

$$y_n = \mathrm{e}^{ax} \cdot a^n$$

$$y_n = a^n \mathrm{e}^{ax}$$

7. Let $y = a^{kx}$
Then, $y_n = a^{kx} \cdot k^n (\log_e a)^k$

$y_n = k^n \cdot a^{kx} \cdot (\log_e a)^{k}$[1]

8. If $y = \sin(ax+b)$, $y_n = a^n \sin\left(ax + b + n \cdot \frac{\pi}{2}\right)$

If $y = \sin ax$, $y_n = a^n \sin\left(ax + n \cdot \frac{\pi}{2}\right)$

If $y = \sin x$, $y_n = \sin\left(x + n \cdot \frac{\pi}{2}\right)$

9. If $y = \cos(ax+b)$, $y_n = a^n \cos\left(ax + b + n \cdot \frac{\pi}{2}\right)$

If $y = \cos ax$, $y_n = a^n \cos\left(ax + n \cdot \frac{\pi}{2}\right)$

If $y = \cos x$, $y_n = \sin\left(x + n \cdot \frac{\pi}{2}\right)$

[1] Examples: If

(i) $y = 7^{3x}$, $y_n = 7^{3x} \cdot (3)^n \cdot (\log_e 7)^3$

(ii) $y = 5^{11x}$, $y_n = 5^{11x}(11)^n(\log_e 5)^{11}$

(iii) $y = \mathrm{e}^{7x}$, $y_n = \mathrm{e}^{7x} \cdot (7)^n \cdot (\log_e \mathrm{e})^7$
$= \mathrm{e}^{7x} \cdot (7)^n (\because \log_e \mathrm{e} = 1)$

17.2 DERIVATIVES OF HIGHER ORDERS: IMPLICIT FUNCTIONS

If y is an implicit function, its higher derivatives are found by differentiating the required number of times the equation connecting x and y, bearing in mind that y and all its derivatives are functions of the independent variable x.

For example, the second derivative of the function y specified by the equation

$$x^2 + y^2 = 1 \tag{1}$$

is found by differentiating equation (i) twice. We get $2x + 2yy' = 0$ or

$$x + yy' = 0 \tag{2}$$

and $(x') + yy'' + y'y' = 0$ or

$$1 + (y')^2 + yy'' = 0 \tag{3}$$

But $y' = -(x/y)$ and $y'' = -(1 + (y')^2)/y$.

$$\therefore\ y'' = -\frac{1 + (-(x/y))^2}{y} = -\frac{x^2 + y^2}{y^3} = -\frac{1}{y^3} \quad \text{[using (1)]}$$

17.3 DERIVATIVES OF HIGHER ORDERS: PARAMETRIC FUNCTIONS

In order to find a derivatived of higher orderd of a function specified by parametric equations, we differentiate the expression of the preceding derivative considering it as a composite function of the independent variable.

Let $x = \phi(t)$ and $y = f(t)$. Then, we have,

$$\frac{dy}{dx} = \frac{f'(t)}{\phi'(t)} = \frac{(dy/dt)}{(dx/dt)}$$

where $dx/dt \neq 0$. Also, the function $x = \phi(t)$ has an *inverse function* $t = \phi^{-1}(x)$.

Furthermore,

$$y'' = \frac{d}{dx}\left[\frac{f'(t)}{\phi'(t)}\right] = \frac{d}{dt}\left[\frac{f'(t)}{\phi'(t)}\right] \cdot \frac{dt}{dx} \quad \left(\text{using the property, } \frac{dy}{dx} = \frac{dy}{dt} \cdot \frac{dt}{dx}\right)$$

$$= \frac{\phi'(t)f''(t) - f'(t) \cdot \phi''(t)}{\left[\phi'(t)\right]^2} \cdot \frac{dt}{dx}$$

From the inverse function $t = \phi^{-1}(x)$, we obtain

$$\frac{dt}{dx} = \frac{1}{\phi'(x)}$$

and arrive at the expression

$$y'' = \frac{\phi'(t)f''(t) - f'(t)\cdot\phi''(t)}{[\phi'(t)]^3}$$

The differentiation of the last relation with respect to x leads to the expression for the third derivative, and so on.

Example (3): Let us find the derivatives y' and y'' of the function specified by the equations $x = a\cos t$ and $y = b\sin t$.

Solution: On differentiating, we obtain

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -a\sin t, \quad \frac{\mathrm{d}^2x}{\mathrm{d}^2t} = -a\cos t$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = b\cos t, \quad \frac{\mathrm{d}^2y}{\mathrm{d}^2t} = -b\sin t$$

$$\therefore\ y' = \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{b\cos t}{-a\sin t} = -\frac{b}{a}\cot t \quad \left[\because\ \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}t}\cdot\frac{\mathrm{d}t}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}t}\Big/\frac{\mathrm{d}x}{\mathrm{d}t}\right]$$

$$y'' = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right) = \frac{(-a\sin t)(-b\sin t) - (b\cos t)(-a\cos t)}{(-a\sin t)^2}\cdot\left(\frac{\mathrm{d}t}{\mathrm{d}x}\right)$$

$$= \frac{ab\sin^2 t + ab\cos^2 t}{a^2\sin^2 t}\cdot\frac{1}{(-a\sin t)} = -\frac{b}{a^2\sin^3 t}$$

$$\therefore\ y' = -\frac{b}{a}\cot t$$

$$y'' = -\frac{b}{a^2\sin^3 t} \quad \textbf{Ans}.$$

17.4 DERIVATIVES OF HIGHER ORDERS: PRODUCT OF TWO FUNCTIONS (LEIBNIZ FORMULA)

It helps us to find the nth derivative of the product of two functions. Let $u(x)$ and $v(x)$ be functions of x, possessing derivatives of nth order, and $y = u\cdot v$. Then,

$$y_n = (uv)_n = {}^nC_0 u_n v_0 + {}^nC_1 u_{n-1} v_1 + {}^nC_2 u_{n-2} v_2 + \cdots + {}^nC_r u_{n-r} v_r + \cdots + {}^nC_n u_0 v_n$$

where,

$${}^nC_r = \frac{n!}{(n-r)!r!}$$

This formula can be formally obtained if we take *Newton's binomial formula* for the expansion of $(u+v)^n$ and then replace the powers of u and v by the derivatives of the corresponding orders of u and v (and put $u_0 = u$, $v_0 = v$). Here, we do not present the general proof of this formula and confine ourselves to considering some examples of its application.[2]

Note (3): When one of the functions in the above theorem is of the form x^m, then we should choose it as (the second function) v, and the other as (the first function) u, because x^m shall have only m derivatives (and not more).

Note (4): From the expression for ${}^nC_r$, we get

$$
\begin{aligned}
{}^nC_1 &= n \\
{}^nC_2 &= \frac{n(n-1)}{2!} = \frac{n(n-1)}{1\cdot 2} \\
{}^nC_3 &= \frac{n(n-1)(n-2)}{3!} = \frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}, \text{ and so on}
\end{aligned}
$$

Example (4): If $y = \mathrm{e}^{ax}x^2$, find y_n.

Solution:

$$
\begin{aligned}
u_0 &= \mathrm{e}^{ax}, & v_0 &= x^2 \\
u_1 &= a\mathrm{e}^{ax}, & v_1 &= 2x \\
u_2 &= a^2\mathrm{e}^{ax}, & v_2 &= 2 \\
u_n &= a^n\mathrm{e}^{ax}, & v_3 &= 0 = v_4 = v_5 = \cdots
\end{aligned}
$$

$$y_n = a^n\mathrm{e}^{ax}x^2 + na^{n-1}\mathrm{e}^{ax}2x + \frac{n(n-1)}{1\cdot 2}a^{n-2}\mathrm{e}^{ax}\cdot 2$$

$$\text{or } y_n = \mathrm{e}^{ax}[a^nx^2 + 2na^{n-1}x + n(n-1)a^{n-2}] \quad \textbf{Ans.}$$

Example (5): Let us compute the 100th derivative of the function $y = x^2 \sin x$.

We have

$$
\begin{aligned}
y_{100} &= (\sin x\cdot x^2)_{100} \\
&= (\sin x)_{100}\cdot x^2 + {}^{100}C_1(\sin x)_{99}(2x) + {}^{100}C_2(\sin x)_{98}(2) \\
&= (\sin x)_{100}\cdot x^2 + 200x(\sin x)_{99} + \frac{100\cdot 99}{2}(\sin x)_{98}(2)
\end{aligned}
$$

All the subsequent terms are omitted here since they are identically equal to zero. Consequently,

$$
\begin{aligned}
y_{100} &= x^2\sin\left(x + 100\frac{\pi}{2}\right) + 200x\sin\left(x + 99\frac{\pi}{2}\right) + 9900\sin\left(x + 98\frac{\pi}{2}\right)^{(3)} \\
&= x^2\sin x - 200x\cos x - 9900\sin x \quad \textbf{Ans.}
\end{aligned}
$$

[2] A rigorous proof of Leibniz formula may be carried out by the method of complete mathematical induction [i.e., by proving that if this formula holds for nth order, it will also hold for the order $(n+1)$].

[3] We know that $\sin[x + (2n)\cdot(\pi/2)] = \sin x$ and $\sin[x + (2n+1)\cdot(\pi/2)] = \cos x$.

Example (6): Differentiate n times the equation

$$x^2\frac{d^2y}{dx^2}+x\frac{dy}{dx}+y=0$$

Here, each term is differentiated n times.

Solution: $D^n(y_2x^2)=y_{n+2}x^2+{}^nC_1y_{n+1}(2x)+{}^nC_2y_n(2)$

$$D^n(y_2x^2)=x^2y_{n+2}+n(2x)y_{n+1}+\frac{n(n-1)}{2!}(2)\cdot y_n \tag{4}$$

$$D^n(y_1x)=x\cdot y_{n+1}+n\cdot y_n \tag{5}$$

$$D^n(y)=+y_n \tag{6}$$

Adding (4), (5), and (6), we get

$$0=x^2\cdot y_{n+2}+(2n+1)\times y_{n+1}+[n(n-1)+n+1]y_n$$
$$0=x^2\cdot y_{n+2}+(2n+1)\times y_{n+1}+[n^2+1]y_n$$

Example (7): If $y=\sin(m\sin^{-1}x)$, then prove $(1-x^2)-xy_1+m^2y=0$ and deduce that

$$(1-x^2)y_{n+2}-(2n+1)\times y_{n+1}-(n^2-m^2)y_n=0$$

Solution: We have $y=\sin(m\sin^{-1}x)$ (7)

$$y_1=\cos(m\sin^{-1}x)\cdot\frac{m}{\sqrt{(1-x^2)}}$$

or $\sqrt{1-x^2}\cdot y_1=m\cos(m\sin^{-1}x)$
or $(1-x^2)\cdot y_1^2=m^2\cos^2(m\sin^{-1}x)$ (on squaring both the sides)
$=m^2[1-\sin^2(m\cdot\sin^{-1}x)]$ [since $\cos^2\theta=1-\sin^2\theta$]
$=m^2[1-y^2]$

$$\therefore 1-x^2\cdot y_1^2=m^2\,[1-y^2] \tag{18}$$

Differentiating both the sides of (18) with respect to x,

$$(1-x^2)2y_1y_2+y_1^2(-2x)=m^2[-2yy_1]$$

or

$$(1-x^2)y_2(2y_1)-x(2y_1^2)=-m^2\cdot y(2y_1)$$

Canceling the factor $(2y_1)$ from both the sides, we get

$$(1-x^2)y_2 - xy_1 + m^2y = 0 \quad (9)\ \text{(Proved)}$$

Now, in order to prove the second relation, we shall differentiate each term of equation (9) n times by Leibniz theorem.

$$D^n(1-x^2)y_2 = y_{n+2}(1-x^2) + ny_{n+1}(-2x) + \frac{n(n-1)}{2!}y_n(-2)$$

$$= y_{n+2}(1-x^2) - ny_{n+1}(2x) - y_n \cdot n(n-1) \quad (10)$$

$$D^n(-xy_1) = y_{n+1}(-x) + {}^nC_1 y_n(-1)$$
$$= -xy_{n+1} - ny_n \quad (11)$$

$$D^n(m^2y) = m^2y_n \quad (12)$$

Adding (10), (11), and (12), we get

$$0 = (1-x^2)y_{n+2} + (-2xn - x)y_{n+1} + [-n(n-1) - n + m^2]y_n$$

$$\text{or } (1-x^2)y_{n+2} - (2n+1)y_{n+1} - (n^2 - m^2)y_n = 0 \quad (4)\ \text{(Proved)}$$

The following results can be easily proved:

If $y = e^{ax} \cdot \sin bx$, then,

$$y_n = (a^2+b^2)^{n/2}\sin(bx + n\alpha)\cdot e^{ax}$$

where $\alpha = \tan^{-1}(b/a)$.

If $y = e^{ax}\sin(bx+c)$, then

$$y_n = (a^2+b^2)^{n/2}\sin(bx + c + n\alpha)\cdot e^{ax}$$

where, $\alpha = \tan^{-1}(b/a)$.

In particular, if $y = e^x \sin x$ (here $a=1$, $b=1$, $c=0$), then

$$y_n = (1^2+1^2)^{n/2}\sin(x+n\alpha)\cdot e^x$$

$$= 2^{n/2}\cdot e^x \cdot \sin(x + n\tan^{-1} 1/1)$$

$$= 2^{n/2}\cdot e^x \cdot \sin(x + n\cdot \pi/4)$$

Similarly, if $y = e^x \cos x$, then

$$y_n = 2^{n/2}\cdot e^x \cdot \cos(x + n\cdot \pi/4)$$

Now, if $y = e^{2x} \cdot \sin x$ (here $a = 2$, $b = 1$, $c = 0$), then

$$\begin{aligned} y_n &= (2^2 + 1^2)^{n/2} \sin(x + n\alpha) \\ &= 5^{n/2} \sin\left(x + n \tan^{-1} \frac{1}{2}\right) \end{aligned}$$

If $y = e^{ax} \cdot \cos bx$, then

$$y_n = (a^2 + b^2)^{n/2} \cos(bx + n\alpha) e^{ax}$$

where $\alpha = \tan^{-1}(b/a)$.

If $y = e^{ax} \cos(bx + c)$, then

$$y_n = (a^2 + b^2)^{n/2} \cdot \cos(bx + c + n\alpha) \cdot e^{ax}$$

where $\alpha = \tan^{-1}(b/a)$.

Note (5): The following material is given here to satisfy the natural curiosity about differentials of higher orders. The reader may find it useful later on.

17.5 DIFFERENTIALS OF HIGHER ORDERS

Consider a function $y = f(x)$, where x is the independent variable. The differential of this function is denoted by

$$dy = f'(x)dx$$

which depends on *two arguments*, namely, *the independent variable x and its differential* dx. Here, it is important to remember that the differential dx of the independent variable x is a magnitude independent of x: for any given value of x, the value of dx can be chosen quite arbitrarily.[(4)]

It means that dy must be looked upon as a function of x alone and that we have the right to speak of the differential of this function. The differential of the differential of a function, that is, $d[df(x)]$, is called the second differential (or the differential of the second order) of the function $f(x)$ and denoted by

$$\begin{aligned} &d^2y : \\ &d^2y = d(dy) \end{aligned}$$

By virtue of the general definition of a differential, we have,

$$d^2y = [f'(x)dx]'dx$$

[(4)] In other words, the differential $f'(x)dx$ is a function of x, but only the first factor [i.e., $f'(x)$] can depend on x. The second factor dx is an increment of the independent variable x and y is independent of the value of the variable dx.

which is a function of x (for an arbitrary but fixed value of dx independent of x). Since dx is independent of x, dx is taken outside the sign of the derivative upon differentiation, and we get

$$d^2y = f''(x)(dx)^2$$

Note (6): When writing the degree of differential, it is common to drop the brackets; in place of $(dx)^2$, we write dx^2, and so on.

$$\therefore\ d^2y = f''(x)dx^2$$

Note (7): To unify the terminology, we call the differential d$f(x)$ $[= f'(x)dx]$ of the function $f(x)$ the differential of the first order (or the first differential).

Similarly, the third differential (or the third-order differential) of a function is the differential of its second differential.

$$d^3y = d(d^2y) = [f''(x)dx^2]dx = f'''(x)dx^3$$

Analogously, for the differential of the nth order, we arrive at the formula

$$d^ny = f^{(n)}(x)dx^n$$

where dx^n is the nth power of dx. Thus, the differential of nth order is equal to the product of the nth derivative with respect to the independent variable by the nth power of the differential of the independent variable.

We have seen (in Chapter 16) that, if $y = f(x)$, then $dy = f'(x)dx$ irrespective of whether the argument x is an independent variable or a function of another argument. [Recall that if $y = f(u)$, where $u = \phi(x)$, then $dy = f'(u)\phi'(x)dx = f'(u)du$]. In the general case, this property (i.e., the invariance property of the first derivative) is not possessed by the differentials of higher orders.

Indeed, suppose that x is no longer an independent variable as before, but a function of a new independent variable t [i.e., $x = \phi(t)$]. Then, dx also becomes function of t, and therefore it is not allowable to regard dx as a constant when the first differential is differentiated. This leads to a new expression of d^2y different from the one above. Computing the differential of dy by applying the differentiation rule for a product, we find

$$d^2y = d[f'(x)dx]$$

Now, treating $f'(x)dx$ as a product of functions, we get,

$$\begin{aligned} d^2y &= d[f'(x)]dx + f'(x)d(dx) \\ &= [f''(x)]dx + f'(x)d^2x \\ &= f''(x)dx^2 + f'(x)d^2x^{(5)} \end{aligned}$$

Observe that there appears the additional term $f'(x)d^2x$. If x is an independent variable, the first term is retained but the second one vanishes, since,

$$d^2x = (x)''dx^2 = 0 \cdot dx^2 = 0$$

(5) We must distinguish between the terms dx^2 and d^2x:

$$dx^2 = (dx)(dx); \quad d^2x = (x)''dx^2 = 0 \cdot dx^2$$

The expression for the third differential in the case when x depends on t is still more complicated. Thus, when finding a higher order differential, we take into account the nature of the function and distinguish between the cases when it is an independent variable or depends on some other variable.(6)

17.6 RATE OF CHANGE OF A FUNCTION AND RELATED RATES

In Chapter 9, we have discussed at length the concept of rate of change of a function $y\ [= f(x)]$ with respect to the independent variable x and invented the definition of derivative of a function. There, we have clarified the distinction between the average rate of change and the instantaneous rate (or the actual rate) of change of a function. Also, we were convinced through examples that in certain situations, the *instantaneous rate* of change is more significant than the *average rate* of change of a function.

In calculus, we are fundamentally concerned with the actual rate of change of a function with respect to the change in the variable on which it depends.

Furthermore, if both x and y are varying with t (i.e., both x and y are functions of t), then

$$\begin{aligned}\frac{dy}{dt} &= \frac{dy}{dx}\cdot\frac{dx}{dt} \quad \text{(by chain rule)}\\ &= f'(x)\cdot\frac{dx}{dt}\end{aligned}$$

Thus, the rate of change of one variable can be calculated if the rate of change of the other (related) variable is known.

For example, when a spherical balloon is inflated, its radius r, volume v, and surface area s grow simultaneously with time t. Thus, r, v, and s are all functions of t, but each of them could also be considered as a function of any one of the remaining variables, since all of them are interrelated. One might be interested in computing the following:

$\frac{dr}{dt}$ = rate of increase of radius per unit increase in time (at the instant when say $r = 8$ cm)

$\frac{ds}{dt}$ = rate of increase of surface area per unit increase in time (at the instant when say $r = 6$ cm)

$\frac{dv}{dt}$ = rate of increase of volume per unit increase in time (at the instant when say $r = 10$ cm)

$\frac{ds}{dr}$ = rate of increase of surface area per unit increase in radius (at the instant when say $r = 25$ cm)

$\frac{dv}{dr}$ = rate of increase of volume per unit increase in radius (at the instant when say $r = 25$ cm)

(6) *Mathematical Analysis* (English translation) by A.F. Bermant and I.G. Aramanovich (pp. 155–173), Mir Publishers, Moscow, 1975.

The d-notation helps us remember which rate of change we are interested in. In many rate-of-change problems, we can find the time rate of change of a quantity Q if we know the time rate of change of one or more related quantities.

Let us consider some examples.

Example (8): If a spherical balloon is inflated at the rate of $10\,\text{cm}^3/\text{s}$, how fast is the radius of the balloon increasing when the radius is 5 cm.

Solution: Let $V =$ volume of the (spherical) balloon.

$$\frac{\mathrm{d}v}{\mathrm{d}t} = 10\,\text{cm}^3/\text{s} \tag{13}$$

The (geometrical) relation connecting the variable is

$$v = \frac{4}{3}\pi r^3$$

$$\therefore \frac{\mathrm{d}v}{\mathrm{d}r} = 4\pi r^2 \tag{14}$$

To compute $(\mathrm{d}r/\mathrm{d}t)$ at $r = 5$.

How do we compute this?

To compute the desired rate, we may consider either the rate whose value is known (here it is $\mathrm{d}v/\mathrm{d}t$) or the rate that is obtained from the relation connecting the variables (here it is $\mathrm{d}v/\mathrm{d}r$). Furthermore, to compute the desired rate, we write

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}v}{\mathrm{d}r} \cdot \frac{\mathrm{d}r}{\mathrm{d}t} \quad \text{(by chain rule)}$$

Observe that $\mathrm{d}r/\mathrm{d}t$ appears in this relation, so that we get

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\mathrm{d}v/\mathrm{d}t}{\mathrm{d}v/\mathrm{d}r}$$

or

$$\frac{\mathrm{d}v}{\mathrm{d}r} = \frac{\mathrm{d}v}{\mathrm{d}t} \cdot \frac{\mathrm{d}t}{\mathrm{d}r} \quad \text{(by chain rule)}$$

Observe that, in this relation, $\mathrm{d}r/\mathrm{d}t$ does not appear, but $\mathrm{d}t/\mathrm{d}r$ appears. However, in view of this definition of derivative as a ratio of differentials, we can write

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\mathrm{d}v/\mathrm{d}t}{\mathrm{d}v/\mathrm{d}r}$$

which is the same expression as obtained above. Now we get

$$\frac{dr}{dt} = \frac{10}{4\pi r^2}$$

$$\therefore \left.\frac{dr}{dt}\right|_{r=5} = \left.\frac{10}{4\pi r^2}\right|_{r=5} = \frac{10}{4\pi(5)^2} = \frac{1}{10\pi}$$

Therefore, when the radius is 5 cm, the radius is increasing at the rate of $\frac{1}{10\pi}$ cm/s Ans.

Note (8): It is necessary to write down the units of the rate computed.

Remark: Note that for radius = 10 cm, dr/dt = 1/40π. It may be observed here that the balloon is inflated at a constant rate (10 cm^3/s), but the rate at which its radius increases is not constant. In fact, dr/dt keeps on decreasing with time (why?).

In the related rate problems, all the variables are interrelated and so are their time rates. If two or more equations connect the variables involved, then we can compute the desired rates (e.g., in this case, dv/dr, ds/dr, and dv/ds) by obtaining the required rate(s) from the right equation.

The important step is to connect the available rate(s) suitably so that the desired rate gets into the relation. Then, by using the available data and the derived data, we can easily compute the desired rate.

Example (9): If v denotes the volume of a sphere and s its surface area, find the rate of change of v with respect to s, when the radius of the sphere is 2 cm.

Solution: Let r = radius (of the sphere),

$$v = \text{volume of the sphere} = \frac{4}{3}\pi r^3 \tag{15}$$

$$\text{and } s = \text{surface area of the sphere} = 4\pi r^2 \tag{16}$$

To find dv/ds, when r = 2 cm.

We have,

$$\frac{dv}{dr} = 4\pi r^2 \quad [\text{from (15)}]$$

$$\text{and } \frac{ds}{dr} = 8\pi r \quad [\text{from (16)}]$$

$$\text{Now, } \frac{dv}{ds} = \frac{dv/dr}{ds/dr} = \frac{4\pi r^2}{8\pi r} = \frac{r}{2}$$

$$\therefore \left.\frac{dv}{ds}\right|_{\text{at } r=2} = \left.\frac{r}{2}\right|_{\text{at } r=2} = \frac{2}{2} = 1\ \text{cm}^3/\text{cm}^2 \qquad \textbf{Ans.}$$

Example (10): An edge of a variable cube is increasing at the rate of 3 cm/s. How fast is the volume of the cube increasing when the edge is 10 cm long?

Solution: Let edge of the (variable) cube = x cm.

$$\therefore \text{ Volume of the cube } v = x^3 \tag{17}$$

Given dx/dt = 3 cm/s.
To find

$$\frac{dv}{dt}, \text{ when } x = 10\,\text{cm} \tag{18}$$

From (17), we easily get

$$\frac{dv}{dx} = 3x^2 \tag{19}$$

Now,

$$\frac{dv}{dx} = \frac{dv}{dt} \cdot \frac{dt}{dx} = \frac{dv/dt}{dx/dt}$$

$$\therefore \frac{dv}{dt} = \frac{dv}{dx} \cdot \frac{dx}{dt}$$

$$\therefore \frac{dv}{dt} = (3x^2) \cdot (3)$$

$$\therefore \left.\frac{dv}{dt}\right|_{\text{at } x=10} = 3(10)^2 \cdot 3 \quad [\text{by using (18) and (19)}]$$

$$= 900\,\text{cm}^3/\text{s} \qquad \textbf{Ans}.$$

Example (11): The radius of a spherical balloon increases at the rate of 4 cm/s. Find the rate at which its volume increases when its radius is 5 cm.

Solution: Let r = radius (of the spherical balloon) and

$$v = \text{volume} = \frac{4}{3}\pi r^3 \tag{20}$$

Also,

$$\frac{dr}{dt} = 4\,\text{cm/s} \tag{21}$$

To find dv/dt when $r = 5$.

Note that from (20), we easily get the following rate:

$$\frac{dv}{dr} = 4\pi r^2 \tag{22}$$

Now, to get dv/dt, we write

$$\frac{\mathrm{d}v}{\mathrm{d}r} = \frac{\mathrm{d}v}{\mathrm{d}t} \cdot \frac{\mathrm{d}t}{\mathrm{d}r} = \frac{\mathrm{d}v/\mathrm{d}t}{\mathrm{d}r/\mathrm{d}t}$$

$$\therefore \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}v}{\mathrm{d}r} \cdot \frac{\mathrm{d}r}{\mathrm{d}t} = (4\pi r^2) \cdot (4)$$

$$\therefore \left.\frac{\mathrm{d}v}{\mathrm{d}t}\right|_{r=5} = 4\pi(5)^2 \cdot (4) = 400\pi \text{ cm}^3/\text{s} \qquad \textbf{Ans.}$$

Example (12): A stone is dropped into a quiet lake and waves move in circles at a speed of 4 cm/s. At the instant when the radius of the circular wave is 10 cm, how fast is the enclosed area increasing?

Solution: Let r = radius of a circle and A = area of the circle.

$$A = \pi r^2 \tag{23}$$

Given

$$\frac{\mathrm{d}r}{\mathrm{d}t} = 4 \text{ cm/s} \tag{24}$$

To find dA/dt when $r = 10$ cm.

From (23), we get

$$\frac{\mathrm{d}A}{\mathrm{d}r} = 2\pi r \tag{25}$$

We write

$$\frac{\mathrm{d}A}{\mathrm{d}r} = \frac{\mathrm{d}A}{\mathrm{d}t} \cdot \frac{\mathrm{d}t}{\mathrm{d}r}$$

$$\therefore \frac{\mathrm{d}A}{\mathrm{d}t} = \frac{\mathrm{d}A}{\mathrm{d}r} \cdot \frac{\mathrm{d}r}{\mathrm{d}t}$$

$$\therefore \frac{\mathrm{d}A}{\mathrm{d}t} = (2\pi r) \cdot (4) \cdot \left[\begin{array}{l} \because \dfrac{\mathrm{d}A}{\mathrm{d}r} = 2\pi r \\ \dfrac{\mathrm{d}r}{\mathrm{d}t} = 4 \text{ (given)} \end{array}\right.$$

$$\text{or } \frac{\mathrm{d}A}{\mathrm{d}t} = 8\pi r$$

$$\therefore \left.\frac{\mathrm{d}A}{\mathrm{d}t}\right|_{r=10} = 2\pi(10) \cdot (4) = 80\pi \text{ cm}^2/\text{s} \qquad \textbf{Ans.}^{(7)}$$

The enclosed area is increasing at the rate of 80π cm^2/s when $r = 10$ cm.

(7) We solve this equation for dA/dt, using the fact that dt/dr = 1/(dr/dt).

Note (9): At $r = 15$ cm, $\mathrm{d}A/\mathrm{d}t = 120\pi\ \mathrm{cm}^2/\mathrm{s}$ and at $r = 5$ cm, $\mathrm{d}A/\mathrm{d}t = 40\pi\ \mathrm{cm}^2/\mathrm{s}$.

Example (13): If the volume of a sphere increases at the rate of $25\ \mathrm{cm}^3/\mathrm{s}$, find the rate of increase of its surface area at the instant when its radius is 10 cm.

Solution: Let $r =$ radius of the (changing) sphere at any instant.

$$\therefore \text{ Volume of sphere } v = \frac{4}{3}\pi r^3 \tag{26}$$

and

$$\text{surface area } s = 4\pi r^2 \tag{27}$$

Given

$$\frac{\mathrm{d}v}{\mathrm{d}t} = 25\ \mathrm{cm}^3/\mathrm{s} \tag{28}$$

To find $\mathrm{d}s/\mathrm{d}t$ when $r = 10$ cm.
From (26) and (27), we get

$$\frac{\mathrm{d}v}{\mathrm{d}r} = 4\pi r^2 \quad [\text{from } (26)] \tag{29}$$

$$\text{and} \quad \frac{\mathrm{d}s}{\mathrm{d}r} = 8\pi r \quad [\text{from } (27)] \tag{30}$$

To compute $\mathrm{d}s/\mathrm{d}t$, we write

$$\frac{\mathrm{d}s}{\mathrm{d}r} = \frac{\mathrm{d}s}{\mathrm{d}t} \cdot \frac{\mathrm{d}t}{\mathrm{d}r}$$

$$\therefore \frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\mathrm{d}s}{\mathrm{d}r} \cdot \frac{\mathrm{d}r}{\mathrm{d}t}$$

$$\therefore \frac{\mathrm{d}s}{\mathrm{d}t} = 8\pi r \cdot \frac{\mathrm{d}r}{\mathrm{d}t} \tag{31}$$

Now, we are required to find the value of $\mathrm{d}r/\mathrm{d}t$, which we can find by using only (26) (why?).[(8)]

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}v}{\mathrm{d}r} \cdot \frac{\mathrm{d}r}{\mathrm{d}t} = 4\pi r^2 \cdot \frac{\mathrm{d}r}{\mathrm{d}t}$$

$$\therefore 25 = 4\pi r^2 \cdot \frac{\mathrm{d}r}{\mathrm{d}t} \tag{32}$$

[(8)] It is given that $(\mathrm{d}v/\mathrm{d}t) = 25\ \mathrm{cm}^3/\mathrm{s}$. Now $(\mathrm{d}v/\mathrm{d}t) = (\mathrm{d}v/\mathrm{d}r)\cdot(\mathrm{d}r/\mathrm{d}t)$ [where $(\mathrm{d}v/\mathrm{d}r) = 4\pi r^2$].

$$\frac{dr}{dt} = \frac{25}{4\pi r^2}$$

$$\therefore \frac{ds}{dt} = \frac{2}{r} \cdot (25) = \frac{50}{r} \text{ cm}^2/\text{s}$$

$$\left.\frac{ds}{dt}\right|_{r=10} = 5 \text{ cm}^2/\text{s} \qquad \textbf{Ans.}$$

Example (14): Sand is pouring from a pipe at the rate of $10\,\text{m}^3/\text{s}$. The falling sand forms a cone on the ground in such a way that the height of the cone is always twice the radius of the base. Find the rate at which the height of the sand cone is increasing when sand in the pile is 8 m high.

Solution: Given

$$\frac{dv}{dt} = 10\,\text{m}^3/\text{s} \tag{33}$$

$$\text{Height of the sand cone } h = 2r \text{ (always)} \tag{34}$$

$$\begin{aligned}\text{Volume of cone } v &= \frac{1}{3}\pi r^2 h = \frac{1}{3}\pi\left(\frac{h}{2}\right)^2 \cdot h \\ &= \frac{1}{12}\pi h^3\end{aligned} \tag{35}$$

To find dh/dt when $h = 8$ m.
From (35), we get

$$\frac{dv}{dh} = \frac{3}{12}\pi h^2 = \frac{1}{4}\pi h^2 \tag{36}$$

Note that, the value of dv/dt is given at (33). Hence, we express this rate in a way such that the desired rate dh/dt gets involved in the relation.
We write,

$$\frac{dv}{dt} = \frac{dv}{dh} \cdot \frac{dh}{dt} = \left(\frac{1}{4}\pi h^2\right) \cdot \frac{dh}{dt} \qquad \left[\because \frac{dv}{dh} = \frac{1}{4}\pi h^2\right]$$

$$\text{or } 10 = \frac{1}{4}\pi h^2 \cdot \frac{dh}{dt} \qquad \left[\because \frac{dv}{dt} = 10\,\text{m}^3/\text{s}\right]$$

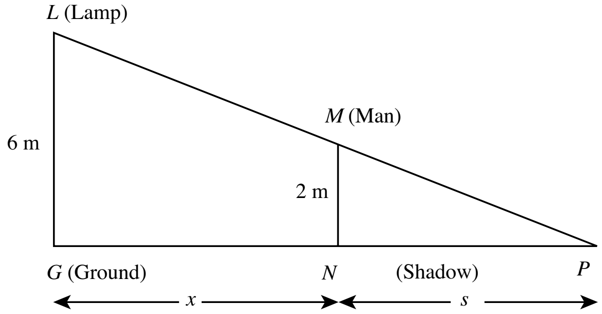
$$\therefore \frac{dh}{dt} = \frac{40}{\pi h^2}$$

$$\therefore \left.\frac{dh}{dt}\right|_{\text{at } h=8} = \frac{40}{\pi(8)^2} = \frac{5}{8\pi} \text{ m/s} \qquad \textbf{Ans.}$$

The units of the final result must be mentioned carefully.

Example (15): A man of height 2 m walks on a level road at a (uniform) speed of 5 km/h, away from a lamppost 6 m high. Find the rate at which the length of his shadow is increasing.

Solution: The figure given below reflects the situation stated in the problem.



Let $x =$ distance between the lamppost and the man at any given instant t.

Then, $s =$ length of the shadow of the man at the instant t.

Here, it is important to note that the length s of the shadow is related to the distance x from the lamppost. Hence, we must express the length s in terms of the length x. Also, note that we have $\mathrm{d}x/\mathrm{d}t = 5$ km/h and we have to find $\mathrm{d}s/\mathrm{d}t$. Since the triangles GPL and NPM are similar, we have

$$\frac{NP}{GP} = \frac{NM}{GL}$$

$$\text{or} \quad \frac{s}{x+s} = \frac{2}{6} = \frac{1}{3}$$

$$\text{or} \quad 3s = x + s \quad \text{or} \quad 2s = x$$

$$\therefore 2\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\mathrm{d}x}{\mathrm{d}t}$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{5}{2} \quad \left[\because \frac{\mathrm{d}x}{\mathrm{d}t} = 5\right]$$

Thus, the length of the shadow increases at the rate of 2.5 km/h **Ans.**

Example (16): The height of an inverted cone is 10 cm and radius of its circular base is 5 cm. Water is poured into it at the rate of 1.5 cm^3/s. Find the rate at which the level of water in the cone is rising when the depth is 4 cm.

Solution: At any time t, let the height of water level $= h$ and radius of cone at $h = r$. We have

$$\frac{r}{h} = \frac{5}{10} = \frac{1}{2} \quad \therefore r = \frac{h}{2} \tag{37}$$

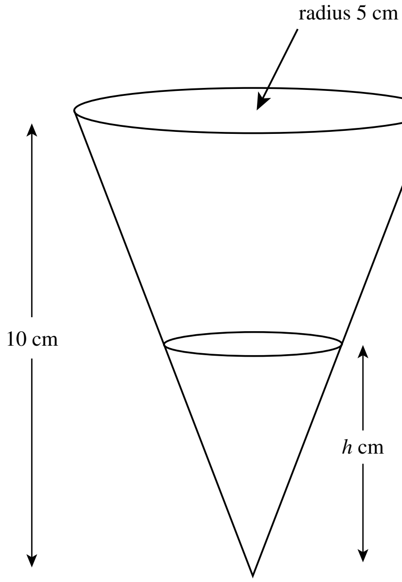
Line of approach:

(i) When water is poured into the cone at the (constant) rate of 1.5 cm^3/s, we can say that the rate of increase in volume of water in the cone is

$$(dv/dt) = 1.5 = (15/10) = (3/2)\ cm^3/s.$$

(ii) We have to find the rate at which the level of water in the cone is rising when $h = 4$ cm, that is, to find dh/dt when $h = 4$.

(iii) Let V = volume of water in the cone at any instant t.



$$\therefore\ V = \frac{1}{3}\pi r^2 h = \frac{\pi}{3}\left(\frac{h}{2}\right)h = \frac{\pi}{12}h^3$$

Since the value of

$$\frac{dV}{dt} = \frac{\pi h^2}{4}\frac{dh}{dt} \tag{38}$$

$$\therefore\ \frac{dh}{dt} = \left[\because \frac{dV}{dt} = \frac{3}{2}\right]$$

$$\therefore\ \frac{dh}{dt}\ (\text{at}\ h = 4) = 4\frac{3}{2}\cdot\frac{1}{\pi\cdot 16} = \frac{3}{8\pi}\ \text{cm/s}$$

$\therefore$ Rate of increase of water level $= 3/8\pi$ cm/s **Ans**.

Example (17): A ladder 5 m long is leaning against a wall. The bottom of the ladder is pulled along the ground away from the wall at the rate of 2 m/s. How fast is its height on the wall decreasing when the foot of the ladder is 4 m away from the wall?

Solution: At any time t, let

(i) the bottom of the ladder be at a distance x m from the wall and

(ii) the height at which the top of the ladder touches the wall be y m.
Then,

$$x^2 + y^2 = (5)^2 = 25 \tag{39}$$

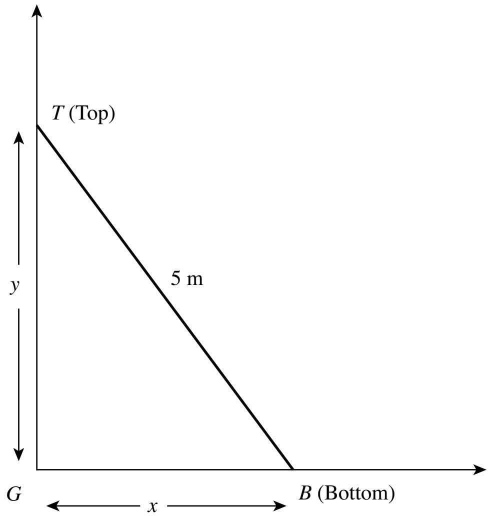


FIGURE 17.1

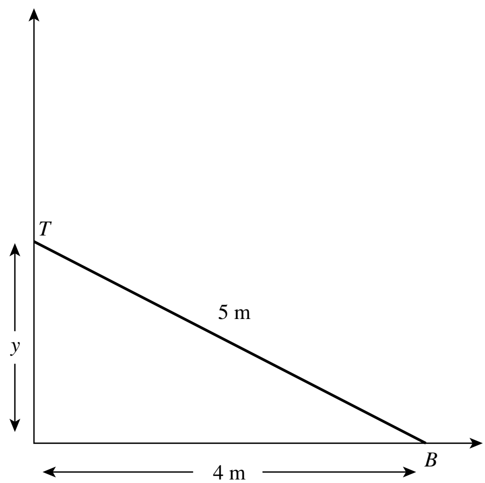


FIGURE 17.2

Note that

(a) Bottom B of the ladder is pulled away along the ground at the rate of 2 m/s.

$$\therefore \frac{dx}{dt} = 2 \tag{40}$$

(b) When the bottom B is 4 m away from the wall, we have

$$y^2 + 4^2 = 5^2$$
$$y^2 = 25 - 16 = 9$$
$$\therefore y = 3\text{ m}$$

(c) We have to find dy/dt when $y = 3$ and $x = 4$.
From (39) we have

$$2x\frac{dx}{dt} + 2y\frac{dy}{dt} = 0, \text{ or } \frac{dy}{dt} = -\frac{x}{y}\frac{dx}{dt} \left(\text{where } \frac{dx}{dt} = 2\right)$$
$$\therefore \frac{dy}{dt} = -\frac{4}{3}\cdot(2) = -\frac{8}{3}$$

Therefore, the end of the ladder comes down at the rate of $-(8/3)$ m/s.
[Negative sign in $-(8/3)$ tells that upper end of the ladder slides downward.]

Important Note (10):
In the preceding example, it is essential to draw Figure 17.1 that represents the situation at any instant. If we had tried to find the rate of slippage from Figure 17.2 that represents the situation only at a particular time t_1, then we would not have been able to obtain a relationship between the rates dx/dt and dy/dt. In particular, x does not appear in this figure. However, this figure is needed to find the value of y for the given value of x at the time t_1. These related values of x and y are then used in part c.

The procedure for solving related rate problems includes the following steps:

Step (1): Write down the available information in a convenient order.

(i) Decide what rates of change are given and express these data in Leibniz notation: $dv/dt = 50\text{ cm}^3/\text{s}$ or $dx/dt = 3$ m/s and so on.
(ii) Write down, if any geometric relation connects the variables involved. Examples: $V = (4/3)\pi r^3$, or $S = 4\pi r^2$, or $x^2 + y^2 = 64$, and so on.
(iii) Write down the derivatives of the quantities involved with respect to the relevant independent variable(s).

Example 18: $\frac{dv}{dr} = 4\pi r^2, \quad \frac{ds}{dr} = 8\pi r$

$$2x + 2y\frac{dy}{dx} = 0 \quad \therefore \frac{dy}{dx} = -\frac{x}{y}$$

(iv) Decide what rate of change is desired and express it in Leibniz notation.

To find dh/dt at $h = 5$, dy/dt when $y = 3$, and so on.

Step (2): If necessary, draw a picture and express the available details therein. Such a figure may be needed for correcting the variables involved [refer to solved examples (9) and (10)] [note down what changes and what does not].

Step (3): Express the given rate of change from step (1.i) in the form of a product of rates of change (applying the chain rule) ensuring that the desired rate of change appears in the relation.

For example, suppose we are given the value of $\mathrm{d}v/\mathrm{d}t$ and we have to compute $\mathrm{d}r/\mathrm{d}t$ when $r = 5$. Then we express

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}v}{\mathrm{d}r} \cdot \frac{\mathrm{d}r}{\mathrm{d}t} \text{ (by chain rule)} \tag{41}$$

Here, $\mathrm{d}v/\mathrm{d}t$ is given and the expression for $\mathrm{d}v/\mathrm{d}r$ is available from step (1.iii). Hence, equation (41) can be solved for $\mathrm{d}r/\mathrm{d}t$ and its value can be computed for any value of r.

Note (11): Sometimes, it may happen that we have to compute the rate $\mathrm{d}s/\mathrm{d}r$, whereas relation (41) involves the rate $\mathrm{d}r/\mathrm{d}s$, which is the reciprocal of the desired rate. In such cases, we can express it in the desired form by transferring it to the other side of the equation and solve the equation for the desired rate.

Note (12): We may also express the relation from step (1.iii) in a suitable form, (applying the chain rule) ensuring that the desired rate appears in the relation. Then, by using the available information, we can obtain the desired rate.

Exercises

(1) The edge of a cube is increasing at the rate of 5 cm/s. How fast is the volume of the cube increasing when the edge is 12 cm long?

Ans. $2160\ \mathrm{cm^3/s}$

(2) A stone is dropped into a quiet lake and waves move outwards in circles at the speed of 4 cm/s. At the instant when the radius of the circular wave is 10 cm, how fast is the enclosed area increasing?

Ans. $80\pi\ \mathrm{cm^2/s}$

(3) A man 1.8 m high walks away from a lamppost at the rate of 1.2 m/s. If the height of the lamppost is 4.5 m, find the following:

(i) The rate at which the length of his shadow increases.
Ans. 0.8 m/s

(ii) The rate at which the tip of shadow is moving.
Ans. 2 m/s

(4) Sand is poured from a pipe at the rate of $12\ \mathrm{cm^3/s}$. The falling sand forms a cone on the ground in such a way that the height of the cone is always one-sixth of the radius of base. Find how fast the height of the sand cone is increasing when the height is 4 cm.

Ans. $\frac{1}{48\pi}\ \mathrm{cm/s}$

18 Applications of Derivatives in Studying Motion in a Straight Line

18.1 INTRODUCTION

Various problems in kinematics can be solved with the use of the derivative. Let a *particle move along a straight line so that its distance s from some fixed point is a function of time.* We express this by writing,

$$s = f(t)$$

Then, the velocity $v = \mathrm{d}s/\mathrm{d}t$ and the acceleration "a" $= \mathrm{d}^2 s/\mathrm{d}t^2$.

One particular example of motion in a straight line is the motion of a falling body under gravity. The *acceleration of a falling body due to gravity* has been calculated as $g = 32$ ft/s^2 or 9.8 m/s^2, towards, the center of Earth. In this chapter, we will use *differentiation* to compute velocity and acceleration of a moving object in some practical situations.

18.2 MOTION IN A STRAIGHT LINE

Example (1): A particle is moving *in a straight line* according to the formula $s = 4t^3 + 2t^2$, where s is the distance traveled in meters and t is in seconds. Find the velocity and acceleration of the particle after 4 s.

Solution: Given,

$$s = 4t^3 + 2t^2 \tag{1}$$

$$\text{Velocity,} \quad v = \frac{\mathrm{d}s}{\mathrm{d}t} = 12t^2 + 4t \tag{2}$$

$$\text{Acceleration,} \quad a = \frac{\mathrm{d}v}{\mathrm{d}t} = 24t + 4 \tag{3}$$

(Note that v and a are both functions of t.)

Applications of derivates 18-Motion in a straight line (including motion under gravity), circular motion and angular velocity. Applications in geometry.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

Velocity, when $t = 4\,\text{s}$, is obtained from equation (2) by putting $t = 4$. We get,

$$V(4) = 12(4)^2 + 4(4) = 192 + 16 = 208\,\text{m/s}$$

Acceleration, when $t = 4\,\text{s}$, is obtained from equation (3) and by putting $t = 4$ we get

$$a(4) = 24(4) + 4 = 100\,\text{m/s}^2$$

Thus,

$$v = 208\,\text{m/s}$$

$$\text{and } a = 100\,\text{m/s}^2 \qquad \textbf{Ans}.$$

Example (2): A particle is moving in a straight line according to the formula $s = t^3 - 9t^2 + 3t + 1$, where s is measured in meters and t in seconds. When the velocity is -24 m/s, find the acceleration.

Solution: We have

$$s = t^3 - 9t^2 + 3t + 1 \tag{4}$$

$\therefore$ The velocity v is given by

$$v = \frac{\mathrm{d}s}{\mathrm{d}t} = 3t^2 - 18t + 3 \tag{5}$$

If v is equal to (-24), we have

$$\therefore \quad 3t^2 - 18t + 3 = -24$$

$$\therefore \quad 3t^2 - 18t + 27 = 0$$

$$\text{or} \quad t^2 - 6t + 9 = 0$$

$$\text{or} \quad (t-3)^2 = 0$$

$$\therefore \quad t = 3\,\text{s}$$

Thus, we get that the velocity is -24 m/s at $t = 3$ s. Now we have to find the acceleration of the particle at $t = 3$ s. The acceleration is given by,

$$a = \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right) = \frac{\mathrm{d}^2 s}{\mathrm{d}t^2} = 6t - 18 = 6(t-3)$$

$$\therefore \text{ At } t = 3\,\text{s}, \quad a(3) = 6(3-3) = 0\,\text{m/s}^2 \qquad \textbf{Ans}.$$

Example (3): The distance s in meters described by a particle in t seconds is given by $s = Ae^t + (B/e^t)$. *Show that the acceleration of the particle at time t is equal to the distance traveled by it up to time t.*

Solution: We have,

$$s = Ae^t + (B/e^t) = Ae^t + Be^{-t} \tag{6}$$

Differentiating both sides of (6) w.r.t. t,

$$\frac{ds}{dt} = Ae^t - Be^{-t} \tag{7}$$

Differentiating once again, we get

$$\text{Acceleration} = \frac{d^2s}{dt^2} = Ae^t + Be^{-t} \tag{8}$$

Comparing (6) and (8), we observe that

$$\frac{d^2s}{dt^2} = s$$

In other words, the *numerical value* of acceleration at time t is the same as the number representing the distance traveled up to the instant t.

Example (4): A particle moves in a straight line such that $s = A\cos(Kt + \theta)$. Find the velocity at any time t and *show that the acceleration "a" is proportional to s.*

Solution: We have, $s = A\cos(Kt + \theta)$.

$$\therefore\ v = \frac{ds}{dt} = -KA\sin(Kt + \theta)$$

and

$$a = \frac{dv}{dt} = -K^2A\cos(Kt + \theta) = -K^2s$$

Thus, $a \propto (-)s \quad [\because\ K^2 \text{ is a constant}]$.

Such a motion in which *the acceleration is proportional to the displacement and is directed in its opposite direction is termed simple harmonic motion.* It is a to-and-fro motion about a central point and is always directed toward the central point.

Exercise (1):

Q1 A particle is moving in a *straight line*. If the law of motion is $s = t^3 - 6t^2 + 9t - 4$, where s is measured in meters, then find

(i) its *displacement* and *acceleration* when velocity is 0 m/s.
(ii) its *displacement* and *velocity* when acceleration is 0 m/s^2.

Ans.

(i) Displacement(s) $= 0$, Acceleration $= 6\,\text{m/s}$
(ii) Displacement $= -2\,\text{m}$, Velocity $= -3\,\text{m/s}$

Q2 A particle is moving in a straight line, where its position s in meters is a function of time t in seconds, given by $s = t^3 + at^2 + bt + c$, where a, b, c are constants.

It is known that at $t = 1$ s, the position of the particle is given by $s = 7$ m, velocity is 7 m/s, and acceleration is 12 m/s^2. Find the values of a, b, c.

Ans. $a = 3$, $b = -2$, $c = 5$

18.2.1 Motion Under Gravity

Motion of a falling body under gravity is a particular instance of motion in a straight line. The acceleration of the falling body due to gravity is called the *acceleration due to gravity* and is generally denoted by "g".[(1)]

Example (5): A stone is thrown vertically upward. It moves according to the formula $s = 490t - 4.9t^2$, where s is in centimeters and t in seconds. Find the maximum height attained by the stone.

Solution: We have

$$s = 490t - 4.9t^2 \tag{9}$$

$$\therefore \quad \frac{ds}{dt} = 490 - 9.8t \tag{10}$$

At the maximum height, the velocity of the stone will be zero.

That is, $\frac{ds}{dt} = 0$

$$\text{or} \quad 490 - 9.8t = 0$$

$$\Rightarrow \quad 9.8t = 490$$

$$\Rightarrow \quad t = \frac{490}{9.8} = 50\,\text{s}$$

Hence, putting $t = 50$ s in (9), we get

$$\textit{maximum height } s = 490 \times 50 - 4.9(50)^2$$
$$= 24{,}500 - 12{,}250 = 12{,}250\,\text{m} \qquad \textbf{Ans.}$$

(1) It is useful to recall the following formulas for free fall near the Earth's surface:

(1) $s = 0.5\,gt^2$,	s = distance,	t = time,	g = gravitational constant
(2) $s = 16\,gt^2$,	s = feet,	t = seconds,	$g = 32$ ft/s^2
(3) $s = 490t^2$,	s = centimeters,	t = seconds,	$g = 980$ cm/s^2
(4) $s = 4.9t^2$,	s = meters,	t = seconds,	$g = 9.8$ m/s^2

Example (6): A ball is thrown vertically upward. The height of the ball from the ground after t seconds is h feet, given by the equation $h = 80t - 16t^2$. Find

(i) the *time interval* when it reaches the ground.
(ii) its *velocities* after (a) 1 s and (b) 3 s. Discuss about the signs of these velocities.
(iii) *the velocity* by which the ball was thrown.
(iv) *the time when the ball was just at rest.*

Solution: We have,

$$h = 80t - 16t^2 \qquad (11)$$

$$\therefore \text{ Velocity, } v = \frac{dh}{dt} = 80 - 32t \qquad (12)$$

(i) It reaches the ground where $h = 0$.

$\therefore$ From equation (11), we get

$$0 = 80t - 16t^2$$

That is, $16t\,(t - 5) = 0$

$$\therefore \; t = 0\,\text{s} \quad \text{or} \quad t = 5\,\text{s}$$

Note (1): The value $t = 0$ s *shows that initially the ball was on the ground.*

Once the ball is thrown up, it must come back on the ground after 5 s. **Ans.**

(ii) (a) When $t = 1$ s, we get from equation (12)

$$v = 80 - 32(1) = 48\,\text{ft/s} \qquad \textbf{Ans.}$$

(b) when $t = 3$ s, $v = 80 - 96 = -16$ ft/s **Ans.**

Explanation for negative sign of velocity at (b) above.

The positive sign in velocity shows that the ball is going upward, while the negative sign shows that the ball is falling down (which is the motion in the opposite direction).

(iii) Initially when the ball was thrown up, $t = 0$ s.
Therefore, by equation (12), we have

$$v = 80 - 32(0) = 80\,\text{ft/s} \qquad \textbf{Ans.}$$

(iv) When the ball just comes to rest, $v = 0$. Hence, by equation (12) we get

$$0 = 80 - 32t$$

$$\therefore \; t = \frac{80}{32} = 2.5\,\text{s} \qquad \textbf{Ans.}$$

18.3 ANGULAR VELOCITY

This is another important concept. When a particle moves along the circumference of a circle, the central angle θ, measured from some fixed direction, is a function of time t.

Definition: We define *angular velocity ω as the rate of change of θ with respect to time t* and write,

$$\omega = \frac{d\theta}{dt}$$

Likewise, angular acceleration α is denoted by

$$\alpha = \frac{d}{dt}\left(\frac{d\theta}{dt}\right) = \frac{d^2\theta}{dt^2}$$ [2]

Example (7): A particle P moves around the circumference of the circle with constant angular velocity. Find v_x, v_y, a_x and a_y.

Solution: Let the equation of the circle be given in parametric form as follows:

$$x = r\cos\theta, \quad y = r\sin\theta$$

$$\text{Then, } v_x = -r\sin\theta\frac{d\theta}{dt} = -r\omega\sin\theta \qquad \textbf{Ans.}$$

$$v_y = r\cos\theta\frac{d\theta}{dt} = r\omega\cos\theta \qquad \textbf{Ans.}$$

$$a_x = \frac{d}{dt}(v_x) = \frac{d}{dt}[-r\omega\sin\theta] = -r\omega\frac{d}{dt}(\sin\theta)$$

$$= -r\omega\cos\theta\frac{d\theta}{dt} = -r\omega^2\cos\theta \qquad \textbf{Ans.}$$

and

$$a_y = \frac{d}{dt}(r\omega\cos\theta) = -r\omega^2\sin\theta \qquad \textbf{Ans.}$$

18.4 APPLICATIONS OF DIFFERENTIATION IN GEOMETRY

(a) *Slope*: We have seen in Chapter 9 that the *slope of the tangent line* to the curve $y = f(x)$ at any point $(x, f(x))$ is given by

$$y' = f'(x) = \frac{dy}{dx} = m = \tan\alpha,$$

where α *is the angle of inclination* of the tangent line.

[2] We also speak of *angular velocity* and *angular acceleration* of a vector OP drawn from the origin O to a point P as P moves along a curve. However, we shall not discuss about it at this point.

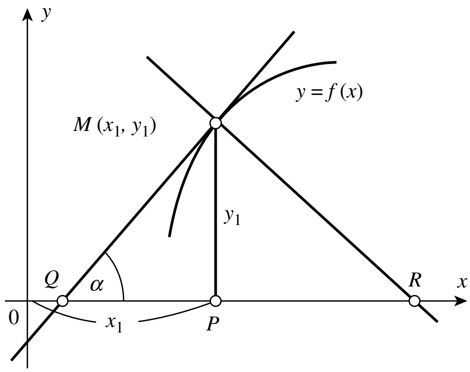


FIGURE 18.1

(b) *The equations of a tangent and of a normal to the curve*:
Let us consider a curve whose equation is

$$y = f(x)$$

On this curve, we take a point $M(x_1, y_1)$ (Figure 18.1) and write the equation of the tangent line to the given curve at the point *M*, *assuming that this tangent line is not parallel to the axis of ordinate*. We write its equation in the *point-slope form*, by expressing the *slope m of the tangent line at M*. We have

$$\frac{y - y_1}{x - x_1} = m$$

$$\therefore \quad y - y_1 = m(x - x_1)$$

This is the equation of a straight line with slope m and passing through the point M (x_1, y_1) (Figure 18.1). For the tangent line in question, $m = f'(x)$ is evaluated at the point (x_1, y_1), so that we have (the numerical value) $m = f'(x_1)$. Thus, the *equation of the tangent* at (x_1, y_1) is given by

$$y - y_1 = f'(x_1)\,(x - x_1) \tag{13}$$

Now, let us consider the *equation of the normal* at $M(x_1, y_1)$

Definition: The normal to a curve, at a given point, is *a straight line passing through the given point and perpendicular to the tangent at that point.*

From the definition of a normal, it follows that its slope $= (-1/m) = -1/f'(x_1)$.

Hence, *the equation of a normal to the curve* $y = f(x)$ at a point $M(x_1, y_1)$ (Figure 18.1) is given by

$$\begin{aligned} y - y_1 &= -\frac{1}{m}(x - x_1) \\ &= -\frac{1}{f'(x_1)}(x - x_1) \end{aligned} \tag{14}$$

Example (8): Find (a) the equation of the tangent and (b) the equation of the normal to the curve $y^2 = 5x - 1$ at the point (1, −2).

Solution: The equation of the tangent will be of the form

$$(y+2) = m(x-1)$$

where $m = y'$ evaluated at (1, −2).

We have,

$$y^2 = 5x - 1$$
$$\therefore \quad 2y\,y' = 5$$
$$\therefore \quad y' = \frac{5}{2y} = \frac{5}{2(-2)} = -\frac{5}{4}$$

Hence, the equation to the normal becomes

$$y+2 = -\frac{5}{4}(x-1)$$
$$y+2 = \frac{4}{5}(x-1) \qquad \textbf{Ans.}$$

18.4.1 More Definitions

(Refer to Figure 18.1 for the following definitions.)

(i) T = length of the tangent (i.e., the length of segment QM of the tangent *between the point of tangency and the x-axis*).
(ii) S_T = length of the subtangent (i.e., the segment QP, *which is the projection of the tangent on x-axis*).
(iii) N = length of the normal (i.e., the segment MR is called the length of the normal).
(iv) S_N = length of the subnormal (i.e., the segment RP, *which is the projection of the normal RM on x-axis*).

Let us find the quantities T, S_T, N, and S_N for the curve $y = f(x)$ with reference to the point $M(x_1, y_1)$ on the curve. From Figure 18.1, it will be seen that,

$$QP = |y_1 \cot\alpha| = \left|\frac{y_1}{\tan\alpha}\right| = \left|\frac{y_1}{y'_1}\right|$$

Therefore,

$$S_T = \left|\frac{y_1}{y'_1}\right| = \left|\frac{y_1}{m}\right|$$

(*Here,* $m = y'_1$, *which stands for the derivative* y', *evaluated at* (x_1, y_1), *and this will be applicable all throughout the text.*)

and $$T = \sqrt{MP^2 + QP^2} = \sqrt{y_1^2 + y_1^2/y_1'^2} = \sqrt{y_1^2(y_1'^2/y_1'^2) + (y_1^2/y_1'^2)}$$

$$= \sqrt{(y_1^2/y_1'^2)(y_1'^2 + 1)}$$

$$= \left|\frac{y_1}{y_1'}\sqrt{y_1'^2 + 1}\right| = \left|\frac{y_1}{m}\sqrt{1 + m^2}\right| \quad \left[\because \sqrt{x^2} = |x|\right]$$

Further, it is clear from Figure 18.1 that

$$PR = S_N = |y_1 \tan\alpha| = |y_1 y_1'| = |y_1 m|$$

(Note that $\angle PMR = \alpha$ (why?))
$\therefore$ In right-angled triangle MPR, $\tan\alpha = PR/y_1$.

And so,

$$S_N = |y_1 y_1'| = |y_1 m|$$

Now,

$$MR = N = \sqrt{MP^2 + PR^2} = \sqrt{y_1{}^2 + (y_1 y_1')^2}$$

$$= \sqrt{y_1{}^2 + y_1{}^2 y_1'^2}$$

$$= \left|y_1\sqrt{1 + y_1'^2}\right| = \left|y_1\sqrt{1 + m^2}\right|$$

It is convenient to remember the above formulas in the following order:

$$S_T = \left|\frac{y_1}{y_1'}\right| = \left|\frac{y_1}{m}\right|$$

$$S_N = |y_1 y_1'| = |y_1 m|$$

$$T = \left|\frac{y_1}{y_1'}\sqrt{y_1'^2 + 1}\right| = \left|\frac{y_1}{m}\sqrt{1 + m^2}\right|$$

$$N = |y_1\sqrt{1 + y_1'^2}| = |y_1\sqrt{1 + m^2}|$$

where m stands for y'[or dy/dx], evaluated at the point of tangency (and in case of parametric curves it stands for the given value of the parameter).

Note (2): These formulas are derived on the assumption that $y_1 > 0$, $y_1' > 0$. However, they hold in the general case as well.

Example (9): Find the lengths of the tangent and the subnormal to the curve

$$y = x^5 - 2x + 3 \quad \text{at } (1, 2)$$

Solution: We have,

$$y = x^5 - 2x + 3$$
$$y' = 5x^4 - 2$$

Hence, $m = y' = f'(1) = 5(1)^4 - 2 = 3$

$$\text{Length of the tangent} = T = \left|\frac{y_1\sqrt{1+m^2}}{m}\right| = \left|\frac{2\sqrt{1+(3)^2}}{3}\right| = \frac{2}{3}\sqrt{10}$$

And length of the subnormal $= S_N = |m\,y_1| = 3 \times 2 = 6$

Example (10): Find

(i) the equations of the tangent and the normal;
(ii) the lengths of the tangent and the subtangent; and
(iii) the lengths of the normal and subnormal for the ellipse.

$$x = a\cos t, \quad y = b\sin t \tag{15}$$

at the point $M(x_1, y_1)$ for which $t = \pi/4$ (see Figure 18.2).

Solution: From equation (15), we find,

$$\frac{dx}{dt} = -a\sin t, \quad \frac{dy}{dt} = b\cos t$$
$$\therefore \frac{dy}{dx} = \frac{dy/dt}{dx/dt} = -\frac{b}{a}\cot t$$

Therefore,

$$\left(\frac{dy}{dx}\right)_{t=\pi/4} = -\frac{b}{a} \quad \left[\because \cot\frac{\pi}{4} = 1\right]$$

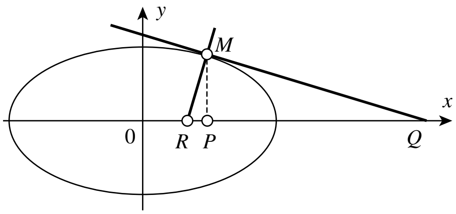


FIGURE 18.2

To find the coordinates of the point of tangency, that is, $M(x, y)$, we put $t = \pi/4$ in equation (15) and obtain

$$x = a\cos\frac{\pi}{4} = \frac{a}{\sqrt{2}} \qquad [\text{denote it by } x_1]$$

$$\text{and } y = b\sin\frac{\pi}{4} = \frac{b}{\sqrt{2}} \qquad [\text{denote it by } y_1]$$

$\therefore$ The equation of the tangent at $M(x_1, y_1)$ is given by

$$y - y_1 = -\frac{b}{a}(x - x_1) \quad \left[\because \left(\frac{dy}{dx}\right) = -\frac{b}{a}\right]$$

$$\text{or} \quad y - \frac{b}{\sqrt{2}} = -\frac{b}{a}\left(x - \frac{a}{\sqrt{2}}\right)$$

$$\text{or} \quad y - \frac{b}{\sqrt{2}} = -\frac{bx}{a} + \frac{b}{\sqrt{2}}$$

$$\therefore \quad y + \frac{bx}{a} = \sqrt{2}\,b$$

$$\therefore \quad ay + bx = \sqrt{2}\,ab$$

$$\text{or} \quad bx + ay - \sqrt{2}\,ab = 0$$

The equation of the normal is

$$y - \frac{b}{\sqrt{2}} = \frac{a}{b}\left(x - \frac{a}{\sqrt{2}}\right)$$

$$\text{or} \quad y - \frac{ax}{b} = \frac{b}{\sqrt{2}} - \frac{a^2}{b\sqrt{2}}$$

Multiplying both sides by $b\sqrt{2}$, we get

$$yb\sqrt{2} - ax\sqrt{2} = b^2 - a^2$$

$$\text{or} \quad \sqrt{2}(by - ax) = b^2 - a^2$$

$$\text{or} \quad -(ax - by)\sqrt{2} = -a^2 + b^2$$

$$\text{or} \quad (ax - by)\sqrt{2} - a^2 + b^2 = 0$$

The *lengths of the subtangent and subnormal*

$$S_T = \left|\frac{y_1}{y_1'}\right| = \left|\frac{y_1}{m}\right|$$

where m stands for y_1' or the derivative $\mathrm{d}y/\mathrm{d}x$ (obtained from the given parametric equations $x = a\cos t$, $y = b\sin t$), and evaluated at $t = \pi/4$.

$$S_T = \left|\frac{b/\sqrt{2}}{-b/a}\right| = \frac{a}{\sqrt{2}} = \frac{\sqrt{2}a}{2}$$

$$S_N = |y_1 y_1'| = |y_1 m| = \frac{b}{\sqrt{2}}\left(-\frac{b}{a}\right) = \frac{b^2}{\sqrt{2}a}$$

The lengths of the tangent and the normal are

$$T = \left|\frac{y_1}{y_1'}\sqrt{y_1'^2+1}\right| = \left|\frac{y_1}{m}\sqrt{1+m^2}\right| = \left|\frac{b/\sqrt{2}}{-b/a}\right|\sqrt{1+\left(-\frac{b}{a}\right)^2}$$

$$= \left|\frac{-a}{\sqrt{2}}\sqrt{\frac{a^2+b^2}{a^2}}\right| = \frac{\sqrt{a^2+b^2}}{\sqrt{2}}$$

$$N = \left|y_1\sqrt{1+y_1'^2}\right| = \left|y_1\sqrt{1+m^2}\right|$$

$$= \left|\frac{b}{\sqrt{2}}\sqrt{1+\left(-\frac{b}{a}\right)^2}\right|$$

$$= \left|\frac{b}{a\sqrt{2}}\sqrt{a^2+b^2}\right|$$

18.4.2 Angle Between Two Curves

The angle between two curves will be the angle between the tangents at the point of their intersection. Accordingly, this angle is given by the formula,

$$\tan\theta_{12} = \frac{m_2 - m_1}{1 + m_1 m_2} \tag{16}$$

where m_1 and m_2 are, respectively, the slopes of the curves (1) and (2), at the point of their intersection, and θ_{12} is the angle measured counterclockwise, from the tangent to the curve 1 to the tangent to curve 2 (see Figure 18.3).

Note (3): In Chapter 4, we have shown that the angle θ between *two nonvertical lines* is given by

$$\tan\theta = \frac{m_2 - m_1}{1 + m_1 m_2}$$

where m_1 and m_2 are, respectively, the slopes of the lines l_1 and l_2.

Once the point of intersection of two curves is known, the derivatives $f_1'(x)$ and $f_2'(x)$, evaluated at that point will give m_1 and m_2 for the equation (16).

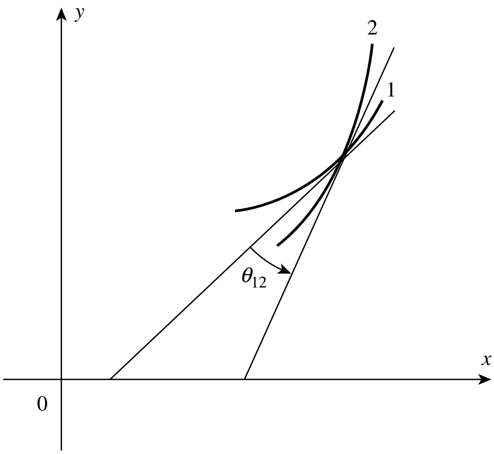


FIGURE 18.3

Example (11): Find the angles at which the following curves intersect *in the first quadrant*

$$\textbf{(i)}\ x^2 + y^2 = 9 \quad \text{and} \quad \textbf{(ii)}\ y^2 = 8x$$

Solution: The point of intersection is determined by solving the two equations simultaneously. This yields,

$$\begin{aligned} x^2 + 8x - 9 &= 0 \\ \text{or} \quad x^2 + 9x - x - 9 &= 0 \\ \text{or} \quad x(x+9) - 1(x+9) &= 0 \\ \text{or} \quad (x+9)\,(x-1) &= 0 \\ \therefore\ x = 1, -9 & \end{aligned}$$

Note (4): *The point of intersection in the first quadrant is obtained by putting $x = 1$ (in any of the equations) and getting $y = \pm\, 2\,\sqrt{2}$, of which $y = 2\sqrt{2}$ is needed for us*. Therefore, *the point of intersection in question* has the coordinates $(1,\ 2\sqrt{2})$.

Now, differentiating (i), we get

$$2x + 2y \cdot y' = 0$$

$$\therefore\ y' = -\frac{x}{y} \quad \therefore\ m_1 = -\frac{1}{2\sqrt{2}} = -\frac{\sqrt{2}}{4}$$

And by differentiating (ii), we get

$$2y \cdot y' = 8 \quad \therefore\ y' = \frac{4}{y} \quad \therefore\ m_2 = \frac{4}{2\sqrt{2}} = \frac{4\sqrt{2}}{4} = \sqrt{2}$$

Therefore,

$$\tan\theta_{12} = \frac{m_2 - m_1}{1 + m_1 m_2} = \frac{\sqrt{2} + \sqrt{2/4}}{1 + \left(-\sqrt{2}/4\right)\sqrt{2}} = \frac{5\sqrt{2}/4}{1/2} = \frac{5}{2}\sqrt{2}$$

and the angle of intersection is given by $\theta_{12} = \tan^{-1}(5/2)\sqrt{2}$.

Example (12): Show that

(i) $x^2 - xy + y^2 - 3 = 0$ and
(ii) $x + y = 0$, intersect at right angles.

Solution: The points of intersection are readily found to be (1, –1) and (–1, 1).

For (i),

$$2x - (xy' + y\cdot 1) + 2yy' = 0$$
$$2x - y + y'(2y - x) = 0$$
$$\therefore\ y' = \frac{-(2x - y)}{2y - x} = \frac{2x - y}{x - 2y}$$

and at either point, $m_1 = 1$.

For (ii), $y' = -1$ so that we have $m_2 = -1$. Since $m_1 = -1/m_2$, hence the angle of intersection is 90°, that is, the tangent lines (and hence the curves) intersect at right angles.

(Note that $\tan\theta_{12} = (m_2 - m_1)/(1 + m_1 m_2) = (-1 - 1)/(1 + (-1)) = -2/0$.)

18.5 SLOPE OF A CURVE IN POLAR COORDINATES

We know that in rectangular coordinates $\mathrm{d}y/\mathrm{d}x$ represents the slope of the curve $y = f(x)$, but *in polar coordinates* $\mathrm{d}\rho/\mathrm{d}\theta$ *does not represent the slope of the curve.*

$$\rho = f(\theta) \tag{17}$$

It merely represents *the rate of change of the radius vector* ρ *with respect to angle* θ. In order to determine the slope of the curve $\rho = f(\theta)$, we use the following relations between rectangular coordinates (x, y) and polar coordinates (ρ, θ). These are

$$\left.\begin{aligned} x &= \rho\cos\theta \\ y &= \rho\sin\theta \end{aligned}\right\} \tag{18}$$

Equation (18) is a *parametric equation of the given curve*, the parameter being the polar angle θ. (Note that ρ is a function of θ.)

If we denote by φ the angle formed by the tangent to the curve at some point $M(\rho, \theta)$ with the positive x-axis, we will have,

$$\tan\varphi = \frac{\mathrm{d}y/\mathrm{d}\theta}{\mathrm{d}x/\mathrm{d}\theta}$$

Now,

$$\frac{\mathrm{d}x}{\mathrm{d}\theta} = \frac{\mathrm{d}\rho}{\mathrm{d}\theta}\cos\theta - \rho\sin\theta$$

and

$$\frac{dy}{d\theta} = \frac{d\rho}{d\theta}\sin\theta + \rho\cos\theta \quad \text{[from equation (2)]}$$

$$\therefore \quad \text{Slope} = \frac{dy}{dx} = \frac{(d\rho/d\theta)\sin\theta + \rho\cos\theta}{(d\rho/d\theta)\cos\theta - \rho\sin\theta} = \frac{\rho'\sin\theta + \rho\cos\theta}{\rho'\cos\theta - \rho\sin\theta}$$

Dividing the numerator and denominator both by $\rho'\cos\theta$, we get,

$$\text{Slope} = \frac{\tan\theta + \rho/\rho'}{1 - \tan\theta\,(\rho/\rho')}, \quad \text{where } \rho' = \frac{d\rho}{d\theta} \tag{19}$$

With this formula we can readily find *the slope of the curve whose equation is given in the polar coordinates.*

Example: Find the slope of the curve $\rho = 2 - \cos\theta$.

(a) At any point (b) at $\theta = \pi/4$.

Solution: The given curve is

$$\rho = 2 - \cos\theta \tag{20}$$

(a) We know that slope of the curve (1) at any point is given by

$$\text{Slope} = \frac{\tan\theta + \rho/\rho'}{1 - \tan\theta\,(\rho/\rho')} \tag{21}$$

We have $\rho' = \sin\theta$ [from (20)].
Now,

$$\tan\theta + \frac{\rho}{\rho'} = \tan\theta + \frac{2-\cos\theta}{\sin\theta}$$

and

$$1 - \tan\theta\frac{\rho}{\rho'} = 1 - \tan\theta\frac{2-\cos\theta}{\sin\theta}$$

$$\therefore \quad \text{Slope} = \frac{\sin\theta\tan\theta + 2 - \cos\theta}{2(\sin\theta - \tan\theta)} \qquad \textbf{Ans}. \tag{22}$$

(b) Slope of the curve at the point where polar angle $\theta = \pi/4$ is obtained by putting $\theta = \pi/4$ in (19).

$$\therefore \quad \text{Slope}\left(\text{at } \theta = \frac{\pi}{4}\right)$$

$$= \frac{\sqrt{2}/2 + 2(-\sqrt{2}/2)}{\sqrt{2} - 2} = \frac{2}{\sqrt{2} - 2} \qquad \textbf{Ans}.$$

18.5.1

An important angle to consider is the angle μ between the radius vector and the tangent *measured counter clockwise from the radius vector to the tangent.*

Now,

$$\text{slope} = \tan\varphi = \tan(\theta+\mu) = \frac{\tan\theta + \tan\mu}{1+\tan\theta\tan\mu} \tag{23}$$

Comparing (19) and (23), we find,

$$\tan\mu = \frac{\rho}{\rho'} \tag{24}$$

18.5.2

The geometric meaning of the derivative of the radius vector ρ with respect to the polar angle θ $[\rho' = d\rho/d\theta]$ (See Calculus by Thomas/Finney, Fig10.37, Page 592, where the angle ψ must be identified as μ.).

From equation (24) we have $\tan\mu = \rho/\rho'$ or $\rho' = \rho\cot\mu$.

Thus, the derivative of the radius vector with respect to the polar angle $(\rho' = d\rho/d\theta)$ is equal to the length of the radius vector multiplied by the cotangent of the angle μ between the radius vector and the tangent to the curve at the given point.

18.5.3 The Angle Between Two Curves in Polar Coordinates

In view of the definition of the angle μ, discussed above, the angle between two curves is given by $\theta_{12} = \tan(\mu_1 - \mu_2)$, where θ_{12} the angle measured counterclockwise from curve (1) to curve (2) is given by

$$\theta_{12} = \frac{\tan\mu_2 - \tan\mu_1}{1+\tan\mu_2\tan\mu_1} \tag{25}$$

Formula (24) is used to evaluate $\tan\mu_1$and $\tan\mu_2$.

19a Increasing and Decreasing Functions and the Sign of the First Derivative

19a.1 INTRODUCTION

In Chapter 6, we have discussed *increasing* and *decreasing* functions on an interval. The distinction between an *increasing and nondecreasing function*, and that between *decreasing and nonincreasing function* are also clarified there.(1)

In this section, we shall discuss the *increasing and the decreasing portions of the graph* and the point(s) at which the increasing (or decreasing) portion of the graph enters into the decreasing (or increasing) portion. For convenience, we revise the definitions of increasing (decreasing) functions as introduced in Chapter 6.

Definition: A function $y = f(x)$ is said to be increasing on an interval I if to greater values of $x \in$ I there correspond greater values of the function. Similarly, a function is said to be decreasing on I, if to greater values of x there correspond smaller values of the function. Analytically, we can define increasing and decreasing functions as follows:

19a.1.1 Increasing and Decreasing Functions on Interval "I"

Definition: Let I be an *open interval*, contained in the domain of a real-valued function. Then, f is said to be

(a) *increasing* on I, if $x_1 < x_2$ in $I \Rightarrow f(x_1) < f(x_2)$ for all $x_1, x_2 \in I$.

(b) *decreasing* on I, if $x_1 < x_2$ in $I \Rightarrow f(x_1) > f(x_2)$ for all $x_1, x_2 \in I$.

Note (1): From the above definitions, it is clear that by the term increasing function, we mean the strictly increasing function and, similarly, the term decreasing function stands for the strictly decreasing function.

Applications of derivatives 19a-Increasing and decreasing functions, and the sign of the first derivative horizontal tangents and local maximum/minimum values of functions Concavity, points of inflection, and the sign of the second derivative.

(1) Recall that when the graph of a function has a horizontal portion, added to an increasing (or decreasing) function, it becomes a nondecreasing (or nonincreasing) function.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

Note (2): It must be noted that the notion of *increasing* and *decreasing* functions are always defined in terms of increasing x. Thus, as we move from left to right along the graph of a function,

- in the case of an increasing function, the height of the graph continuously increases, and
- in the case of a decreasing function, the height of the graph continuously decreases.

19a.1.2

In this chapter, we will use *differentiation* to find out whether a function is *increasing* or *decreasing* or neither.

The derivative of a function $y = f(x)$ is the rate at which y changes with respect to x. It defines the slope of the function's graph at x and allows us to estimate how much y changes when we change x by a small amount. These concepts were discussed in Chapter 16. However, it is useful to revise the process of computing approximate changes in the value of a function $y = f(x)$ when the independent variable x is changed by an small amount. The following example makes it clear.

Example (1): Consider the function,

$$y = x^3 \tag{1}$$

so that we have,

$$y' = 3x^2 \tag{2}$$

Equation (2) tells us that at $x = 1$, $y' = 3(1)^2 = 3$ and similarly at $x = 2$, $y' = 3(2)^2 = 12$.

These calculations tell us that if we change the value of x by a small amount, say 0.2 units at $x = 1$, then the value of y $[= x^3]$ will approximately change three times, that is, by 0.6 units. In other words, the height of the graph will be more (approximately) by 0.6 units at the point $x = 1.2$ than at the point $x = 1$. Similarly, the height of the graph at the point $x = 2.2$ will be more by 2.4 units (i.e., 12 times of 0.2) than at $x = 2.0$, and so on.

19a.1.3

If a function $y = f(x)$ *has a derivative* at a point x_0, then we know that f is a continuous at x_0. Accordingly, if a function has a derivative over an interval, then it is continuous over the interval. In other words, the graph of a differentiable function is without any break.

We can gain even more information about the graph of a *differentiable function* if we know where its derivative is positive, negative, or zero. We shall also see where the graph is rising, is falling, and has a horizontal tangent.(2)

19a.1.4

Refer to Figure 19a.1 showing the graph of a function f for all x in the *closed interval* $[x_1, x_7]$ on which $f(x)$ is continuous. This figure shows that as a point moves along the curve from A to B, the function values increase as x increases, and that as a point moves along the curve from B to C, the function values decrease as x increases. We say, then, that $f(x)$ is increasing on the closed interval $[x_1, x_2]$ and that $f(x)$ is decreasing on the closed interval $[x_2, x_3]$.

(2) In this section, we shall learn how the sign of the first derivative helps in deciding the increasing (or decreasing) nature of a function. Later on, the signs of the first and the second derivatives will be used in determining extreme values (i.e., maximum and minimum values) of functions. In fact, the signs of the first and the second derivatives together tell us how the graph of a function is shaped.

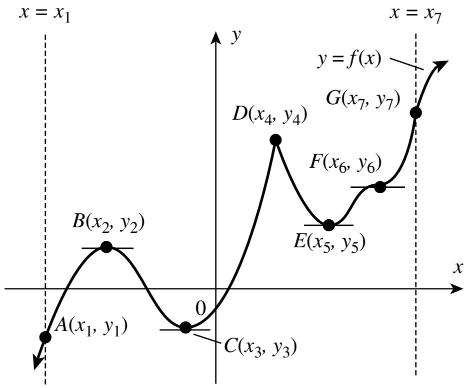


FIGURE 19a.1

Thus, the function of Figure 19a.1 is increasing on the closed intervals $[x_1, x_2]$; $[x_3, x_4]$; $[x_5, x_6]$; $[x_6, x_7]$; $[x_5, x_7]$ and it is decreasing on the closed intervals $[x_2, x_3]$; $[x_4, x_5]$. Let us see what is happening geometrically.

We observe that when the slope of the tangent line (to the curve) is positive, the function is increasing, and when it is negative, the function is decreasing. We know that the slope of the tangent line to the curve $y = f(x)$ at a point is represented by the derivative $f'(x)$. Our observation tells us that when $f'(x) > 0$, the function is increasing as x increases; and when $f'(x) < 0$, the function is decreasing as x increases. Later on, we shall prove that these conclusions (drawn from our observations) are true. We also observe the following:

(i) At the point(s) of transition (between the rising and falling portions of the curve), there is a horizontal tangent line that means $f'(x) = 0$ at such points. In Figure 19a.1, these points (on the curve) are $B(x_2, y_2)$, $C(x_3, y_3)$, and $E(x_5, y_5)$.

(ii) At the point of transition $D(x_4, y_4)$, no unique tangent line exists (we say that the derivative does not exist at x_4). Note that the point D on the graph is a sharp point (or corner point).

(iii) At the point $F(x_6, y_6)$, the horizontal tangent line exists, we say that the rate of change of the function at $x = x_6$ is *zero*, but it is not a transition point of the curve, since the function increases throughout the interval $[x_5, x_7]$ as x increases and the point $F(x_6, x_6)$ lies in this interval.

From Figure 19a.1, we have gained useful information about the increasing/decreasing portions of the graph, *points of transition*, and the existence of horizontal tangent lines at certain points on the graph of a function. Now consider the following examples:

Example (2): The function $y = x^2$ decreases on $(-\infty, 0)$, where $y' = 2x < 0$. It increases on $(0, \infty)$, where $y' = 2x > 0$. At $x = 0$, the point of transition $y' = 0$, and the curve has a horizontal tangent (Figure 19a.2).

Example (3): The function $f(x) = \tan x$ increases on $(-(\pi/2), (\pi/2))$ and $((\pi/2), (3\pi/2))$, where $f'(x) = \sec^2 x = (1/\cos^2 x) > 0$. The graph of $y = \tan x$ increases on $-(\pi/2) < x < (\pi/2)$ and on $(\pi/2) < x < (3\pi/2)$ (Figure 19a.3).

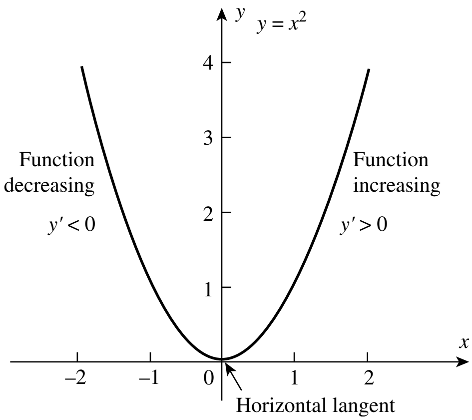


FIGURE 19a.2

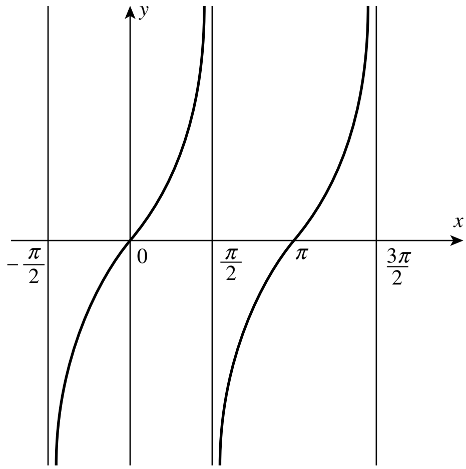


FIGURE 19a.3

On these intervals, the function does not decrease anywhere. Thus, the graph of this function does not have a transition point. (Note that, f is not defined at $x = \pi/2$, but this fact is not important here.)

Example (4): The function $y = 1/x^2$ increases from left to right on $(-\infty, 0)$, where $y' = (-2/x^3) > 0$ and decreases from left to right on $(0, \infty)$, where $y' = (-2/x^3) < 0$.

The *derivative* y' $(=(-2/x^3))$ is not defined at $x = 0$, which is the point of transition.

Furthermore, note that the point of transition $(x = 0)$ does not lie on the graph of $y = 1/x^2$. In fact, the function of $y = 1/x^2$ itself is not defined at $x = 0$ (Figure 19a.4).

Note (2): The above examples show that a function may increase over one interval and decrease over another. They also suggest that we can speak of a function increasing or decreasing at a point.

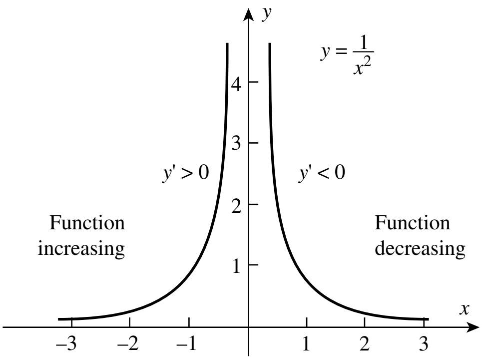


FIGURE 19a.4

Remark: It is clear that if a function is increasing (or decreasing) in an interval, then it is definitely increasing (or decreasing) at every point in that interval. However, it is useful to clarify (through a simple definition) what it means when we say that a function is increasing (or decreasing) at a point x_0 in the domain of the function.

19a.1.5

Definition: A function $f(x)$ is said to be increasing at a point $x = x_0$, if there exists a neighborhood $(x_0 - \delta, x_0 + \delta)$ of x_0 such that $f(x) < f(x_0)$ whenever $x < x_0$, and $f(x) > f(x_0)$ whenever $x > x_0$ (Figure 19a.5).

Analogously, a function $f(x)$ is said to be decreasing at a point $x = x_0$, if given some neighborhood of x_0, $f(x) > f(x_0)$ whenever $x < x_0$, and $f(x) < f(x_0)$ whenever $x > x_0$.

In Chapter 20, it is proved that if $y = f(x)$ is differentiable with $f'(x) > 0$ at every point of an interval I, then $f(x)$ increases on I. Similarly, if $f'(x) < 0$ at every point of I, then $f(x)$ decreases on I.

For the time being, we assume these results and record them as the first derivative test for rise and fall.

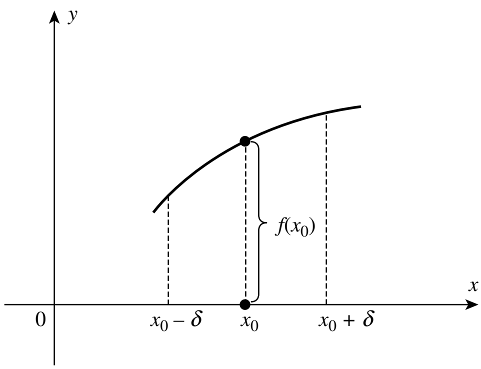


FIGURE 19a.5

19a.2 THE FIRST DERIVATIVE TEST FOR RISE AND FALL

Suppose that $y = f(x)$ has a derivative at every point x of an interval I. Then,

$$\left.\begin{array}{ll}\text{(i)} & f(x) \text{ increases on } I, \text{ if } f'(x) > 0 \text{ for all } x \text{ in } I \\ \text{(ii)} & f(x) \text{ decreaseson } I, \text{ if } f'(x) < 0 \text{ for all } x \text{ in } I\end{array}\right\}^{(3)}$$

In geometric terms, the first derivative test says that a differentiable function increases where the tangent to its graph has a *positive slope* and decreases where the tangent to its graph has a *negative slope*.

(This permits us to judge the increasing or decreasing nature of a function by the sign of its derivative.)

Remark: The first derivative test gives us the sufficient condition for a function to increase (or decrease) in an interval. It is worth mentioning that if $f(x)$ increases on $[a, b]$, then it does not follow $f'(x) > 0$ everywhere in (a, b) as is clear from the following example:

Example (5): The function $f(x) = x^3$ increases on $[-1, 1]$. However, the derivative $f'(x) = 3x^2$ equals the value 0 at $x = 0$. Similarly, the function $g(x) = -x^3$ is a decreasing function on $[-1, 1]$ with $g'(x) = -3x^2$, which equals the value 0 at $x = 0$. [Note that this function increases even at the point $x = 0$, where $f'(x) = 0$.]

The following theorem specifies the sufficient conditions for a function to be increasing or decreasing at a point.

19a.2.1

Theorem: Let $f(x)$ have a derivative $f'(x_0)$ at x_0. If $f'(x_0) > 0$, then $f(x)$ increases at x_0, and if $f'(x_0) < 0$, then $f(x)$ decreases at x_0.

Proof:
Let $f'(x_0) > 0$. Then, by definition of the derivative, we have, $x_0 \in (x_0 - h, x_0 + h)$, for all $h \in R$.

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} > 0$$

This means that, there exists $\delta > 0$, such that for *all* h, $\dfrac{f(x_0 + h) - f(x_0)}{h} > 0$, whenever $0 < |h| < \delta$

It follows that, if $0 < |h| < \delta$, then h and $[f(x_0 + h) - f(x_0)]$ are of the same sign.

Thus, if $h < 0$, then $[f(x_0 + h) - f(x_0)] < 0$, that is, $f(x_0 + h) < f(x_0)$, and if $h > 0$, then $[f(x_0 + h) - f(x_0)] > 0$, that is, $f(x_0 + h) > f(x_0)$.

By definition, this means that $f(x)$ increases at x_0. Using similar reasoning, we can show that if $f'(x_0) < 0$, then $f(x)$ decreases at x_0. (Proved)

Note (4): The conditions specified in the above theorem are not necessary. Note that the function shown in Figure 19a.6 increases at $x = 0$; however, the derivative of this function does not exist at $x = 0$. [Also see the example $f(x) = x^3$ discussed above.]

(3) We shall use these results in our further discussion and draw many useful conclusions. These facts are proved in Chapter 20 under an application of Lagrange's Mean Value Theorem as hinted above.

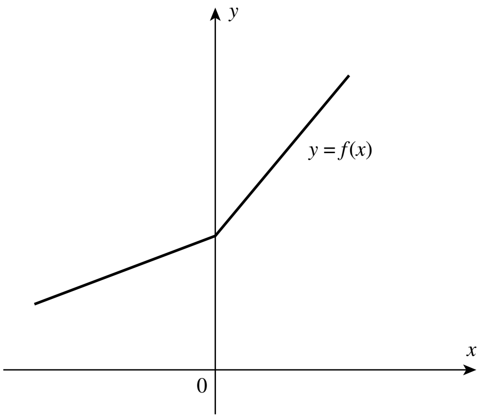


FIGURE 19a.6

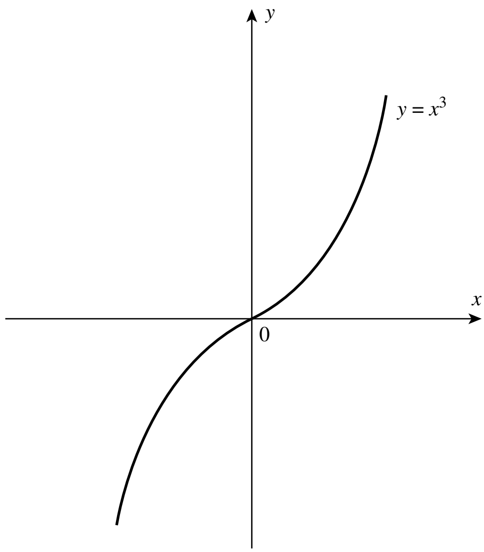


FIGURE 19a.7

The function $f(x) = x^3$ increases at $x = 0$, and its derivative $f'(x) = 3x^2$ vanishes at $x = 0$ (Figure 19a.7).

Note (5): We have seen that a function may increase over one interval and decrease over another. Such intervals are called the intervals of monotonicity, and our interest lies in finding these intervals for a given function.

19a.3 INTERVALS OF INCREASE AND DECREASE (INTERVALS OF MONOTONICITY)

An interval on which the function increases is called the interval of increase, and an interval on which the function decreases is called its interval of decrease. For simple functions, whose

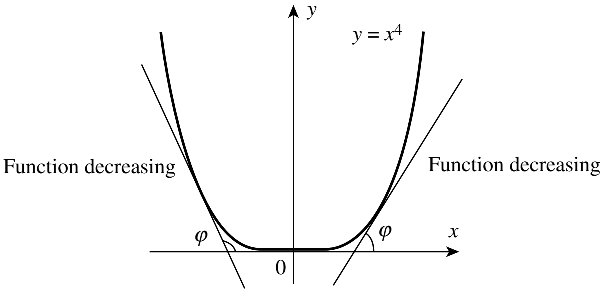


FIGURE 19a.8

graphs are known, these intervals of monotonicity are easily determined. Later on, some techniques will be developed that will make it possible to find the intervals on which a function is monotonic, without requiring to construct its graph. Consider the following example:

Example (6): Determine the domain of increase and decrease of the function $y = x^4$.

Solution: The derivative of y is given by $y' = 4x^3$: for $x > 0$, we have $y' > 0$ and the function increases; for $x < 0$, we have $y' < 0$ and the function decreases (Figure 19a.8).

19a.3.1 Sign of a Continuous Function $f(x)$

Let $y = f(x)$ be a *continuous function*. By the sign of $f(x)$, at any point $x = a$ in its domain, we mean the sign of the value $f(a)$ provided $f(a) \neq 0$. If $f(a) \neq 0$, then either $f(x) > 0$ or $f(x) < 0$.

Our interest lies in solving the inequalities $f'(x) > 0$ and $f'(x) < 0$, which will give us the intervals on which $f(x)$ increases and those on which $f(x)$ decreases, respectively. Now, we shall show how our intuitive knowledge of continuity can be applied to solve a quadratic inequality and other inequalities. But first, we must provide a framework on which to build our technique.

Consider the graph of a continuous function $y = f(x)$ (Figure 19a.9). There is a relationship between the (real) roots of the equation $f(x) = 0$ and the points where the graph of $y = f(x)$ meets the x-axis. These points are called the x-intercepts of the graph. If the graph of f has an intercept $(r, 0)$, then $f(r) = 0$ and so r is a root of the equation $f(x) = 0$.

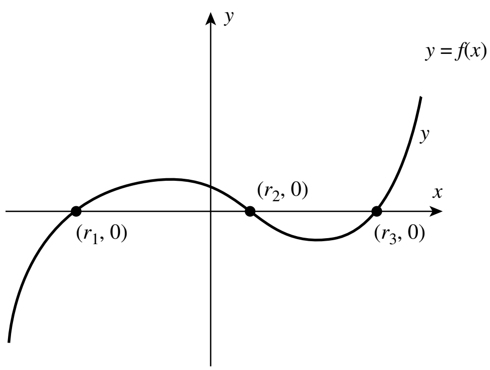


FIGURE 19a.9

Hence, from the graph of $y = f(x)$ in Figure 19a.9, we conclude that r_1, r_2, and r_3 are roots of the equation $f(x) = 0$. On the other hand, if r is any real root of the equation $f(x) = 0$, then $f(r) = 0$ and hence $(r, 0)$ lies on the graph of f. It means that all real roots of the equation $f(x) = 0$ can be represented by the points where the graph of f meets the x-axis.

19a.3.2 Procedure to Solve an Inequality Involving a Polynomial

Now, suppose we have to solve the *quadratic inequality* $x^2 + 3x - 4 > 0$.

We put $f(x) = x^2 + 3x - 4 = (x + 4)(x - 1)$.

Since $f(x)$ is a polynomial, it is continuous everywhere. The roots of the equation $f(x) = 0$ are (-4) and 1. Hence, the graph of $f(x)$ has x-intercepts $(-4, 0)$ and $(1, 0)$. These roots (or to be more precise the x-intercepts) determine three intervals on the real line: $(-\infty, -4)$, $(-4, 1)$, and $(1, \infty)$ (Figures 19a.10 and 19a.11).

Consider the interval $(-\infty, -4)$. Since f is continuous on this interval, we claim that throughout this interval either $f(x) > 0$ or $f(x) < 0$.(4)

We prove this indirectly as follows:

Suppose $f(x)$ did indeed change sign in the interval $(-\infty, -4)$. Then, by continuity of f, there would be a point in $(-\infty, -4)$ where the graph would intersect the x-axis. Suppose this point is c. Then, c would be a root of the equation $f(x) = 0$, so that we should get $f(c) = 0$.

This cannot occur since there is no root of the equation $x^2 + 3x - 4 = 0$ that is less than -4. Hence, $f(x)$ must be (strictly) positive or (strictly) negative on $(-\infty, -4)$ as well as on the other intervals.

Thus, to determine the sign of $f(x)$ on any interval, it is sufficient to determine its sign at any point in the interval. This permits us to select any convenient point in the interval to find the sign of $f(x)$.

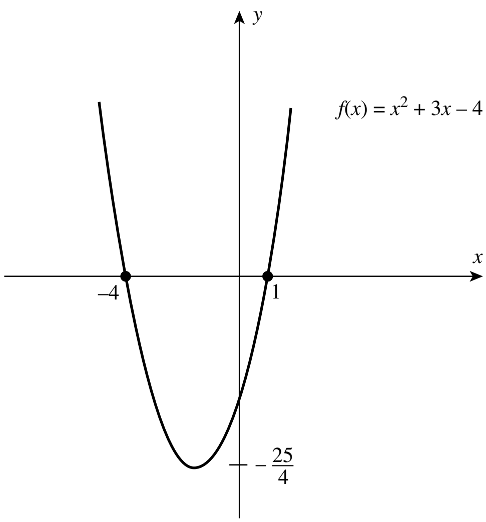


FIGURE 19a.10

(4) The statement $f(x_1) > 0$ tells us that the point $(x_1, f(x_1))$ is above the x-axis. Similarly, the statement $f(x_2) < 0$ tells us that the point $(x_2, f(x_2))$ is below the x-axis.

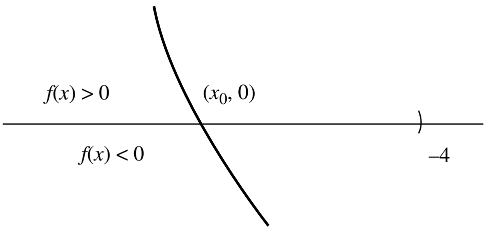


FIGURE 19a.11

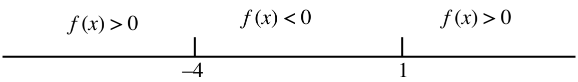


FIGURE 19a.12

For instance, -5 is in $(-\infty, -4)$. Therefore, for $f(x) = x^2 + 3x - 4$, we get $f(-5) = (-5)^2 + 3(-5) - 4 = 6 > 0$. Thus, $f(x) > 0$ on $(-\infty, -4)$. Since 0 is in $(-4, 1)$ and $f(0) = -4 < 0$, $f(x) < 0$ on $(-4, 1)$.

Similarly, 3 is in $(1, \infty)$ and $f(3) = 14 > 0$. Therefore, $f(x) > 0$ on $(1, \infty)$ (Figure 19a.12).

Therefore, we get $f(x) = x^2 + 3x - 4 > 0$ for $x < -4$ and for $x > 1$

(This is the solution of the inequality $x^2 + 3x - 4 > 0$. One must realize the importance of the role played by the concept of continuity.)

Let us review what we have learnt. If we consider a polynomial function, $y = f(x)$, then the roots of the equation $f(x) = 0$ (say x_1, x_2, x_3) represent points on the x-axis. These points determine the intervals $(-\infty, x_1)$, (x_1, x_2), (x_2, x_3), (x_3, ∞) on the real line. If a is any point on an interval and $f(a) > 0$, then the graph of the function must be above the x-axis in that interval. Similarly, if $f(b) < 0$ in an interval, then the graph of the function must be below the x-axis in that interval.

Every polynomial function is continuous and differentiable. Now, our interest lies in differentiable functions whose graphs are definitely continuous. Also, we know that $f'(x)$ represents the slope of the tangent line at any point $(x, f(x))$ of the graph. Therefore, the sign of the first derivative $f'(x)$ tells us all that we need to know about where the curve rises and where it falls. The roots of the equation $f'(x) = 0$ can help us to determine the intervals on which $f'(x) > 0$ and those on which $f'(x) < 0$.

The intervals on which $f'(x) > 0$, the function $f(x)$ increases, and the intervals on which $f'(x) < 0$, the function decreases.

For any differentiable function, we can find (using the first derivative test) the intervals on which the function $f(x)$ increases (or decreases), since the first derivative test is applicable to any differentiable function.

19a.3.3 Practical Method for Finding Intervals of Monotonicity

Example (7): Now, let us use the above technique to determine the intervals in which the function $f(x) = 2x^3 - 3x^2 - 36x + 7$ is (a) increasing and (b) decreasing.

Solution: $f(x) = 2x^3 - 3x^2 - 36x + 7$

$$\therefore \quad f'(x) = 6x^2 - 6x - 36$$
$$= 6(x^2 - x - 6)$$

Putting $f'(x) = 0$, we get $6(x - 3)(x + 2) = 0$. Therefore, $x = 3$ and $x = -2$ are the roots of $f'(x) = 0$.

The points $x = -2$ and $x = 3$ divide the real line into three disjoint intervals, namely, $(-\infty, -2)$, $(-2, 3)$, $(3, \infty)$. In each interval, the sign of $f'(x)$ is determined by the signs of the factors of $f'(x)$.

We have $f'(x) = 6(x - 3)(x + 2)$. Now observe the following:

(i) (-3) is in $(-\infty, -2)$

$$\therefore \quad f'(-3) = 6(-)(-) = (+) \quad \text{and so} \quad f'(x) > 0 \text{ on } (-\infty, -2)$$

Therefore, $f(x)$ is increasing on $(-\infty, -2)$.

(ii) 0 is in $(-2, 3)$

$$\therefore \quad f'(0) = 6(-)(+) = (-) \quad \text{and so} \quad f'(x) < 0 \text{ on } (-2, 3)$$

Therefore, $f(x)$ is decreasing on $(-2, 3)$.

(iii) 4 is in $(3, \infty)$

$$\therefore \quad f'(4) = 6(+)(+) = (+) \quad \text{and so} \quad f'(x) > 0 \text{ on } (3, \infty)$$

Therefore, $f(x)$ is increasing on $(3, \infty)$.

Remark: Note that, it is not necessary that we actually evaluate $f'(-3)$, $f'(0)$, or $f'(4)$. To find the sign of $f'(x)$, we factorize $f'(x)$ and find the sign of each factor. The sign of $f'(x)$ is then obtained by using rules of algebra.

Now, let us investigate the behavior of exponential, trigonometric and logarithmic functions.

Example (8): Prove that the exponential function e^x is increasing throughout its domain, in this case R.

Solution: We know that $(d/dx)(e^x) = e^x$. We also know that, $e^x = 1 + x + (x^2/2!) + (x^3/3!) + \cdots$.

(i) When x is positive, e^x is positive, because

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots > 1$$

(ii) When x is negative, e^x is positive

$$\therefore \quad e^x = \frac{1}{e^{-x}} = \frac{1}{\text{a positive number}} > 0$$

(iii) When x is 0, $e^x = 1 > 0$ (Figure 19a.13).

Therefore, e^x is positive for all values of x.

Since $(d/dx)(e^x) = e^x$ is always positive, it follows that e^x is an increasing function throughout R.

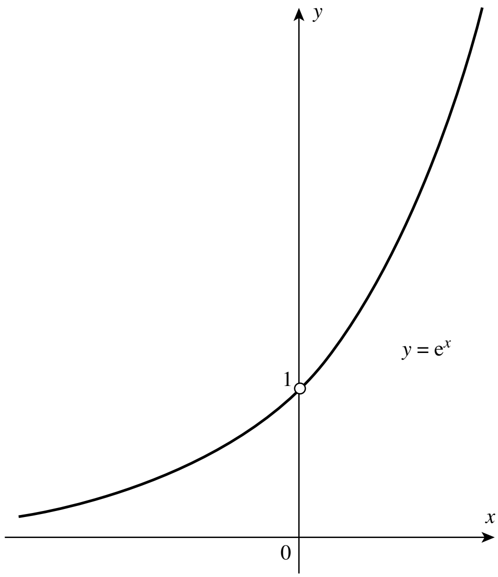


FIGURE 19a.13

Example (9): Prove that the function sin x is increasing in the interval $(0, \pi/2)$ and decreasing in the interval $(\pi/2, \pi)$.

Solution: $(\mathrm{d}/\mathrm{d}x)(\sin x) = \cos x$

We know that cos x is positive on $(0, \pi/2)$ and negative on $(\pi/2, \pi)$. Therefore, sin x is increasing on $(0, \pi/2)$ and decreasing on $(\pi/2, \pi)$.

Remark: If we consider the entire interval $(0, \pi)$, sin x is neither increasing nor decreasing.

Example (10): Consider the function $f(x) = x^2 - x + 1$, $0 < x < 1$. We have,

$$f'(x) = 2x - 1 = 2\left(x - \frac{1}{2}\right)$$

Observe that for $x > 1/2, f'(x)$ is *positive*, but if $x < 1/2$, then $f'(x)$ is *negative*. Therefore, on the interval $(1/2, 1)$, $f(x)$ is increasing, whereas on the interval $(0, 1/2)$, it is decreasing. If we consider the entire interval $(0, 1)$, $f(x)$ is neither increasing nor decreasing.

Example (11): Separate the intervals in which $f(x) = x^3 - 6x^2 + 9x + 5$ is increasing or decreasing.

Solution: We have $f(x) = x^3 - 6x^2 + 9x + 5$

$$\begin{aligned} \therefore \quad f'(x) &= 3x^2 - 12x + 9 \\ &= 3(x^2 - 4x + 3) \\ &= 3(x - 1)(x - 3) \end{aligned}$$

We solve $f'(x) = 0$.

$$\therefore \quad f'(x) = 3(x-1)(x-3) = 0$$

$$\therefore \quad x = 1 \quad \text{and} \quad x = 3$$

These points determine three intervals as $(-\infty, 1)$, $(1, 3)$, and $(3, \infty)$ on the real line. We discuss the behavior of $f(x)$ in these intervals separately. Observe the following:

(i) 1/2 is in $(-\infty, 1)$

$$\therefore \quad f'(1/2) = 3(-)(-) = (+) \text{ and so } f'(x) > 0 \text{ on } (-\infty, 1)$$

Therefore, $f(x)$ is increasing on $(-\infty, 1)$.

(ii) 2 is in $(1, 3)$

$$\therefore \quad f'(2) = 3(+)(-) = (-) \text{ and so } f'(x) < 0 \text{ on } (1, 3)$$

Therefore, $f(x)$ is decreasing on $(1, 3)$.

(iii) 4 is in $(3, \infty)$

$$\therefore \quad f'(4) = 3(+)(+) = (+) \text{ and so } f'(x) > 0 \text{ on } (3, \infty)$$

Therefore, $f(x)$ is increasing on $(3, \infty)$.

Note that, the set of values of x for which $f(x)$ is increasing is $(-\infty, 1) \cup (3, \infty)$.

Also, the set of values of x for which $f(x)$ is decreasing is $(1, 3)$. Ans.

Example (12): Show that the function $f(x) = 3x^3 - 3x^2 + x + 25$ is increasing on R.

Solution: We have, $f(x) = 3x^3 - 3x^2 + x + 25$.

$$\therefore \quad f'(x) = 9x^2 - 6x + 1 = (3x-1)^2$$

Note that, $f'(x)$ is a *perfect square.* At $x = 1/3, f'(x) = 0$, but for all other values of $x, f'(x) > 0$. Geometrically, it means that slope of f at each point is positive except at $(1/3, f(1/3))$, where tangent line is horizontal [since $f'(1/3) = 0$]. As $f'(x)$ does not change sign on the whole real line, it follows that $f(x)$ increases throughout R.

Note (6): It is easy to show that the logarithmic function (i.e., $y = \log x$) is an increasing function, wherever it is defined. Try this.

Example (13): Prove that, $x - 1 > \log x > (x-1)/x, \ \forall x > 1$.

Solution: We shall do this in two steps:

(i) $x - 1 > \log x \quad \forall x > 1$

(ii) $\log x > \dfrac{x-1}{x} \quad \forall x > 1$

(i) Let $f(x) = (x - 1) - \log x$ (3)[5]

$$\therefore \quad f'(x) = 1 - \frac{1}{x}$$

Now, for $x > 1$, $(1/x) < 1$.

$$\therefore \quad 1 - \frac{1}{x} > 0$$

$$\therefore \quad f'(x) = 1 - \frac{1}{x} > 0, \quad \forall\, x > 1$$

Therefore, $f(x)$ is increasing at every $x > 1$.
Next, observe that

$$f(1) = (1 - 1) - \log 1 = 0 - 0 = 0 \text{ from equation (3)}$$

Thus, $f(x)$ is increasing for $x > 1$ and that $f(1) = 0$.
It follows that, $f(x) > 0$, $\forall\ x > 1$.

$$\therefore \quad (x - 1) - \log x > 0. \therefore \quad (x - 1) > \log x \tag{4}$$

(ii) Now, let

$$\begin{aligned} g(x) &= \log x - \frac{x - 1}{x} \\ &= \log x - (1 - \frac{1}{x}) = \log x - 1 + \frac{1}{x} \end{aligned} \tag{5}$$

$$\therefore \quad g'(x) = \frac{1}{x} - \frac{1}{x^2} = \frac{x - 1}{x^2}$$

Observe that, for every $x > 1$,

$$\frac{x - 1}{x^2} > 0$$

$$\therefore \quad g'(x) = \frac{x - 1}{x^2} > 0, \quad \forall\, x > 1$$

Therefore, $g(x)$ is increasing at every $x > 1$.
But,

$$g(1) = \log 1 - \frac{1 - 1}{1} = 0 - 0 = 0$$

[5] We will show that $f(x) = (x - 1) - \log x > 0$, so that $(x - 1) > \log x$.

Thus, $g(x)$ is increasing for $x > 1$ and that $g(1) = 0$.

It follows that, $g(x) > 0, \forall\, x > 1$.

$$\therefore \quad g(x) = \log x - \frac{x-1}{x} > 0, \forall\, x > 1$$

$$\therefore \quad \log x > \frac{x-1}{x}, \forall\, x > 1 \tag{6}$$

From equations (4) and (6), we have

$$1 - x > \log x > \frac{x-1}{x}, \quad \forall\, x > 1. \quad \textbf{Ans.}$$

Example (14): Show that, $\log x < x - 1$ for $0 < x < 1$.

Solution: Let $f(x) = (x - 1) - \log x$, for $0 < x < 1$.

$$\therefore \quad f'(x) = 1 - 0 - \frac{1}{x} = 1 - \frac{1}{x}, \quad \text{for } 0 < x < 1$$

Since $x < 1$, $(1/x) > 1$.

$\therefore \quad f'(x) = 1 - \frac{1}{x} < 0, \quad \text{for } 0 < x < 1$

Therefore, f is decreasing for $0 < x < 1$.

$$\therefore \quad x < 1 \Rightarrow f(x) > f(1)$$

But $f(1) = (1-1) - \log 1 = 0$ [$\log 1 = 0$].

$$\begin{aligned} \therefore \quad x < 1 &\Rightarrow f(x) > 0 \\ &\Rightarrow (x-1) - \log x > 0 \\ &\Rightarrow x - 1 > \log x \\ &\Rightarrow \log x < x - 1, \text{ for } 0 < x < 1. \quad \text{Ans.} \end{aligned}$$

19a.4 HORIZONTAL TANGENTS WITH A LOCAL MAXIMUM/MINIMUM

Consider the graph of a differentiable function $y = f(x)$ shown in Figure 19a.14. We observe that the function increases on (a, c), where $f'(x) > 0$, decreases on (c, d), where $f'(x) < 0$, and increases again on (d, b). The points of transition P and Q on the curve (at $x = c$ and $x = d$, respectively) are marked by horizontal tangents.

If the derivative f' of a function $y = f(x)$ is continuous, then f' can go from negative to positive values only by going through 0. (This is a consequence of the Intermediate Value Theorem for continuous functions stated in Chapter 8.)

The statement $f'(x) = 0$ tells us that the slope of the tangent line at the transition point is 0, which means that at such a transition point, the graph of f has a horizontal tangent.

If f' changes continuously from positive to negative values as x passes from left to right through a point P, then the value of f at c is a local maximum value of f as shown in Figure 19a.14.

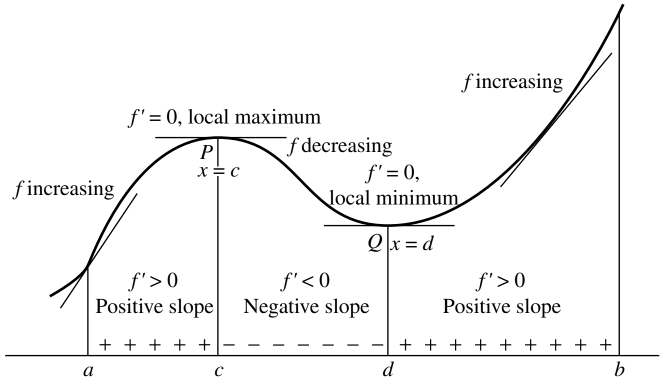


FIGURE 19a.14

That is, $f(c)$ is the largest value the function takes on in the immediate neighborhood of $x = c$. Similarly, if f' changes from negative to positive values as x passes from left to right through a point d, then the value of f at Q is a local minimum value of f. That is, $f(d)$ is the smallest value f takes in the immediate neighborhood of $x = d$.[(6)]

19a.4.1 A Horizontal Tangent Without a Maximum or a Minimum

Suppose $y = f(x)$ has a continuous derivative f' that changes sign as x passes through a point c, then we know that $f'(c) = 0$. However, a change in sign does not always occur when the derivative is zero. The curve may cross its horizontal tangent and keep on rising, as happens in the graph of $y = x^3$ at (0, 0) of (Figure 19a.7).

Similarly, the curve may cross its horizontal tangent and keep on falling, as $y = -x^3$ does at (0, 0). Neither function has a local maximum value or a local minimum value at $x = 0$, through $f'(0) = 0$. This situation arises because the function $y = x^3$ increases on the entire x-axis, and yet the first derivative $y' = 3x^2$ is 0 at $x = 0$. Since the first derivative does not change sign as x passes through the point 0, a local maximum does not exist at $x = 0$. For the same reason, a local minimum does not exist for the function $y = -x^3$ at $x = 0$.

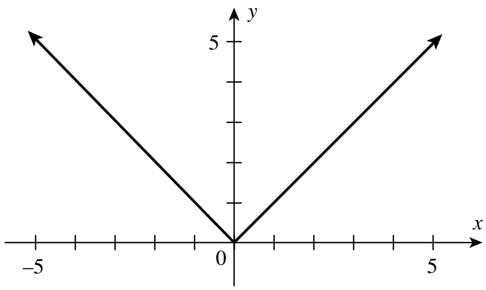
19a.4.2 A Local Maximum or Minimum Without a Horizontal Tangent

We give below an example of a function that is continuous on the interval on which it rises and falls, but the derivative fails to exist at the point of transition. In other words, a maximum or minimum may exist at a point (of transition) without a horizontal tangent.

Example (15): The function $y = |x|$ decreases on $(-\infty, 0)$, where $y' = -1$, and increases on $(0, \infty)$ where $y' = 1$. This function has no derivative at $x = 0$. The transition from negative slope to positive slope (i.e., from falling to rising) takes place at a point $x = 0$, where the derivative fails to exist (Figure 19a.15).

Note (7): We have seen in Example (4) (on page 554) that the function $y = 1/x^2$ rises on the interval $(-\infty, 0)$, and falls on the interval $(0, \infty)$. These intervals are separated by the point

(6) We will give more formal definitions of local maximum and local minimum when we will study the theory of maximum and minimum values of functions in Chapter 19b.



$f(x) = |x|$

FIGURE 19a.15

$x = 0$ that is the transition point. But, this point does not lie on the curve $y = 1/x^2$ (or that this function is not continuous at $x = 0$). Since the function $y = 1/x^2$ is not continuous at the point of transition ($x = 0$), the local maximum value of f does not exist at $x = 0$.

19a.5 CONCAVITY, POINTS OF INFLECTION, AND THE SIGN OF THE SECOND DERIVATIVE

Just as the first derivative gives information about the behavior of a function and its graph, so does the second derivative. In fact, the first and the second derivatives together tell us how the graph of a function is shaped.

Definition: Concave Up and Concave Down: The graph of a differentiable function $y = f(x)$ is concave down on an interval where y' decreases, and concave up on an interval where y' increases. But how do we check this?

If a function $y = f(x)$ has a second derivative as well as a first, we can apply the first derivative test to the (derived) function $f'(= y')$ as follows:

At any point in an interval, if $y'' < 0$, then y' decreases, and if $y'' > 0$, then y' increases in that interval.

We therefore have a test that we can apply to the formula $y = f(x)$ to determine the concavity of its graph. This test is called the second derivative test for concavity.

19a.5.1 The Second Derivative Test for Concavity

The graph of $y = f(x)$ is concave down on an interval where $y'' < 0$ and concave up on an interval where $y'' > 0$.

The idea is that if $y'' < 0$, then y' decreases as x increases and the tangent turns clockwise (Figure 19a.16a). Conversely, if $y'' > 0$, then y' increases as x increases and the tangent turns counterclockwise (Figure 19a.16b).[7]

We have the following definitions:

[7] It is easy to imagine that if y' is decreasing (i.e., the slope is decreasing), then the tangent will turn clockwise so that the graph will be concave down. Similarly, if y' is increasing, then the tangent will turn counterclockwise, so that the graph will be concave up.

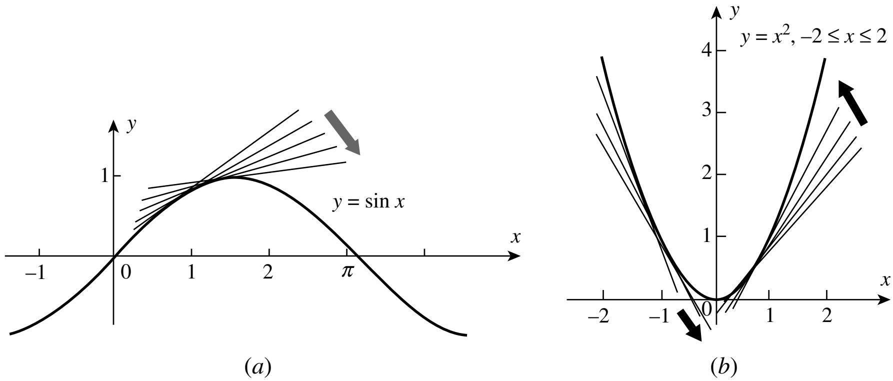


FIGURE 19a.16 (a) Concave down. The tangent turns clockwise as x increases; y' is decreasing. (b) The tangent turns counterclockwise as x increases; y' is increasing.

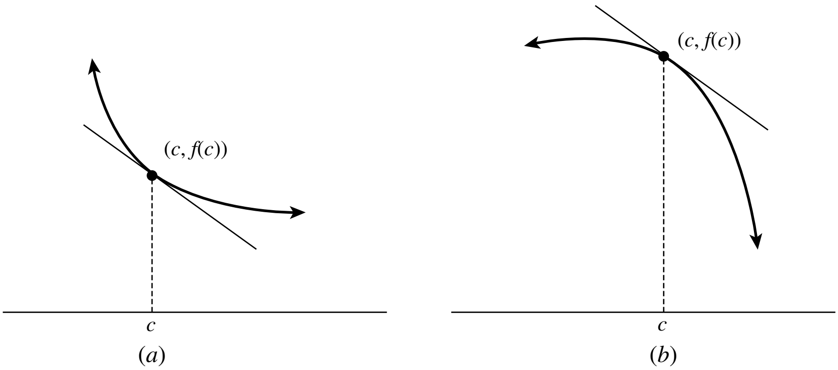


FIGURE 19a.17 (a) A portion of the graph of a function f. Concave upward at the point $(c, f(c))$. (b) A portion of the graph of a function f. Concave downward at the point $(c, f(c))$.

19a.5.1.1 Definition of Concave Upward (at a Point) The graph of a function f is said to be concave upward at the point $(c, f(c))$, if $f'(c)$ exists and if there is an open interval I containing c such that for all values of $x \neq c$ in I, the point $(x, f(x))$ on the graph is above the tangent line to the graph at $(c, f(c))$ (Figure 19a.17a).

19a.5.1.2 Definition of Concave Downward (at a Point) The graph of a function f is said to be concave downward at the point $(c, f(c))$, if $f'(c)$ exists and if there is an open interval I containing c such that for all values of $x \neq c$ in I, the point $(x, f(x))$ on the graph is below the tangent line to the graph at $(c, f(c))$ (Figure 19a.17b).

Example (16): Considering the function defined by $f(x) = x^2$, $f'(x) = 2x$ and $f''(x) = 2$. Thus, $f''(x) > 0$ for all x. Furthermore, because the graph of f, appearing in Figure 19a.18a, is above all of its tangent lines, the graph is *concave upward* at all of its points.(8)

(8) The conclusion that f must always be concave up tells us that the graph of $y = x^2$ must be as shown in Figure 19.18c and not as in Figure 19.18d, for in that situation there are intervals on which the curve is concave down. Thus, the concept of concavity is very useful in sketching curves.

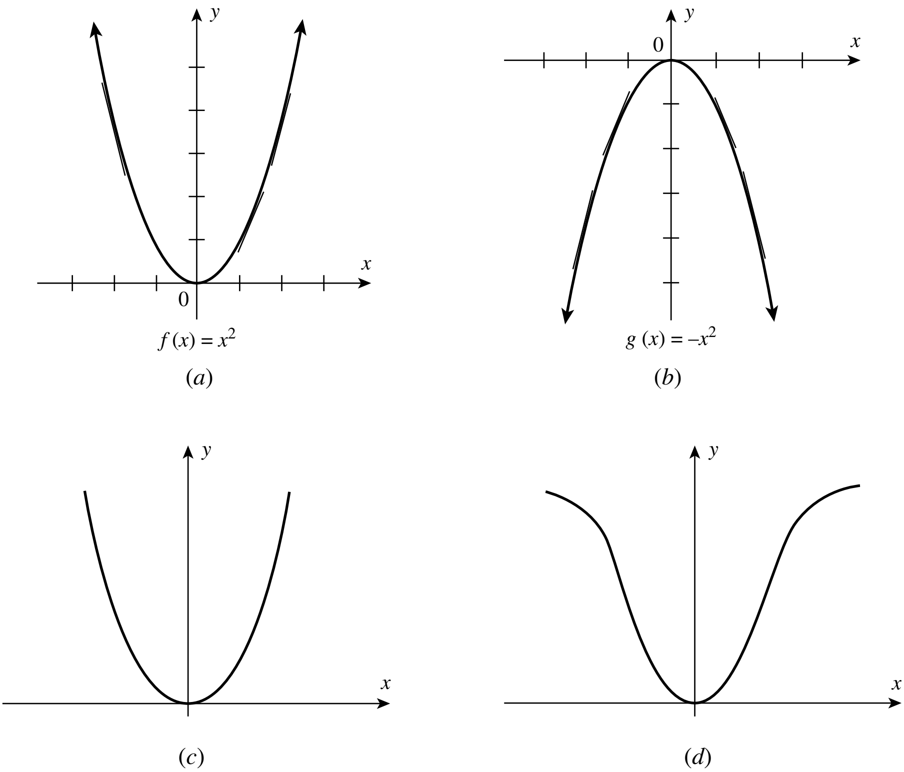


FIGURE 19a.18

Example (17): If g is a function defined by $g(x) = -x^2$, then $g'(x) = -2x$ and $g''(x) = -2$. Hence, $g''(x) < 0$ for all x. Also, because the graph of g, shown in Figure 19a.18b is below all its tangent lines, it is concave downward at all of its points.

Example (18): Consider the curve $y = f(x) = \sin x$, $0 < x < \pi$. We have $f'(x) = \cos x$ and $f''(x) = -\sin x$.

Note that, for any x in $(0, \pi)$, $f''(x) < 0$. Therefore, the curve f is concave down over the interval $(0, \pi)$.

19a.5.2 Point of Inflection

A point on a curve $y = f(x)$ where concavity changes from up to down or vice versa is called a point of inflection. See Figure 19a.19a and b for the point(s) of inflection on the curve.

In view of the above discussion, a point of inflection on a (twice-differentiable) curve is a point where y'' is positive on one side and negative on the other. If y'' is continuous, it implies that y'' must be 0 at a point of inflection.

Example (19): The curve $y = x^3$ has a point of inflection at $x = 0$, where $y'' = 6x$, changes sign, as x increases from negative to positive values (Figure 19a.19c).

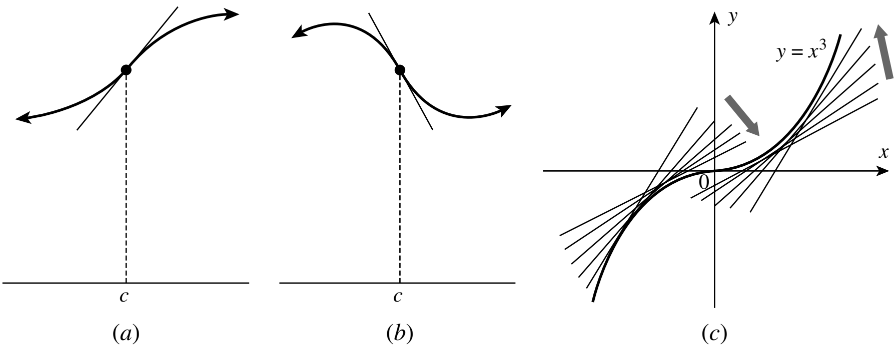


FIGURE 19a.19

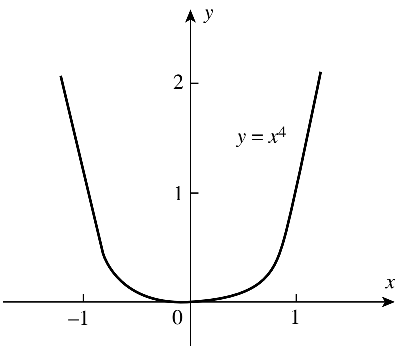


FIGURE 19a.20 The graph of $y = x^4$ has no inflection point at the origin even though $y''(0) = 0$.

Remark: There are functions for which the condition $y'' = 0$ does not confirm the existence of a point of inflection, as can be seen in Example (5). Besides, a point of inflection on a graph may occur where y'' fails to exists, as in Example (6).

Example (20): See Figure 19a.20. The curve $y = x^4$ has no point of inflection at $x = 0$ even though $y'' = 12x^2$ is 0 there. The second derivative does not change sign at $x = 0$ (in fact, y'' is never negative). The curve is *concave up* over the entire x-axis because $y' = 4x^3$ is an increasing function on $(-\infty, \infty)$.

Note (8): In the above example, the second derivative test for concavity is not satisfied (note that $y'' = 12x^2$ is positive for all $x \neq 0$). It follows that the condition $y'' = 0$ is a sufficient condition, it is not a necessary one.

Note (9): Most points of inflection occur at those points where $f''(x) = 0$, but a point of inflection may occur where $f''(x)$ is undefined, as Example (20) shows.

Example (21): In Figure 19a.21, the curve $y = x^{1/3}$ has a point of inflection at $x = 0$ even though the second derivative does not exist here. To see this, let us calculate y'' at $x \neq 0$. We have $y = x^{1/3}$, $y' = (1/3)x^{-(2/3)}$, $y'' = -(2/9)x^{-(5/3)}$.

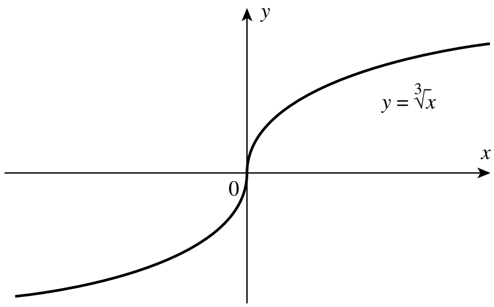


FIGURE 19a.21

For $x < 0, f''(x) > 0$, so that the curve is concave up and $f'(x)$ is increasing. On the other hand, for $x > 0, f''(x) < 0$, so that the curve is concave down and $f'(x)$ is decreasing. Thus, concavity changes as x passes through 0.

Therefore, the point (0, 0) is a point of inflection. However, $f''(x)$ does not exist at $x = 0$. Note that as $x \to 0, f''(x) \to \infty$. Yet the curve is concave up for $x < 0$ (where $y'' > 0$ and y' is increasing) and concave down for $x > 0$ (where $y'' < 0$ and y' is decreasing).

Note (10): It is important to understand clearly that a point of inflection separates a concave down arc from a concave up arc (or vice versa) of a curve. Of course, it is possible that the function under consideration may not be differentiable at the point of inflection, as we have seen in Example (21).

(Note that the tangent line is vertical at $x = 0$.) Now we are in a position to give the following definition.

19a.5.3 Definition of a Point of Inflection

The point $(c, f(c))$ is a point of inflection of the graph of the function f if the graph has a tangent line and if there exists an open interval I containing c, such that, if x is in I, then either

(i) $f''(x) < 0$ if $x < c$, and $f''(x) > 0$ if $x > c$, or
(ii) $f''(x) > 0$ if $x < c$ and $f''(x) < 0$ if $x > c$.

The existence of a point of inflection on different curves is indicated in Figure 19a.22.

Figure 19a.22a illustrates a point of inflection where the sense of concavity changes from downward to upward at the point of inflection.

In Figure 19a.22b, the sense of concavity changes from upward to downward at the point of inflection.

Figure 19a.22c gives another illustration, where the sense of concavity changes from downward to upward at the point of inflection.

Note that, in Figure 19a.22c, the graph has a horizontal tangent line at the point of inflection. Figure 19a.22d illustrates a point of inflection where the sense of concavity changes from upward to downward at the point of inflection. Note that in Figure 19a.22d, the graph has a vertical tangent line at the point of inflection.

Remark: A crucial part of the definition of the point of inflection is that the graph must have a tangent line. This will be clear from the following example.

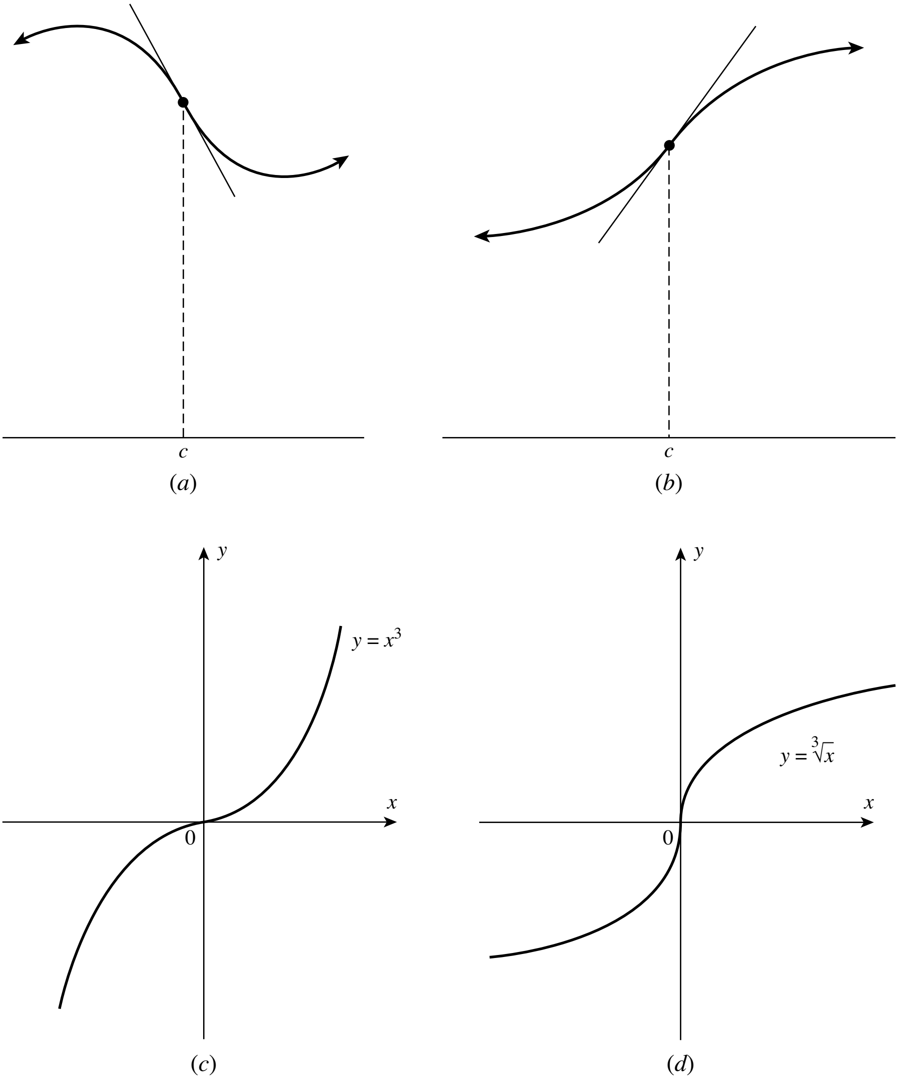


FIGURE 19a.22

Example (22): Consider the function f defined by

$$f(x) = \begin{cases} 4 - x^2, & \text{if } x \leq 1 \\ 2 + x^2, & \text{if } 1 < x \end{cases}$$

The graph of f appears in Figure 19a.23.

$$f(x) = \begin{cases} 4 - x^2, & \text{if } x \leq 1 \\ 2 + x^2, & \text{if } 1 < x \end{cases}$$

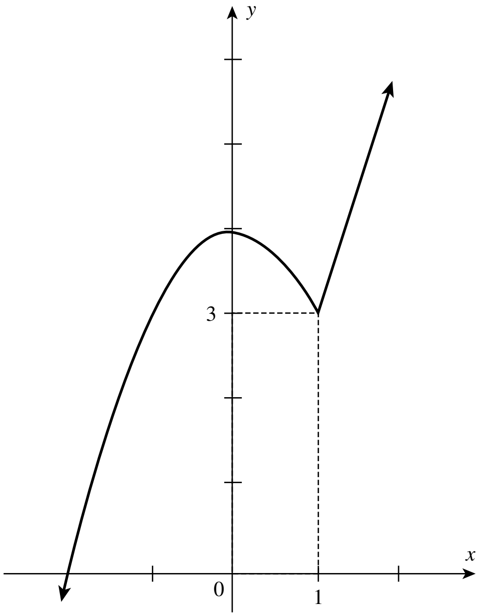


FIGURE 19a.23

Observe that,

$$f''(x) = -2, \quad \text{if } x < 1$$
$$f''(x) = 2, \quad \text{if } x > 1$$

Thus, at the point (1, 3) on the graph, the sense of concavity changes from downward to upward. However, (1, 3) is not a point of inflection because the graph does not have a tangent line.

Example (23): Let us discuss concavity and inflection points for the curve $y = f(x) = x^3 - 3x^2 + 2$.

Solution: We have $f'(x) = 3x^2 - 6x$ and $f''(x) = 6x - 6 = 6(x-1)$.

Now, for $f''(x) = 0$, we get,

$$6(x-1) = 0$$
$$\therefore \quad x = 1$$

We note that if $x < 1$, then,

$$6(x-1) < 0, \text{ that is, } f''(x) < 0$$

Therefore, the curve is concave down, if $x < 1$.
Next, if $x > 1$, then,

$$6(x-1) > 0, \text{ that is, } f''(x) > 0$$

Therefore, the curve is concave up, if $x > 1$.

Since the concavity changes as x increases through 1, the point (1, 0) is an inflection point on the curve.

Note (11): Standard textbooks may be referred to for exercises.

19b Maximum and Minimum Values of a Function

19b.1 INTRODUCTION

An important *application of derivatives* is to determine where a function attains *its maximum and minimum values*. The value of a function *f* at $x = x_0$, denoted by $f(x_0)$, is represented by the *height of its graph* at x_0. Thus, *maximum* and *minimum* values of a function are most *easily imagined in terms of the graph of a function*.

A function *f* has a maximum at the point x_1, if the value of the function at the point x_1 [i.e., $f(x_1)$] is greater than its values at all points of a *certain (small) interval containing the point* x_1. Similarly, we say that a function has a minimum at the point x_2, if the value of the function at the point x_2 [i.e., $f(x_2)$] is less than its values at all points of a *certain (small) interval containing the point* x_2.

In connection with *the above definitions* of *maximum* and *minimum* values of a function, note the following points carefully:

- *One should not think* that the *maximum* and *minimum* of a function are its respective *largest* and *smallest values* over the given interval. At a point of maximum, a function has *the largest value only in comparison* with those values that it has *at all points sufficiently close* to the point of maximum, and the *smallest value only in comparison with* those that it has *at all points sufficiently close* to the minimum point.
- The above discussion suggests that it is more appropriate to identify the *maximum* and *minimum* of a function by the terms *local maximum* and *local minimum*, respectively. The term *local extremum* stands to mean either the local maximum or the local minimum value of the function. (The terms *maxima*, *minima*, and *extrema* are the plurals of maximum, minimum, and extremum values, respectively.)
  To illustrate, consider the Figure 19b.1. Here is a function $y = f(x)$ defined on the interval $[a, b]$, which at $x = x_1$ and $x = x_3$ has a maximum, at $x = x_2$ and $x = x_4$ has a minimum, but the *minimum of the function at* $x = x_4$ is *greater than the maximum of the function* at $x = x_1$.

At this stage, we introduce the following terms and concepts that will be frequently used in this chapter.

Applications of Derivatives 19b-Maxima and minima: theory and problems (Investigating functions with the aid of derivatives for finding extremum values of a function, and the extreme value theorem)

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

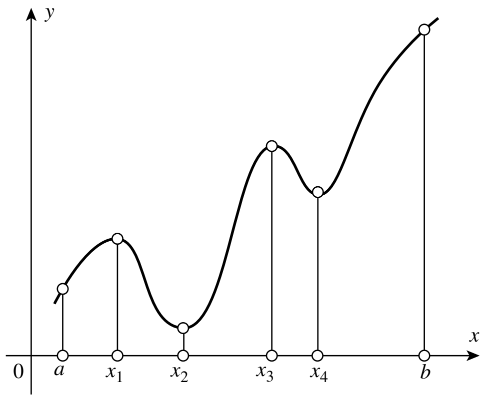


FIGURE 19b.1

- *Absolute Maximum (Minimum) of a Function*: In Figure 19b.1, note that at $x = b$, the value of the function is *greater than any maximum of* the function on the interval $[a, b]$. Thus, *the greatest value* of the function occurs at $x = b$, and similarly, the *smallest value* occurs at $x = x_2$. We say that *the absolute maximum* of *f* is $f(b)$ and *the absolute minimum* is $f(x_2)$.
- *The Points of Extreme Values of a Function*: The points like x_1, x_2, x_3, x_4, and *b at which the extreme values of the function f occur*, are called the points of extremum (or extreme) values of the function. Note that, a function defined on an interval can reach maximum and minimum values *only for the points that lie within the given interval.*[1]

  Our interest lies in finding the *points of extreme values* of a *continuous function* by using *the concept of the derivative*. Once such points are known, it is easy to compute the extreme values of the function and then select the *absolute extreme values, which have practical applications*, as will be clear from some solved examples.

  In the case of some functions, it is not difficult to find the points of extrema without using calculus, but it will be seen that in general it is not possible to find the extreme values without applying differential calculus.[2]

  The knowledge of such points is very useful in sketching the graph of a given function. Besides, these extreme values have many practical applications in widely varying areas such as engineering and various sciences, and so on.

19b.2 RELATIVE EXTREME VALUES OF A FUNCTION

The term *relative extreme values* is frequently used, to stand for local extreme values (including the absolute extreme values), in a broader sense, This is due to the fact that all extreme values of a function can be easily compared for relatively smaller or larger values.

19b.2.1 Classification of Relative Extreme Values of a Function

In Figure 19b.2, we indicate how maxima and minima are classified. The reader may note how the term "*relative extreme value*" is more general than the term "*local extreme value*".

(1) Later on, we will show (through examples) that an open interval may not have any point of extremum.

(2) If it were possible to draw easily the graph of any function accurately then we could easily find the extreme values of the function without using differential calculus. But, we know that this is not so simple.

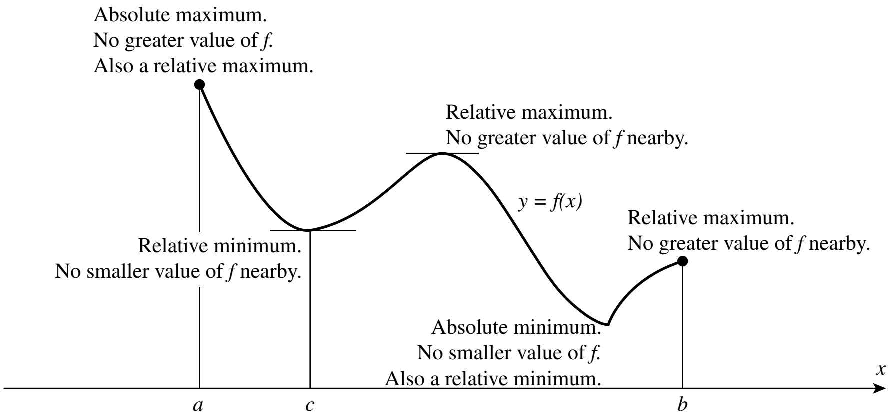


FIGURE 19b.2 Classification of relative extreme values of a function.

(Shortly, it will be seen how the term "relative extreme value" is more useful than the term "local extreme value".)

19b.2.1.1 Definition: Point of Relative Maximum A function f is said to have a *relative (or local) maximum* at $x = c$, if,

$$f(x) \leq f(c)$$

for all values of x in some open interval about c.

If c is an end point of the domain of f, the interval is to be half open, containing c as the end point. The interval might be small or it might be large, but no value of the function in the interval (under consideration) is greater than $f(c)$. *Note that, it is only at* $x = c$ *where* $f(x) = f(c)$. [Even when c is an end point of the interval, we have $\lim_{x \to c} f(x) = f(c)$.]

19b.2.1.2 Definition: Point of Relative Minimum A function f is said to have a *relative (or local) minimum* at $x = c$, if,

$$f(c) \leq f(x)$$

for all values of x in some open interval about c (or half open interval with c as an end point).

Note: The word relative (or local) is used to distinguish such a point from the point of *absolute maximum* (*or absolute minimum*). The precise definition of absolute extrema will be given later in the chapter.

We now give the following simplified definitions:

(i) *Definition of a Relative Maximum Value of a Function*: The function f has a *relative maximum value* at the point "c", if there exists some number $h > 0$, such that,

$$f(c) > f(x), \quad \text{for all } x \in (c - h,\ c + h) \tag{1}$$

The value $f(c)$ is called the *relative* (*or local*) *maximum* value of f.

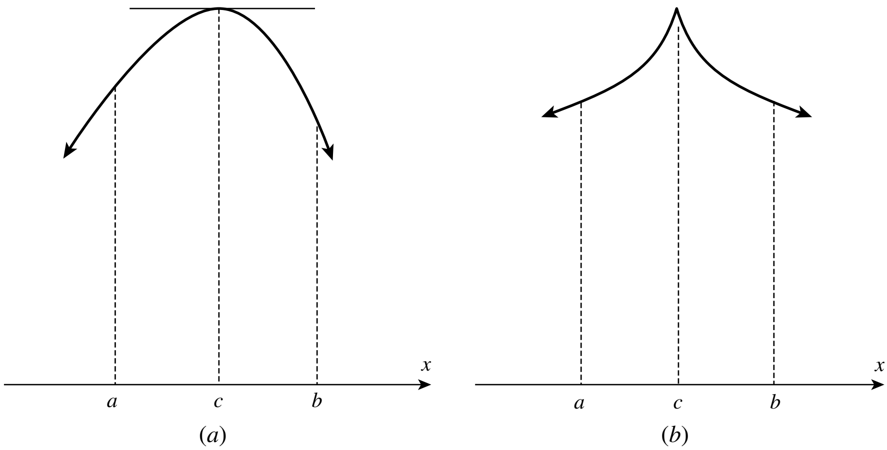


FIGURE 19b.3

Figures 19b.3a and 19b.3b show a portion of the graph of a function having a relative maximum value.

(ii) *Definition of a Relative Minimum Value of a Function*: The function f has a *relative minimum value* at the point "c", if there exists some number $h > 0$, such that,

$$f(c) < f(x), \quad \text{for all } x \in (c-h,\ c+h) \tag{2}$$

The value $f(c)$, in this case, is called the relative (or local) minimum value of f.

Figures 19b.4a and 19b.4b show a portion of the graph of a function having a relative minimum value.

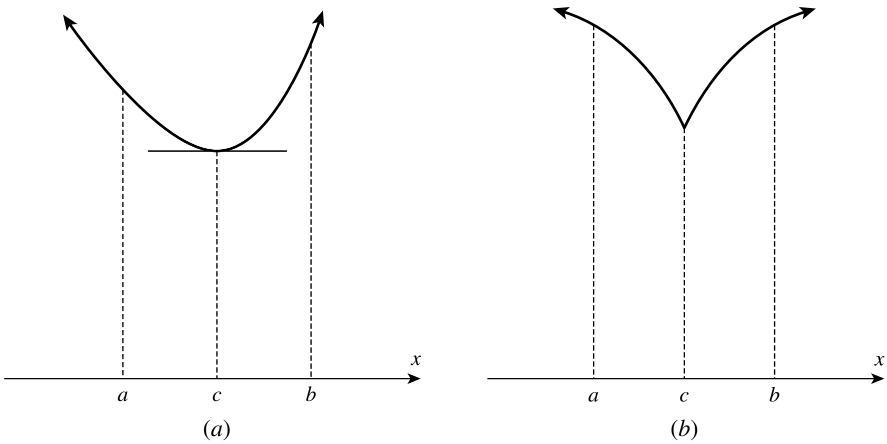


FIGURE 19b.4

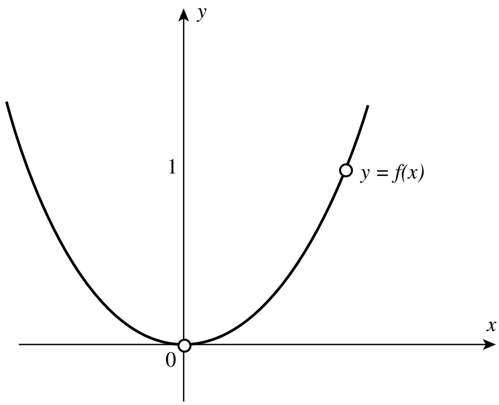


FIGURE 19b.5

Note: From the definitions at (i) and (ii) above, it must be clear that in the case of a *relative maximum* (*or minimum*) of a function *f*, the function must be defined in some open interval $(c - h, c + h)$, wherein, the *strict inequalities* (1) and (2) must hold.

Remark (1): Note that the above inequalities (1) and (2) will be satisfied even when *f* is *not continuous* at $x = c$, *but* $f(c)$ *is defined*. Such a relative extreme value is called "*strict maximum* (minimum) value at a point".

19b.2.1.3 Definition: Strict Maximum (Minimum) Value A function *f* is said to have a *strict maximum* (*minimum*) at the point $x = c$, if there holds the *strict inequality*,

$$f(x) < f(c) \quad [f(x) > f(c)]$$

for all values of x in some open interval $(c - h, c + h)$.

[*Here, we do not assume that f(x) is continuous at* $x = c$.]

Example (1): Consider the function,

$$f(x) = \begin{cases} x^2 & \text{for } x \neq 0 \\ 1 & \text{for } x = 0 \end{cases}$$

Note that, this function *f* is not continuous at $x = 0$, but it has a *relative maximum value at* $x = 0$ (more precisely, *f* has a *strict maximum value*). Here, we can define $f(x)$ in a way such that it has any desired *maximum* (*or minimum*) value at $x = 0$ (Figure 19b.5). [*Of course, in the case of a continuous function, such extreme value(s) cannot be chosen.*]

Remark (2): *When speaking of an extremum at a point* x_0*, we usually mean strict extremum at* x_0*, irrespective of whether the function is continuous at* x_0 *or not. The only important* *requirement is that the value $f(x_0)$ should be defined and be finite.* [Note that strict maximum (minimum) is also a relative maximum (minimum).][3]

The following theorem is used to locate the possible points (numbers) at which a function may have *relative extreme values.*

19b.3 THEOREM A

If (i) $f(x)$ exists *for all values of x in the open interval* (a, b), (ii) f has a relative extremum at c, where $a < c < b$, and (iii) $f'(c)$ exists, then $f'(c) = 0$.

We differ the proof of this theorem for the time being, but let us see what it says, and what it does not say.

In geometric terms, the theorem states that, *if f has a relative extremum at c, and if $f'(c)$ exists,* then the graph of f must have a *horizontal tangent line* at the point $x = c$.

(Observe that this situation prevails for the graphs in Figures 19b.3a and 19b.4a.)

Caution: Note carefully what the theorem says:

It says that $f'(c) = 0$ *at all those interior points* "*c*" *where f has a relative maximum or minimum* and $f'(c)$ *exists*.

- The theorem *does not say what happens if a relative maximum or minimum occurs at a point c where $f'(c)$ is not defined* [i.e., either $f'(c)$ is infinite or the point $(c, f(c))$ on the graph is a sharp point such that no unique tangent line can be drawn at that point].
- Also, *it does not say that f must have a relative maximum or minimum at every point* "*c*" *where $f'(c) = 0$.*

19b.3.1

The converse of theorem (A) does not hold.

It cannot be said that *there definitely exists a relative extremum for every value at which the derivative vanishes.*

Example (2): The function $y = f(x) = x^3$, whose graph appears in Figure 19b.6, has a derivative equal to zero, at $x = 0$.

$$(y')_{x=0} = (3x^2)_{x=0} = 0$$

But at this point, the function has neither a relative maximum nor a relative minimum. Indeed, no matter how close the point x is to zero, we will always have

$$x^3 < 0, \quad \text{when } x < 0$$
$$x^3 > 0, \quad \text{when } x > 0$$

This is the example of a *continuous function* that has derivative at each point in its domain.

(3) It must be noted that the term "relative extremum" is used for the following two comparisons:

(i) For comparing the values of function within a small neighbourhood of the point of extremum.

(ii) For comparing the maximum (or minimum) values of the function to select the absolute maximum (or minimum) values of the function.

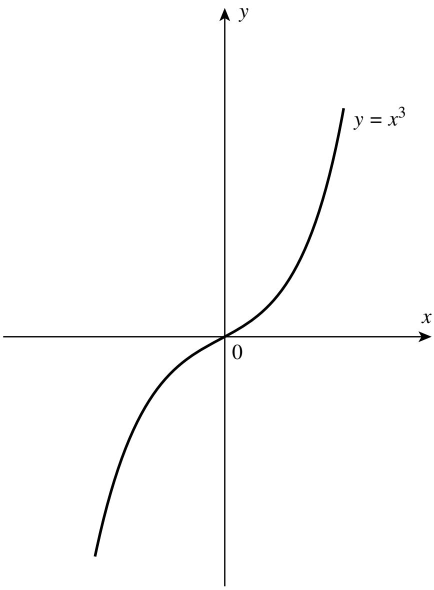


FIGURE 19b.6

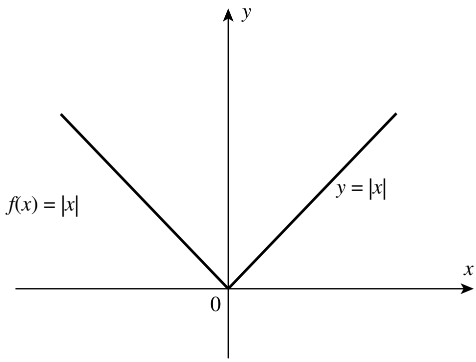


FIGURE 19b.7

Example (3): The function $y = |x|$ has no derivative at the point $x = 0$ (at this point the curve does not have a definite tangent line), but the *function has a relative minimum at this point*: $y = 0$ when $x = 0$. Note that, for any other point x different from zero, we have $y > 0$ (Figure 19b.7).

Example (4): Let the function f be defined by

$$f(x) = \begin{cases} 2x - 1 & \text{if } x \leq 3 \\ 8 - x & \text{if } 3 < x \end{cases}$$

The graph of this function appears in Figure 19b.8, showing that f has a relative maximum value at $x = 3$.

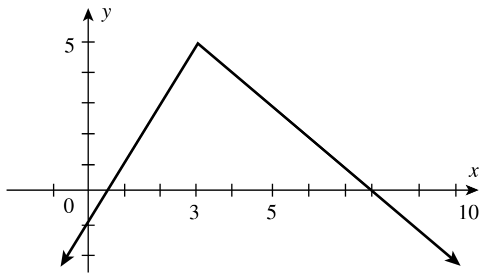


FIGURE 19b.8

The derivative from the left of $x = 3$ is given by $f'(x) = 2$ [note that $f'^{-}(3) = 2$], and the derivative from the right of $x = 3$ is $f'(x) = -1$ [note that $f'^{+}(3) = -1$]. Therefore we conclude *that $f'(3)$ does not exist, but still a relative maximum exists at $x = 3$.*

Note (3): It is possible that *a function f can be defined at a number c where $f'(c)$ does not exist and yet f may not have a relative extremum there*. The following example gives such a function.

Example (5): The function $y = \sqrt[3]{x} = \mathrm{x}^{1/3}$ does not have a derivative at $x = 0$. Since the derivative $y' = f'(x) = (1/3)x^{-2/3}$ approaches infinity as $x \to 0$, we say that $f'(x)$ *does not exist* at $x = 0$.

The domain of *f* is the set of all real numbers. Figure 19b.9 shows the graph of the function.

At this point the function has *neither a relative maximum nor a relative minimum. This is also clear from the fact that $f(0) = 0$, $f(x) < 0$ for $x < 0$ and $f(x) > 0$ for $x > 0$.*

Example (6): Figure 19b.10 shows the graph of $y = f(x) = x^{2/3}$, on $[-2, 3]$.

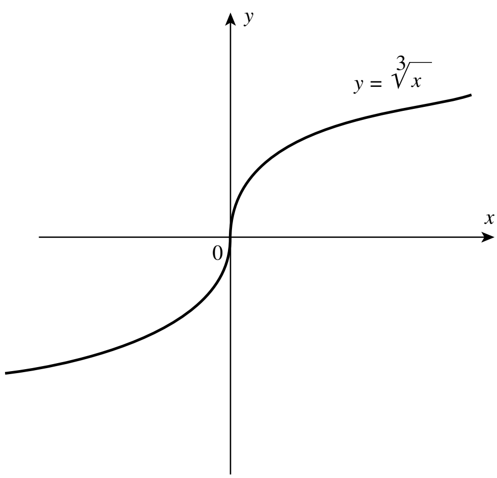


FIGURE 19b.9

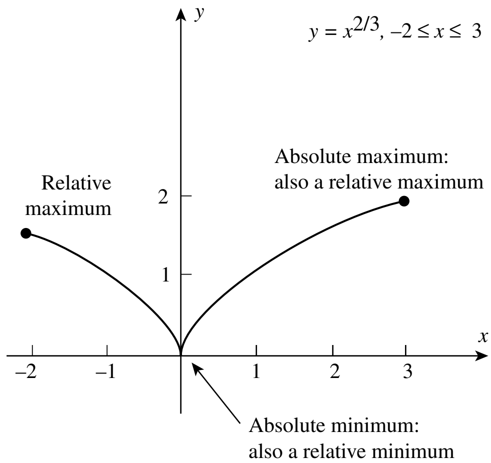


FIGURE 19b.10

The derivative $y' = f'(x) = (2/3)x^{-1/3} = 2/3\sqrt[3]{x}$ *does not exist at* $x = 0$, where y has its *minimum* value of zero. [Observe that, the curve has a vertical tangent at (0, 0), because $\lim_{x\to 0} f'(x) = \infty$.]

A function can have a minimum value at a point where its derivative does not exist. One way this can happen is shown here, where the curve has a vertical tangent at $x = 0$. Another way is shown in Figure 19b.7, where $|x|$ has no tangent at all at $x = 0$. Again, in Figure 19b.8, a relative maximum occurs at $x = 3$, where $f'(x)$ does not exist.

Remark (1): Example (5) shows that when a maximum or minimum occurs at the end of a curve, that exists only over a limited interval, the derivative need not vanish at such a point.

Remark (2): The Examples (2), (3), (4), and (5) demonstrate why the condition "$f'(c)$ exists," must be included in the hypothesis of Theorem A.

In summary, then, if a function f is defined at a number c, *a necessary condition for f to have a relative extremum there is that* $f'(c) = 0$ *or* $f'(c)$ *does not exist*. But, this *condition is not sufficient as we have seen in the above* examples. Before we discuss sufficient conditions for existence of a relative extrema, it is important to define the following terms.

19b.3.2 Definition of Critical Points of $f(x)$

If *c is a number in the domain of the function f*, and *if either* $f'(c) = 0$ or $f'(c)$ *does not exist*, then c is a critical point of f.

Thus, critical points include the *roots of the equation* $f'(x) = 0$, and the *numbers where $f'(x)$ does not exist*. (In particular, the numbers "c" where $\lim_{x\to c} f'(x) \to \pm\infty$ must be carefully checked for extreme values.) Note that, for the function $f(x) = x^{2/3}$, the derivative $f'(x) = (2/3)x^{-1/3}$ does not exist at $x = 0$, but $f(x)$ has its minimum value of zero at $x = 0$ (see Figure 19b.10). In fact, the curve has a vertical tangent at (0, 0). On the other hand, a *different situation exists* for the function, $f(x) = x^{1/3}$. Note that, here again, there is a vertical tangent at (0, 0) [meaning that $f'(x)$ does not exist there], *but no extremum exists at* $x = 0$ (see Example 4, Figure 19b.9).

19b.3.3 Stationary Point(s) of *f*(*x*)

The points where $f'(x) = 0$ are called the *stationary points of f(x)*, since the derivative $f'(x)$ that stands for the rate of change of the function *f*(*x*) (at such points) is zero. (A stationary point is a critical point but does not necessarily have a relative extrema.) Note the difference in terminology: a point of extremum of a function is a point lying on the axis along which the independent variables runs, while a point of inflection (discussed in earlier chapter) is a point lying on the curve itself.[(4)]

Note (4): From what has been said we conclude that *every critical point of a function need not have a relative extremum*. However, if at some point the function attains a relative extremum *then this point is definitely critical*. Therefore, to find the relative extrema of a function we must proceed as follows:

Find all the critical points, and then, investigate separately each critical point, to find out whether the function will have relative maximum (or a minimum) at that point.

Investigation of a function at critical points is based on the following theorem.

19b.4 THEOREM B: SUFFICIENT CONDITIONS FOR THE EXISTENCE OF A RELATIVE EXTREMA—IN TERMS OF THE FIRST DERIVATIVE

Let there be a *function f(x) continuous on some interval containing a critical point x_1 and differentiable at all points of the interval, with the exception, possibly, of the point x_1 itself.*

- If when moving from left to right through this point the derivative changes sign from plus to minus, then, at $x = x_1$, the function has a relative maximum.
- But, if when moving through the point x_1 from left to right, the derivative changes sign from minus to plus, the function has a relative minimum at $x = x_1$.

In other words,

$$\text{if} \quad \text{(a)} \quad \begin{cases} f'(x) > 0 & \text{when } x < x_1 \quad \text{and} \\ f'(x) < 0 & \text{when } x > x_1 \end{cases}$$

then at x_1 the *function has a relative maximum*; but

$$\text{if} \quad \text{(b)} \quad \begin{cases} f'(x) < 0 & \text{when } x < x_1 \quad \text{and} \\ f'(x) > 0 & \text{when } x > x_1 \end{cases}$$

then at x_1 the *function has a relative minimum.*

Note that, the conditions (a) and (b) *must be fulfilled for all values of x that are sufficiently close to x_1* (*i.e., at all points of some sufficiently small neighborhood of the critical point x_1*). Also, the theorem demands that $f'(x)$ may not exist at $x = x_1$ but *f* must be continuous at x_1.

[(4)] For more details, refer to *Mathematical Analysis* by A.F. Bermant and I.G. Aramanovich (pp. 202–203), Mir Publishers.

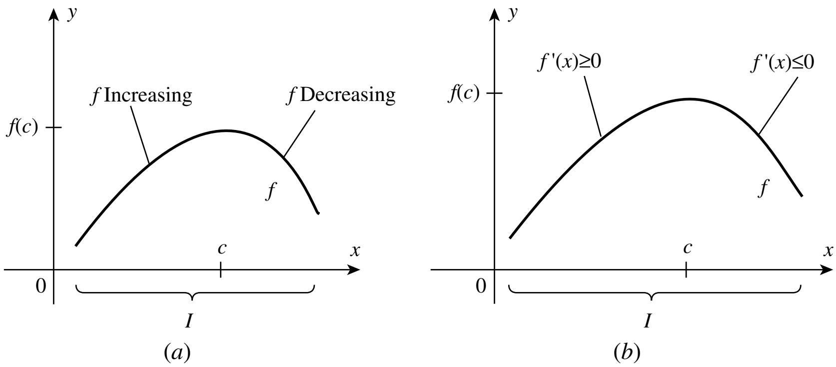


FIGURE 19b.11

Proof: Let $f'(x) > 0$ for $x < x_1$. This means that *on the left of the point x_1, there is an interval of increase of the function* $f(x)$ adjoining the point x_1.

If $f'(x) < 0$ for $x < x_1$, then *on the right of the point x_1, there is an interval of decrease of the function* adjoining the point x_1. *Consequently, x_1 is a point of (relative) maximum* (Figure 19b.11).

Other cases, *when the derivative changes its sign from negative to positive* as x passes through the point x_1 from left to right, are investigated quite similarly.[5]

In other words, if the derivative $f'(x)$ *changes sign as x passes through the point* x_1 (from left to right), the point x_1 is a point of *relative extremum* (Figure 19b.12). (If the derivative changes sign from positive to negative there is a *relative maximum* at the point x_1; if it changes from negative to positive it is *relative minimum* at the point x_1.)

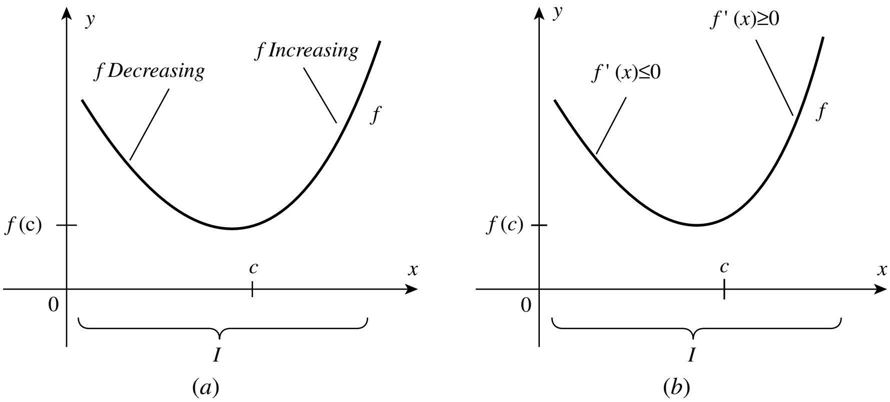


FIGURE 19b.12

[5] The theorem can also be proved by applying Lagrange's mean value theorem (introduced later in Chapter 20). Such a proof is very simple. [See *Differential and Integral Calculus*, Vol. 1 by Piskunove (p. 162), Mir Publishers.]

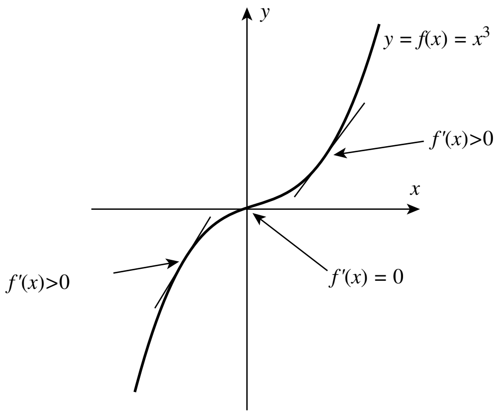


FIGURE 19b.13

It is clear *that if the derivative* $f'(x)$ *does not change sign as x passes through the point* x_1, *there is no relative extremum at the point* x_1. This can be seen from the behavior of the function $y = x^3$ in the vicinity of the point $x = 0$. Earlier, we have already discussed about this function (Figure 19b.6), however, we again give the graph of this function with extra supporting information for more clarity (Figure 19b.13).

Remark (3): For the sufficient conditions given by Theorem B to be *satisfied, it is important that the function f(x) be continuous at* $x = x_1$. It is important to note that *if it is only known that the derivative changes sign at a point, it is impossible to judge upon the existence of a (relative) extremum*, because *it is necessary to know additionally that the function is continuous at that point itself.*

For instance, take the function $y = (1/x^2)$. Its derivative $y = -(2/x^3)$ changes sign as x passes through the point $x = 0$:

$$y' > 0 \text{ for } x < 0 \quad \text{and} \quad y' < 0 \text{ for } x > 0$$

Consequently, *the function increases on the left of* $x = 0$ *and decrease on the right of* $x = 0$. *At the same time* $x = 0$ *is not a point of relative maximum of the function* since 0 is not in the domain of "f". It has an infinite discontinuity at that point (see Figure 19b.14).

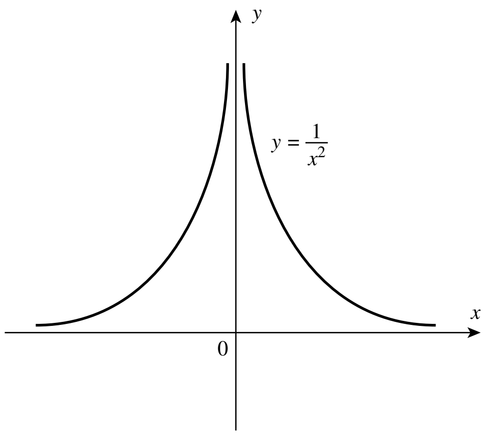


FIGURE 19b.14

19b.4.1 Scheme for Investigating Functions by Means of the Sufficient Condition for (Relative) Extremum in Terms of the First Derivative

The proceeding section permits us to formulate a rule for testing a *differentiable function*, $y = f(x)$, for *relative maximum* and *minimum*.

(1) Find the *first derivative* of the function, that is, $f'(x)$.

(2) Find the critical values of the argument x. To do this:

   (a) Equate the first derivative to zero and find the *real roots* of the equation $f'(x) = 0$.

   (b) Find the values of x at which the derivative $f'(x)$ is not defined.

(3) Let these critical points [obtained from (a) and (b)] be denoted, in an increasing order, as

$$x_1 < x_2 < \cdots < x_n$$

We *split the interval* $[a, b]$, in which the function is considered, into the subintervals.

$$(a,\ x_1),\ (x_1,\ x_2), \ldots,\ (x_{n-1},\ x_n),\ (x_n, b)$$

In Chapter 19a, we have seen that *the sign of $f'(x)$ remains unchanged in each such subinterval*. In other words, the sign of the derivative in each such subinterval may be either positive or negative. Thus, *these subintervals are the intervals of monotonicity of the function*. The sign of the derivative in each subinterval specifies the character of variation of the function in each subinterval.

It is now sufficient to investigate the sign of the derivative on left and right of each critical point x_i. The specification of the change of sign of the derivative (as x passes through the point x_i from left to right) indicates *which of these points give a relative maximum are and which points give a relative minimum.*

Note (5): It may also turn out that some of the points x_i are not points of (relative) extremum. This is the case when the derivative has the same sign in two adjoining subintervals separated by the point x_i (for instance, for the function $y = x^3$ the point $x = 0$ belongs to this type).

(4) The substitution of the critical values $x = x_i$ into $f(x)$ yields the corresponding values of the function:

$$f(x_1) < f(x_2) < \cdots < f(x_n)$$

each of which need not be a relative extremum. This gives us the following table of possible cases:

Sign of Derivative $f'(x)$ When Passing Through Critical Point x_1			
$x < x_1$	$x = x_1$	$x > x_1$	Character of Critical Point
$+$	$f'(x_1) = 0$ or $f'(x_1)$ is not defined	$-$	Point of relative maximum
$-$	$f'(x_1) = 0$ or $f'(x_1)$ is not defined	$+$	Point of relative minimum
$+$	$f'(x_1) = 0$ or $f'(x_1)$ is not defined	$+$	Neither a relative maximum nor a relative minimum. (Function increases throughout)
$-$	$f'(x_1) = 0$ or $f'(x_1)$ is not defined	$-$	Neither a relative maximum nor a relative minimum. (Function decreases throughout)

Note (6): It is possible to *establish another sufficient test for relative extremum* with the aid of the *second derivative* $f''(x)$ of the function $f(x)$ under investigation. It will be found that *the second derivative test (stated below) sometimes proves simpler and more convenient* than the one in the foregoing section.

> In what follows *we assume that in a neighborhood of a given point x_1 the function $f(x)$ itself and its first and second derivatives are continuous.*

19b.5 SUFFICIENT CONDITION FOR RELATIVE EXTREMUM (IN TERMS OF THE SECOND DERIVATIVE)

Theorem C: If the first derivative vanishes at the point x_1 $[f'(x_1) = 0]$ while the second derivative is different from zero $[f''(x_1) \neq 0]$, then x_1 is a point of relative extremum.[6]

Furthermore, if $f''(x_1) < 0$, the point x_1 is a point of relative maximum, if $f''(x_1) > 0$, the point x_1 is a point of relative minimum.

Proof: Let $f'(x_1) = 0$ and $f''(x_1) > 0$.
By the hypothesis, the second derivative is continuous, and therefore its sign is retained in a neighborhood of the point x_1. It follows that the function $f'(x_1)$ increases in this neighborhood because its derivative $(f'(x_1))' = f''(x_1)$ is positive (by assumption).

Further, since $f'(x_1) = 0$, the derivative $f'(x_1)$ assumes values less than $f'(x_1) = 0$ *on the left of the point x_1 and is therefore negative*:

$$f'(x_1) < 0 \quad \text{for } x < x_1$$

Similarly, on the right of the point x_1, its values are greater than $f'(x_1) = 0$, that is, its values are positive:

$$f'(x_1) > 0 \quad \text{for } x > x_1$$

Hence, as x passes through the point x_1 from left to right, *the function $f'(x_1)$ changes sign from negative to positive* and therefore, according to the foregoing test for extremum (in terms of the first derivative), *x_1 is a point of relative minimum of the function $f(x)$.*

An analogous arguments shows that, if $f''(x_1) < 0$ the function $f'(x)$ decreases and changes its sign from positive to negative as x passes through the point x_1, means that x_1 is a point of relative maximum of the function $f(x)$.

Remark (1): If both $f'(x_1) = 0$ and $f''(x_1) = 0$, the second derivative test is inapplicable, *and one should resort to the first derivative test*. For instance, consider the following examples which will make this clear.

[6] It follows that the second derivative test for relative extremum is applicable to a function f, if $f(x_1)$, $f'(x_1)$ and $f''(x_1)$ are continuous at $x = x_1$, and $f''(x_1) \neq 0$.

Example (7): Test for (relative) maximum and minimum the functions

$$y = x^4 \quad \text{and} \quad y = x^3$$

Solution: The first and the second derivatives of the function $y = x^4$ turn into zero at the point $x = 0$. Therefore, the second derivative test is inapplicable while *the first test indicates that there is a (relative) minimum* at that point, since the derivative $y' = 4x^3$ changes sign from negative to positive as x passes through the origin, from left to right.

At the same time, the function $y = x^3$ whose first and second derivatives also vanish at the point $x = 0$, has no relative extremum at the origin. This is so, because *its first derivative does not change sign as x passes through the point x = 0.*

Example (8): Test the following function for (relative) maximum and minimum

$$f(x) = 1 - x^4$$

Solution:

(1) Find the critical points:

$$f'(x) = -4x^3$$
$$-4x^3 = 0 \text{ gives } x = 0$$

(2) Determine the sign of the second derivative at $x = 0$.

$$f''(x) = -12x^2$$
$$\text{Now, } [f''(x)]_{x=0} = 0$$

It is thus impossible here to determine the character of the critical point by means of the sign of the second derivative (Figure 19b.15).

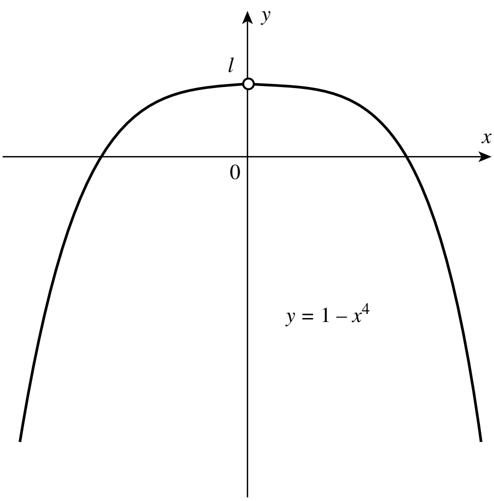


FIGURE 19b.15

(3) Investigate the *character of the critical point by the first derivative test*

$$[f'(x)]_{x<0} > 0, \quad [f'(x)]_{x>0} < 0$$

Consequently, at $x = 0$, the function has a (relative) maximum, namely $[f'(x)]_{x=0} = 1$.

Remark (2): Whenever the second derivative test is applicable, it proves extremely convenient, *since, it does not require the determination of the sign of the function $f'(x)$ at points different from the point at which the given function is tested for relative extremum, and this makes it possible to judge upon the existence of the relative extremum by the sign of the function $f''(x)$ at the same point.*

Now, we give below two examples in which it is tedious to obtain $f''(x)$ or inconvenient to calculate it. In such cases, checking the change of sign of $f'(x)$ gives a quicker result in classifying the *critical values.*

Example (9): Show that the function

$$x^5 - 5x^4 + 5x^3 - 1$$

has a maximum when $x = 1$, a minimum when $x = 3$, and neither when $x = 0$.

Solution: Let $y = f(x) = x^5 - 5x^4 + 5x^3 - 1$

$$\therefore \ \frac{dy}{dx} = f'(x) = 5x^4 - 20x^3 + 15x^2$$

$$= 5x^2(x^2 - 4x + 3)$$
$$= 5x^2[x^2 - 3x - x + 3]$$
$$= 5x^2[x(x-3) - 1(x-3)]$$
$$= 5x^2(x-3)(x-1) \qquad (3)$$

$$\text{and} \quad f''(x) = 20x^3 - 60x^2 + 30x = 10x[2x^2 - 6x + 3] \qquad (4)$$

For *critical values* of $f(x)$, we must have $dy/dx = f'(x) = 0$.

That is, $5x^2(x - 3)(x - 1) = 0$, which gives $x = 0$, 1, and 3 (i.e., *the critical values*) at which $f(x)$ may have possible maxima or minima or neither.

When $x = 1$, we have from (4),

$$f''(x) = 10[2-6+3] = -10 < 0$$
$$\therefore \ f(x) \text{ is maximum for } x = 1$$

When $x = 3$, we have

$$f''(x) = 10(3)[2(3)^2 - 6(3) + 3]$$
$$= 30[18 - 18 + 3] = 90 > 0$$
$$\therefore \ f(x) \text{ is minimum for } x = 3$$

Finally, *when $x = 0$*, we get

$$f''(x) = 10x(2x^2 - 6x + 3) = 0$$

Hence, the test fails. Therefore, we use the first derivative test [i.e., to find whether $f'(x)$ changes sign as x increases through $x = 0$].

$$\text{We have } f'(x) = 5x^2(x-1)(x-3)$$

when x is *slightly less than zero*, we have from (3),

$$f'(x) = (5)(+)(-)(-) = +$$

and when x is *slightly greater than zero*,

$$f'(x) = (5)(+)(-)(-) = +$$

Thus, we observe that $f'(x)$ does not change sign as x increases through 0. Hence, the function $f(x)$ is neither maximum nor minimum at $x = 0$.

$\therefore$ The critical value $x = 0$ is *a point of inflection* on the curve, its coordinates being $(0, -1)$.

Note (7): The maximum and minimum values of the function, on putting $x = 1$ and $x = 3$, respectively, in $f(x) = x^5 - 5x^4 + 5x^3 - 1$, are 0 and -28. One might check and convince himself that $f(x)$ cannot have value greater than zero and less than -28 for any value of x.

Example (10): Show that the function

$$f(x) = \frac{(x+1)^2}{(x+3)^3}$$

has a maximum value 2/27 and a minimum value zero.

Solution: Let $y = f(x) = \dfrac{(x+1)^2}{(x+3)^3}$

To find $dy/dx = f'(x)$*, it is convenient to take logarithms first.*

Thus, $\log_e y = 2\log_e(x+1) - 3\log_e(x+3)$

Differentiating both sides w.r.t. x, we get

$$\frac{1}{y}\cdot\frac{dy}{dx} = \frac{2}{x+1} - \frac{3}{x+3}$$

$$= \frac{2(x+3) - 3(x+1)}{(x+1)(x+3)} = \frac{(3-x)}{(x+1)(x+3)}$$

$$\therefore \frac{dy}{dx} = y\cdot\frac{(3-x)}{(x+1)(x+3)} = \frac{(x+1)^2}{(x+3)^3}\cdot\frac{(3-x)}{(x+1)(x+3)}$$

$$= \frac{(x+1)(3-x)}{(x+3)^4} \tag{5}$$

Now, observe that it is tedious to obtain $d^2y/dx^2 = f''(x)$. Therefore, we choose to check the change of sign of $f'(x)$, as x increases through the *critical values* [i.e., $x = -1$, $x = 3$, which are in the domain of $f'(x)$]. Note that $dy/dx = 0$, when $x = -1, 3$.

First, consider $x = -1$

when x is *slightly less than* -1, we see from (1), that

$$\frac{dy}{dx} = \frac{(x+1)(3-x)}{(x+3)^4} = \frac{(-)(+)}{+} = -$$

and when x is *slightly greater than* -1, then

$$\frac{dy}{dx} = \frac{(+)(+)}{+} = +$$

Thus, dy/dx *changes sign from negative to positive as* x *increases through* -1.

$\therefore$ $y = f(x)$ is *minimum* for $x = -1$, and this minimum value obtained by putting $x = -1$ in the expression for $y = f(x)$, is zero.

Next, consider $x = 3$

when x is *slightly less than* 3, we have, from (5)

$$\frac{dy}{dx} = \frac{(x+1)(3-x)}{(x+3)^4} = \frac{(+)(+)}{+} = +$$

and when x is *slightly greater than* 3, we get

$$\frac{dy}{dx} = \frac{(+)(-)}{+} = -$$

so that dy/dx changes sign from positive to negative as x *increases* through 3.

$\therefore$ $y = f(x)$ is maximum for $x = 3$ and its maximum value, on putting $x = 3$ in the expression for y, is

$$\frac{(3+1)^2}{(3+3)^3} = \frac{2 \times 2 \times 2 \times 2}{6 \times 6 \times 6} = \frac{2}{27} \qquad \text{Ans.}$$

Note (7): In the above example, the derivative $dy/dx = ((x+1)(3-x))/(x+3)^4$ is not defined at $x = -3$, *hence* $x = -3$ *is a critical value that must be investigated for existence of extrema*. But we also observe that the function $y = (x+1)^2/(x+3)^3$ is not continuous at $x = -3$, since, y is not defined for $x = -3$. In other words, $x = -3$ is not in the domain of the function and so this point is not to be considered for extreme values.

Example (11): We will show that the maximum value of $(1/x)^x$ is $e^{1/e}$.

Solution: Let $y = (1/x)^x$

$$\therefore \log_e y = x \log_e(1/x) = x \log_e(x)^{-1}$$

$$= -x \log_e x$$

$$\therefore \frac{1}{y} \cdot \frac{dy}{dx} = -\left[x \cdot \frac{1}{x} + \log_e x\right]$$

$$= -[1 + \log_e x]$$

$$\therefore \frac{dy}{dx} = -y[1 + \log_e x]$$

$$\text{or } \frac{dy}{dx} = -\left(\frac{1}{x}\right)^x (1 + \log_e x) \qquad (6)$$

Equating dy/dx to zero, we obtain $(1 + \log_e x) = 0$

$$\therefore \quad \log_e x = -1 \quad \therefore \quad x = e^{-1} = 1/e$$

Note that 1/e is a positive number less than 1.

(Now we have to investigate the critical value $x = 1/e$ for existence of extreme value.)

When x is slightly less than 1/e (which means that the value of $\log_e x$ is slightly toward -2 from -1), $(1 + \log_e x) < 0$ and we have, from (6)

$$\frac{dy}{dx} = -\left(\frac{1}{x}\right)^x (1 + \log_e x) = (-)(+)(-) = +$$

and when x is slightly greater than 1/e (the value of $\log_e x$ is slightly toward 0 from -1), $(1 + \log_e x) > 0$ and we have,

$$\frac{dy}{dx} = -\left(\frac{1}{x}\right)^x (1 + \log_e x) = (-)(+)(+) = -$$

Thus, dy/dx changes sign from positive to negative as x *increases* through the value 1/e. Hence y is maximum for $x = 1/e$, and this *maximum* value is given by

$$\left(\frac{1}{1/e}\right)^{1/e} = e^{1/e} \qquad \textbf{Ans.}$$

19b.6 MAXIMUM AND MINIMUM OF A FUNCTION ON THE WHOLE INTERVAL (ABSOLUTE MAXIMUM AND ABSOLUTE MINIMUM VALUES)

We are frequently concerned with a function defined on a given interval, and we wish to find the *largest or smallest value of the function on the interval*. These intervals can be either closed, open, or closed on one end and open at the other. We now give the precise definitions of the absolute extreme values of a function.

(a) *Definition of an Absolute Maximum Value on an Interval*: The function f has an *absolute maximum value on an interval if there is some number c in the interval* such that

$$f(c) \geq f(x), \quad \text{for all } x \text{ in the interval}$$

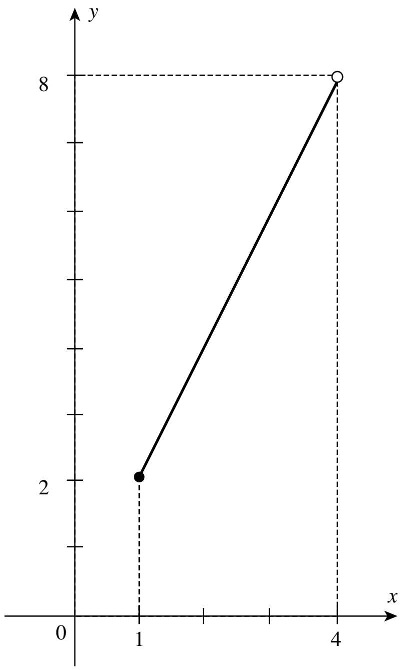
The number $f(c)$ is then *the absolute maximum value* of f on the interval.

(b) *Definition of an Absolute Minimum Value on an Interval*: The function f has an *absolute minimum value on an interval if there is some number c in the interval* such that

$$f(c) \leq f(x), \quad \text{for all } x \text{ in the interval}$$

The number $f(c)$ is then *the absolute minimum value* of f on the interval. (If a function has either an *absolute maximum value* or *an absolute minimum value* on an interval, then the function is said to have an *absolute extremum* on that interval.)

A function may or may not have an absolute extremum on a particular interval. In each of the following examples, a function and an interval are given, and we find the *absolute extrema* of the function on the interval, if there is any.



$f(x) = 2x$, for $1 \leq x < 4$

FIGURE 19b.16

Example (12): Consider the function defined by $f(x) = 2x$

The graph of f on $[1, 4)$ is sketched in Figure 19b.16, This function has the minimum value of 2 on $[1, 4)$. There is *no maximum value* of f on $[1, 4)$ because $\lim_{x\to 4} f(x) = 8$, but $f(x)$ is always less than 8 on the interval.

On the other hand, let us imagine the graph of the function

$$f(x) = 2x \quad \text{for } 1 < x \leq 4$$

It has the maximum value of 8 on $(1, 4]$ but there is no minimum value of the function. Again, the function $f(x) = 2x$ defined on $(1, 4)$ *has neither maximum value nor minimum value.*

Example (13): Consider the function defined by $f(x) = -x^2$

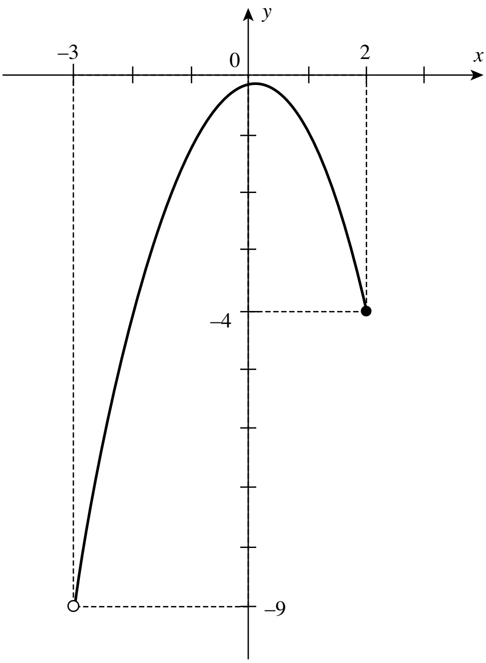
The graph of f on $(-3, 2)$ appears in Figure 19b.17. This function has an *absolute maximum value* on $(-3, 2)$. There is *no absolute minimum value* of f on $(-3, 2]$ because $\lim_{x\to -3^+} = -9$ but $f(x)$ is *always greater than* -9 *on the given interval.* Of course, there *is relative minimum value* of -4 at $x = 2$.

Note that, in this example a relative minimum occurs at an end points of the interval.

Example (14): Consider the function defined by $f(x) = x^2$, $x \in [0, \infty)$

This function has the *absolute minimum value* of 0, at $x = 0$.

It *does not have the absolute maximum value, since the function can attain any positive value.* (Here, f is defined on an *unbounded interval.*)



$f(x) = -x^2, x \in (-3, 2)$

FIGURE 19b.17

Further, the function,

$$f(x) = x^2 \quad 1 \leq x < 3$$

has the absolute minimum value at 1, but there is *no absolute maximum value*. (Why?) *On the other hand, the function,*

$$f(x) = x^2 \quad 1 \leq x \leq 3$$

has both *the absolute minimum value of 1 at* $x = 1$ *and the absolute maximum value of 9 at* $x = 3$.

In contrast, consider the function,

$$f(x) = \begin{cases} x^2 & \text{for } 1 \leq x < 3 \\ 5 & \text{for } x = 3 \end{cases}$$

Here, f has the *absolute minimum* value of 1 at $x = 1$, but there is *no absolute maximum value. (Why?)*

Note that, for $x = 3, f(x) = 5$; but *there are infinitely many points less than 3, in the interval* $[1, 3]$, for which $f(x) > 5$. However, *it is not possible to choose a single point at which f has the maximum value*. Also, note that this function is *defined on the closed interval* $[1, 3]$, *but it is discontinuous at* $x = 3$, *which is an end point of the interval.*

Remark: This example shows that a function defined on a closed interval may not attain the absolute extremum, *if it is discontinuous*, anywhere in the interval—including the end point. On the other hand, there are examples showing that a discontinuous function defined on an open interval may have both an absolute maximum and an absolute minimum value.

The above examples suggest that we can be much more precise about *possible extreme values* if the function f is continuous and the domain S is a closed interval. The *extreme value theorem* answers the existence question for some of the problems that come up in practice.

19b.6.1 The Extreme Value Theorem

If the function f is *continuous* on the *closed interval* $[a, b]$, then f has an *absolute maximum value and an absolute minimum value on* $[a, b]$.[7]

Note the key words; *f is required to be continuous and the set S is required to be a closed interval.*

Remark: The extreme value theorem states that the continuity of a function on a closed interval is a *sufficient condition to guarantee* that the function *has both an absolute maximum value and an absolute minimum value* on the interval. However, *it is not a necessary condition*. For example, the function whose graph appears in Figure 19b.18, has an absolute maximum value at $x = c$ and an absolute minimum value at $x = d$, even through the function is discontinuous on the open interval $(-1, 1)$.

An absolute extremum of a function continuous on a closed interval must be either a relative extremum or a function value at an end point of the interval.

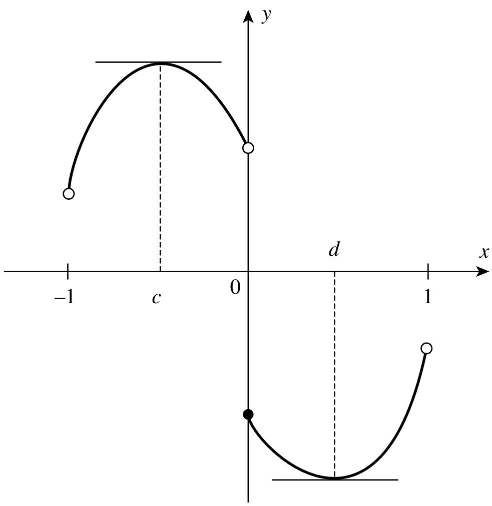


FIGURE 19b.18

[7] Though this theorem is intuitively obvious, a rigorous proof is quite difficult. The proof of this theorem can be found in an advanced calculus text. One such reference is *Calculus with Analytic Geometry* (Alternate Edition) by Robert Ellis and Denny Gulik, HBT Publication.

Since a necessary condition for a function to have a relative extremum at a number "c" is for c to be a critical number, the absolute maximum value and the absolute minimum value of continuous function f on a closed interval $[a, b]$ can be determined by the following procedure:

(1) Find the function values at the critical numbers of f on (a, b).
(2) Find the values $f(a)$ and $f(b)$.
(3) The largest of the values from steps 1 and 2 is the absolute maximum value, and the smallest of the values is the absolute minimum value.

Exercise (1)

Q1. Test for maximum and minimum of the function $y = x^6$

Ans. The function has minimum at $x = 0$.

Q2. Test for maximum and minimum of the function $y = (x - 1)^3$

Ans. The function has neither a maximum nor a minimum.

Q3. To find the greatest and the least values of $x^3 - 18x^2 + 96x$, in the interval $[0, 9]$

Ans. The greatest value $= 160$ and the least value $= 0$.

Q4. To find the greatest and the least values of $3x^4 - 2x^3 - 6x^2 + 6x + 1$ in the interval $[0, 2]$

Ans. The greatest value $= 21$, and the least value $= 1$.

Q5. Prove that x^x has minimum value at $x = 1/e$, and the minimum value is $(1/e)^{1/e}$

Q6. Find the maximum value of $\frac{\log x}{x}$

Ans. Maximum value $= 1/e$.

Q7. Prove that the maximum value of $\sin x + \cos x$ is $\sqrt{2}$

It is possible to give a step-by-step procedure for solving word problems concerning maximum and minimum. *Of these steps the most important step is to express the quantity (to be maximized or minimized) as a function f of the other quantity. We now proceed to discuss such applied problems.*

19b.7 APPLICATIONS OF MAXIMA AND MINIMA TECHNIQUES IN SOLVING CERTAIN PROBLEMS INVOLVING THE DETERMINATION OF THE GREATEST AND THE LEAST VALUES

By using techniques that we learnt (for finding where a function attains its maximum and minimum (i.e., extreme) values), we can examine situations in science, business, and economics that require determining the value of a variable, which will maximize or minimize a function.

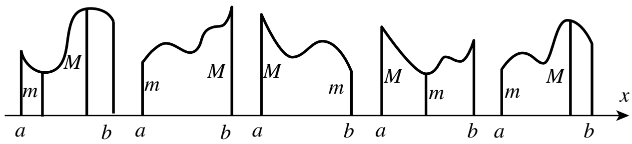


FIGURE 19b.19

In examining such situations, we are concerned with the problems in which the solution is an absolute extremum of a function.

Of course, the extreme value theorem assures us that a function continuous on a closed interval $[a, b]$ has both *an absolute maximum value* and *an absolute minimum value* on the interval.

However, we have seen the graph of a function (Figure 19b.18), which has *an absolute maximum value* at $x = c$ and *an absolute minimum value* at $x = d$, even though the function is *discontinuous on the* open interval $(-1, 1)$. It is important to remember that *the greatest value* M (the least value m) of the function on the interval $[a, b]$ *is either one of its relative maximum (minimum) values or an endpoint value. Some of the possible cases are shown in* Figure 19b.19.

It is also clear that when a function $y = f(x)$ is *monotonic*, in a closed interval $[a, b]$, its *greatest value is f(b) and the least value is f(a), if the function increases and conversely, the greatest value is f(a) and the least value is f(b), if the function decreases.*

It often occurs that a given function has *only one point of extremum in an interval.* In this case, the value of the function at that point is the greatest (an absolute maximum) on the interval in the case of relative maximum, and the least (an absolute minimum) in the case of relative minimum. Thus, we can deal with applications involving absolute extremum, even when the extreme value theorem cannot be employed. The following theorem is sometimes useful to determine if a relative extremum is an absolute extremum.

Theorem D: Suppose the function f is continuous on the interval I containing the number c. If $f(c)$ is a *relative extremum of f on I* and c is the only number in I for which f has a *relative extremum*, then $f(c)$ is *an absolute extremum of f on I.*[(8)]

Suppose there are two magnitude connected by a functional relationship, and it is required to find the value of one of them (*belonging to an interval that can be finite or infinite*) for which the other magnitude assumes its *least* or *greatest* value among all the possible values. *To solve such a problem we must find the expression of the function describing the relationship between the magnitudes in question and then determine* the *least* or the *greatest* value of this function on the given interval.

Example (15): Let us determine the *least* length l of the fence enclosing a rectangular plot of land with given area s adjoining a wall.

Solution: Denoting by x one of the sides of the rectangular plot of land with given area s adjoining a wall (see Figure 19b.20) we readily obtain,

$$s = x(l - 2x) \quad \text{where } l = 2x + s/x \tag{7}$$

The problem now reduces to finding the least value of this function as x ranges from 0 to ∞.

[(8)] We accept this theorem without proof. For proof of Theorem "D", refer to *The Calculus – 7 of a Single Variable* by Louis Lethold (p. 288), Harper Collins.

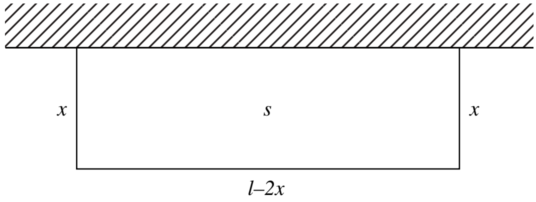


FIGURE 19b.20

Note from (7) that $l \to \infty$ for both $x \to 0$ and $x \to \infty$, *the least value of the function l must be among the minimum values of l, for some x in the interval* $(0, \infty)$.

We find the derivative,

$$\frac{dl}{dx} = 2 - \frac{s}{x^2}$$

Now, $dl/dx = 0$ gives $2 - s/x^2 = 0$ or $x = \sqrt{s/2}$

It follows that in the interval in question there is only one *stationary point* $x = \sqrt{s/2}$ at which the function has an extremum.

Now, the second derivative,
$d^2l/dx^2 = 2s/x^3$ is *positive*, for *any positive value of x*, we get that I has the minimum value at $x = \sqrt{s/2}$, and it is given by

$$l_{\min} = 2\sqrt{\frac{s}{2}} + \frac{s}{\sqrt{s/2}} = \frac{s+s}{\sqrt{s/2}} = \frac{2s}{\sqrt{s/2}} = 2\sqrt{2s}$$

This relation tells us that the length of any fence enclosing a rectangular plot of land with a given area s adjoining a wall cannot be less than $2\sqrt{2s}$, and it is equal to this only when the smaller side of the rectangle (which is equal to $x = \sqrt{s/2} = (1/2)\sqrt{2s}$) is half the greater side [which is equal to $(l - 2x) = (2\sqrt{2s} - 2(1/2)\sqrt{2s} = \sqrt{2s})$].

Thus, in these circumstances, the most economical fence is the one whose greater side is twice the smaller one.

Example (16): Divide a *positive integer N* into two parts such that their product is maximum.

Solution: Let one part of N be x

$$\therefore \text{ The other part } = (N - x)$$

Let the product of these parts be denoted by y. Then, we have

$$\begin{aligned} y &= x(N-x) \\ &= Nx - x^2 \\ \therefore \frac{dy}{dx} &= N - 2x \end{aligned} \tag{8}$$

Now, $dy/dx = 0$ gives $N - 2x = 0$

$$\therefore x = N/2$$

Thus, y is extremum for $x = N/2$. *We check the sign change of* dy/dx *when x passes through N/2 from the left to right.* When x is (*slightly*) *less than N/2*, we have from (1), dy/dx is *positive*. When x is *slightly more than N/2*, we have dy/dx is negative. Thus, the sign of dy/dx changes from positive to negative. *Therefore, y has maximum value when* $x = N/2$.

[If desired, *second-derivative test could be done*. Note that $d^2y/dx^2 = -2$ (which is a negative number). Hence, y has a maximum value when $x = N/2$.]

Remark: Product of *two equal parts* of a *positive integer* (these equal parts may be positive integers or positive rational numbers) gives the maximum product.

Note (8): If it is desired to partition a given positive number into *any number of equal parts*, then the largest product is obtained when each part is as close as possible to e (e = 2.718) (see Chapter 13a for the properties of the number "e").

Example (17): An agency agreed to conduct a tour for a group of 50 people at a rate of Rs. 400/- each. In order to secure more tourists, the agency agreed to deduct Rs. 5/- from the cost of the trip, for each additional person joining the group. What number of tourists would give the agency maximum gross receipts? (It was specified that 75 was the upper practical limit for the size of the group).

Solution: Just imagine that four people were to join the group, the reduction in the cost of the tour per person would be Rs. 20/-. If 10 people joined, the reduction in cost per person would be Rs. 50/- (for the entire group).

If we represent by x the number of additional tourists, the reduction will be Rs. $5x$ per person [so that cost of the tour for each person would be Rs. $(400 - 5x)$].

$$\text{Thus, cost of tour (for each person)} = \text{Rs. } (400 - 5x)$$

$$\text{and} \quad \text{Number of tourists} = 50 + x$$

$$\therefore \text{ Gross receipts of the company} = (400 - 5x)(50 + x)$$

Let us denote the gross receipts by the symbol y.

$$\therefore \; y = \text{Rs. } (400 - 5x)(50 + x)$$

$$\text{or } y = 2000 - 250x + 400x - 5x^2$$

$$y = 2000 + 150x - 5x^2 \text{ (It is desired that y should be maximum.)}$$

To find the *maximum gross receipt, we use the technique of finding the derivative and equating the result to zero.*

$$\frac{dy}{dx} = 150 - 10x$$

$$150 - 10x = 0$$

$$10x = 150$$

$$\therefore \; x = 15$$

Thus, for $x = 15$, y will have *extremum* value for y.

Now, there are two methods to check whether $x = 15$ will give maximum or minimum receipts. One is *to check the change of sign of dy/dx when x increases through the number 15*, and *the other is to check whether* d^2y/dx^2 *is negative or positive*.

If x is slightly less that 15, the sign of dy/dx is positive and for x more than15, dy/dx is negative. Thus, the sign of dy/dx changes from positive to negative at $x = 15$. Thus, y will have a maximum value for $x = 15$. (Also $d^2y/dx^2 = -10$ that is negative. Hence, y has maximum value at $x = 15$.)

Thus, with any of the above techniques, it is easily shown that for $x = 15$, y has the maximum value. Accordingly, if there are 15 additional tourists then the gross receipts will be maximum. Thus, the number of tourists in the group should be 65.

Note (10): Check the above conclusion by varying the number of tourists and computing the gross receipts.

Example (18): If two *real numbers* x and y are such that $x > 0$ and $xy = 1$, then find the *minimum* value of $x + y$.

Solution: It is given that

$$xy = 1,\ x > 0 \tag{9}$$

$$\Rightarrow y = \frac{1}{x} \text{(obviously, } y > 0)$$

$$\text{Let} f(x) = x + y \tag{10}$$

$$\text{or} f(x) = x + \frac{1}{x} \tag{11}$$

$$\therefore\ f'(x) = 1 - \frac{1}{x^2} \tag{12}$$

$$\text{and}\quad f''(x) = 0 + 2x^{-3} = \frac{2}{x^3} \tag{13}$$

For $f(x)$ to be minimum, $f(x) = 0$

$$\Rightarrow 1 - \frac{1}{x^2} = 0 \quad \text{[from (12)]}$$

$$\Rightarrow \frac{1}{x^2} = 1$$

$$\Rightarrow x^2 = 1 \Rightarrow x = +1, -1$$

But $x > 0$ given, therefore $x = -1$ is not acceptable.

Putting $x = 1$ in equation (13), we get

$$f''(x) = \frac{2}{1^3} = 2, \quad \text{which is positive}$$

Hence, $f(x)$ is minimum at $x = 1$ and the minimum value of $f(x)$ is obtained from equation (11).

$$f(1) = 1 + \frac{1}{1} = 2 \qquad \text{Ans.}$$

Example (19): A manufacturer of baby food wishes to package his product in cylindrical metal cans, each of which has to contain a certain volume V_0 of baby food. Let us find the ratio of the height of the can to its radius, in order to minimize the amount of metal, assuming that the ends and side (i.e., cylindrical portion of the can) are made from metal of the same thickness.

Solution: We wish to find a relationship between the height and the base radius of the right-circular can in order for the total surface area to be an absolute minimum for a fixed volume. Therefore, we consider the volume V_0 of the can a constant.

$$\begin{aligned} &\text{Let radius of the can} = r, \quad (r > 0) \\ &\text{and height of the can} = h, \quad (h > 0) \end{aligned}$$

Then, volume of baby food container in each can is given by

$$V_0 = \pi r^2 h \tag{14}$$

and the surface area of circular portion $= 2\pi rh$

Now, the total surface area of can consists of *two circular disks* at the ends and the *cylindrical portion*.

∴ The total surface area of each can given by

$$S = 2\pi r^2 + 2\pi rh \tag{15}$$

Because V_0 is constant, we could solve equation (14) for either r or h, in terms of the other and substitute in (15), which will give us S as a function of one variable.

$$\text{From (14), we get } h = \frac{V_0}{\pi r^2} \tag{16}$$

$$\begin{aligned} \therefore\ S &= 2\pi r^2 + 2\pi r \frac{V_0}{\pi r^2} \\ S &= 2\pi r^2 + 2 \cdot \frac{V_0}{r} \\ \text{or } S &= 2\pi r^2 + 2 \cdot \frac{V_0}{r} \end{aligned} \tag{17}$$

Now, for S to be minimum, we obtain from (17), and equate it to zero.

$$\begin{aligned} &\frac{\mathrm{d}s}{\mathrm{d}r} = 4\pi r - \frac{2V_0}{r^2} \\ &\text{or } \frac{\mathrm{d}s}{\mathrm{d}r} = 2\left(2\pi r - \frac{V_0}{r^2}\right) \\ &\text{Now, } \frac{\mathrm{d}s}{\mathrm{d}r} = 0, \text{ gives } 2\pi r = \frac{V_0}{r^2} \\ &\therefore\ r^3 = \frac{V_0}{2\pi} \end{aligned} \tag{18}$$

Step (1): The manufacturer wishes to minimize the surface area S of the can.

Step (2): We have found that the surface area of each can is given by equation (15)

Note that, r and h are not independent of each other. Since we have chosen r as the independent variable, then S depends on r; also, h depends on r. We have found only one candidate in equation (18), that is, the value of r which is related to h, and our interest lies in ratio h/r, which should make ds/dr zero.[9]

Recall from equation (16) that

$$h = \frac{V_0}{\pi r^2}$$

$$\therefore \quad \frac{h}{r} = \frac{V_0}{\pi r^3} = \frac{V_0}{\pi(V_0/2\pi)} \qquad \left[\therefore \; r^3 = \frac{V_0}{2\pi}\right]$$

$$\frac{h}{r} = \frac{V_0}{\pi} \times \frac{2\pi}{V_0} = 2 \qquad \text{Ans.}$$

Note (11): This example illustrates the practical importance of *extremum problems*. For a cylindrical can of minimal surface area and containing a given volume, we should have $h = 2r$, that is, its height should equal the diameter.

Example (20): A square sheet of tin, "a" cm on a side, is to be used to make an open top box by cutting a small square of tin from each corner and bending up the sides. How large a square should be cut from each corner so that the box has as large a volume as possible?

Solution:

Let the side of the square cut from each corner be x cm. Then, the volume of the (open) box in cubic centimeter is given by (Figure 19b.21)

$$\begin{aligned} v(x) &= x(a-2x)^2, \qquad 0 \le 2x \le a \\ \text{i.e., } v(x) &= x(a-2x)^2, \quad 0 \le x \le a/2 \end{aligned} \qquad (19)^{(10)}$$

[Note (12): It is clear from (1) that $v(x) = 0$, when $x = 0$ or when $x = a/2$, therefore, maximum volume of $v(x)$ must occur at a value of x between 0 and $a/2$.]

The function in (19) has a derivative at every such point, and hence the extremum occurs at an interior point of $[0, a/2]$ where $v'(x) = 0$.

From equation (1), we get,

$$\begin{aligned} v(x) &= a^2x - 4ax^2 + 4x^3 \\ \therefore \quad v'(x) &= a^2 - 8ax + 12x^2 \quad [12a^2 = (-6a)\cdot(-2a) \\ &= a^2 - 6ax - 2ax + 12x^2 \\ &= a(a-6x) - 2x(a-6x) \\ &= (a-6x)(a-2x) \end{aligned}$$

so that $y' = 0$ when $x = a/6$ or $x = a/2$.

[9] Note that, the equation (18) (i.e., $r^3 = V_0/2\pi$) helps in deciding the relation between h and r.

[10] The restrictions placed on the length x in equation (14) are due to the fact that one can neither cut a negative amount of material from a corner nor cut away more than the total amount present.

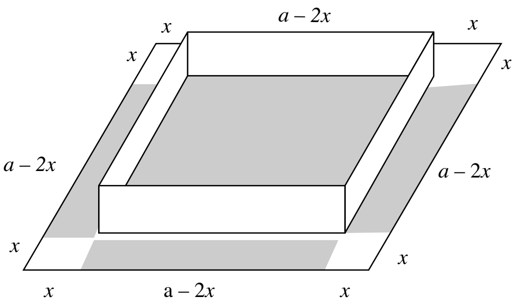


FIGURE 19b.21

Of these, only $x = a/6$ lies in the interior of $[0, a/2]$. Therefore, the maximum or minimum occurs at $x = a/6$.

Now, by checking the sign change of y' (when x increases through $a/6$) or by finding the sign of $y''\ (a/6)$, it can be easily shown that y has maximum value for $x = a/6$.

Thus, each corner square should have dimensions $a/6$ by $a/6$ to produce a box of maximum volume. Ans.

Remark: Note that we have solved a general problem for *making boxes of maximum volume, from any given square sheet*.

Exercise (2)

Q1. The sum of two positive numbers is 20. Find the numbers

(i) if their product if maximum;
(ii) if the sum of their squares is minimum.

Ans. $x = 10$ and $y = 10$.

Q2. Show that the perimeter of the rectangle of given area is minimum if it is a square.

Q3. Divide 100 into two parts such that the sum of the twice of first part and square of second is minimum.

Ans. 99, 1.

Q4. The two sides of a rectangle are $2x$ and $(!5 - 2x)$ units, respectively. For what value of x, the area of rectangle will be maximum?

Ans. $\dfrac{14}{4}$

Q5. Find the two positive numbers whose product is 64 and sum is minimum

Ans. 8, 8.

Q6. A wire of length 28 m is to be cut into two pieces, one of the piece is to be made into a square and the other into a circle. Where should the wire be cut so that the combined area is minimum?

Ans. $112/(\pi + 4)$ from one end. Hence length from second end $= 28\pi/(\pi + 4)$.

20 Rolle's Theorem and the Mean Value Theorem (MVT)

20.1 INTRODUCTION

One of the most important theorems in calculus is the *Mean Value Theorem* (MVT), which is used to prove many theorems of both differential and integral calculus, as well as other subjects, such as numerical analysis. MVT is said to be the midwife of calculus—not very attractive or glamorous by itself, but often helping to deliver other theorems that are of major significance. *The proof of the Mean-Value Theorem* is based on a special case of it known as *Rolle's Theorem*, which we discuss first.

The French mathematician Michel Rolle (1652–1719) proved that *if f is a function continuous on a closed interval* $[a, b]$ *and differentiable on the open interval* (a, b), and if $f(a)$ and $f(b)$ *both equal zero, then there is at least one number c between a and b at which* $f'(c) = 0$.

In the statement of this theorem, there are three conditions, which must be satisfied for the theorem to hold. By way of illustrations, we shall show that all the three conditions in Rolle's theorem are important and if they are violated, the theorem may not hold.[1]

First, let us see what this means geometrically. Figure 20.1 shows the graph of a function f satisfying the conditions in the preceding paragraph.

We see intuitively that there is at least one point on the curve between the points $(a, 0)$ and $(b, 0)$ at which *the tangent line is parallel to the x-axis*; that is, the slope of the tangent line is *zero*. This situation is illustrated in this figure at the point P.

Note that, the x-coordinate of P is c such that $f'(c) = 0$.

Note (1): The function, whose graph appears in Figure 20.1, is not only differentiable on the open interval (a, b) but is also differentiable at the end points of the interval. However, the intuitive feeling that *f should be differentiable at the end points is not necessary for the graph to have a horizontal tangent line at some point in the interval*. Figure 20.2 illustrates this.

Applications of derivatives: 20-Rolle's theorem and mean value theorem (MVT)

[1] Some authors state Rolle's Theorem by relaxing the condition (iii) to read it as $f(a) = f(b)$, thus not requiring that both $f(a)$ and $f(b)$ should be necessarily zero. In fact, the condition $f(a) = f(b)$ is *more general* than the condition $f(a) = f(b) = 0$. Thus, while verifying whether Rolle's theorem is applicable for a specific function, it is enough to check whether $f(a) = f(b)$, instead of requiring $f(a) = f(b) = 0$.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

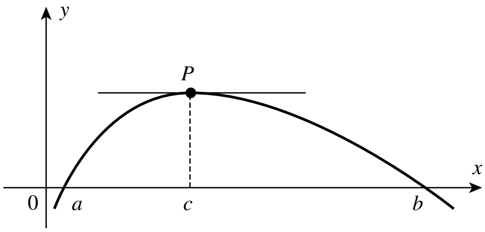


FIGURE 20.1

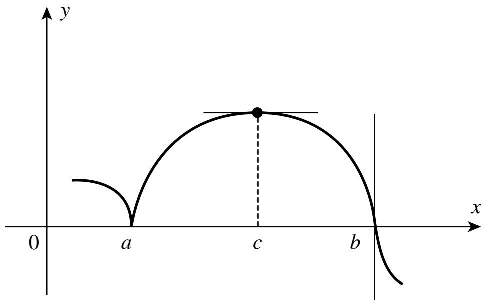


FIGURE 20.2

We see in Figure 20.2 that the function is not differentiable at a and b; however, there is a horizontal tangent line at the point where $x = c$, and c is between a and b.[(2)]

Note (2): *It is necessary that the function should be continuous at the end points of the interval to guarantee a horizontal tangent line at an interior point*. Figure 20.3 shows the graph of a function continuous on the interval (a, b) but discontinuous at b. Observe that, *the function is differentiable on the open interval (a, b), and the function values are zero at both a and b*. However, *there is no point* at which the graph has a horizontal tangent line.

Note (3): *The condition that f(x) be differentiable in (a, b) is reasonable because the conclusion of Rolle's theorem is about the vanishing of the derivative.*[(3)]

Note (4): *The condition $f(a) = f(b)$ cannot be eliminated from Rolle's Theorem*. For example, if $f(x) = x$, then $f'(x) = 1$, for all x, in any open interval (a, b). *This implies that $f'(c) \neq 0$ for all c in (a, b)*. Thus, without meeting the condition $f(a) = f(b)$, we cannot conclude that $f'(c) = 0$ at some c in (a, b). In fact, *the conclusion of Rolle's Theorem is applicable to a curve that rises and falls smoothly*.

[(2)] An example of this type is given by the function $y = f(x) = \sqrt{1 - x^2}$, in the interval $[-1, 1]$. This function represents the upper half of a circle, with its center at the origin and having radius 1. Observe that $f(-1) = f(1) = 0$, f is continuous in $[-1, 1]$ and the derivative $f'(x) = x/\sqrt{1 - x^2}$ exists in $(-1, 1)$, though it does not exist at the end points of the interval $[-1, 1]$. We have $f'(0) = 0$ and "0" lies in the interval $(-1, 1)$. Thus, Rolle's theorem is valid in this case. Note that, differentiability at the end points of the closed interval is not needed.

[(3)] From this assumption it follows that f is continuous in (a, b). But, it must be remembered that the continuity of the function f at both the end points a and b of the interval $[a, b]$ is also necessary and that this requirement cannot be dropped, as already emphasized in the Note (2) above.

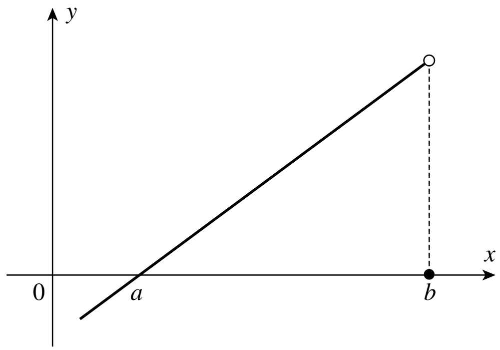


FIGURE 20.3

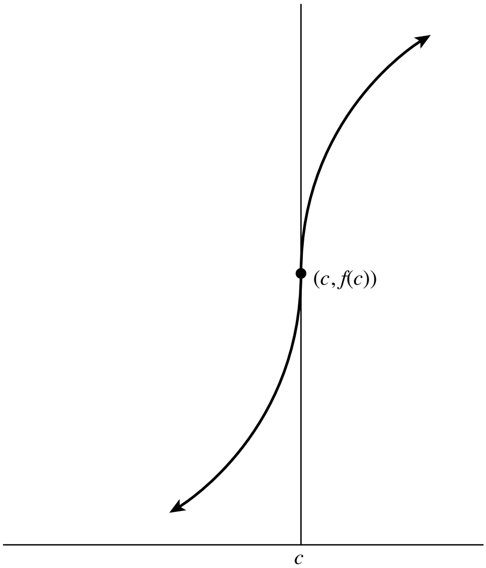
A point on a continuous curve where the derivative does not exist is called a *slopeless point*. This can happen under two situations:

(i) The graph of f has a vertical tangent line at some point c in (a, b) (Figure 20.4).

(ii) The graph of f has a sharp turn (or corner) at some point c in (a, b) (Figure 20.5).

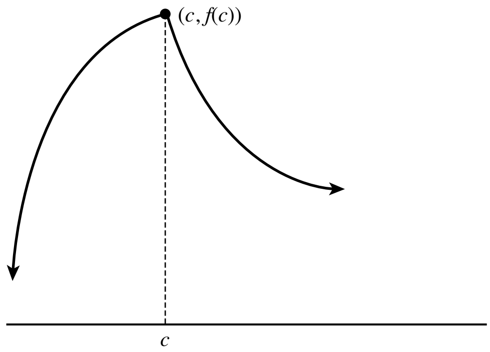
If a function f is not differentiate at some point in (a, b) then there may not be any point x in (a, b) at which $f'(x) = 0$.

Note (5): Rolle's Theorem guarantees only the existence of *at least one point c* in (a, b) for which $f'(c) = 0$. Of course, there may be *more such points* in (a, b), for which the derivative of f is zero. This is illustrated geometrically in Figure 20.6.



f is not differentiable at $c \in (a, b)$.
f is continuous at c.

FIGURE 20.4



f is not differentiable at $c \in (a, b)$.
f is continuous at c.

FIGURE 20.5

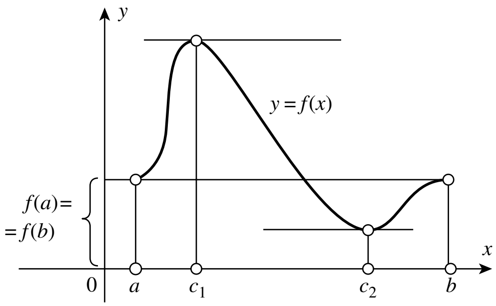


FIGURE 20.6

Observe that there is a horizontal tangent line at the point where $x = c_1$ and also at the point $x = c_2$, such that $f'(c_1) = 0$ and $f'(c_2) = 0$. *The theorem does not define the location of c in* (a, b) but states that c must lie somewhere within (a, b).

Now, we state and prove Rolle's Theorem.

20.2 ROLLE'S THEOREM (A THEOREM ON THE ROOTS OF A DERIVATIVE)

Let f be a function such that

(i) it is *continuous* on the *closed interval* $[a, b]$;

(ii) it is *differentiable* on the *open interval* (a, b); and

(iii) it vanishes at the end points $x = a$ and $x = b$ [i.e., $f(a) = 0$ and $f(b) = 0$]. Then there is a number c in the open interval (a, b), such that $f'(c) = 0$.

[The number c is called a root of the function $\phi(x)$ if $\phi(c) = 0$.]

Proof: Since the function $f(x)$ is *continuous* on the interval $[a, b]$, it has a *maximum M* and a *minimum m* on that interval. We consider two cases.

Case (1): If $M = m$, the function $f(x)$ is *constant*, which means that *for all values of x, it has a constant value f(x)* $= m$. But then *at any point of the interval* $f'(x) = 0$, and the theorem is proved.

Case (2): Suppose $M \neq m$. Then at least one of these numbers is *not equal to zero*. For the sake of definiteness, let us assume that $M > 0$ and that the function takes on its maximum value at $x = c$, so that $f(c) = M$. Here, it is important to note that c is *not equal either to a or to b, since it is given that* $f(a) = 0, f(b) = 0$. Since $f(c)$ is the maximum value of the function, it follows that $f(c + h) - f(c) \leq 0$, *under both situations* when $h > 0$ and when $h < 0$.

Accordingly, we get the following inequalities.

If h is positive, we have,

$$\frac{f(c+h) - f(c)}{h} \leq 0$$

On the other hand, if h is negative,

$$\frac{f(c+h) - f(c)}{h} \geq 0$$

But, f is *differentiable* in (a, b), which means that the *derivative* at $x = c$ *exists.*

Therefore, upon passing to the limit as $h \to 0$, we get

$$\lim_{h \to 0} \frac{f(c+h) - f(c)}{h} = f'(c) \leq 0 \qquad (\text{when } h > 0) \tag{1}$$

and

$$\lim_{h \to 0} \frac{f(c+h) - f(c)}{h} = f'(c) \geq 0 \qquad (\text{when } h < 0) \tag{2}$$

But, $f'(c)$ is *unique*. This is possible if $f'(c) = 0$, which we get on comparing (1) and (2). Consequently, *there is a point c inside the interval* $[a, b]$ at which the derivative $f'(x) = 0$. This establishes the theorem.

Note (6): *The converse of Rolle's Theorem is not true*. That is, *if a function f defined on* $[a, b]$ *is such that* $f'(c) = 0$, *with* $a < c < b$, *then we cannot conclude that the conditions (i), (ii,) and (iii) of the theorem must hold.*

Now, we consider some examples to understand Rolle's Theorem better.

Example (1): Given $f(x) = 4x^3 - 9x$. Verify that the three conditions of the hypothesis of Rolle's Theorem are satisfied for each of the following intervals:

$$\left[-\frac{3}{2}, 0\right], \quad \left[0, \frac{3}{2}\right], \quad \text{and} \quad \left[-\frac{3}{2}, \frac{3}{2}\right]$$

Then find a suitable choice for c in each of these intervals for which $f'(c) = 0$.

Solution: Given

$$f(x) = 4x^3 - 9x$$
$$\therefore \; f'(x) = 12x^2 - 9$$

Because $f'(x)$ exists for all values of x, f is *differentiable* on $(-\infty, \infty)$. Thus, conditions (i) and (ii) of Rolle's Theorem hold on any interval. *To determine the intervals on which the condition* *(iii) holds*, we find the values of x for which $f(x) = 0$. If $f(x) = 0$, it means that

$$4x\left(x^2 - \frac{9}{4}\right) = 0$$
$$\therefore \quad x = 0, \quad x = \pm\frac{3}{2}$$
$$\text{That is, } x = -\frac{3}{2},\ x = 0,\ x = \frac{3}{2}$$

Therefore, at $a = -(3/2)$ and $b = 0$, we have $f(x) = 0$. Therefore, Rolle's Theorem holds on $[-(3/2), 0]$. Similarly Rolle's Theorem holds on $[0, (3/2)]$ and $[-(3/2), (3/2)]$. To find *suitable values for c*, we set $f'(x) = 0$ and get

$$12x^2 - 9 = 0 \qquad \therefore\ 4x^2 - 3 = 0$$
$$\therefore \quad x = \pm\frac{1}{2}\sqrt{3}$$

Therefore, in the interval $[-(3/2), 0]$ a suitable choice for c is $-(1/2)\sqrt{3}$. In the interval $[0, (3/2)]$, we take $c = \frac{1}{2}\sqrt{3}$. In the interval $[-(3/2), (3/2)]$ there are two possibilities for c: either $-(1/2)\sqrt{3}$ or $(1/2)\sqrt{3}$.

Example (2): Consider the continuous function

$$y = f(x) = \sqrt[3]{x^2} = x^{2/3}, \quad x \in [-1, 1]$$

It assumes equal values at the end points of the interval $[-1, 1]$. However, its derivative $f'(x) = 2/3\sqrt[3]{x^2}$ does not vanish anywhere. In this example, the condition of differentiability is violated at the point $x = 0$, which lies in the interval $(-1, 1)$. Note that, the derivative does not exist at $x = 0$ (since there is a vertical tangent at $x = 0$). (See Figure (20.8)).

Example (3): Verify the conditions of Rolle's Theorem for the function

$f(x) = \log(x^2 + 2) - \log 3$ on $[-1, 1]$, and find the value of c where the derivative vanishes.

Solution:

(i) Since *logarithmic function* and a *constant function* both are continuous functions, hence their sum given by

$$f(x) = \log(x^2 + 2) - \log 3$$

is continuous on $[-1, 1]$.

(ii) $f'(x) = (1/(x^2 + 2)) \cdot (2x) = 2x/(x^2 + 2)$, which exists for all x. Thus, the function $f(x)$ is *differentiable in the open interval* $(-1, 1)$.

(iii) $f(-1) = \log(1 + 2) - \log 3 = 0$ and

$$f(1) = \log(1 + 2) - \log 3 = 0$$
$$\therefore\ f(-1) = f(1)$$

Thus, $f(x)$ satisfies all the conditions of Rolle's Theorem.

$\therefore$ There must exist *at least one value* "c" of x, in $(-1, 1)$ for which $f'(c) = 0$.

$$\text{Now } f'(c) = \frac{2c}{c^2+2} = 0 \Rightarrow c = 0 \in (-1, 1)$$

Hence, Rolle's Theorem is verified for the given function. **Ans.**

Example (4): Verify the conditions of Rolle's Theorem for the function $y = f(x) = e^{1-x^2}$, $x\varepsilon$ $[-1, 1]$ and find c for which $f'(c) = 0$.

Solution:

(i) The function $f(x) = e^{1-x^2}$ is an exponential function of x and hence it is continuous on $[-1, 1]$.

(ii) $f'(x) = -2x \cdot e^{1-x^2}$, which exists in the open interval $(-1, 1)$.

(iii) $f(-1) = e^0 = 1$, and $f(1) = e^0 = 1$. Thus, $f(-1) = f(1)$.

Since $f(x)$ satisfies all the conditions of Rolle's Theorem, there must exist *at least one value c* of x in $(-1, 1)$, for which $f'(c) = 0$.

We have,
$$f'(c) = -2(c) \cdot e^{1-c^2}$$
$$= -2c \cdot e^{1-c^2}$$

Hence,
$$f'(c) = 0 \rightarrow c = 0$$

Observe that, the number $c = 0$ *lies in the open interval* $(-1, 1)$.

Hence, Rolle's Theorem is verified for the given function. **Ans.**

Example (5): Discuss whether Rolle's Theorem is applicable for the function $y = f(x) = |x|$ on $[-1, 1]$.

Solution: We have $f(x) = |x|$, $x\varepsilon$ $[-1\ 1]$. In the given interval,

$$\text{by definition} \quad f(x) = |x| = \begin{cases} x & \text{if } 0 \le x \le 1 \quad \text{i.e., } x \in [0, 1] \\ -x & \text{if } -1 \le x < 0 \quad x \in [-1, 0) \end{cases}$$

$$\text{Now, } f'(0) = \lim_{x \to 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x \to 0} \frac{|x| - 0}{x - 0} = \lim_{x \to 0} \frac{|x|}{x}$$

$$\therefore \quad f'(0) = 1 (x > 0)$$
$$= -1 (x < 0)$$

We note that

(i) $f(x)$ is *continuous* in the closed interval $[-1, 1]$

(ii) $f'(x) = \begin{cases} 1 & \text{if } x \in (0, 1] \\ -1 & \text{if } x \in [-1, 0) \end{cases}$

Thus, $f'(x)$ *does not exist* at $x = 0$.

(iii) $f(-1) = -(-1) = 1$, and $f(1) = 1$

$\therefore$ $f(-1) = f(1) = 1$

Observe $f(x)$ is not differentiable at $x = 0$ in the open interval $(-1, 1)$. Thus, function $|x|$ *does not satisfy the condition (ii) of the Rolle's Theorem*. Hence, the conclusion of Rolle's Theorem is not applicable for $|x|$. Therefore, there is *no point c* in $(-1, 1)$ at which $f'(c) = 0$.

Example (6): Verify the conditions of Rolle's Theorem for the function

$$f(x) = \sin x + \cos x - 1 \quad \text{on } [0, \pi/2]$$

Solution:

(i) The function $f(x) = \sin x + \cos x - 1$ is continuous on the $[0, \pi/2]$.

(ii) $f'(x) = \cos x - \sin x$. Obviously, $f'(x)$ exists in the open interval $[0, \pi/2]$.

(iii) $f(0) = 0 + 1 - 1 = 0$

$$\text{and} \quad f(\pi/2) = 1 + 0 - 1 = 0$$

Thus, $f(0) = f(\pi/2)$

Thus, all the three conditions of Rolle's Theorem are satisfied. Accordingly, there must exist at least one value c of x in the open interval $(0, \pi/2)$, at which $f'(c) = 0$.

$$\text{Now, } f'(c) = \cos c - \sin c = 0$$

$$\therefore \quad \cos c = \sin c \Rightarrow \tan c = 1$$

$$\Rightarrow c = \pi/4 \qquad \textbf{Ans.}$$

Example (7): Consider the function

$$y = f(x) = \begin{cases} x & \text{for } 0 \le x < 1 \\ 0 & \text{for } x = 1 \end{cases}$$

Observe that, for the given interval $[0, 1]$, $f(0) = 0$ and $f(1) = 0$. Also, $f(x)$ is differentiable at every point in $(0, 1)$. It is clearly seen that $f'(x) = 1$ at all the points of the interval $(0, 1)$, but there is no point in $(0, 1)$ at which it turns into zero, because this function is *discontinuous* at the end point ($x = 1$) of the interval $[0, 1]$. *This example also emphasizes the requirement of continuity at the end points of the closed interval.* [Now refer to Note (2) and Figure 20.3.]

20.2.1 Geometric Conclusion of Rolle's Theorem

Rolle's Theorem says (geometrically) that a curve that rises and falls (without *any breaks or slopeless points*) must have leveled off in the mean time.

20.2.2 Dynamic Face to Rolle's Theorem

When a ball is thrown up vertically at instant $t = a$ (say), and returns at $t = b$, there is an instant c, between a and b at which the ball stops momentarily, that is, it has zero velocity.

20.2.3 A Useful Interpretation of Rolle's Theorem

Rolle's theorem gives us the fact that if a *polynomial has n distinct zeros, its derivative has at least* $n - 1$ *distinct zeros*.

Consider a *polynomial equation* $f(x) = 0$ *where* $f(x)$ *satisfies the conditions of Rolle's Theorem*, and let $x_1, x_2, x_3, \ldots, x_n$ be the roots of the equation. Then, by Rolle's Theorem, the equation $f'(x) = 0$ has the roots $c_1, c_2, c_3, \ldots, c_{n-1}$, one or more of which lie in between the roots of $f(x) = 0$, that is, $x_1 < c_1 < x_2 < c_2 < x_3 \cdots < x_{n-1} < c_{n-1} < x_n$.

[We studied maxima and minima (exterma) of a function in earlier Chapter 19b and found that $c_1, c_2, c_3, \ldots, c_{n-1}$ are the points of relative extrema.]

20.3 INTRODUCTION TO THE MEAN VALUE THEOREM

If $y = f(x)$ is *continuous* at each point of $[a, b]$ and *differentiable* at each point of (a, b), then there is *at least one number* c between a and b, at which,

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

[From the statement of the MVT, *note that, it is not necessary for the function "f" to be differentiable at the end points* $x = a$ *and* $x = b$.] As mentioned earlier, the proof of the MVT is based on Rolle's Theorem, which is powerful in its own right and is a special case of the MVT as we will see.

The MVT has only two conditions that are in common with those of Rolle's theorem, which has an additional condition $f(a) = f(b)$ to be satisfied. *The MVT does not require the condition* $f(a) = f(b)$ *to be satisfied.* With its two conditions, it asserts the existence of a number c (somewhere) in (a, b) at which

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

We know that the condition $f(a) = f(b)$ cannot be dropped from Rolle's Theorem. Note that, if $f(a) = f(b) = 0$, then the end points of the graph must lie on the x-axis (see Figure 20.1). However, if $f(a) = f(b) \neq 0$, then the end points of the graph are on some line which is parallel to x-axis (see Figure 20.7). Thus, *when all the conditions of the Rolle's Theorem are satisfied,*

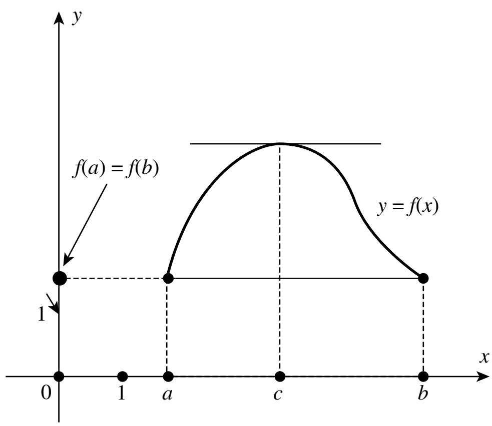


FIGURE 20.7

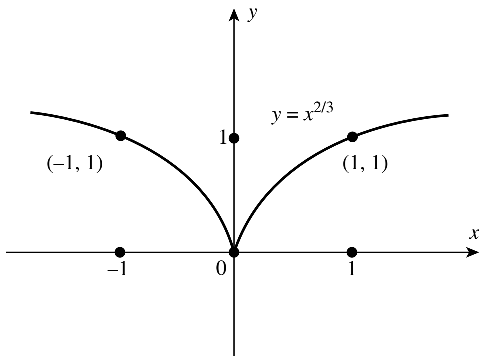


FIGURE 20.8

we can conclude that somewhere on the graph, the tangent line is parallel to the x-axis (see Figure 20.7).

The MVT states that, if the line joining *the end points of a (smooth) curve is not parallel to the x-axis* [since, $f(a) \neq f(b)$], *then there is at least one point on the curve where the tangent line is parallel to the line joining the points (a, f(a)) and (b, f(b)) on the curve.* Suppose, we call the line joining any two points of the curve as a *chord of the curve*. Then, our improved statement of the Mean Value Theorem reads as under:

Given a chord of a smooth curve, there is at least one point on the curve where the tangent line is parallel to this chord (see Figures 20.9 and 20.10).

To be more specific, consider a function, $y = f(x)$, and let $A(a, f(a))$ and $B(b, f(b))$ be two points on its graph, which rises and falls, without any breaks or slopeless points. We may assume that $a < b$. Then, the MVT asserts the existence of a number c, $(a < c < b)$ such that the tangent at $(c, f(c))$ to the graph of f, is parallel to the chord joining $A(a, f(a))$ and $B(b, f(b))$. *To visualize this we make use of coordinate geometry.*

We know that *two nonvertical lines are parallel if and only if they have the same slope.* Here, the slope of the line joining the points $A(a, f(a))$ and $B(b, f(b))$ is given by $(f(b) - f(a))/(b - a)$. Also, the *slope of the tangent at the point* $(c, f(c))$ is given by $f'(c)$. *If this tangent line is to be parallel to the chord AB*, then

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

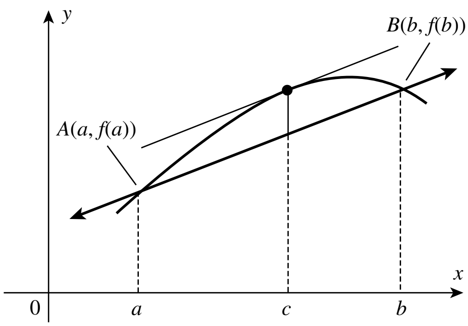


FIGURE 20.9

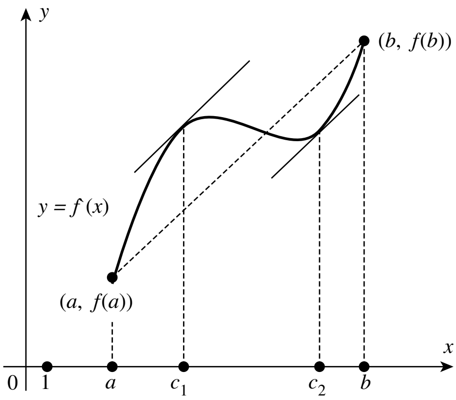


FIGURE 20.10

The MVT states that such a number c necessarily exists in (a, b). We now state the Mean Value Theorem and prove it using Rolle's Theorem.

20.3.1 The Mean Value Theorem

Let f be a function, such that,

(i) It is continuous on the closed interval $[a, b]$.
(ii) It is differentiable on the open interval (a, b).
Then, there is a number "c" in the open interval (a, b), such that,

$$f'(c) = \frac{f(b) - f(a)}{b - a} \tag{3}$$

Proof: Consider the number Q, defined by the equation

$$\frac{f(b) - f(a)}{b - a} = Q \tag{4}$$

We will show that $Q = f'(c)$, $c \in (a, b)$.

From the equation (4), we get,

$$f(b) - f(a) = Q(b - a) \tag{5A}$$

$$\text{or } f(b) - f(a) - Q(b - a) = 0 \tag{5B}$$

We introduce *an auxiliary function F* that allows us to simplify the proof by using Rolle's Theorem. To obtain this auxiliary function, we write x for b in equation (5B), and denote the expression on left-hand side by $F(x)$.

$$F(x) = f(x) - f(a) - Q(x - a) \tag{6}$$

It is easy to show that the function $F(x)$ satisfies all the conditions of Rolle's Theorem.

(i) $F(x)$ is continuous on $[a, b]$ since it is the sum of f and a linear function, both of which are continuous there.

(ii) $F(x)$ is differentiable on (a, b), because f is differentiable on (a, b).

(iii) $F(a) = 0$ and $F(b) = 0$ [using (5B)].

Therefore, by Rolle's Theorem, there is a number "c" in the open interval (a, b), such that,

$$F'(c) = 0 \tag{7}$$

Now, from equation (6), we have

$$\begin{aligned} F(x) &= f(x) - f(a) - xQ + aQ \\ \therefore \quad F'(x) &= f'(x) - Q \qquad [\text{for all } x \in (a, b)] \\ \therefore \quad F'(c) &= f'(c) - Q \qquad [\text{since } c \in (a, b)] \\ \text{or} \quad 0 &= f'(c) - Q \qquad [\text{using } (7)] \\ \therefore \quad Q &= f'(c) \\ \text{or} \quad \frac{f(b) - f(a)}{b - a} &= f'(c) \end{aligned}$$

This establishes the theorem.

Note (7): The Mean Value Theorem discussed above is due to J.L. Lagrange (1736–1813), an outstanding French Mathematician and astronomer, hence it is also known as Lagrange's Mean Value Theorem.

20.3.2 The Geometric Significance of the Function F(*x*)

We write the equation of the chord AB (Figure 20.11), taking into account that its slope is $(f(b) - f(a))/(b - a) = Q$, and that it passes through the point $(a, f(a))$:

$$y - f(a) = Q(x - a)$$

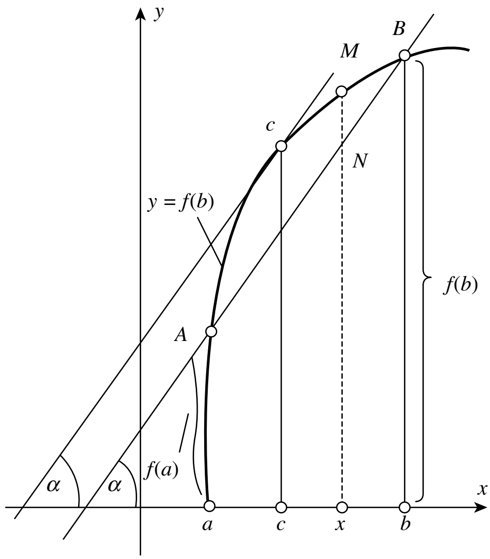


FIGURE 20.11

which gives,

$$y = f(a) + Q(x - a)$$

$$\text{But } F(x) = f(x) - [f(a) + Q(x - a)]$$

Thus, for each value of x, $F(x)$ is equal to the *difference between the ordinates of the curve* $y = f(x)$ and the chord $y = f(a) + Q(x - a)$, for points with the same abscissa. In other words, $F(x)$ represents the length of the segment MN for each $x \in [a, b]$.

Now, we consider some examples on MVT.

Example (8): Let $f(x) = \frac{1}{3}x^3 + 2x$. Find a number c in (0, 3) such that $f'(c) = \frac{f(3) - f(0)}{3 - 0}$

Solution: We have $f(3) = \frac{1}{3}(3)^3 + 2(3) = 15$ and $f(0) = 0$.

$$\therefore \quad \frac{f(3) - f(0)}{3 - 0} = \frac{15 - 0}{3 - 0} = 5$$

We search for a number c in (0, 3) such that $f'(c) = 5$. But, $f'(x) = x^2 + 2$.

Thus, c must satisfy $f'(c) = c^2 + 2 = 5$.

Therefore, $c^2 = 3$, so that $c = \pm\sqrt{3}$. But, c must be in (0, 3), we therefore conclude that $c = \sqrt{3}$. **Ans.**

Example (9): Test whether Lagrange's MVT holds for $f(x) = x - x^3$ in the interval (−2, 1) and if so, find the appropriate value of c.

Solution: Here, $a = -2$, $b = 1$ and $f(x) = x - x^3$.

$$\therefore \ f(a) = f(-2) = -2 - (-8) = 6, \qquad f(b) = f(1) = 0$$

$$\therefore \quad \frac{f(b) - f(a)}{b - a} = \frac{0 - 6}{1 - (-2)} = -2 \qquad (8)$$

Now,

$$f'(x) = 1 - 3x^2 \qquad (9)$$

If the MVT holds for the given function, then the number c must satisfy the equation,

$$I - 3x^2 = -2$$

$$\therefore \ 3c^2 = 3 \quad c = 1 \text{ or } -1$$

Here, $c = -1$ lies within (−2, 1).

$\therefore$ LMVT holds for the given function. **Ans.**

20.3.3 A Closer Look at the Mean Value Theorem

The Mean Value Theorem (Lagrange's Theorem), proved above, involves first derivatives. Hence, it is called the MVT for first derivatives. In fact, it is known as the Fundamental Mean

Value Theorem and is one of the *most powerful tools in calculus. It is employed very often in proving other important theorems.*[4]

The *conclusions* of the MVT are *intuitively appealing* and its *hypotheses* are naturally expected. Further, its applications will be found marvelously tangible (i.e., fitting with experience). The student will find in the MVT, an ever present tool just waiting to be applied, both in proving theorems and in solving problems.

The adjective "*mean*" carries both the notions "*between*" and "*average*", each of which gives a significant clue to the basic idea in the theorem. What the MVT does, is single out a derivative value that plays the role of an average derivative value, and this derivative value is attained at a point strictly between the end points of the interval domain of the function.

Consider a *continuous function*,

$$f\colon [a, b] \to R$$

which is differentiable at every point of the open interval (a, b).[5]

What the MVT does is to identify the difference quotient $(f(b) - f(a))/(b - a)$ with the derivative $f'(x)$ evaluated at a point (say) c, lying strictly between a and b.

That is,

$$\left.\begin{array}{ll} (1) & \dfrac{f(b) - f(a)}{b - a} = f'(c) \\ & \text{or equivalently} \\ (2) & f(b) - f(a) = f'(c)(b - a) \end{array}\right\} \quad \text{where } a < c < b$$

It must be clearly understood once and for all that *the location of c is not really pinpointed*; we only know that it lies somewhere inside an open interval (a, b). But, the interesting fact is *the mere knowledge that c is a mean point* (i.e., it lies strictly between the end points of the interval) shows the real *power behind the theorem and its applications*. Of course, the exact location of c can be found in some cases (as we have seen in some solved examples) but in general, it is never needed in any application.

Many important concepts in mathematics are based on *Existence Theorems*, the MVT being one of them. Some other examples of existence theorems are the Intermediate Value Theorem (IVT) and Extreme Value Theorem (EVT), both pertaining to continuous functions defined on closed and bounded intervals. Without going into the proof of these theorems, we indicate why the property each guarantees is practically useful.

(I) *Intermediate Value Theorem* [Already introduced in Chapter 8, but again repeated here for ready reference]: Let f be continuous on the *closed* and *bounded interval* $[a, b]$ and let y be any number between $f(a)$ and $f(b)$. Then, *there exists a number c between a and b for which f(c) = y.*

[4] There is another MVT for second derivatives that generalizes the MVT for first derivative and sets the stage for further generalization, namely, Taylor's Theorem, one of the most remarkable achievements in mathematics. Here, it may be emphasized that the method and the steps in proving the above LMVT, using Rolle's Theorem, is very important since similar steps are required to be taken to establish the MVTs for higher derivatives, using the MVT for first derivatives.

Besides, there is a theorem known as Generalized Mean Value Theorem (Cauchy's Theorem) that is useful for evaluating limits of indeterminant forms [i.e., the limit(s) of ratios of two functions $f(x)$ and $\phi(x)$ approaching the forms of the type $0/0$, ∞/∞ as $x \to 0$ (or $x \to \infty$)].

[5] In other words, the graph of f is tied to the end points $(a, f(a))$ and $(b, f(b))$ and has neither breaks nor slopeless points (in particular, no sharp points) anywhere between a and b (see Figure 20.12).

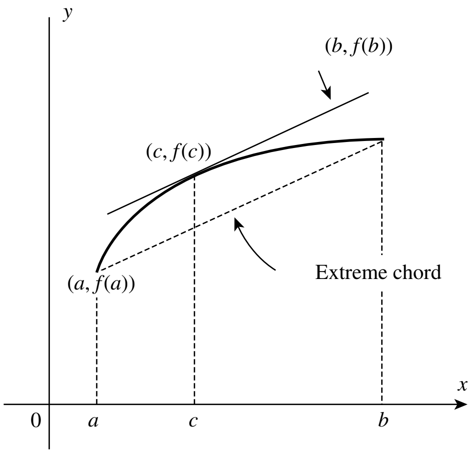


FIGURE 20.12

The IVT says that, if f is continuous on $[a, b]$, *then the range of f contains not just f(a) and f(b) but everything in between*. This means that the graph of a continuous function f is unbroken. In other words, it means that enroute from $(a, f(a))$ to $(b, f(b))$, the *graph of f crosses every horizontal line* at one (or more number of points) between $y = f(a)$ and $y = f(b)$.

(II) *Extreme Value Theorem*: Let f be continuous on the closed and bounded interval $[a, b]$. Then, f assumes both *a maximum value and a minimum value* somewhere on $[a, b]$.

The EVT guarantees that, if f is *continuous on the closed interval* $[a, b]$, then f attains both a *maximum* and a *minimum* somewhere therein.

[Both hypotheses of the EVT—that f be *continuous* and that the *interval be closed* are necessary. If either fails, f need not assume a maximum or a minimum.]

Note (8): Each of these existence theorems guarantees *the existence of at least one point in the domain* with some desirable property. Neither theorem states where in $[a, b]$ these points may fall or how many (such points) there may be. It will be found that these theorems, (besides having theoretical importance) have a lot of practical utility.

20.3.4 Some Aspects of the Conclusion of the MVT Expressed by its Formulas

(a) *Geometric Aspect*: In view of the formula $(f(b) - f(a))/(b - a) = f'(c)$, the slope of the extreme chord of a graph is attained by the tangent line at some mean point on the graph.(6)

Example (10): Find the tangent line to the graph of $f(x) = x^3$, which is parallel to the chord joining (1, 1) to (2, 8) and has its point of contact between the given points.

(6) That is in the family of all tangent lines whose points of contact lie between two given points on the curve, there is at least one tangent line parallel to the chord joining the two given points.

Solution: Slope of the chord joining (1, 1) and (2, 8) is given by,

$$\frac{f(2)-f(1)}{2-1} = \frac{2^3-1^3}{2-1} = 7$$

Also, $f'(x) = 3x^2$.

If c is the point in the interval (1, 2) for which $f'(c) = 7$, then we have $3c^2 = 7$ or $c = \pm(\sqrt{7}/\sqrt{3}) = \pm((\sqrt{7}\cdot\sqrt{3})/3) = \pm(\sqrt{21}/3)$ of which only $c = (\sqrt{21}/3)\varepsilon(1,2)$.

$\therefore$ the desired point of contact is $\left((1/3)\cdot\sqrt{21}, (7/9)\cdot\sqrt{21}\right)$, so that the tangent line is $\left\{(x,y)\colon y = 7\left(x-(1/3)\sqrt{21}\right)+(7/9)\sqrt{21}\right\}$. **Ans.**

(b) *Kinematic Aspect*: Let $f(x)$ be a position function (of a moving object) with the time interval $[a, b]$ as its domain. *Then, $f'(x)$ is the velocity function and $f'(c)$ is a mean velocity, if $c\ \varepsilon\ (a, b)$.*

Thus, the equality at the formula $f(b)-f(a) = f'(c)(b-a)$ says that the displacement $f(b)-f(a)$ can be obtained as a product $f'(c)\cdot(b-a)$, where $f'(c)$ is the mean velocity and $(b-a)$ is the time interval.(7)

Example (11): A vehicle has a *quadratic position function* $f(t) = at^2 + bt + d$, where a, b, and d are any real numbers. Show that *over any interval of motion, the average velocity is attained at the mid point.*

Solution: Let $[t_1, t_2]$ be any time interval of motion. Suppose c is the instant in this interval at which $f'(t)$ is average, then we have

$$\frac{f(t_2)-f(t_1)}{t_2-t_1} = f'(c)$$

$$\text{or} \quad \frac{a\left(t_2^2-t_1^2\right)+b(t_2-t_1)}{t_2-t_1} = 2ac+b, \quad [\text{where } f'(t) = 2a+b]$$

$$\text{or} \quad a(t_2+t_1)+b = 2ac+b$$

$$\text{or} \quad c = \frac{1}{2}(t_2+t_1) \qquad \textbf{Ans.}$$

(c) *Formula of Finite Increments*: The relation, $f(b)-f(a) = f'(c)(b-a)$, $c\in(a,b)$, in the Mean Value Theorem is known as *the formula of increments. It states that the increment of a differentiable function on an interval is equal to the product of the derivative of the function at an intermediate point by the increment of the independent variable.*

The formula of finite increments makes it possible to find *the exact expression for the increment of a function in terms of the increment of the argument and the value of the derivative at an interior point of the interval.* It has significant theoretical importance and lies in the foundation of the proofs of a number of important theorems.

(7) In other words, if an object moves with varying velocity, then during motion of the object, a velocity is attained, which if it is applied as a uniform velocity (which is a constant velocity) for the same time interval, then the same displacement will be achieved.

Thus, if a car traveled 120 km in 2 h, then it must have traveled 60 km/h at some instant during motion. Of course, it is assumed that the car traveled throughout the interval.

Note (9): It is important that the reader thinks of the MVT whenever he sees a difference of functional values. That is, whenever the difference $f(b) - f(a)$ turns up, the reader should think of replacing it by the product $f'(c)\,(b - a)$ with the knowledge that c lies strictly between a and b.

(d) From the equation,

$$f(b) - f(a) = f'(c)(b - a) \tag{10A}$$

in the MVT, where $c \in (a, b)$, we get,

$$f(b) = f(a) + f'(c)(b - a) \tag{10B}$$

Here, the functional value $f(a)$ may be looked upon as *an approximation for f(b) with the error measured by a mean-derivative multiple of the deviation of b from a.*[8]

If we think of b as an independent variable on $[a, b]$, we can write (10B) in the form

$$f(x) = f(a) + f'(c)(x - a) \tag{10C}$$

valid for the interval $[a, x]$, with $c \in (a, x)$. The right-hand side of (10C) looks *like the linear approximation of f near a.*

If $f'(x)$ is continuous and c is close to a (as it will have to be if x is close to a), then, $f'(c)$ is close to $f'(a)$, and (10C) gives

$$f(x) \approx f(a) + f'(a)(x - a) \tag{10D}$$

which is the linear approximation of f near a.

Note (10): In Chapter 16, we produced and used linearizations without knowing exactly how good they were. Now, with an extended version of the MVT for the second derivative (to be studied later in Chapter 22), we shall see that the error in (10D) is proportional to $(x - a)^2$. Therefore, if $(x - a)$ is small the error will be very small.[9]

20.3.5 Alternate Form of the MVT

For the closed interval $[a, b]$, if we write $b = a + h$ (where h is a positive number) then the above interval becomes $[a, a + h]$ where h denotes the length of the interval and we have $h = b - a$.

Also, the number c lies between a and $a + h$, so that we write

$$a < c < a + h$$

$$\text{or} \quad 0 < c - a < h$$

$$\text{or} \quad 0 < \frac{c - a}{h} < 1$$

$$\text{or} \quad 0 < \theta < 1 \quad \text{where} \quad \theta = \frac{c - a}{h}$$

$$\therefore \quad c = a + \theta h$$ [10]

[8] Equation (10B) may also be looked upon as $f(a + h) = f(a) + f'(c)\cdot h$, where $b = (a + h)$.

[9] Later on, when we extend the Mean Value Theorem to Taylor's formula, we will be able to express $f(x)$ by extremely accurate polynomial approximations for a large class of functions that have derivatives of all orders.

[10] The number c that lies between a and $(a + h)$ is greater than a by some fraction of h. Here, θ is a proper fraction (i.e., $0 < \theta < 1$) and so we get $c = a + \theta h$.

Substituting this value of c in $(f(b)-f(a))/(b-a) = f'(c)$, we get

$$f(b) = f(a) + (b-a)f'(a+\theta h)$$
$$\text{or}\quad f(a+h) = f(a) + h\cdot f'(a+\theta h)$$

which is an alternate form of the MVT.

20.4 SOME APPLICATIONS OF THE MEAN VALUE THEOREM

The Mean Value Theorem is one of the most important results in Calculus. It is employed very often in proving other important theorems that may or may not be related to one another.

In Chapter 19a, we stated as the first derivative test for rise and fall, the fact that, a differentiable $f(x)$ increases on intervals where $f'(x) > 0$ and decreases on intervals where $f'(x) < 0$. *This fact can now be deduced from the Mean Value Theorem in the following way.*

(I) *Monotonicity Theorem*: Suppose that f is continuous on $[a, b]$ and differentiable on (a, b). If $f'(x) > 0$ throughout (a, b), then f is an increasing function on $[a, b]$. If $f'(x) < 0$ throughout (a, b), then f is decreasing on $[a, b]$.

In either case, f is one to one.

Proof: Let x_1 and x_2 be any two numbers in $[a, b]$ such that $x_1 < x_2$.

Applying the MVT to f on $[x_1, x_2]$,

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(c)$$

for some c between x_1 and x_2

$$\text{or}\quad f(x_2) - f(x_1) = f'(c)(x_2 - x_1) \tag{11}$$

The sign of the right-hand side of (11) is the same as the sign of $f'(c)$, because $(x_2 - x_1)$ is positive.

Therefore, $f(x_2) > f(x_1)$, if $f'(x)$ is positive on (a, b), that is, f is increasing, and $f(x_2) < f(x_1)$ if $f'(x)$ is negative on (a, b), that is, f is decreasing.

In either case, $x_1 \neq x_2$ implies that $f(x_1) \neq f(x_2)$, so f is one to one. *Hence proved.*

(II) *Constant Function Theorem*: Let f be continuous on a closed interval $[a, b]$. If $f'(x) = 0$ *for each point* x of (a, b), then f is constant on $[a, b]$.

Proof: Let x_1 and x_2 be arbitrary numbers in $[a, b]$, with $x_1 < x_2$. By the MVT, there is a number c in (x_1, x_2), such that,

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(c) \tag{12}$$

By assumption, $f'(c) = 0$, and thus (12) reduces to

$$f(x_2) - f(x_1) = 0$$
$$\therefore\quad f(x_2) = f(x_1)$$

Since x_1 and x_2 are arbitrary, it follows that f assigns the same value at any two points in $[a, b]$, so f is constant on $[a, b]$. (Proved)

Note (11): The above theorem states that the only functions whose derivatives are equal to zero throughout an interval, are the functions that are constant on that interval. *The most significant result implied by the above theorem is the following theorem, which gives the structure of functions having the same derivatives over an interval.*

(III) *Constant Difference Theorem*: Let f and g be continuous on a closed interval $[a, b]$. If $f'(x) = g'(x)$, for each point x of (a, b), then $f(x) - g(x)$ is *constant* on $[a, b]$. In other words, there is a constant C such that $f(x) = g(x) + C$, for all x in $[a, b]$.

Proof: We define a function $\phi(x) = f(x) - g(x)$, $x \varepsilon [a, b]$[11]
Differentiating both sides, we get,

$$\begin{aligned}\phi'(x) &= [f(x) - g(x)]' \\ &= f'(x) - g'(x) \quad \text{on } (a, b)\end{aligned}$$

But, it is given that $f'(x) = g'(x)$ for all x in (a, b).

$$\therefore \text{ we get} f'(x) = 0$$

$\therefore$ By the constant function theorem, it follows that,

$$\phi(x) = \phi(x_1) = \phi(x_2) = \cdots = C(\text{say}) \text{ for all } x_1,\ x_2 \text{ in } [a,\ b].$$

$$\therefore\ f(x) - g(x) = C \text{ or} f(x) = g(x)\, C. \text{ (Proved)}$$

Note (12): The above theorem says that the only way two functions can have identical rates of change on an interval is that their values should differ by some *fixed constant* on the interval.

For example, we know that for $f(x) = x^3, f'(x) = 3x^2$.

Therefore, if $g(x)$ is any differentiable function whose rate of change with respect to x is $3x^2$, that is, if $dg(x)/dx = 3x^2$ then $g(x) = x^3 + C$, for some constant C.

In determining that $g(x) = x^3 + C$, we say that we have determined g up to a constant.

If f is a function defined on an interval I, then any function F such that $F'(x) = f(x)$, for each x in I is called an *antiderivative* of f (since f is the derivative of F on I).

Thus, on any given interval, x^5 is an antiderivative of $5x^4$, x^2 is an antiderivative of $2x$, $\sin x$ is an antiderivative of $\cos x$, and e^{2x} is an antiderivative of $(1/2)e^{2x}$.

Once we know a single antiderivative F of a given function f, then all other antiderivatives can be ascertained by adding constants to F. It follows that on any interval I, the only antiderivatives of $3x^2$ are the functions of the form $x^3 + C$, the only antiderivatives of $\cos x$ are functions of the form $\sin x + C$, and so on.

Techniques for determining functions from their rates of change are extremely important in science and engineering. These techniques are discussed in Part II of this book.

Exercise

Verify the conditions of Rolle's Theorem for the following functions on their respective intervals and find c, if any, for which $f'(c) = 0$.

[11] The key to the proof is to show that the difference function $\phi(x) = f(x) - g(x)$ has derivative equal to zero on (a, b).

Q1. $y = f(x) = x^3 - 4x(-\infty < x < +\infty)$.

Q2. $y = f(x) = x^2(1-x)^2$ in $[0, 1]$.

Q3. $y = f(x) = 1 - \sqrt[3]{x^2} = 1 - x^{2/3}$.

Q4. It is given that for the function $y = f(x) = x^3 - 6x^2 + ax + b$ on [1,3]. Rolle's Theorem holds with $c = 2 + (1/\sqrt{3})$. Find the values of a and b.

Q5. On the curve $y = x^2$, find a point at which the tangent is parallel to the chord joining (0, 0) and (1, 1).

Q6. Verify MVT for the function $f(x) = (x-1)(x-2)(x-3)$ in [0, 4].

Q7. Find a point on the graph of $y = x^3$, where the tangent is parallel to the chord joining (1, 1) and (3, 27).

Q8. By use of MVT prove that $|\tan^{-1} x_2 - \tan^{-1} x_1| \leq |x_2 - x_1| \forall\, x_1, x_2$.

Q9. Using MVT prove that $x/(1+x) < \ln(1+x) < x,\ x > -1$.

Note (13): The solutions to these problems are available in Appendix C.

21 The Generalized Mean Value Theorem (Cauchy's MVT), L' Hospital's Rule, and their Applications

21.1 INTRODUCTION

The mean value theorem (MVT), also known as Lagrange's mean value theorem (LMVT), is the *fundamental mean value theorem that deals with a single function f(x). Augustin L. Cauchy discovered another mean value theorem* that uses two functions, $f(x)$ and $\phi(x)$, instead of one. It is known as the *generalized mean value theorem*, which is elegantly used in proving a rule, known as *L'Hospital's rule, which extends our ability to calculate limits*. We state and prove the *generalized mean value theorem.*

21.2 GENERALIZED MEAN VALUE THEOREM (CAUCHY'S MVT)

Theorem: If $f(x)$ and $\phi(x)$ are *two functions* such that

(i) $f(x)$ and $\phi(x)$ are *continuous on the closed interval* $[a, b]$;
(ii) $f(x)$ and $\phi(x)$ are *differentiable on the open interval* (a, b);
(iii) for all x *in the open interval* (a, b), $\phi'(x) \neq 0$
then, there exists a number $c \in (a, b)$, such that,

$$\frac{f(b)-f(a)}{\phi(b)-\phi(a)} = \frac{f'(c)}{\phi'(c)} \tag{1}$$

Proof: Let us denote by Q, the number $(f(b)-f(a))/(\phi(b)-\phi(a))$, so that we have

$$\frac{f(b)-f(a)}{\phi(b)-\phi(a)} = Q \tag{2A}$$

Now, we show that, $\phi(b)-\phi(a) \neq 0$, that is, $\phi(b) \neq \phi(a)$.

Assume $\phi(b) = \phi(a)$. Note that, with this assumption, ϕ satisfies all the conditions of Rolle's theorem. Hence, there exists some number c in (a, b), such that $\phi'(c) = 0$.

Further applications of derivatives: 21-The generalized mean value theorem (Cauchy's MVT), L' Hospital's rule and its applications in calculating limits of various indeterminate forms

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

But, condition (iii) of the hypothesis of the theorem demands that, for all x in (a, b), $\phi'(x) \neq 0$. Therefore, the above assumption leads to a contradiction. Hence, the assumption $\phi(b) = \phi(a)$ is false. Consequently, $\phi(b) - \phi(a) \neq 0$.[(1)]

From (2A), we get,

$$f(b) - f(a) - Q[\phi(b) - \phi(a)] = 0 \tag{2B}$$

Let us construct an *auxiliary function F(x)* defined by[(2)]

$$F(x) = f(x) - f(a) - Q[\phi(x) - \phi(a)] \tag{3}$$

Observe that,

1. $F(a) = 0$ and $F(b) = 0$
   (Note that, $F(a) = 0$ *from the definition of the function F(x)* and $F(b) = 0$ from the definition of the number Q. When we write b for x in (3), $F(b)$ becomes the LHS of (2B), which equals zero.)
2. $F(x)$ is continuous on $[a, b]$ since $f(x)$ and $\phi(x)$ both are continuous on $[a, b]$.
3. $F(x)$ has a derivative $F'(x)$ at every point in (a, b), since every term on the right-hand side of (3) has a derivative in (a, b). *Thus, the function F(x) satisfies all the hypotheses of Rolle's theorem on the interval* $[a, b]$. We, therefore, conclude that there exists a number $x = c$ between a and b such that $F'(c) = 0$.

By differentiating (3) both sides, we get,

$$F'(x) = f'(x) - Q\phi'(x),$$

$$\therefore\ F'(c) = f'(c) - Q\phi'(c) = 0, \quad [\text{since } F'(c) = 0]$$

$$\therefore\ Q = \frac{f'(c)}{\phi'(c)}, \quad c \in (a, b)$$

By substituting the value of Q in equation (2A), we get,

$$\frac{f(b) - f(a)}{\phi(b) - \phi(a)} = \frac{f'(c)}{\phi'(c)}$$

which is the *desired formula.*

Note (1): Observe that, if we take $\phi(x) = x$, then, we have $\phi(b) = b$, $\phi(a) = a$ and $\phi'(x) = 1$. Using these values in the above formula, it may be noted that the conclusion of Cauchy's MVT

[(1)] We get the same conclusion by applying LMVT to ϕ as follows: Because ϕ satisfies both conditions in the hypothesis of LMVT, there is a number c in (a, b) such that $\phi'(c) = [\phi(b) - \phi(a)]/(b - a)$. But, if $\phi(b) = \phi(a)$, (by assumption) we get $\phi'(c) = 0$, which contradicts the condition (iii) of Cauchy's MVT. Hence, $\phi(b) - \phi(a) \neq 0$. [Recall that, Rolle's Theorem is a special case of LMVT.]

[(2)] Note that, the auxiliary function $F(x)$ has been obtained by replacing b by x in (2B). It means that we are treating b as an independent variable, with c as a point lying in between a and x. This is justified, since both $f(x)$ and $g(x)$ are continuous on $[a, b]$, which means that x can vary from a to any $b \in [a, b]$. This permits us to replace b by x. Later on, it will be noted that we use the same understanding while extending the MVT to Taylor's formula in Chapter 22.

becomes the conclusion of Lagrange's MVT. *Thus, Lagrange's MVT is a special case of Cauchy's MVT. This justifies the name, generalized, mean value theorem.*

Note (2): Cauchy's MVT *cannot be proved by a simple term-by-term division of the relations expressing LMVT for the functions f and ϕ*, since in this case we would get (after canceling out $(b-a)$) the formula

$$\frac{f(b)-f(a)}{\phi(b)-\phi(a)}=\frac{f'(c_1)}{\phi'(c_2)}$$

in which $a < c\subset 1 < b$ and $a < c\subset 2 < b$.

This is obviously not the result of Cauchy's MVT (since, generally, $c_1 \neq c_2$).

21.2.1 Geometrical Interpretation of Cauchy's MVT

Now, we will show that *Cauchy's MVT* can be given the *same geometrical interpretation* as in the case of *Lagrange's MVT*. For this purpose, let us consider a curve in the x–y plane with parametric equations, $x=\phi(t)$ and $y=f(t)$. As the parameter t runs through the interval, say $[t_1, t_2]$, the variable point (x, y) describes a curve in the x–y plane, whose initial and final points are, respectively, $(\phi(t_1), f(t_1))$ and $(\phi(t_2), f(t_2))$.

The slope of the chord connecting these points is given by the ratio:

$$\frac{f(t_2)-f(t_1)}{\phi(t_2)-\phi(t_1)}$$

The derivative of y (regarded as a parametrically represented function of x) with respect to x is given by,

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\mathrm{d}y/\mathrm{d}t}{\mathrm{d}x/\mathrm{d}t}=\frac{f'(t)}{\phi'(t)}$$

Consequently, we get,

$$\frac{f(t_2)-f(t_1)}{\phi(t_2)-\phi(t_1)}=\frac{f'(c)}{\phi'(c)}, (t_1 < c < t_2)$$

Note that, the LHS of this equation expresses the slope of the chord subtending an arc, whereas the RHS represents the slope of the tangent line drawn at some intermediate point c of the arc.

Note (3): Rolle's theorem, the mean value theorem (i.e., Lagrange's MVT), and the generalized MVT (i.e.. Cauchy's MVT) imply that there exists some "middle point" $c \in (a, b)$ at which some of the named relations are true. For this reason, *all these theorems are collectively named the mean value theorems for derivatives.*

21.3 INDETERMINATE FORMS AND L'HOSPITAL'S RULE

In Chapter 7, we stated the *quotient rule for computing limits*:

$$\lim_{x\to x_0}\frac{f(x)}{g(x)}=\frac{\lim\limits_{x\to x_0} f(x)}{\lim\limits_{x\to x_0} g(x)}$$

provided the *limits on the right-hand side exist* and $\lim_{x \to x_0} g(x) \neq 0$. However, there are examples of quotient functions that have a limit, even though

$$\lim_{x \to x_0} g(x) = 0.$$

Perhaps the simplest example of such a function is x/x, with $x \to 0$. Thus, we have,

(i) $\lim_{x \to 0} \frac{x}{x} = 1$

(ii) Another example of this type is

$$\lim_{x \to 2} \frac{x^2 - 3x + 2}{x^2 + x - 6} = \lim_{x \to 2} \frac{(x-2)(x-1)}{(x-2)(x+3)}$$
$$= \lim_{x \to 2} \frac{x-1}{x+3} = \frac{1}{5}$$

(iii) A less trivial example is

$$\lim_{x \to 0} \frac{\sin x}{x} = 1$$

While the limits at (i) and (ii) are computed by *cancellation of factors*, you may recall that *an intricate geometric argument led to the conclusion at* (iii) above.

Note that, all these limits have a *common feature*. In each case, *a quotient is involved in which both numerator and denominator have 0 as their limits.*

In all these cases, *the quotient rule does not apply* since it requires that the *limit of the denominator be different from* 0. However, as we have seen, these limits may exist. Of course, we cannot use the quotient rule cannot determine them.

Note that, except for canceling factors, where possible, *we have so far no systematic method for evaluating limits of quotients in which both the numerator and the denominator have* "0" *as their limits.*

L'Hospital's rule provides an *extremely simple and convenient method for evaluating the limits of such quotients.*

21.3.1 Indeterminate Form 0/0 and Evaluating its Limit

Let the functions $f(x)$ and $g(x)$ be defined in a neighborhood of a point $x = x_0$ and let $f(x_0) = 0 = g(x_0)$. Then the ratio $f(x)/g(x)$ *is not defined for* $x = x_0$, *but may have a very definite meaning for values of* $x \neq x_0$. *Hence, we can raise the question of searching for the limit of this ratio as* $x \to x_0$. Evaluating limit(s) of this type is usually *known as evaluating indeterminate form of the type* $0/0$.[3]

This form gives an explicit connection between derivatives and limits that lead to the indeterminate form $0/0$. This rule stands in two forms, namely, the "*first form*" and the "*stronger form*", both discovered independently. While the first form follows from a simple

[3] There *are seven Indeterminate Forms* viz. $\frac{0}{0}$, $\frac{\infty}{\infty}$, $0.\ \infty$, $\infty - \infty$, 0^0, ∞^0 and 1^∞. All these forms can be brought to the indeterminate form, 0/0, by suitable arrangement and so we shall first discuss this form. [Later on, it will be shown why other possible symbols like $\frac{0}{\infty}$, $\frac{\infty}{0}$, $\infty + \infty$, $\infty \cdot \infty$, 0^∞ and ∞^∞ cannot be considered as indeterminate forms.]

observation, *the stronger form is based on Cauchy's MVT.* The first form states as given in the following theorem.

21.3.2 Theorem

Suppose that f and g satisfy the following conditions:

(i) $f(a) = g(a) = 0$
(ii) $f'(a)$ and $g'(a)$ exist, and that
(iii) $g'(a) \neq 0$ then,

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}$$

Thus, if f and g satisfy certain condition (as stated above), then $\lim_{x \to a} f(x)/g(x)$ equals the ratio $f'(a)/g'(a)$ of derivatives, where $g'(a) \neq 0$.

Remark: Since nothing is said about the location of "a" in the domain (common interval) of f and g, we conclude that it can be anywhere in the interval. Now the question is: What can we do about the limit $\lim_{x \to a} f(x)/g(x)$ if in the ratio $f'(a)/g'(a)$, $g'(a) = 0$?[(4)]

We get the answer to this question from the "stronger form" of L'Hospital's rule.

21.3.3 Statement of the Stronger Form of L'Hospital's Rule

Roughly speaking, the *stronger form* of L'Hospital's rule says if $f(a) = g(a) = 0$, then,

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}$$

provided the derivatives and limit on the right-hand side exist.

Note (4): *In fact, it is a stronger form of the rule we call L'Hospital's rule for evaluating the limit in the indeterminate form* $0/0$. *It says that whenever the rule gives* $0/0$, *we can apply it again, repeating the process until we get a different result.* (It will be found that this rule is useful in determining limits of all types of indeterminate forms.)

Here, we restate that the proof of L'Hospital's rule is based on Cauchy's MVT (that we have already proved) and that *this rule cannot be proved from the "first form"*.

At a glance, both the forms might appear identical, but the distinction between the two can be easily observed. Though our interest lies only in the "*stronger form*" of the rule, it is useful to observe and enjoy the approach leading to the proof of the first form of the rule and then study carefully the stronger form.

Recall that derivatives (at a point) are calculated using the limit

$$f(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \tag{4}$$

Also, note that this limit always produces the indeterminate form $0/0$.

(4) For example, consider the limit $\lim_{x \to 0} (x - \sin x)/x^3 = \lim_{x \to 0} (x - \sin x)'/(x^3)' = \lim_{x \to 0} (1 - \cos x)/(3x^2) = 0|_{x=0}$. The first form of L'Hospital's rule does not tell us what the limit is because the derivative of $g(x) = x^3$ is zero at $x = 0$.

From the limit at (4), we get an idea *to use derivatives to calculate limits that lead to indeterminate forms* 0/0.

For example, consider $\lim_{x \to 0} \sin x/x$. Note that $\lim_{x \to 0} \sin x = \sin 0 = 0$.

$$\therefore \lim_{x \to 0} \frac{\sin x}{x} = \lim_{x \to 0} \frac{\sin x - \sin 0}{x - 0} = \frac{\mathrm{d}}{\mathrm{d}x}(\sin x)\Big|_{x=0} = \cos 0 = 1$$

Now, we state and prove both forms of L'Hospital's rule.

21.4 L'HOSPITAL'S RULE (FIRST FORM)

Suppose that f and g satisfy the following conditions:

(i) $f(a) = g(a) = 0$
(ii) $f'(a)$ and $g'(a)$ exist, and that
(iii) $g'(a) \neq 0$ then,

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)}$$

Proof: Working backward, from $f'(a)$ and $g'(a)$, which are themselves limits, we have

$$\frac{f'(a)}{g'(a)} = \frac{\lim_{x \to a}(f(x) - f(a))/(x-a)}{\lim_{x \to a}(g(x) - g(a))/(x-a)} = \lim_{x \to a} \frac{f(x) - f(a)}{g(x) - g(a)} = \lim_{x \to a} \frac{f(x) - 0}{g(x) - 0} = \lim_{x \to a} \frac{f(x)}{g(x)}$$

Therefore, we have,

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)} \tag{5}$$

provided that $f'(a)$ and $g'(a)$ exist and that $g'(a) \neq 0$.

Note (5): Having observed the limitation of the *first form*, we give below the statement of the *stronger form* of L'Hospital's rule. Here, we wish to analyze the statement in detail so that whenever we restate it as a theorem, its proof can be easily understood. (This approach is important since its proof depends on Cauchy's MVT, demanding careful attention.)

21.4.1 Theorem: L'Hospital's Rule (Stronger Form)

Suppose that $f(x_0) = g(x_0) = 0$ and that *the functions f and g are both differentiable on an open interval (a, b) that contains the point* x_0.

Suppose also, *that $g'(x) \neq 0$ at every point in (a, b), except possibly at x_0*. Then,

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)}$$

provided the limit on the right-hand side exists.

Before providing the proof of this theorem, the following points (from the hypotheses) must be carefully noted:

(i) The function *f* and *g both are differentiable on some open interval* (a, b), which contains a point x_0, and that $f(x_0) = g(x_0) = 0$.

(ii) The point x_0 *can be anywhere in (a, b).*

(iii) $g'(x) \neq 0$ *at every point in (a, b), except possibly at* x_0.

(iv) The existence of $\lim_{x \to x_0} f'(x)/g'(x)$ *implies that both $f'(x)$ and $g'(x)$ exist in at least a small interval $(x_0, x]$, wherein $g'(x) \neq 0$.*

Since $f'(x)$ and $g'(x)$ both exist in $(x_0, x]$, it follows that *f* and *g* both are continuous in this interval, *but we do not know whether they are continuous at* x_0. (This observation is important.)

(v) *L'Hospital's theorem says nothing about the limits of f(x) and g(x) as* $x \to x_0$, but the values $f(x_0)$ and $g(x_0)$ are given to be zero. This suggests that by defining $\lim_{x \to x_0^+} f(x) = 0$ and $\lim_{x \to x_0^+} g(x) = 0$ we can make both *f* and *g* (right) continuous at x_0. *This step is very important* since we can now say that both these functions satisfy the hypotheses of Cauchy's MVT on the closed interval $[x_0, x]$.

(Whenever we say that a function is *continuous* on *a closed interval* $[a, b]$, we mean that it is continuous in the open interval (a, b), *right continuous* at a and *left continuous* at b.)

Note (6): Applying the same logic, as in step (v) above, we can make *f* and *g* satisfy the hypotheses of Cauchy's MVT on the closed interval $[x, x_0]$.

Remark: Observe that the functions *f* and *g* satisfying the conditions of L'Hospital's theorem *can be made to satisfy the conditions of Cauchy's MVT,* by defining (or redefining) them suitably *so that they become continuous at any point $x_0 \in (a, b)$,* where $g'(x_0)$ could vanish though $g'(x) \neq 0$ at any other point in (a, b).[5]

Also, note that *this way of defining f and g does not affect the limit,* $\lim_{x \to x_0} f(x)/g(x)$ *for the reason given at (v) above.*

Now, we are in a position to prove L'Hospital's rule (stronger form), but before attempting *to prove it*, we illustrate it through the following solved examples.

Example (1): Use L'Hospital's rule to show that

$$\lim_{x \to 0} \frac{\sin x}{x} = 1 \text{ and } \lim_{x \to 0} \frac{1 - \cos x}{x} = 0$$

Solution: Recall that we worked pretty hard to demonstrate these two facts in Chapter 11. After *noting that both limits have the* $0/0$ *form, we can now establish the desired results in two lines.*

[5] It must be clear that, at most, the value $g'(x_0)$ could be equal to zero, but for any other point $x \in (a, b)$, $g'(x) \neq 0$.

By L'Hospital's rule,

$$\lim_{x\to 0}\frac{\sin x}{x} = \lim_{x\to 0}\frac{(\sin x)'}{(x)'} = \lim_{x\to 0}\frac{\cos x}{1} = 1$$

$$\lim_{x\to 0}\frac{1-\cos x}{x} = \lim_{x\to 0}\frac{(1-\cos x)'}{(x)'} = \lim_{x\to 0}\frac{\sin x}{1} = 0$$

Example (2): Find $\lim_{x\to 2}(x^2-3x+2)/(x^2+x-6)$.

Solution: This limit has the $0/0$ form, so by L'Hospital's rule,

$$\lim_{x\to 2}\frac{x^2-3x+2}{x^2+x-6} = \lim_{x\to 2}\frac{2x-3}{2x+1} = \frac{1}{5} \quad \textbf{Ans}.$$

Recall that this limit was handled earlier by the method of factoring.
Of course, we get the same answer either way.

Example (3): Find $\lim_{x\to 0}\tan 2x/\log_e(1+x)$.

Solution: Both the numerator and the denominator have limit 0. Hence,

$$\lim_{x\to 0}\frac{\tan 2x}{\log_e(1+x)} = \lim_{x\to 0}\frac{2\sec^2 2x}{1/(1+x)} = \frac{2}{1} = 2 \quad \textbf{Ans}.$$

Caution: Note that, to apply L'Hospital's rule to f/g, we divide the derivative of f by the derivative of g. *Do not fall into the trap of taking the derivative of the ratio* f/g. The quotient to use is f'/g', not $(f/g)'$.

Now we proceed to prove L'Hospital's theorem.

21.5 L'HOSPITAL'S THEOREM (FOR EVALUATING LIMITS(S) OF THE INDETERMINATE FORM 0/0.)

Theorem: Suppose that,

$$f(x_0) = g(x_0) = 0$$

and that *the functions f and g are both differentiable on an open interval* (a, b) that contains point x_0.

Suppose also *that $g'(x) \neq 0$ at every point in (a, b), except possibly at x_0*. Then,

$$\lim_{x\to x_0}\frac{f(x)}{g(x)} = \lim_{x\to x_0}\frac{f'(x)}{g'(x)} \tag{6}$$

provided the limit on the right-hand side exists.

(To prove the theorem, we must find a closed interval in (a, b) on which both the functions f and g are continuous.)

Proof: We first establish equation (6) for the case $x \to x_0^+$. The method needs almost no change to apply to $x \to x_0^-$, and the combination of these two cases establishes the result.

It is given that $f'(x)$ and $g'(x)$ exist on (a, b) and that $g'(x) \neq 0$, *at every point in (a, b), except possibly at $x_0 \in (a, b)$. Of course, the location of x_0* (in (a, b)) *is not known.*

Suppose that x lies to the right of x_0 (so that $a < x_0 < x < b$). Then, $g'(x) \neq 0$ on $(x_0, x]$. Since $f'(x)$ and $g'(x)$ exist on (a, b), *it follows that f and g are both continuous on* $(x_0, x]$. But, it is also given that $f(x_0) = 0$ and $g(x_0) = 0$. Hence, f and g both can be made *(right) continuous at* x_0 by defining

$$\lim_{x \to x_0^+} f(x) = 0 \quad \text{and} \quad \lim_{x \to x_0^+} g(x) = 0^{(6)}$$

This permits us to say that the functions f and g both satisfy the hypotheses of Cauchy's MVT on the closed interval $[x_0, x]$. This produces a number c between x_0 and x such that

$$\frac{f'(c)}{g'(c)} = \frac{f(x) - f(x_0)}{g(x) - g(x_0)} \tag{7}$$

But $f(x_0) = g(x_0) = 0$, so that,

$$\frac{f'(c)}{g'(c)} = \frac{f(x)}{g(x)} \tag{8}$$

As x approaches x_0, *c approaches x_0 because it lies between x and x_0.*

Therefore,

$$\lim_{x \to x_0^+} \frac{f(x)}{g(x)} = \lim_{c \to x_0^+} \frac{f'(c)}{g'(c)} = \lim_{x \to x_0^+} \frac{f'(x)}{g'(x)} \tag{9}$$

This establishes L'Hospital's rule for the case where *x approaches x_0 from the right*. The case where x approaches x_0 from the left is proved by applying Cauchy's MVT on the closed interval $[x, x_0]$, $x < x_0$.

Finally, by combining these two cases, we get the desired result:

$$\left.\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)} \right\} \tag{10}$$

provided the limit on the right-hand side exists.

Proved

Remark (1): The conclusion of L'Hospital's rule (for $0/0$ forms) is that

$$\left.\begin{array}{l} \displaystyle\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)} \\ \text{provided } \displaystyle\lim_{x \to x_0} \frac{f'(x)}{g'(x)} \text{ exists} \end{array}\right\} \tag{11}$$

(6) Note that, in this way we simply make f and g *right continuous* at x_0 in $[x_0, x]$. In any case, these definitions do not affect the limit $\lim_{x \to x_0} f(x)/g(x)$ since the limit does not depend on whether f and g are defined at x_0.

The statement (11) (representing L'Hospital's rule above) allows, *under some specific conditions* (which must be satisfied by f and g), *to replace a limit of a quotient of functions by a limit of a quotient of their derivatives, which is sometimes easier to compute.*[7]

Remark (2): Just because we have an elegant rule does not mean that we should use it indiscriminately. L'Hospital's rule does not apply when either the numerator or the denominator has a finite nonzero limit. Hence, we apply L'Hospital's rule as long as we still get the form $0/0$ at $x = x_0$. To make sure at every stage whether the rule applies, we reflect our observation on the right-hand side of each step to indicate whether the expression is in the form [0/0], [still 0/0], or [not 0/0], as indicated in the following solved examples. To find $\lim_{x \to x_0} f(x)/g(x)$, by L'Hospital's rule, we proceed to differentiate *f*(x) and *g*(x) as long as we still get the form $0/0$ at $x = x_0$. We stop differentiating when either the numerator or the denominator has a nonzero limit. L'Hospital's rule does not apply when either the numerator or the denominator has a finite nonzero limit.

Example (4): Find $\lim_{x \to 0}(x - \sin x)/x^3$.

Solution: $\lim_{x \to 0} \dfrac{x - \sin x}{x^3}$ $\quad \left[\text{This is in the form} \left[\dfrac{0}{0}\right]\right]$

$$= \lim_{x \to 0} \frac{1 - \cos x}{3x^2} \quad \left[\text{still } \frac{0}{0}\right]$$

$$= \lim_{x \to 0} \frac{\sin x}{6x} \quad \left[\text{still} \frac{0}{0}\right] = \lim_{x \to 0} \frac{\cos x}{6}$$

(At this stage, we stop differentiation and evaluate the limit.)

$$\therefore \quad \lim_{x \to 0} \frac{\cos x}{6} = \frac{1}{6} \qquad \textbf{Ans}.$$

Example (5): Find $\lim_{x \to 0}(1 - \cos x)/(x^2 + 3x)$.

Solution: $\lim_{x \to 0} \dfrac{1 - \cos x}{x^2 + 3x} = \lim_{x \to 0} \dfrac{(1 - \cos x)'}{(x^2 + 3x)'}$

$$= \lim_{x \to 0} \frac{\sin x}{2x + 3} \quad \left[\text{not} \frac{0}{0}\right]$$

(At this stage, we stop differentiation and evaluate the limit.)

$$= \frac{0}{3} = 0 \quad \text{(Right)}$$

[7] In practice, the functions we deal with (in this book) satisfy the hypotheses of L 'Hospital's Rule.

If we continue to differentiate in an attempt to apply L'Hospital's rule once more, we get the wrong result as follows:

$$\lim_{x\to 0}\frac{1-\cos x}{x^2+3x} = \lim_{x\to 0}\frac{\sin x}{2x+3} = \lim_{x\to 0}\frac{\cos x}{2} = \frac{1}{2}$$

which is wrong.

Note (7): In applying L'Hospital's rule, *we may reach a point where one of the derivatives is zero at* $x = x_0$ *and the other is not.* Then, the limit of the fraction is either zero as in Example (5) or infinity as in Examples (6) and (7).

Example (6):

$$\lim_{x\to 0}\frac{\sin x}{x^2} \quad \left[\frac{0}{0}\right]$$

$$\lim_{x\to 0}\frac{(\sin x)'}{(x^2)'} = \lim_{x\to 0}\frac{\cos x}{2x} = \infty$$

Example (7):

$$\lim_{x\to 0}\frac{1-\cos x}{x-\sin x}, \quad \left[\frac{0}{0}\right]$$

$$\lim_{x\to 0}\frac{\sin x}{1-\cos x}, \quad \left[\frac{0}{0}\right]$$

$$\lim_{x\to 0}\frac{\cos x}{\sin x} = \infty$$

Remark (3): Recall that, we have already defined infinity "∞" as a limit (in Chapter 7b) though it does not represent a real number. Accordingly, in view of Remark (2), L'Hospital's rule remains valid when the *ratio of the derivatives tends to infinity*, as we have seen in Examples (6) and (7). In other words, if $\lim\limits_{x\to x_0} f'(x)/g'(x) = \infty$, then it follows that $\lim\limits_{x\to x_0} f(x)/\, g(x) = \infty$.

It is easy to justify this. If $\lim\limits_{x\to x_0} g'(x) = 0$, but $\lim\limits_{x\to x_0} f'(x) \neq 0$, then the theorem is applicable to the reciprocal ratio $g(x)/f(x)$, which tends to 0 as $x \to x_0$. Hence, the ratio $f(x)/g(x)$ tends to infinity.

Remark (4): L'Hospital's rule $\lim\limits_{x\to x_0} f(x)/g(x) = \lim\limits_{x\to x_0} f'(x)/g'(x)$ holds also for the case where the functions $f(x)$ and $g(x)$ are not defined at $x = x_0$, but $\lim\limits_{x\to x_0} f(x) = 0$, $\lim\limits_{x\to x_0} g(x) = 0$.

In order to reduce this case to the earlier considered case, we redefine the functions f(x) and g(x) at point $x = x_0$, *so that they become continuous at point* x_0. To do this, it is sufficient to put $f(x_0) = \lim\limits_{x\to x_0} f(x) = 0, g(x_0) = \lim\limits_{x\to x_0} g(x) = 0$.

Note that redefining f and g in this way does not affect the limit $\lim\limits_{x\to x_0} f(x)/g(x)$,(since the limit (at $x \to x_0$) *does not depend on whether the functions f(x) and g(x) are defined at* $x = x_0$).

Note (8): L'Hospital's rule (for $0/0$ form) is also applicable if $\lim\limits_{x\to\infty} f(x) = 0$ and $\lim\limits_{x\to\infty} g(x) = 0$.

Indeed, putting $x = 1/t$, we see that $t \to 0$ as $x \to \infty$ and therefore from the statement $\lim\limits_{x\to\infty} f(x) = 0$, it follows that $\lim\limits_{t\to 0} f(1/t) = 0$ and similarly from the statement $\lim\limits_{x\to\infty} g(x) = 0$ we get $\lim\limits_{t\to 0} g(1/t) = 0$.

Applying the L'Hospital's rule to the ratio $f(1/t)/g(1/t)$, we find

$$\lim_{x\to\infty}\frac{f(x)}{g(x)} = \lim_{t\to 0}\frac{f(1/t)}{g(1/t)} = \lim_{t\to 0}\frac{f'(1/t)(-1/t^2)}{g'(1/t)(-1/t^2)} = \lim_{t\to 0}\frac{f'(1/t)}{g'(1/t)} = \lim_{x\to\infty}\frac{f'(x)}{g'(x)}$$

which is what we wanted to prove.

Example (8): Evaluate $\lim\limits_{x\to\infty}\sin(k/x)/(1/x)$.

$$\lim_{x\to\infty}\frac{\sin(k/x)}{1/x} = \lim_{x\to\infty}\frac{k\cos(k/x)(-1/x^2)}{-1/x^2} = \lim_{x\to\infty} k\cos\frac{k}{x} = k \quad \text{Ans.}$$

(If, for $x \to x_0$ (or for $x \to \infty$), both $f(x)$ and $g(x)$ simultaneously tend to infinity, then L'Hospital's rule remains valid, but the proof becomes more sophisticated and we do not treat it here.)

Thus, L'Hospital's rule is extended to state as follows:

If $\lim\limits_{x\to\infty} f(x) = \infty$, $\lim\limits_{x\to\infty} g(x) = \infty$, and $\lim\limits_{x\to\infty} f'(x)/g'(x)$ exist, then $\lim\limits_{x\to\infty}\dfrac{f(x)}{g(x)} = \lim\limits_{x\to\infty}\dfrac{f'(x)}{g'(x)}$

Warning: L'Hospital's rule *can be applied only when an indeterminate form is reduced to the form* $0/0$ *or* ∞/∞ *(since it is proved only for these forms). We therefore emphasize that L'Hospital's rule must not be applied to compute* $\lim\limits_{x\to a} f(x)/g(x)$ *unless the quotient* $f(a)/g(a)$ *is an indeterminate form* $0/0$ *or* ∞/∞. To illustrate note that

$$\lim_{x\to 0}\frac{x^2}{\cos x} = \frac{0}{1} = 0$$

In this case, the limit of the denominator is 1, which is a nonzero real number. Hence, we cannot apply L'Hospital's rule in this case (see Remark (2)). If we apply the rule in such cases, the result may be incorrect. Let us see what happens if we apply L'Hospital's rule in this case:

$$\lim_{x\to 0}\frac{x^2}{\cos x} = \lim_{x\to 0}\frac{2x}{-\sin x} = \lim_{x\to 0}\frac{2}{-\cos x} = \frac{2}{-1} = -2$$

which is wrong.

Note (9): When L'Hospital's rule is used repeatedly, *it is advisable to perform beforehand all possible simplifications of the given expression, for instance, to cancel the common factors and to use the limits already known.*

Example (9):

$$\lim_{x\to 0}\frac{\sqrt{1+x}-1-(x/2)}{x^2} \quad \left[\frac{0}{0}\right]$$

$$= \lim_{x\to 0}\frac{(1/2)(1+x)^{-1/2}-1/2}{2x} \quad \left[\text{still}\,\frac{0}{0}\right]$$

$$= \lim_{x\to 0}\frac{(-1/4)(1+x)^{-3/2}}{2} = -\frac{1}{8} \qquad \textbf{Ans.}$$

Example (10):

$$L = \lim_{x\to 0} \frac{\frac{1}{1+x^2}\frac{1}{\sqrt{1-x^2}}}{\left(\frac{1}{\cos^2 x}\right) - \cos x} \quad \left[\frac{0}{0}\right]$$

$$\frac{d}{dx}[1-\cos x + x\sin x] = \sin x + x\cos x + \cos x$$

$$= \quad 2\sin x + x\cos x$$

$$= \lim_{x\to 0}\frac{2\ \sin\ x + \text{xcosx}}{\sin\ x} \quad \left[\frac{0}{0}\right]$$

$$\frac{d}{dx}[2\sin x + x\cos x] = 2\cos x + x\sin x + \cos x$$

$$= 3\cos x - x\sin x$$

$$= \lim_{x\to 0}\frac{3\cos x - x\sin x}{\cos x} = 3 \qquad \textbf{Ans.}$$

Example (11):

$$\lim_{x\to 0}\frac{\tan^{-1}x - \sin^{-1}x}{\tan x - \sin x} \quad \left[\frac{0}{0}\right]$$

Solution: Let the above limit be denoted by L, then by applying L'Hospital's rule we get

$$L = \lim_{x\to 0}\frac{(1/1+x^2) - 1/\sqrt{1-x^2}}{(1/\cos^2 x) - \cos x} \quad \left[\frac{0}{0}\right] \tag{12}$$

In order to express this limit in simplified form, it is useful to consider the Numerator (Nr) and the Denominator (Dr) in the expression (12) as follows:

$$\text{Nr} = \frac{\sqrt{1-x^2} - (1+x^2)}{(1+x^2)\sqrt{1-x^2}} = \frac{\sqrt{1-x^2} - 1 - x^2}{(1+x^2)\sqrt{1-x^2}} \quad \text{and}$$

$$\text{Dr} = \frac{1-\cos^3 x}{\cos^2 x} = \frac{(1-\cos x)(1+\cos x + \cos^2 x)}{\cos^2 x}$$

$$\therefore\ L = \lim_{x\to 0}\frac{\sqrt{1-x^2} - 1 - x^2}{1-\cos x}\frac{\cos^2 x}{(1+\cos x+\cos^2 x)(1+x^2)\sqrt{1-x^2}}$$

In this expression, note that the limit of the second factor is 1/3; therefore, on *applying L'Hospital's rule to the first factor, we get*,

$$L = \frac{1}{3}\lim_{x\to 0}\frac{(\sqrt{1-x^2} - 1 - x^2)'}{(1-\cos x)'} = \frac{1}{3}\lim_{x\to 0}\frac{(1/2)(1-x^2)^{-1/2}(2x) - 2x}{\sin x}$$

$$L = \frac{1}{3}\lim_{x\to 0}\frac{(-2x/2\sqrt{1-x^2})-2x}{\sin x} = -\frac{1}{3}\lim_{x\to 0}\frac{x\Big((1/\sqrt{1-x^2})+2\Big)}{\sin x}$$

$$= -\frac{1}{3}\lim_{x\to 0}\frac{x}{\sin x}\left(\frac{1}{1-x^2}+2\right) = \left(-\frac{1}{3}\right)\cdot 1\cdot(1+2) = -1 \qquad \textbf{Ans.}$$

Note (10): *Evaluating indeterminate forms of the type* ∞/∞, $\infty\cdot 0$, and $\infty-\infty$.

Sometimes, *when we try to evaluate a limit as* $x \to a$ (by substituting $x = a$), we get an ambiguous expression like ∞/∞, $\infty\cdot 0$, or $\infty-\infty$, instead of $0/0$.

In more advanced books, it is proved that L'Hospital's rule applies both to the indeterminate form ∞/∞ *and to* $0/0$. Shortly, we will show that expressions such as $\infty\cdot 0$ (or $0\cdot\infty$) and $\infty-\infty$ can be easily expressed in the form ∞/∞ or $0/0$.

(As for the remaining indeterminate forms, we shall show, through solved examples, how they can be brought to these forms.)

21.6 EVALUATING INDETERMINATE FORM OF THE TYPE ∞/∞

We now consider the question of the limit of a ratio of functions $f(x)$ and $g(x)$ approaching infinity as $x \to a$ (or as $x \to \infty$).

Theorem: Let the functions $f(x)$ and $g(x)$ *be continuous and differentiable for* all $x \neq a$ in the neighborhood of the point a and the derivative $g'(x)$ does not vanish. Furthermore, let $\lim_{x\to a} f(x) = \infty$, $\lim_{x\to a} g(x) = \infty$, and let there be a limit $\lim_{x\to a} f'(x)/g'(x) = L$, which may be either a finite number or $-\infty$ or $+\infty$, then there is a limit $\lim_{x\to a} f(x)/g(x)$, and

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \lim_{x\to a}\frac{f'(x)}{g'(x)} = L^{(8)}$$

(In the notation $x \to a$, "a" may be either finite or infinite.)

Note (11): A rigorous proof of the above theorem is *quite difficult, but there is an intuitive way of seeing that the result has to be true*. It is important to analyze and assign some logical meaning to the symbol ∞/∞.

Imagine that $f(t)$ and $g(t)$ represent the positions of two cars on t-axis at time t. *These two cars* (the f-car and the g-car) *are on an endless journey with respective velocities* $f'(t)$ and $g'(t)$. Now, if $\lim_{x\to\infty} f'(t)/g'(t) = L$, *then ultimately the f-car travels about L times as fast as the g-car*. It is therefore reasonable to say that in the long run, the f-car will travel about L times (the distance) as what is traveled by the g-car; that is, $\lim_{x\to\infty} f(t)/g(t) = L$. Thus, a meaning has been assigned to the expression $\infty/\infty = L$, *wherein L may be finite or infinite.*

We do not call this a proof, *but it gives a logical meaning to the limit of a ratio of functions,* which take the form ∞/∞.

[8] The proof of this theorem is available in *Differential and Integral Calculus* by N. Piskunove (vol. I, 140–143), Mir Publishers, Moscow, 1974.

Example: (12):

$$\lim_{x\to\pi/2}\frac{\tan x}{\tan 3x}\quad\left[\frac{\infty}{\infty}\right]$$

$$=\lim_{x\to\pi/2}\frac{\sec^2 x}{3\sec^2 3x}=\lim_{x\to\pi/2}\frac{1/\cos^2 x}{3/\cos^2 3x}\quad\left[\text{Still}\frac{\infty}{\infty}\right]$$

$$=\lim_{x\to\pi/2}\frac{1}{3}\frac{\cos^2 3x}{\cos^2 x}\quad\left[\text{Still}\frac{\infty}{\infty}\right]$$

$$=\lim_{x\to\pi/2}\frac{1}{3}\frac{2\cos 3x\cdot(-\sin 3x)\cdot 3}{2\cos x(-\sin x)}$$

$$=\lim_{x\to\pi/2}\frac{1}{3}\frac{2\cdot 3\cdot\cos 3x\sin 3x}{2\cos x\sin x}$$

$$=\lim_{x\to\pi/2}\frac{\cos 3x}{\cos x}\lim_{x\to\pi/2}\frac{\sin 3x}{\sin x}\left[\text{note that }\lim_{x\to\pi/2}\frac{\cos 3x}{\cos x}\text{is of the form}\frac{0}{0}\right]$$

$$=\lim_{x\to\pi/2}\frac{-3\sin 3x}{-\sin x}\frac{(-1)}{(1)}\quad\left[\because\quad \sin 3\frac{\pi}{2}=-1\quad\text{and}\quad\sin\pi/2=1\right]$$

$$=3\frac{(-1)}{(1)}\frac{(-1)}{(1)}=3$$ **Ans.**

Example (13):

$$\lim_{x\to\infty}\frac{ax^2+b}{cx^2-d}\quad\left[\frac{\infty}{\infty}\right]$$

$$=\lim_{x\to\infty}\frac{2ax}{2cx}=\frac{a}{c}$$ **Ans.**

Example (14): Find $\lim_{x\to\infty} e^x/x$.

$$\lim_{x\to\infty}\frac{e^x}{x}=\lim_{x\to\infty}\frac{(e^x)'}{(x)'}=\lim_{x\to\infty}\frac{e^x}{1}=\infty$$ **Ans.**

Example (15):

$$\lim_{x\to\infty}\frac{x}{e^x}$$

This is of the form $[\infty/\infty]$.
Applying L'Hospital's rule, we get

$$\lim_{x\to\infty}\frac{x}{e^x}=\lim_{x\to\infty}\frac{1}{e^x}=0$$

Note (12): *Generally, for any integer* $n > 0$

$$\lim_{x\to\infty}\frac{x^n}{e^x} = \lim_{x\to\infty}\frac{nx^{n-1}}{e^x}$$
$$= \lim_{x\to\infty}\frac{n(n-1)(n-2)\ldots 1}{e^x} = 0$$

Obviously, *for any real number* $k > 0$, $\lim_{x\to\infty} x^k/e^x = 0$.

Example (16): Show that, if a is any positive real number, then,

$$\lim_{x\to\infty}\frac{\ln x}{x^a} = 0$$

Solution: Both $\ln x$ and x^a tend to ∞ as $x \to \infty$. Hence, by the application of L'Hospital's rule,

$$\lim_{x\to\infty}\frac{\ln x}{x^a} \quad \left[\frac{\infty}{\infty}\right]$$
$$= \lim_{x\to\infty}\frac{1/x}{ax^{a-1}} = \lim_{x\to\infty}\frac{1}{ax^a} = 0 \qquad \textbf{Proved}$$

Remark (5): Examples (15) and (16) *imply something that is worth mentioning. In Example (15) for large x,* e^x *grows faster than any constant power of x,* while in Example (16) $\ln x$ *grows slower than any constant power of x.*

For example, e^x grows faster than x^{100}, and $\ln x$ grows slower than $\sqrt[100]{x}$.

The following chart offers additional illustration indicating how some common functions grow.[(9)]

x	10	100	1000
ln x	2.3	4.6	6.9
$\sqrt{x}$	3.2	10	31.6
$x \ln x$	23	46.1	6908
x^2	100	10,000	10^6
e^x	10^4	10^{43}	10^{434}

Example (17): Find $\lim_{x\to\infty} e^x/x^2$. This is of the form $[\infty/\infty]$.

Solution:

$$\lim_{x\to\infty}\frac{e^x}{x^2} = \lim_{x\to\infty}\frac{e^x}{2x} \quad \left[\text{Still}\,\frac{\infty}{\infty}\right]$$
$$= \lim_{x\to\infty}\frac{e^x}{2} = \infty$$
$$\therefore \lim_{x\to\infty}\frac{e^x}{x^2} = \infty \qquad \textbf{Ans.}$$

(9) One may check that $e^{10} \approx 22026 = 2\,(11013) \approx 2.10^4$. For convenience, we may write, $e^{10} \approx 10^4$. Similarly, $e^{100} \approx 10^{43}$ and $e^{1000} \approx 10^{243}$. These approximate values give us an idea about how fast the function e^x grows.

Example (18):

$$\lim_{x\to\infty}\frac{x-2x^2}{3x^2+5x} \quad \left[\frac{-\infty}{\infty}\right]$$
$$= \lim_{x\to\infty}\frac{1-4x}{6x+5} \quad \left[\text{still}\,\frac{-\infty}{\infty}\right]$$
$$= \lim_{x\to\infty}\frac{-4}{6}=\frac{2}{3} \qquad \textbf{Ans.}$$

Remark (6): Once again, note that the formulas

$$\lim_{x\to x_0}\frac{f(x)}{g(x)} = \lim_{x\to x_0}\frac{f'(x)}{g'(x)} \tag{13}$$

and

$$\lim_{x\to\infty}\frac{f(x)}{g(x)} = \lim_{x\to\infty}\frac{f'(x)}{g'(x)} \tag{14}$$

hold only if *the limit on the right-hand side (which may be finite or infinite) exists.* It may happen that the limit on the left exists, while there is no existing limit on the right. *If this happens, we say that L'Hospital's rule is not applicable to such a ratio.*

Example (19):

$$\lim_{x\to\infty}\frac{x+\sin x}{x}, \quad \left[\frac{\infty}{\infty}\right]$$

Indeed,

$$\lim_{x\to\infty}\left(1+\frac{\sin x}{x}\right) = 1$$

In view of Remark (6), we should not apply L'Hospital's rule in this case.

On the other hand, if we apply the rule to this ratio, the ratio of derivatives, on simplification gives,

$$\frac{(x+\sin x)'}{(x)'} = \frac{1+\cos x}{1} = 1+\cos x$$

and therefore, we get,

$$\lim_{x\to\infty}\frac{x+\sin x}{x} = \lim_{x\to\infty}(1+\cos x)$$

which does not approach any limit. It constantly oscillates between 0 and 2.

Note that, this example does not contradict L'Hospital's rule. Simply, L'Hospital's rule is not applicable to this case as mentioned above. Another example of this type follows.

Example (20): Let $f(x) = x^2 \sin 1/x$ and $g(x) = x$. Then,

$$\lim_{x \to 0} \frac{f(x)}{g(x)} = \lim_{x \to 0} \frac{x^2 \sin(1/x)}{x}, \quad \left[\frac{0}{0}\right]$$
$$= \lim_{x \to 0} x \sin\left(\frac{1}{x}\right) = 0$$

On the other hand, *the quotient of derivatives*

$$\frac{f'(x)}{g'(x)} = x^2 \cos\left(\frac{1}{x}\right)\left(-\frac{1}{x^2}\right) + 2x \sin\left(\frac{1}{x}\right)$$
$$\text{or} \quad \frac{f'(x)}{g'(x)} = 2x \sin\left(\frac{1}{x}\right) - \cos\frac{1}{x}$$

has no limit as $x \to 0$.

Note (13): The two examples (19) and (20), given above, tell us that *from the existence of* $\lim_{x \to x_0} f(x)/g(x)$ *it does not necessarily follow that* $\lim_{x \to x_0} f'(x)/g'(x)$ exists. Of course, there are examples in which $\lim_{x \to x_0} f(x)/g(x)$ and $\lim_{x \to x_0} f'(x)/g'(x)$, both give the same answer, as in the next two examples.

Example (21):

$$\lim_{x \to 0} \frac{x - 2x^2}{3x^2 + 5x} = \lim_{x \to 0} \frac{1 - 4x}{6x + 5} = \frac{1}{5} \qquad \textbf{Ans.}$$

Example (22):

$$\lim_{x \to \infty} \frac{x - 2x^2}{3x^2 + 5x} = \lim_{x \to \infty} \frac{1 - 4x}{6x + 5} \quad \left[\frac{\infty}{\infty}\right]$$
$$= \lim_{x \to \infty} \frac{-4}{6} = -\frac{2}{3} \qquad \textbf{Ans.}$$

Note (14): If the conditions of L'Hospital's theorem are satisfied on the interval $(a - \delta, a)$ (or on $(a, a + \delta)$), then L'Hospital's rule is applicable to computation of the limit of $f(x)/g(x)$ as $x \to a^-$ (or as $x \to a^+$). In the following two examples, *we consider such one-sided limits.*

Example (23):

$$\lim_{x \to 0^+} \frac{\ln \sin ax}{\ln \sin bx} \quad \left[\frac{-\infty}{-\infty}\right]$$
$$= \lim_{x \to 0^+} \frac{(\ln \sin ax)'}{(\ln \sin bx)'}$$
$$= \lim_{x \to 0^+} \frac{\frac{a \cos ax}{\sin ax}}{\frac{b \cos bx}{\sin bx}} = \frac{a}{b} \lim_{x \to 0^+} \left[\frac{\cos ax}{\cos bx} \frac{\sin bx}{\sin ax}\right]$$
$$= \lim_{x \to 0^+} \left[\frac{\cos ax}{\cos bx} \frac{\sin bx}{b} \frac{a}{\sin ax}\right]$$
$$= 1.1.1 \; [a > 0, b > 0]$$
$$= 1 \qquad \textbf{Ans.}$$

Example (24):

$$\lim_{x \to 0^+} \frac{\ln x}{1/x} \quad \left[\frac{-\infty}{\infty}\right]$$

$$= \lim_{x \to 0^+} \frac{(\ln x)'}{(1/x)'} = \lim_{x \to 0^+} \frac{1/x}{-1/x^2} = \lim_{x \to 0^+} (-x) = 0 \quad \textbf{Ans.}$$

Remark (7): Even if L'Hospital's rule applies, it may not help us, as examples (25) and (26) suggest (of course, with a proper understanding and approach such problems can be easily solved).

Example (25): Find $\lim\limits_{x \to \infty} e^{-x}/x^{-1}$.

Solution:

$$\lim_{x \to \infty} \frac{e^{-x}}{x^{-1}} \quad \left[\frac{0}{0}\right]$$

$$= \lim_{x \to \infty} \frac{-e^{-x}}{-x^{-2}} = \lim_{x \to \infty} \frac{e^{-x}}{x^{-2}} = \dots$$

Clearly, we are only complicating the problem.

A better approach is to do a bit of algebra first, as follows:

$$\lim_{x \to \infty} \frac{e^{-x}}{x^{-1}} = \lim_{x \to \infty} \frac{x}{e^x}, \quad \left[\frac{\infty}{\infty}\right]$$

$$= \lim_{x \to \infty} \frac{(x)'}{(e^x)'} = \lim_{x \to \infty} \frac{1}{e^x} = 0 \qquad \textbf{Ans.}$$

Which form is more convenient $0/0$ *or* ∞/∞?

We know that L'Hospital's rule applies to indeterminate forms of the type $0/0$ and ∞/∞. Also, we can easily convert form $0/0$ to ∞/∞ and vice versa. We may choose any of these *forms, depending on which is easier to handle as far as the differentiation is concerned.*

Example (26): Find $\lim\limits_{x \to 0^+} \ln x/\cot x$.

Solution: As $x \to 0^+$, $\ln x \to -\infty$ and $\cot x \to \infty$. So L'Hospital's rule applies.

$$\lim_{x \to 0^+} \frac{\ln x}{\cot x} = \lim_{x \to 0^+} \frac{(\ln x)'}{(\cot x)'} = \lim_{x \to 0^+} \left[\frac{1/x}{-\operatorname{cosec}^2 x}\right], \quad \left[\frac{\infty}{-\infty}\right]$$

This is still indeterminate as it stands, but it may be observed that *if we apply L'Hospitals rule again, it will only make things worse.* On the other hand, *if we rewrite the bracketed expression as follows, the situation is simplified.*

$$\frac{1/x}{-\operatorname{cosec}^2 x} = -\frac{\sin^2 x}{x} = -\sin x \frac{\sin x}{x}$$

Thus,

$$\lim_{x \to 0^+} \frac{\ln x}{\cot x} = \lim_{x \to 0^+} \left[-\sin x \frac{\sin x}{x} \right] = (0) \times 1 = 0 \qquad \textbf{Ans.}$$

L'Hospital's rule consists of several versions on the theme of using derivatives to evaluate limits of quotients. However, it is useful to have an overall view of the L'Hospital's rule, stated in simplified language in the next section.

21.7 MOST GENERAL STATEMENT OF L'HOSPITAL'S THEOREM

Theorem: Let $f(x)$ and $g(x)$ be two functions tending simultaneously to zero or infinity as $x \to u$ (or as $x \to \infty$). *If the ratio of their derivatives has a limit (finite or infinite)*, the ratio of the functions possesses a limit that is equal to the limit of the ratio of the derivatives:

$$\lim_{x \to u} \frac{f(x)}{g(x)} = \lim_{x \to u} \frac{f'(x)}{g'(x)}$$

Here, u may stand for a, a^-, a^+, $-\infty$, or $+\infty$.

Note (15): (Historical Note) L'Hospital's rule should actually be called "Bernoulli's rule" because it appears in a correspondence from Johann Bernoulli to L'Hospital. L'Hospital and Bernoulli had made an agreement under which L'Hospital paid Bernoulli a monthly fee for solutions to certain problems, with the understanding that Bernoulli would tell no one of the arrangement. As a result, the rule described in the above theorem first appeared in L'Hospitals 1696 treatise. It was only recently discovered that the rule, its proof, and relevant examples all appeared in a 1694 letter from Bernoulli to L'Hospital.

21.8 MEANING OF INDETERMINATE FORMS

Certain limit problems have been classified *as indeterminate forms*. In fact, *the term indeterminate form is used to say that the result is not obvious. We classify them as follows:*

(i) Indeterminate Limit Problems of the Form $0/0$ and $\pm\infty/\infty$ (Quotient Forms): Consider the Limits,

$$\lim_{x \to 1} \frac{(x-1)^2}{x-1} = 0, \qquad \lim_{x \to 1} \frac{2(x-1)^2}{x-1} = 2, \quad \text{and} \quad \lim_{x \to 1} \frac{(x-1)^2}{(x-1)^4} = \infty$$

These examples show that *one could define* $0/0$ *to be 0, 2, or* ∞ *with equal justification*. It is for this reason that one does not attempt to define $0/0$. This expression is *an example of an indeterminate form* (see Chapter 1).

Next, consider the limit problems in the form ∞/∞:

$$\lim_{x \to 2^+} \frac{1/(x-2)}{2/(x-2)} = \frac{1}{2} \quad \text{and} \quad \lim_{x \to 2^+} \frac{1/(x-2)}{3/(x-2)^2} = 0$$

where *both numerator and denominator* in each limit approach ∞ as x approaches 2. These examples suggest that we should consider ∞/∞ to be an indeterminate form.

Note that $0/\infty$ is *not an indeterminate form*, for if $\lim_{x\to a} f(x) = 0$; $\lim_{x\to a} g(x) = \infty$, then $\lim_{x\to a} f(x)/g(x) = 0$.

Also, $2/0$ *is not an indeterminate form, for if* $\lim_{x\to a} f(x) = 2$ and $\lim_{x\to a} g(x) = 0$, then $\lim_{x\to a} f(x)/g(x)$ *is always undefined, and the quotient* $f(x)/g(x)$ *becomes large in absolute value as x approaches a. Thus,* $0/\infty$ *and* $\infty/0$*, are both not indeterminate forms.*

(ii) Indeterminate Limit Problems in Product Forms [$0 \cdot \infty$ or $\infty \cdot 0$]: *Consider the limits* $\lim_{x\to 1^+} (x-1)[3/g(x)] = 3$ *and* $\lim_{x\to 1^+} [2/(x-1)](x-1)^2 = 0$.
These examples show that we should consider $\underline{0} \cdot \infty$ *(or* $\infty \cdot 0$*) to be an indeterminate form.*
Note that the product ($\infty \cdot \infty$) *is not an indeterminate product form (why ?).*

(iii) Indeterminate Sum and Difference [$(-\infty) + \infty$ (or $\infty - \infty$)]: Consider the limits

$$\lim_{x\to a^+} \left(\frac{1}{(x-a)} - \frac{1}{(x-a)}\right) = \lim_{x\to a^+} \left[\frac{1-1}{(x-a)}\right] = 0$$

and

$$\lim_{x\to a^+} \left(\frac{1}{x-a} - \frac{1+2a-2x}{x-a}\right) = \lim_{x\to a^+} \left(\frac{1}{x-a} - \frac{1}{x-a} + \frac{2(a-x)}{x-a}\right) = \lim_{x\to a^+} \left[\frac{2(x-a)}{(x-a)}\right] = 2$$

These examples show that $-(\infty)+\infty$ and $\infty - \infty$ should be considered to be of indeterminate form.

Note that, *the sum* $\infty + \infty$ *is not an indeterminate form (why?).*

(iv) The Indeterminate Exponential Forms [$0 \cdot 0$, $1 \cdot \infty$, $1 - \infty$, $\infty \cdot 0$]: Indeterminate exponential forms arise from expressions of the type $\lim_{x\to a} f(x)^{g(x)}$.

Recall that we have defined the exponential r^s *for all s only* $r > 0$.

Hence, we assume that $f(x) > 0$ for $x \neq a$.

Since *the logarithm function is continuous and is the inverse of the exponential function,* we see that,

$$\lim_{x\to a} \ln\left[f(x)^{g(x)}\right] = b \Rightarrow \ln\left[\lim_{x\to a} f(x)^{g(x)}\right]$$

so that we can write $\lim_{x\to a} f(x)^{g(x)} = e^b$.
(Recall that $\ln x$ ($= \log_e x$) $= b$ means $e^b = x$.)

Note that, the exponential forms 0^∞ and ∞^∞ *are not indeterminate exponential forms* (why?).

Remark (8): We have seen that the following seven symbols, (1) $0/0$, (2) ∞/∞, (3) $0 \cdot \infty$, (4) $\infty - \infty$, (5) 0^0, (6) ∞^0, and (7) 1^∞, *represent indeterminate forms. Though there are many other possibilities symbolized by, for example,* $0/\infty$, $\infty/0$, $\infty + \infty$, $\infty \cdot \infty$, 0^∞, *and* ∞^∞, *they are not indeterminate forms because in all these cases the result is easily guessed*

without any confusion. (This is so because in all these cases the forces are in collusion, not in competition.)
Consider the following example pertaining to the exponential form.

Example (27): Find $\lim\limits_{x \to 0^+} \sin x^{\cot x}$.

Solution: We might call this 0^∞ form, *but it is not an indeterminate form.* Note that, sin x is approaching zero, *and raising it to the exponent cot x, an increasing large number, serves to make it approach zero faster*. Thus,

$$\lim_{x \to 0^+} \sin x^{\cot x} = 0$$

21.9 FINDING LIMITS INVOLVING VARIOUS INDETERMINATE FORMS (BY EXPRESSING THEM IN THE FORM 0/0 OR ∞/∞)

A limit corresponding to an indeterminate form is usually computed by trying to convert the problem to a limit corresponding to *the indeterminate quotient form* $0/0$ *or* ∞/∞. Once this is done, we can *usually* determine the correct limit. Of course, there may be unusual situations as in Examples (25) and (26).

Note (16): There is a helpful way to remember how to convert a $0 \cdot \infty$-type problem to a $0/0$-type problem. We write,

$$0 \cdot \infty = \frac{0}{1/\infty} = \frac{0}{0}$$

(Note that, *the above statement is mathematically wrong, but it is quite helpful* if we agree to remember it in this way. Similarly, the statement $\infty/\infty = (1/\infty)/(1/\infty) = 0/0$ enables us to convert a ∞/∞-type problem to a $0/0$-type problem. Now, we proceed *to solve some problems, wherein conversion to the form* $0/0$ *or* ∞/∞ *is involved.*

21.9.1 Indeterminate Product Forms

Example (28): The limit $\lim_{x \to \infty} x \sin(1/x)$ leads to the form $\infty \cdot 0$, but we can change it to the form 0/0 by writing $x = 1/t$ and letting $t \to 0$. Thus, we have,

$$\lim_{x \to \infty} x \sin\frac{1}{x} = \lim_{t \to 0} \frac{1}{t} \sin t \quad [\infty \cdot 0]$$

$$= \lim_{t \to 0} \frac{\sin t}{t} \quad \left[\frac{0}{0}\right]$$

∴ By applying L Hospital s rule, we get,

$$L = \lim_{t \to 0} \frac{\cos t}{1} = 1 \qquad \textbf{Ans.}$$

Example (29) Find $\lim\limits_{x \to 0^+} x \log_e x$.

Solution: Observe that $\lim\limits_{x \to 0^+} x = 0$ and $\lim\limits_{x \to 0^+} \log_e x = -\infty$.
Therefore, the given limit is of the form $0 \cdot (-\infty)$.

However, we can transform it into the indeterminate form ∞/∞ by rewriting,

$$\lim_{x \to 0^+} x \log_e x = \lim_{x \to 0^+} \frac{\log_e x}{1/x}$$

This is of the form $[-\infty/\infty]$.
$\therefore$ By applying L'Hospital's rule, we get,

$$\lim_{x \to 0^+} \frac{\log_e x}{1/x} = \lim_{x \to 0^+} \frac{1/x}{-1/x^2} = \lim_{x \to 0^+} -\left(\frac{x^2}{x}\right)$$

$$= \lim_{x \to 0^+} (-x) = 0$$

$$\therefore \lim_{x \to 0^+} x \log_e x = 0 \qquad \textbf{Ans.}$$

Example (30): Find $\lim_{x \to \pi/2} (\tan x \ln \sin x)$.

Solution: We have,

$$\lim_{x \to \pi/2} \ln \sin x = 0 \quad \left[\because \lim_{x \to \pi/2} \sin x = 1\right]$$

and

$$\lim_{x \to \pi/2} \tan x = \infty$$

Therefore, the given limit (L, say) is of the form $0 \cdot \infty$. We can rewrite it in the form $0/0$ by simply changing tan x to 1/cot x. Thus,

$$\lim_{x \to \pi/2} (\tan x \ln \sin x) = \lim_{x \to \pi/2} \frac{\ln \sin x}{\cot x} \quad \left[\text{This is of the form} \frac{0}{0}\right]$$

By applying L'Hospital's rule, we get,

$$L = \lim_{x \to \pi/2} \frac{(1/\sin x)\cos x}{-\operatorname{cosec}^2 x} = \lim_{x \to \pi/2} (-\cos x \sin x) = 0$$

21.9.2 Indeterminate Sum and Difference Form

Now, we consider the type $(\infty - \infty)$.

Example (31): Find $\lim_{x \to 0} [(1/\sin x) - (1/x)]$.

Solution: If $x \to 0^+$, then sin $x \to 0^+$ and $1/\sin x \to +\infty$, while $1/x \to +\infty$. The expression $[(1/\sin x) - (1/x)]$ formally becomes $+\infty - (+\infty)$, *which is indeterminate.* On the other hand, if $x \to 0^-$, then $1/\sin x \to -\infty$ and $1/x \to -\infty$, so that $(1/\sin x) - (1/x)$ becomes $-\infty + \infty$, *which is also indeterminate.*

We may also write,

$$\left(\frac{1}{\sin x}\right) - \left(\frac{1}{x}\right) = \frac{x - \sin x}{x \sin x}$$

which is of the form $0/0$.

Thus,

$$\begin{aligned}\lim_{x\to 0}\left(\frac{1}{\sin x}-\frac{1}{x}\right) &= \lim_{x\to 0}\frac{x-\sin x}{x\sin x} \quad \left[\frac{0}{0}\right]\\ &= \lim_{x\to 0}\frac{1-\cos x}{x\cos x+\sin x} \quad \left[\text{still}\,\frac{0}{0}\right]\\ &= \lim_{x\to 0}\frac{\sin x}{-x\sin x+2\cos x} = 0 \qquad \textbf{Ans}.\end{aligned}$$

21.9.3 Indeterminate Exponential Forms

Now, we consider the indeterminate forms 0^0, ∞^0, and 1^∞.

The trick for these forms is *to not consider the original expression, but rather, its logarithm*. We first take the logarithm of the given expression and then determine the limit of that logarithm. Finally, *from this limit, we find the limit of the original function, which is allowable because of the continuity of the logarithmic function.*

Instead of the detailed theoretical analysis of the techniques used for evaluating such limits, we show through examples how this reduction is performed practically.

Example (32): Find $\lim\limits_{x\to 0^+} x^x$.[10]

Solution: The given limit has the indeterminate form 0^0.
Let $y = x^x$.

(In order to find the $\lim\limits_{x\to 0^+} x^x$, we first take the logarithm of the given expression, as suggested above.)

Taking logarithms on both sides, we get

$$\begin{aligned}\lim_{x\to 0^+}\ln y &= \lim_{x\to 0^+}\ln x^x\\ &= \lim_{x\to 0^+}[x\ln x]\end{aligned}$$

This has the form $[0(-\infty)]$.

$$= \lim_{x\to 0^+}\left[\frac{\ln x}{1/x}\right] \quad \left[\frac{-\infty}{\infty}\right]$$

By applying L'Hospital's rule, we get

$$\begin{aligned}\lim_{x\to 0^+}\ln y &= \lim_{x\to 0^+}\left[\frac{1/x}{-1/x^2}\right]\\ &= \lim_{x\to 0^+}(-x) = 0\end{aligned}$$

But we have to find $\lim y$ *not* $\lim(\ln y)$. Also, we know that $\lim(\ln y) = \ln(\lim y)$. Therefore, we write, $\ln(\lim y) = 0$.

In other words, it means that,

$$\lim y = e^0 = 1$$

[10] It is reasonable that we consider only a one-sided limit, when $x \to 0$ through positive values of x.

or

$$\lim_{x \to 0^+} x^x = 1 \qquad \textbf{Ans.}$$

Example (33): Show that $\lim_{x \to \infty} ((1) + (1/x))^x = \mathrm{e}$.

Solution: As in Example 5, we first find *the limit of the logarithm of the expression on the left.* In other words, we are finding a number b such that

$$\lim_{x \to \infty} \ln\left(1 + \frac{1}{x}\right)^x = b$$

(Our answer for the original limit will be e^b.) We find that,

$$\lim_{x \to \infty} \ln\left(1 + \frac{1}{x}\right)^x = \lim_{x \to \infty} x \ln\left(1 + \frac{1}{x}\right) = \lim_{x \to \infty} \frac{\ln(1 + 1/x)}{1/x}$$

This expression is now prepared for applying L'Hospital's rule because

$$\lim_{x \to \infty} \ln\left(1 + \frac{1}{x}\right) = 0 = \lim_{x \to \infty} \frac{1}{x}$$

As a result,

$$\lim_{x \to \infty} \frac{\ln(1 + 1/x)}{1/x} = \lim_{x \to \infty} \frac{[1/(1 + 1/x)](-1/x^2)}{-1/x} = \lim_{x \to \infty} \frac{1}{1 + 1/x} = \frac{1}{1 + 0} = 1$$

Thus, $b = 1$, so that,

$$\lim_{x \to \infty} \left(1 + \frac{1}{x}\right)^x = \mathrm{e}^b = \mathrm{e}^1 = \mathrm{e}$$

Remark: By applying L'Hospital's rule, we get

$$\lim_{x \to \infty} \left(1 + \frac{1}{x}\right)^x = \mathrm{e}$$

Thus, e *could also be defined by means of the above limit*, as is done in some textbooks.

Exercise

Q1. Find $\lim_{x \to 1} ((x/x - 1) - (1/\log_e x))$. **Ans.** $\frac{1}{2}$

Q2. $\lim_{x \to 0} x^n \log_e x$ **Ans.** 0

Q3. $\lim_{x \to 0} \cos x^{1/x^2}$ **Ans.** $1/\sqrt{\mathrm{e}}$

Q4. $\lim_{x \to 0} (\cot x)^{1/\log_e x}$ **Ans.** $1/\mathrm{e}$

Q5. Find $\lim_{x \to (\pi/2)^-} (\tan x)^{\cos x}$. **Ans.** 1

Q6. Find $\lim_{x \to 0^+} (x+1)^{\cot x}$. **Ans**. e

We give below the solution to the above exercise for the convenience of the readers.

Q1. Find $\lim_{x \to 1} ((x/x-1) - (1/\log_e x)), \quad [\infty - \infty]$.

Solution:

$$\lim_{x \to 1} \left(\frac{x}{x-1} - \frac{1}{\log_e x} \right)$$

$$= \lim_{x \to 1} \frac{x \log_e x - x + 1}{(x-1)\log_e x}, \quad \left[\frac{0}{0}\right]$$

$$(x \log_e x - x + 1)' = x\frac{1}{x} + \log_e x - 1 = \log_e x$$

$$[(x-1)\log_e x]' = \frac{x-1}{x} + \log_e x = 1 - \frac{1}{x} + \log_e x$$

$$= \lim_{x \to 1} \frac{\log_e x}{(x-1)/x + \log_e x} = \lim_{x \to 1} \frac{1/x}{1/x^2 + 1/x} = \frac{1}{2}$$ **Ans**.

Q2.

$$\lim_{x \to 0} x^n \log_e x \quad [0 \cdot (-\infty)]$$

Solution:

$$= \lim_{x \to 0} \frac{\log_e x}{1/x^n} \quad \left[\frac{-\infty}{\infty}\right]$$

$$= \lim_{x \to 0} \frac{1/x}{-n/x^{n+1}}$$

$$= \lim_{x \to 0} \frac{-x^n}{n} = 0$$

$$\therefore \lim_{x \to 0} x^n \log_e x = 0$$ **Ans**.

Q3. To find $\lim_{x \to 0} (\cos x)^{1/x^2} \quad [1^\infty]$.

Solution: Let $\lim_{x \to 0} (\cos x)^{1/x^2} = A$.

Taking the logarithm, we get

$$\log_e \left[\lim_{x \to 0} (\cos x)^{1/x^2} \right] = \log_e A$$

$$\text{or} \quad \log_e A = \lim_{x \to 0} \left[\log_e (\cos x)^{1/x^2} \right]^{(11)}$$

$$= \lim_{x \to 0} \frac{\log_e(\cos x)}{x^2}, \quad \left[\frac{0}{0}\right]$$

(11) This is permitted because of the continuity of the logarithmic function.

Now, applying L'Hospital's rule, we get,

$$\log_e A = \lim_{x\to 0} \frac{(1/\cos x)(-\sin x)}{2x} = \lim_{x\to 0} \frac{-\tan x}{2x} \quad \left[\text{still } \frac{0}{0}\right]$$

$$= \lim_{x\to 0} \frac{-1/\cos^2 x}{2} = -\frac{1}{2}$$

$$\therefore\ A = e^{-1/2} = 1/\sqrt{e}$$

or $\lim_{x\to 0} (\cos x)^{1/x^2} = 1/\sqrt{e}$ **Ans.**

Q4. To find $\lim_{x\to 0} (\cot x)^{1/\log_e x}$ $[\infty^0]$.

Solution: Let $\lim_{x\to 0} (\cot x)^{1/\log_e x} = A$.

Taking the logarithm, we get,

$$\log_e\left[\lim_{x\to 0} (\cot x)^{1/\log_e x}\right] = \log_e A$$

$$\text{or} \quad \log_e A = \lim_{x\to 0} \left[\log_e(\cot x)^{1/\log_e x}\right]$$

$$= \lim_{x\to 0} \left[\frac{\log_e(\cot x)}{\log_e x}\right], \quad \left[\frac{\infty}{-\infty}\right]$$

By applying L'Hospital's rule, we get,

$$\log_e A = \lim_{x\to 0} \frac{(1/\cot x)(-\operatorname{cosec}^2 x)}{1/x}$$

$$= \lim_{x\to 0} \frac{-\left(\dfrac{1}{\cot x}\right)\left(\dfrac{1}{\sin^2 x}\right)}{1/x}$$

$$= \lim_{x\to 0} \frac{-x}{\cos x \sin x} \quad \left[\frac{0}{0}\right]$$

$$= \lim_{x\to 0} \frac{-1}{\cos x \cos x + \sin x(-\sin x)} = \lim_{x\to 0} \frac{-(\sin x/\cos x \cdot 1/\sin^2 x)}{1/x}$$

$$\therefore\ \log_e A = \lim_{x\to 0} \frac{-1}{\cos^2 x - \sin^2 x} = -1$$

$$\therefore\ A = e^{-1} = 1/e$$

$\therefore \lim_{x\to 0} (\cot x)^{1/\log_e x} = 1/e$ **Ans.**

Q5. Find $\lim_{x\to (\pi/2)^-} (\tan x)^{\cos x.}$

Solution: This has the indeterminate form ∞^0.

$$\begin{aligned}\text{Put } y &= (\tan x)^{\cos x}.\\ \therefore\ \ln y &= \cos x \ln \tan x\\ &= \frac{\ln \tan x}{\sec x}\end{aligned}$$

Then,

$$\begin{aligned}\lim_{x\to(\pi/2)^-} y &= \lim_{x\to(\pi/2)^-} \frac{\ln \tan x}{\sec x}\\ &= \lim_{x\to(\pi/2)^-} \frac{(1/\tan x)\sec^2 x}{\sec x \tan x}\\ &= \lim_{x\to(\pi/2)^-} \frac{\sec x}{\tan^2 x}\\ &= \lim_{x\to(\pi/2)^-} \frac{\cos x}{\sin^2 x} = 0\end{aligned}$$

Therefore,

$$\lim_{x\to(\pi/2)^-} y = e^0 = 1 \qquad \textbf{Ans.}$$

Q6. Find $\lim\limits_{x\to 0^+} (x+1)^{\cot x}$.

Solution: This takes the indeterminate form.

$$\begin{aligned}\text{Let } y &= (x+1)^{\cot x}\\ \therefore\ \ln y &= \cot x \ln(x+1)\end{aligned}$$

By applying L'Hospital's rule for the $0/0$ form,

$$\begin{aligned}\lim_{x\to 0^+} \ln y &= \lim_{x\to 0^+} \frac{\ln(1+x)}{\tan x}\\ &= \lim_{x\to 0^+} \frac{[1/(1+x)]}{\sec^2 x} = 1\end{aligned}$$

Thus, $\lim_{x\to 0^+} \ln y = 1$, [where $y = (x+1)^{\cot x}$]

Our interest lies in computing $\lim\limits_{x\to 0^+} \ln y$ [not $\lim\limits_{x\to 0} \ln y$].

Now,

$$\lim_{x\to 0^+} \ln y = \ln[\lim_{x\to 0^+} \ln y]$$

$$\therefore \text{ We write } \ln[\lim_{x\to 0^+} \ln y] = 1$$

$$\therefore \lim_{x\to 0^+} y = \mathrm{e}^1 = \mathrm{e} \qquad \textbf{Ans.}$$

22 Extending the Mean Value Theorem to Taylor's Formula: Taylor Polynomials for Certain Functions

22.1 INTRODUCTION

In Chapter 20, we have introduced the Mean Value Theorem, which says that if a function f is continuous on an interval $[a, b]$ and differentiable on (a, b), then,

$$\frac{f(b) - f(a)}{b - a} = f'(c) \tag{1A}$$

or

$$f(b) = f(a) + f'(c) \cdot (b - a) \tag{1B}$$

for some c between a and b.

Here, $f(b)$ is expressed in terms of $f(a)$ and $f'(c)$, $(b - a)$ being the length of the interval (a, b). Since x can vary from a to any value $b \notin [a, b]$, we may think of b as an independent variable. This permits us to replace b by x and rewrite (1B) in the following form.

$$f(x) = f(a) + f'(c)(x - a) \tag{1C}$$

In this new formula (1C), we think of x as an independent variable on $[a, b]$ and the number c lies in the interval between a and x. (The equation of the Mean Value Theorem is often stated in this form.)

Note (1): The right-hand side of (1C) *looks like the linear approximation of f near a*. If f' is *continuous* and c is close to a (as it will have to be if x is close to a), then $f'(c)$ is *close to $f'(a)$* and (1C) gives,

$$f(x) \approx f(a) + f'(a)(x - a) \tag{1D}$$

which is the linear approximation of f near a.

22-Extending the MVT to Taylor's formula (Taylor polynomials approximating certain functions Taylor's formula for polynomials and arbitrary functions)

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

We have studied linear approximations in Chapter 16 without knowing how good they were. Now, with an extended version of MVT for second derivative(s), we shall see that the error in (1D) is proportional to $(x-a)^2$. Therefore, if $(x-a)$ is small, the error will be very small.

Note (2): At this stage, the reader may read the proof of MVT and note carefully *how the auxiliary function F(x) is defined there. In the process of extending the MVT for second derivative(s)* (which generalizes the MVT for the first derivative(s) and sets the stage for further generalization), *it is important to study carefully the steps involved in the proof. In particular, the way of defining the auxiliary function F(x)* (which satisfies the hypotheses of Rolle's Theorem) is important. Now, we proceed to state and prove the MVT for second derivatives.

22.2 THE MEAN VALUE THEOREM FOR SECOND DERIVATIVES: THE FIRST EXTENDED MVT

Let f be a (real) function defined on $[a, b]$, such that,

(i) f and f' are continuous on $[a, b]$ (from the statement f' is continuous on $[a, b]$, it follows that f' exists on $[a, b]$).
(ii) f' is differentiable on (a, b).

Then, there exists a number c_2 between a and b, such that,

$$f(b) = f(a) + f'(a)(b-a) + \frac{f''(c_2)}{2}(b-a)^2 \tag{2A}$$

Proof: Let a number K be defined by

$$f(b) = f(a) + f'(a)(b-a) + K(b-a)^2 \tag{2B}$$

or

$$f(b) - f(a) - f'(a)(b-a) - K(b-a)^2 = 0 \tag{2C}$$

Note (3): The significance of equation (2A) is not the fact that some number K satisfies the equation (2B), *but the fact that the value of K defined by (2B) is actually given by*

$$K = \frac{f''(c_2)}{2} \tag{3}$$

for some point c_2 in the interval between a and b.

Therefore, given that K is the number that satisfies (2B), *we will show that K must satisfy (3) for some number c_2 between a and b.*

Note (4): Now, our interest lies in obtaining a function $f(x)$, which must satisfy equation (2B). Since, the independent variable x varies in $[a, b]$, we can say that x varies from a to any point b in

$[a, b]$. This is equivalent to looking at b as an independent variable in $[a, b]$. Accordingly, equation (2B) says that when $x = b$ we can write (2B) in the following form:

$$f(x) = f(a) + f'(a)(x - a) + K(x - a)^2 \tag{4}$$

Note that, when $x = b$, the function $f(x)$ and the function $f(a) + f'(a)(x - a) + K(x - a)^2$ have the same value [see equation (2B)].[1]

Also, these two functions have the same value when $x = a$ [namely, $f(a)$], as can be easily checked.

We now define a new function $F(x)$, being the difference of the above two functions. Then, generally $F(x)$ must be different from zero, for all values of x, other than $x = a$ and $x = b$. This also means that $F(x)$ is *not, constant function*. Thus,

$$\begin{aligned} F(x) &= f(x) - [f(a) + f'(a)(x - a) + K(x - a)^2] \\ &= f(x) - f(a) - f'(a)(x - a) - K(x - a)^2 \end{aligned} \tag{5}$$

is different from zero.
Now, observe that,

(a) $F(x)$ *is continuous* on $[a, b]$, (because f, $(x - a)$ and $(x - a)^2$ are continuous on $[a, b]$),
(b) $F(x)$ is *differentiable* on (a, b) for the same reason, and
(c) $\left.\begin{aligned} F(a) &= 0 \quad \text{[by (4)]} \\ F(b) &= 0 \quad \text{[by (2C)]} \end{aligned}\right\}$ [2]

Thus, the function $F(x)$ *satisfies all the conditions of Rolle's Theorem on the interval* $[a, b]$. Therefore, $F'(x) = 0$ at some point c_1 between a and b. Thus, we have,

$$F'(c_1) = 0, \; a < c_1 < b \tag{6}$$

Now, from equation (5), we obtain the derivative,

$$F'(x) = f'(x) - f'(a) - 2K(x - a) \tag{7}$$

and for $x = c_1$, we get,

$$F'(c_1) = f'(c_1) - f'(a) - 2K(c_1 - a) = 0 \quad \text{[using (6)]}$$

Next, we observe that *the function* $F'(x)$ *satisfies all the hypotheses of Rolle's Theorem on the interval* $[a, c_1]$ as follows:

- $F'(c_1) = 0$ [from (6)],
- $F'(a) = 0$ [from (7)],
- $F'(x)$ is continuous on $[a, c_1]$ and differentiable on (a, c_1), because both $f'(x)$ and $(x - a)$ are.

[1] In fact, this is so because we have defined $f(x)$ based on requirement of equation (2B).
[2] Note that, $F(a) = 0$ is a result of cancellation of differences in (5), while $F(b) = 0$ follows from (2C) due to the way K is defined.

Therefore, the derivative $F''(x)$ *must be zero at some point* c_2, between a and c_1 (*and hence between a and b*).

Now, by differentiating (7) we get,

$$F''(x) = f''(x) - 2K, \text{ and for } x = c_2, \text{ we get}$$

$$0 = f''(c_2) - 2K \quad [\because \; F''(x) = 0 \text{ at some point } c_2 \in (a,\ c_1)]$$

$$\therefore \; K = \frac{F''(c_2)}{2}$$

Substituting this value of K in (2B), we get the result that we wanted to prove.

22.2.1 Linear Approximations

We are now in a position to calculate the error in the linear approximation defined at equation (1D). We begin by regarding b as an independent variable in equation (2A) (as indicated in the note above) and rewriting this equation in the following form

$$f(x) = f(a) + f'(x-a) + \frac{f''(c_2)}{2} \cdot (x-a)^2 \ldots \tag{8}$$ [3]

with the understanding that c_2 lies in between a and x.

From equation (1D) we get the linear approximation

$$f(x) \approx f(a) + f'(a)(x-a), \tag{9}$$

valid on the interval from a to x with an error of

$$\mathrm{e}_1(x) = \frac{f''(c_2)}{2} \cdot (x-a)^2 \tag{10}$$

If f'' is continuous on the closed interval from a to x, then it has a maximum value on the interval and $\mathrm{e}_1(x)$ satisfies the inequality

$$|\mathrm{e}_1(x)| \le \frac{\max|f''|}{2}(x-a)^2 \tag{11}$$

where *max refers to the interval joining a and x.*

This inequality gives an upper bound for the error on the interval from a to x that *is of practical value in many cases*. To get more accuracy in a linear approximation, we add a quadratic term, as given in equation (8).

[3] This equation holds for $x < a$ as well as for $x > a$. [Since the proof and the theorem remain valid if one refers to "the interval with end points a and b" rather than explicitly to $[a, b]$ or (a, b). *Calculus and Analytic Geometry* (Sixth Edition) by Thomas/Finney, Remark at page 239.]

For more details refer to *Calculus and Analytic Geometry* (Sixth Edition) by Thomas/Finney.

22.2.2 Quadratic Approximations

We can get a good estimate of the error in *quadratic approximations by extending the Mean Value Theorem one step further in the following way.*

22.2.3 The Mean Value Theorem for Third Derivatives: The Second Extended MVT

Theorem: If f, f', and f'' are *continuous* on $[a, b]$ and f'' is also *differentiable* on (a, b), then there exists a number c_3 between a and b for which

$$f(b) = f(a) + f'(a)(b-a) + \frac{f''(a)}{2}(b-a)^2 + \frac{f'''(c_3)}{6}(b-a)^3 \tag{12}$$

[For the proof of this theorem, refer to *Calculus and Analytic Geometry* (Sixth Edition) by Thomas/Finney.]

In applications, we usually write this equation with x in place of b, i.e., we write,

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \frac{f'''(c_3)}{6}(x-a)^3 \tag{13}$$

with the understanding that c_3 lies between a and x.

From (13), we get the *quadratic approximation,*

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 \tag{14}$$

valid on the interval between a and x.[(4)]

Note that, the first two terms on the right-hand side of (14) give the standard linear approximation of f. To get the quadratic approximation, we have only to add the quadratic term without changing the linear part.

If $f'''(x)$ is continuous on the closed interval from a to x, then it has a maximum value on the interval, and we can write,

$$|e_2(x)| \le \max\left|\frac{f'''(x)}{6}\right|(x-a)^3 \tag{15}$$

The extended mean value theorems are special cases of a theorem called *Taylor's Theorem, which holds for any natural number n*. The *most convenient statement of the theorem* for our purpose is the following theorem.

[(4)] In the quadratic approximation (14), the error is given by $e^2(x) = \frac{f'''(c_3)}{6}(x-a)^3 = \frac{f'''(c_3)}{3!}(x-a)^3$, which is the fourth term on the right-hand side of (13).

22.3 TAYLOR'S THEOREM

If f and its first n derivatives $f', f'', \ldots\ldots, f^{(n)}$ are *continuous* on $[a, b]$ and if $f^{(n)}$ is *differentiable* on (a, b), then there exists a number c_{n+1} between a and b such that

$$f(b) = f(a) + f'(a)(b-a) + \frac{f''(a)}{2!}(b-a)^2 + \frac{f'''(a)}{3!}(b-a)^3 + \ldots + \frac{f^{(n)}(a)}{n!}(b-a)^n + \frac{f^{(n+1)}(c_{n+1})}{(n+1)!}(b-a)^{n+1} \quad (16)$$

Equation (16) provides *extremely accurate polynomial approximations for a large class of functions that have derivatives of all orders*. (We do not prove this theorem.)

Note (5): Through Taylor's Theorem, calculus provides a remarkably powerful and *general method of estimating the values of certain differentiable functions with any prescribed degree of accuracy.*

22.4 POLYNOMIAL APPROXIMATIONS AND TAYLOR'S FORMULA

While *values of polynomial functions* can be found by performing a finite number of additions and multiplications, other functions such as the *exponential, logarithmic, and trigonometric functions* cannot be evaluated as easily. We show in this section that many such functions can be *approximated by polynomials*, and that *the polynomial instead of the original function, can be used for computations when the difference between the actual function value and the polynomial approximation is sufficiently small.* Various methods can be employed to approximate a given function by polynomials. One of the most widely used involves Taylor's Formula [equation (16)].

We shall consider the functions e^x, $\log_e x$ (i.e., $\ln x$), $\sin x$, $\cos x$, which occur frequently. Their values are available in mathematical tables. Also, many calculators and computers have been programmed to produce their values on demand. *Where do the values in the tables come from?* By and large these numbers come from calculating *partial sums of power series, which are in fact polynomials*. We, therefore, begin with the definition of a *Power Series*.

22.4.1 Definition: Power Series

A power series is a functional series of the following form:

$$f(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + a_3(x - x_0)^3 + \ldots + a_n(x - x_0)^n + \ldots$$

whose terms are the products of constant factors $a_0, a_1, a_2, \ldots, a_n, \ldots$ by integral powers of the difference $(x - x_0)$.

The constants $a_0, a_1, a_2, \ldots, a_n \ldots\ldots$ are called the *coefficients of the power series*. Unless otherwise stated, the *coefficients will be assumed to be real*. In particular, if $x_0 = 0$, *the power series is arranged in ascending powers of* x:

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_n x^n + \ldots \quad (17)$$

We shall confine ourselves to studying series of the type (17) because any power series can be brought to this type with the aid of the substitution, $x - x_0 = X$.

Note: For the sake of convenience, we shall call $a_n \cdot x^n$ the nth term of the power series (17), *although it occupies the* $(n+1)^{th}$ *place. The constant term* a_0 *of the series will be referred to as its zeroth term.*

The simplest example of the power series is the *geometric series*:

$$1 + x + x^2 + \ldots + x^n + \ldots$$

The whole theory of power series is based on the following fundamental theorem.

Abel's Theorem: If the power series (17) converges at a point $x_0 \neq 0$, *it is absolutely convergent in the interval* $(-|x_0|, |x_0|)$, that is, for every value of x satisfying the condition $|x| < |x_0|$.

22.4.2 Definition: Interval of Convergence

The number R such that power series (17) is convergent for all x satisfying the condition $|x| < R$, and divergent for all x satisfying the inequality $|x| > R$ is called *the radius of convergence* of the power series. The interval $(-R, R)$ is referred to as *the interval of convergence*.

22.4.3 Properties of Power Series (Without Proof)

1. The sum of a power series is a *continuous function in the interval of convergence* of the series; that is, the function

$$s(x) = a_0 + a_1 x + a_2 x^2 + \ldots a_n \cdot x^n + \ldots (-R < x < R) \text{ is continuous.}^{(5)}$$

2. A power series can be *integrated term wise within its interval of convergence*. The new (integrated) series possesses the *same radius of convergence as the original series*.
3. Every power series is infinitely *term wise differentiable inside its interval of convergence*. The *radii of convergence* of the differentiated series remain the same as that of the original series.

Now, we show *how a power series can arise when we wish to approximate a function*

$$y = f(x) \tag{18}$$

by a sequence of the polynomials $f_n(x)$ of the form

$$f_n(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \tag{19}$$

(5) The geometric series: $1 + x + x^2 + \ldots + x^n + \ldots$ converges for $|x| < 1$, *to the rational function*, $f(x) = 1/(1-x)$, which represents the sum of geometric series. The function $1/(1-x)$ is defined and is continuous everywhere except at the point $x = 1$ *but it only serves as the sum of the series* for $|x| < 1$. For $|x| \geq 1$, the series is divergent and *it is senseless to speak of its sum*.

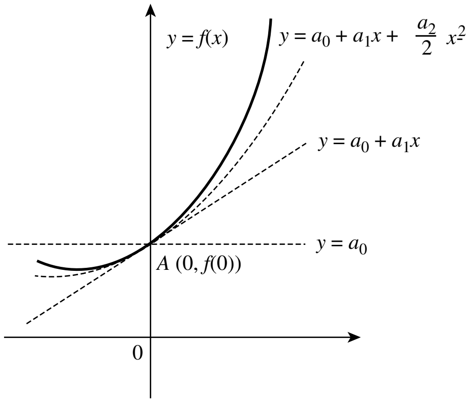


FIGURE 22.1

If a *function f(x)* can be represented as the sum of a power series, we say that it *is expanded into the power series*. The existence of such an expansion is extremely important since it makes *it possible to replace (approximately) the given function by the sum of the first several terms of the power series, which is in fact a polynomial.*

The computation of the values of the given function then reduces to the computation of the values of the polynomial, which can be achieved with the aid of the simplest arithmetical operations.

We immediately face the following questions:

- What does it mean for a function f to approximate another function g?
- How can we choose a "*good*" *polynomial approximation?*
- How *are derivatives involved?*

We shall be interested, at least at first, in making the approximation for the values of x near 0, because we want the term $a_n x^n$ to decrease as n increases. Hence, we focus our attention on a portion of the curve $y = f(x)$ near the point $A(0, f(0))$, as shown in Figure 22.1.

(1) The graph of the polynomial $f_0(x) = a_0$ of degree zero will pass through $(0, f(0))$ if we take

$$a_0 = f(0).$$

(2) The graph of the polynomial $f_1(x) = a_0 + a_1 x$ *will pass through* $(0, f(0))$ *and have the same slope as the given curve at that point*, if we choose,

$$a_0 = f(0) \text{ and } a_1 = f'(0).$$

(3) The graph of the polynomial $f_2(x) = a_0 + a_1 x + a_2 x^2$, *will pass through* $(0, f(0))$ *and have the same first and second derivative as the given curve at that point,*

$$\text{if } a_0 = f(0),\ a_1 = f'(0),\ \text{and } a_2 = \frac{f''(0)}{2}$$

(4) In general, the polynomial,

$$f_n(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$$

which we choose to approximate $y = f(x)$ near $x = 0$, is the one *whose graph passes through* $(0, f(0))$ *and whose first n derivatives match the derivatives of f(x) at x = 0.*

The constant, linear, and quadratic approximations (which are respectively of zeroth-, first-, and second-order approximations) may be looked upon as "made to order" approximations to f. In each case, the "*order*" refers to the number of derivatives of f with which a particular approximation agrees at 0. An even better agreement is possible if polynomials of degree 3, 4, and higher are used to approximate f.

Suppose we take the nth degree polynomial $f_n(x)$ [given at equation (4) above] to approximate $y = f(x)$ near $x = 0$. Then, our task is to find the coefficients a_0 to a_n. This is surprisingly easy. The key idea is that the coefficients a_i are closely related to the derivatives of $f_n(x)$ at $x = 0$. To match the derivative of $f_n(x)$ to those of $y = f(x)$ at $x = 0$, we equate the corresponding derivatives at $x = 0$, and obtain the values of $a_0, a_1, \ldots, a_n$ in terms of the derivatives of $y = f(x)$.

To see how this may be done, we write down the polynomial $f_n(x)$ and its derivatives as follows:

$$f_n(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots\ldots + a_nx^n$$

$$f'_n(x) = a_1 + 2a_2x + 3a_3x^2 + \ldots\ldots + na_nx^{n-1}$$

$$f''_n(x) = 2a_2 + 3 \cdot 2a_3x + \ldots\ldots + n(n-1)a_nx^{n-2}$$

$$f_n^{(n)}(x) = n(n-1)(n-2)\ldots(1) \cdot a_n = (n!)a_n.$$

But the first n derivatives of the (approximating) polynomial $f_n(x)$ is required to match with the corresponding derivatives of the given function $y = f(x)$ at $x = 0$. Therefore, we put $x = 0$ in $f_n(x), f'_n(x), \ldots, f_n^{(n)}(x)$, and equate their values respectively with $f'(0), f''(0), \ldots, f^{(n)}(0)$. We obtain,

$$f_n(0) = a_0 = f(0)$$

$$f'_n(0) = a_1 = f'(0)$$

$$f''_n(0) = 2a_2 = f''(0) \qquad \therefore a_2 = \frac{f''(0)}{2!}$$

$$f'''_n(0) = 3 \cdot 2a_3 = f'''(0) \qquad \therefore a_3 = \frac{f'''(0)}{3!}$$

$$f_n^{(n)}(0) = n!\,a_n = f^{(n)}(0) \qquad \therefore a_n = \frac{f^{(n)}(0)}{n!}$$

Thus, the required polynomial, which approximates $f(x)$ at $x = 0$ is given by

$$f_n(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \ldots + \frac{f^{(n)}(0)}{n!}x^n \tag{20}$$

The polynomial at (20) is called the n^{th} *degree Taylor polynomial* of f at $x = 0$, after the name of the English mathematician Brook Taylor (1685–1731), the author of an early calculus book (published in 1717).

Here onwards, we shall denote the Taylor polynomial by $P_n(x)$ instead of $f_n(x)$. Thus, the Taylor polynomial approximating the given function $y = f(x)$ at $x = 0$, given at (20) above will be written as:

$$P_n(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \ldots + \frac{f^{(n)}(0)}{n!}x^n \tag{21}$$

Remark: In a Taylor polynomial $[P_n(x)]$ of degree n, all Taylor polynomials of degree 0 to n appear and thus there can be at most $(n+1)$ terms.

Note (7): The advantages of approximating a function $f(x)$ with a polynomial $g(x)$ are given below:

- If f is complicated, poorly understood or otherwise inconvenient, then it is useful to replace f with a simpler, better-behaved, and better-understood polynomial g.
- Polynomial functions are simple, convenient, well understood, and easy to use, so it is natural to use them to approximate more complicated functions.

Example (1): Let $f(x) = e^x$. Find a formula for the nth Taylor polynomial of f. Also, find $P_n(1)$ and compute $P_5(1)$.

Solution: The given function and its derivatives are

$$\begin{aligned} f(x) &= e^x & \therefore\ f(0) &= e^0 = 1 \\ f'(x) &= e^x & \therefore\ f'(0) &= e^0 = 1 \\ f''(x) &= e^x & \therefore\ f''(0) &= e^0 = 1 \\ f^{(n)}(x) &= e^x & \therefore\ f^{n}(0) &= e^0 = 1 \end{aligned}$$

$$\therefore\ f(0) = 1, \quad f'(0) = 1, f''(0) = 1, \ldots\ldots, f^{(n)}(0) = 1$$

Using the formula for $P_n(x)$, we get,

$$\begin{aligned} P_n(x) &= f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \ldots\ldots \frac{f^{(n)}(0)}{n!}x^n \\ &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots\ldots + \frac{x^n}{n!} \end{aligned}$$

The graph of the function $y = e^x$ and graphs of three approximating polynomials,

(A) a straight line $1 + x$

(B) a parabola $1 + x + \frac{x^2}{2}$

(C) a cubic curve $1 + x + \frac{x^2}{2} + \frac{x^3}{6}$

are shown in Figure 22.2.

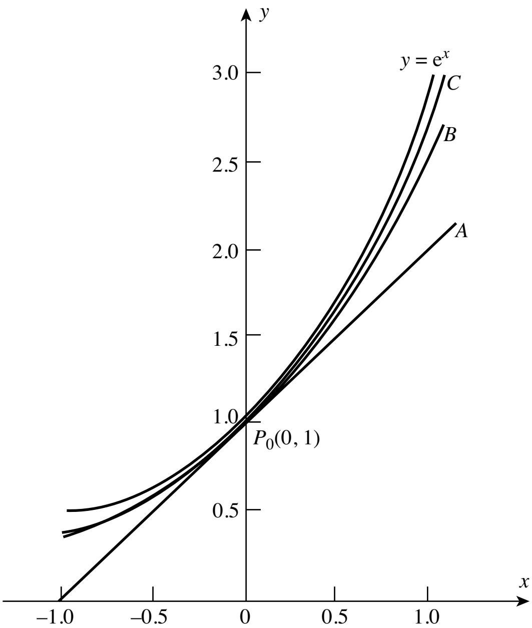


FIGURE 22.2

For $x = 1$,

$$P_n(1) = 1 + 1 + \frac{1^2}{2!} + \frac{1^3}{3!} + \ldots\ldots + \frac{1^n}{n!} = \mathrm{e}$$

[Note that, $f(x) = \mathrm{e}^x \approx P_n(x)\ \therefore\ f(1) = \mathrm{e} \approx P_n(1)$]
For $n = 5$ (i.e., by considering first six terms) we obtain, for $x = 1$

$$P_5(1) = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} = 2.71667.$$

Since $P_n(x)$ represents an approximation to $f(x) = \mathrm{e}^x$, we should examine how well $P_5(1)$ approximates $f(1) = \mathrm{e}$. The value of e is 2.71828 (accurate to six digits) and $P_5(1) \approx 2.71667$. So $P_5(1)$ approximates e with an error of about 0.00161.

Example (2): Let $f(x) = \log_e(1 + x)$. Find a formula for the nth Taylor polynomial of $f(x)$, and then calculate $P_6(1)$.

Solution: First, we calculate the derivatives of the function $f(x)$ and compute their values at $x = 0$:

$$f(x) = \log_e(1 + x) \quad \therefore f(0) = 0$$

$$f'(x) = \frac{1}{1 + x} \qquad \therefore f'(0) = 1$$

$$f''(x) = \frac{-1}{(1+x)^2} \qquad \therefore\ f''(0) = -1$$

$$f^{(3)}(x) = \frac{(-1)(-2)}{(1+x)^3} = \frac{(-1)^2 2!}{(1+x)^3} \qquad \therefore\ f^{(3)}(0) = 2!$$

$$f^{(4)}(x) = \frac{(-1)^3 3!}{(1+x)^4} \qquad \therefore\ f^{(4)}(0) = -3!$$

In general, for $k \geq 1$,

$$f^{(k)}(x) = \frac{(-1)^{k-1}(k-1)!}{(1+x)^k} \qquad \therefore\ f^{(k)}(0) = (-1)^{k-1}(k-1)!$$

Consequently,

$$\begin{aligned} P_n(x) &= f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \ldots\ldots + \frac{f^{(n)}(0)}{n!}x^n \\ &= 0 + x - \frac{1}{2!}x^2 + \frac{2!}{3!}x^3 + \ldots + \frac{(-1)^{n-1}\cdot(n-1)!}{n!} \\ &= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots + \frac{(-1)^{n-1}}{n}x^n \end{aligned}$$

We conclude that

$$P_6(1) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} = \frac{37}{60} \approx 0.616667$$

We expect $P_n(x)$ to approximate $f(x)$.

Since, the value of $f(1) = \log_e 2 = 0.693147$ (accurate to six digits), and $P_6(1) = 0.616667$, we find that $P_6(1)$ approximates $\log_e 2$ with an error of about 0.07648.

Example (3): Let $f(x) = \frac{1}{1-x}$. Find a formula for the nth Taylor polynomial of f, and compute $P_n(2)$.

Solution: The derivatives of f and their values at $x = 0$ are obtained as follows:

$$\text{We have, } f(x) = \frac{1}{12x} = (1-x)^{-1} \qquad \therefore\ f(0) = 1$$

$$f'(x) = (-1)(1-x)^{-2}(-1) = \frac{1}{(1-x)^2} \quad \therefore\ f'(0) = 1$$

$$f''(x) = (-2)(1-x)^{-3}(-1) = \frac{2!}{(1-x)^3} \quad \therefore\ f''(0) = 2!$$

$$f^{(3)}(x) = \frac{3!}{(1-x)^4} \qquad \therefore\ f^{(3)}(0) = 3!$$

In general,

$$f^{(k)}(x) = \frac{k!}{(1-x)^{k+1}} \qquad \therefore\ f^{(k)}(0) = k!$$

As a result,

$$\begin{aligned} P_n(x) &= f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \ldots\ldots + \frac{f^{(n)}(0)}{n!}x^n \\ &= 1 + x + x^2 + \ldots + x^n. \end{aligned}$$

$$\therefore P_n(2) = 1 + 2 + 2^2 + \ldots + 2^n.$$

Example (4): Let $f(x) = \cos x$. Find a formula for the nth Taylor polynomial of f at $x = 0$.

Solution: The cosine and its derivatives are

$$\begin{aligned} &f(x) = \cos x, && f'(x) = -\sin x \\ &f''(x) = -\cos x, && f^{(3)}(x) = \sin x \\ &f^{(2k)}(x) = (-1)^k \cos x, f^{(2k+1)}(x) = (-1)^{k+1} \sin x \end{aligned}$$

But, when $x = 0$, the cosines are 1 and the sines are 0, so that,

$$f^{(2k)}(0) = (-1)^k, \qquad f^{(2k+1)}(0) = 0$$

Thus, $f(0) = 1, \qquad f'(0) = 0$

$$\begin{aligned} &f''(0) = -1,\ f^{(3)}(0) = 0 \\ &f^{(4)}(0) = 1,\ \ f^{(5)}(0) = 0 \\ &f^{(6)}(0) = -1, f^{(7)}(0) = 0 \text{ and so on.} \end{aligned}$$

The Taylor polynomials for $f(x) = $ *cos x have only even powered terms*. Thus, for $n = 2k$, we have

$$\begin{aligned} P_{2k}(x) &= f(0)x^0 + \frac{f^{(2)}(0)}{2!}x^2 + \frac{f^{(4)}(0)}{4!}x^4 + \frac{f^{(6)}(0)}{6!}x^6 + \ldots + \frac{f^{(2n)}(0)}{n!}x^{2n} \\ &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots + (-1)^k \frac{x^{2k}}{(2n)!} \end{aligned}$$

We have, $\cos x \approx 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} \ldots$, *in the vicinity of the point* $x = 0$.

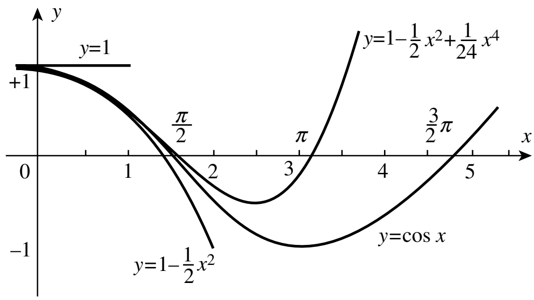


FIGURE 22.3

Figure 22.3 shows the graph of function $y = \cos x$ and those of approximating polynomials $y = 1$, $y = 1 - (x^2/2!)$, and $y = 1 - (x^2/2!) + (x^4/24)$ *in the neighborhood of origin.*

Example (5): Let $f(x) = \sin x$. Find a formula for the nth Taylor polynomial of f at $x = 0$.

Solution: It is easy to show that the Taylor polynomial for $y = \sin x$ has only odd powered terms. In the neighborhood of the origin, we have,

$$\sin x \approx x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!}$$

Figure 22.4 illustrates the graph of the function $y = \sin x$ and the graphs of the approximating Taylor polynomials $y = x$, $y = x - (x^3/6)$, and $y = x - (x^3/6) + (x^5/120)$.

22.4.4 The Maclaurin Series for $f(x)$

The degrees of the Taylor polynomials of a given function f(x) are limited by the degree of differentiability of the function at x = 0. But, if $f(x)$ has derivatives of all orders at the origin, then it is natural to ask *whether for a fixed value of x, the values of these approximating polynomials converge to f(x) as $n \to \infty$?*

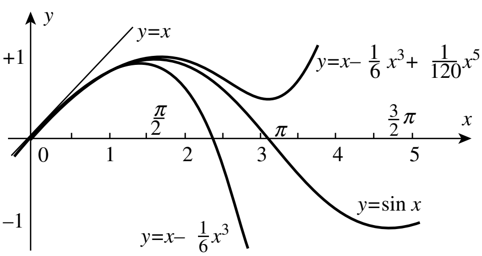


FIGURE 22.4

Now, these (approximating) polynomials are precisely the partial sums of a series, known as *the Maclaurin series for f(x)*, given by

$$f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \ldots + \frac{f^{(n)}(0)}{n!}x^n + \ldots \tag{22}$$

Thus, the question just posed is equivalent to asking *whether the Maclaurin series for f converges to f(x) as a sum?* It certainly has the correct value $f(0)$, at $x = 0$, *but how far away from* $x = 0$ *may we go and still have convergence?* And if the series does converge away from $x = 0$, *does it still converge to f(x)?*

The graphs in Figures 22.2–22.4 are encouraging, and it can be shown that normally a *Maclaurin series converges to its function in an interval about the origin*. For many functions, this interval is the entire x-axis.

22.5 FROM MACLAURIN SERIES TO TAYLOR SERIES

Now, suppose we are interested in approximating a function $f(x)$ near a point $x = a$ (instead of near a point $x = 0$) *in an interval I*, then we write our approximating polynomial $P_n(x)$ in powers of $(x - a)$, as follows:

$$P_n(x) = a_0 + a_1(x - a) + a_2(x - a)^2 + \ldots + a_n(x - a)^n \tag{23}$$

where the constants $a_0, a_1, a_2, \ldots, a_n$ are to be determined. Obviously, the function $f(x)$ is assumed to have all derivatives upto the $(n + 1)^{\text{th}}$ order (inclusive) in the interval I.

Since the polynomial $P_n(x)$ and its first n derivatives must agree with the given function $f(x)$ and its corresponding derivatives at $x = a$, the following conditions should be satisfied:

$$\left.\begin{aligned} P_n(a) &= f(a) \\ P'_n(a) &= f'(a) \\ P''_n(a) &= f''(a) \ \ldots \\ P_n^{(n)}(a) &= f^{(n)}(a) \end{aligned}\right\} \tag{24}$$

Let us first find the derivative of $P_n(x)$

$$\left.\begin{aligned} P'_n(x) &= a_1 + 2a_2(x - a) + 3a_3(x - a)^2 + \ldots + na_n(x - a)^{n-1} \\ P''_n(x) &= 2 \cdot 1a_2 + 3 \cdot 2a_3(x - a) + \ldots + n(n - 1)a_n(x - a)^{n-2} \\ P_n^{(3)}(x) &= 3 \cdot 2 \cdot 1a_3 + \ldots + n(n - 1)(n - 2)a_n(x - a)^{n-3} \ldots \\ P_n^{(n)}(x) &= n(n - 1)(n - 2) \ldots 2 \cdot 1a_n \end{aligned}\right\} \tag{25}$$

Putting $x = a$, in equations (23) and (25), we get

$$\left.\begin{aligned} P_n(a) &= a_0 = f(a) \\ P'_n(a) &= a_1 = f'(a) \\ P''_n(a) &= 2!a_2 = f''(a) \\ P_n^{(3)}(a) &= 3!a_3 = f^{(3)}(a)\ldots \\ P_n^{(n)}(a) &= n! = f^n(a) \end{aligned}\right\} \begin{aligned} &\text{[Using (23) and (24)]} \\ &\text{[Using (24) and (25)]} \end{aligned}$$

$$\left.\begin{aligned} \therefore\ \ a_0 = f(a),\ a_1 = f'(a),\ a_2 = \frac{1}{2!}f''(a) \\ a_3 = \frac{1}{3!}f^{(3)}(a), \ldots \frac{1}{n!}f_n(a). \end{aligned}\right\}$$

Substituting in (23), the values of a_0, a_1, a_2, ..., a_n, we get the required polynomial:

$$P_n(x) = f(a) + \frac{(x-a)}{1!}f'(a) + \frac{(x-a)^2}{2!}f''(a) + \frac{(x-a)^3}{3!}f^{(3)}(a) + \ldots \frac{(x-a)^n}{n!}f^{(n)}(a) \quad (24)$$

The right-hand side of (24) represents an approximation of *f* near "*a*" by the Taylor polynomial $P_n(x)$. *It is called the Taylor series expansion of f about a.* There are two things to notice here:

First is that the Maclaurin series are Taylor series with $a = 0$ and

Second, a function cannot have a Taylor series expansion about $x = a$, unless it has finite derivatives of all orders at $x = a$.

For instance, $f(x) = \log_e x$ does not have a Maclaurin series expansion, since the function itself, (to say nothing of its derivatives), does not have a finite value at $x = 0$. *On the other hand, it does have a Taylor series expansion in powers of $(x - 1)$, since $\log_e x$ and all its derivatives are finite at $x = 1$.*

Note: In Examples (1) and (2) the value of the Taylor polynomial provided a reasonable approximation to the corresponding value of the given function. Indeed, we found that if $f(x) = e^x$, then

$$|f(1) - P_5(1)| = |e - P_5(1)| \approx 0.00161$$

and if $f(x) = \log_e(1 + x)$, then

$$|f(1) - P_6(1)| = |\log_e 2 - P_6(1)| \approx 0.07648$$

By contrast, if $f(x) = 1/(1-x)$ [as in Example (3)], then for $n \geq 1$,

$$\begin{aligned} |f(2) - P_n(2)| &= \left|\frac{1}{1-2} - \left(1 + 2 + 2^2 + \ldots + 2^n\right)\right| \\ &= \left|-1 - \left(1 + 2 + 2^2 + \ldots + 2^n\right)\right| \\ &\geq 2^n \end{aligned}$$

Consequently $P_n(2)$ is not reasonably close to $f(2)$ for any value of n. In fact, the larger n becomes, the worse $P_n(2)$ approximates $f(2)$.

22.6 TAYLOR'S FORMULA FOR POLYNOMIALS

Consider the n^{th} degree polynomial

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots\ldots + a_nx^n$$

where $a_0, a_1, a_2, a_3, \ldots, a_n$ are constant coefficients. We can express $f(x)$ as the expansion in powers of $(x - a)$, with *some coefficients*, where "a" is an arbitrary number. We know that the Taylor polynomial $P_n(x)$ approximating $f(x)$ at $x = a$ is given by

$$f(x) = P_n(x) = f(a) + \frac{(x-a)}{1!}f'(a) + \frac{(x-a)^2}{2!}f''(a) + \frac{(x-a)^3}{3!}f^{(3)}(a) + \ldots + \frac{(x-a)^n}{n!}f^{(n)}(a) \tag{27}$$

The expression on the right-hand side of (27) is called Taylor's formula for the polynomial $f(x)$. Note that, this formula is a partial sum of a series

$$f(a) + \frac{(x-a)}{1!}f'(a) + \frac{(x-a)^2}{2!}f''(a) + \frac{(x-a)^3}{3!}f^{(3)}(a) + \ldots \frac{(x-a)}{n!}f^{(n)}(a) + \ldots$$

called *Taylor series expansion of f about $x = a$.*

If we put $a = 0$ in (27), we get the Taylor polynomial $P_n(x)$ approximating $f(x)$ at $x = 0$, we write,

$$f(x) = P_n(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \ldots\ldots \frac{f^{(n)}(0)}{n!}x^n \tag{28}$$

The expression on the right-hand side of (28) [which is a specific case of Taylor's formula for $f(x)$] is called Maclaurin's formula for $f(x)$. Note that, this formula is a partial sum of a series

$$f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \ldots\ldots \frac{f^{(n)}(0)}{n!}x^n + \ldots$$

called Maclaurin series for $f(x)$ about $x = 0$.

Example (6): Expand the polynomial

$$f(x) = x^2 - 3x + 2 \quad \text{in powers of } x \text{ and in power of } x - 1.$$

Solution: Applying Maclaurin's formula, we get,

$$f(x) = x^2 - 3x + 2 \Rightarrow f(0) = 2$$
$$f'(x) = 2x - 3 \Rightarrow f'(0) = -3$$
$$f''(x) = 2 \Rightarrow f''(0) = 2$$

and
$$f(x) = 2 - \frac{3}{1!}x + \frac{2}{2!}x^2$$
$$= 2 - 3x + x^2$$

Thus, the expansion of $f(x)$ in powers of x is identical to $f(x)$ itself.

To expand $f(x)$ in powers of $(x-1)$, we apply Taylor's formula and get

$$f(x) = x^2 - 3x + 2 \Rightarrow f(1) = 0$$
$$f'(x) = 2x - 3 \Rightarrow f'(0) = -1$$
$$f''(x) = 2 \Rightarrow f''(0) = 2$$

and
$$f(x) = 0 - 1(x-1) + \frac{2}{2!}(x-1)^2$$
$$= -(x-1) + (x-1)^2$$

Notice that Taylor's formula gives the value of $f(x)$ at any point x, provided that the values of $f(x)$ and all its derivatives at some point a are known.

Example (7): Express the polynomial

$$f(x) = 2x^3 - 9x^2 + 11x - 1$$

as a polynomial in $(x-2)$.

Solution: We wish to write $f(x)$ in the form (B) with $a = 2$, and to do so we must compute the derivatives of f at 2:

$$f(x) = 2x^3 - 9x^2 + 11x - 1$$
$$\therefore\ f(2) = 2 \cdot 8 - 9 \cdot 4 + 11 \cdot 2 - 1$$
$$= 16 - 36 + 22 - 1 = 1$$

$$f'(x) = 6x^2 - 18x + 11 \qquad \therefore\ f'(2) = -1$$
$$f''(x) = 12x - 18 \qquad \therefore\ f''(2) = 6$$
$$f^{(3)}(x) = 12 \qquad \therefore\ f^{(3)}(2) = 12$$
$$f^{(n)}(x) = 0 \text{ for } n \geq 4 \qquad \therefore\ f^{(n)}(2) = 0 \text{ for } n \geq 4.$$

We know that

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3, \text{ where } a = 2$$

$$\therefore \; f(x) = 1 + (-1)(x-2) + \frac{6}{2!}(x-2)^2 + \frac{12}{3!}(x-2)^3$$

$$= 1 - (x-2) + 3(x-2)^2 + 2(x-2)^3$$

Although the form of the polynomial just obtained looks quite different from the given polynomial, both polynomials represent the same function.

Example (8): Arrange

$$7 + (x+2) + 3(x+2)^3 + (x+2)^4 - (x+2)^5 \text{ in powers of } x.$$

Solution:

Let $f(x) = 7 + (x+2) + 3(x+2)^3 + (x+2)^4 - (x+2)^5$

$\therefore \; f'(x) = 1 + 9(x+2)^2 + 4(x+2)^3 - 5(x+2)^4$

$\therefore \; f''(x) = 18(x+2) + 12(x+2)^2 - 20(x+2)^3$

$f^{(3)}(x) = 18 + 24(x+2) - 60(x+2)^2$

$f^{(4)}(x) = 24 - 120(x+2)$

$f^{(5)}(x) = -120$

$f^{(n)}(x) = 0, \text{ for } n \geq 6$

$\therefore \; f(0) = 7 + 2 + 24 + 16 - 32 = 17$

$f'(0) = -11, f''(0) = -76$

$f^{(3)}(0) = -174, f^{(4)}(0) = -216$

$f^{(5)}(0) = -120, f^{(6)}(0) = 0$, and so on.

Hence,

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2!}f''(0) + \frac{x^3}{3!}f^{(3)}(0) + \frac{x^4}{4!}f^{(4)}(0) + \frac{x^5}{5!}f^{(5)}(0)$$

$$= 17 + x(-11) + \frac{x^2}{2!}(-76) + \frac{x^3}{3!}(-174) + \frac{x^4}{4!}(-216) + \frac{x^5}{5!}(-120)$$

$$= 17 - 11x - 38x^2 - 29x^3 - 9x^4 - x^5$$ **Ans.**

22.7 TAYLOR'S FORMULA FOR ARBITRARY FUNCTIONS

Now, consider a non-polynomial function, $f(x)$ defined at $x = a$ and which has finite derivatives of all orders at $x = a$ (Figure 22.5).

Let us denote by $R_n(x)$ *the difference between the values of the given function $f(x)$ and the constructed polynomial $P_n(x)$, that is:*

$$R_n(x) = f(x) - P_n(x),$$
$$\therefore \ f(x) = P_n(x) + R_n(x)$$

or, in the expanded form,

$$f(x) = f(a) + \frac{(x-a)}{1!}f'(a) + \frac{(x-a)^2}{2!}f''(a) + \ldots + \frac{(x-a)^n}{n!}f^{(n)}(a) + R_n(x) \tag{29}$$

$R_n(x)$ is called the *remainder*.

The value of $R_n(x)$ tells us how well $P_n(x)$ approximates $f(x)$. The smaller $R_n(x)$ is, the better $P_n(x)$ approximates $f(x)$. For those values of x, for which the remainder $R_n(x)$ is small, the polynomial $P_n(x)$ yields an approximate representation of the function $f(x)$.

Thus, formula (29) enables one *to replace the function $y = f(x)$ by the polynomial $y = P_n(x)$, to an appropriate degree of accuracy assessed by the value of the remainder $R_n(x)$.*

Our next problem is to evaluate the quantity $R_n(x)$ for various values of x. Let us write the remainder in the following form

$$R_n(x) = \frac{(x-a)^{n+1}}{(n+1)}Q(x) \tag{30}$$

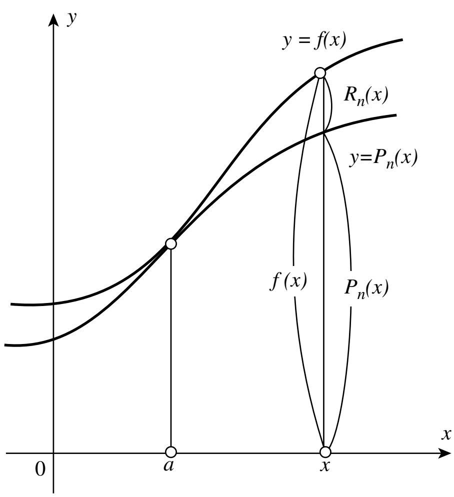


FIGURE 22.5

where $Q(x)$ *is a certain function to be defined*, and accordingly we rewrite (29).

$$f(x) = f(a) + \frac{(x-a)}{1!}f'(a) + \frac{(x-a)^2}{2!}f''(a) + \frac{(x-a)^3}{3!}f^{(3)}(a) + \ldots + \frac{(x-a)^n}{n!}f^{(n)}(a) + \frac{(x-a)^{n+1}}{(n+1)!}Q(x) \quad (31)$$

For fixed x and a, the function $Q(x)$ *has a definite value*. Let us denote it by Q. Let us further examine *the auxiliary function of t* (t lying between a and x):

$$F(t) = f(x) - f(t) - \frac{(x-t)}{1!}f'(t) - \frac{(x-t)^2}{2!}f''(t) - \frac{(x-t)^3}{3!}f^{(3)}(t) - \ldots - \frac{(x-t)^n}{n!}f^{(n)}(t) - \frac{(x-t)^{n+1}}{(n+1)!}Q \quad (32)$$

where Q has the value defined by the relation (31); here we consider a and x to be definite (fixed) numbers.

We find the derivative $F'(t)$ (with respect to t) from equation (32)[6]

$$F'(t) = -f'(t) + f'(t) - \frac{(x-t)}{1}\cdot f''(t) + \frac{2(x-t)}{2!}\cdot f''(t) - \frac{(x-t)^2}{2!}f'''(t) + \ldots - \frac{(x-t)^{n-1}}{(n-1)!}f^n(t) + \frac{n(x-t)^{n-1}}{n!}f^n(t) - \frac{(x-t)^n}{n!}f^{(n+1)}(t) + \frac{(n+1)(x-t)^n}{(n+1)!}Q \quad (33)$$

Note that, except the last two terms, all the terms on the right-hand side of (33) get cancelled.

[6] Derivative of a few terms from (32) are worked out below:

$$-\left[\frac{x-t}{1}f'(t)\right]' = -\left[\frac{x-t}{1}\cdot f''(t) + f'(t)(-1)\right] = f'(t) - \frac{x-t}{1}f''(t)$$

$$-\left[\frac{(x-t)^2}{2!}f''(t)\right]' = -\left[\frac{(x-t)^2}{2!}\cdot f'''(t) + \frac{2(x-t)(-1)}{2!}f''(t)\right] = \frac{2(x-t)}{2!}\cdot f''(t) - \frac{(x-t)^2}{2!}f'''(t)$$

$$-\left[\frac{(x-t)^n}{n!}f^n(t)\right]' = -\left[\frac{n(x-t)^{n-1}(-1)}{n!}f^n(t) + \frac{(x-t)^n}{n!}f^{n+1}(t)\right] = \frac{(x-t)^{n-1}}{(n-1)!}f^n(t) - \frac{(x-t)^n}{n!}f^{n+1}(t)$$

$$-\left[\frac{(x-t)^{n+1}}{(n+1)!}Q\right]' = -\left[\frac{(n+1)(x-t)^n(-1)}{(n+1)!}Q\right] = \frac{(x-t)^n}{n!}Q - \frac{(x-t)^{n+1}}{n+1!}\cdot(0) \quad \left[\because \quad \frac{d}{dt}(Q) = 0\right]$$

∴ On cancelling, we get,

$$F'(t) = -\frac{(x-t)^n}{n!} f^{(n+1)}(t) + \frac{(x-t)^n}{(n)!} Q \tag{34}$$

Thus, the function $F(t)$ [at (32) above] has a derivative at all points t lying near the point with abscissa a ($a \leq t \leq x$), when $a < x$, and $a \geq t \geq x$ when $a > x$.

It will be further noted that, on the basis of (31),

$$\left. \begin{aligned} F(x) &= 0 \\ F(a) &= 0 \end{aligned} \right\} [7]$$

Therefore, Rolle's Theorem is applicable to the function $F(t)$ and consequently, there exists a value $t = c$ lying between a and x such that $F'(c) = 0$.

Therefore, on the basis of relation (34), we get,

$$-\frac{(x-c)^n}{n!} f^{(n+1)}(c) + \frac{(x-c)^n}{(n)!} Q = 0$$

and from this, we get,

$$Q = f^{(n+1)}(c)$$

Substituting this expression into (30), we get,

$$R_n(x) = \frac{(x-a)^{n+1}}{(n+1)} f^{(n+1)}(c)$$

This is the *Lagrange form* of the remainder.

Since c lies between x and a, it may be represented in the following form.

$$c = a + \theta(x-a)$$

where θ is a number lying between 0 and 1, (i.e., $0 < \theta < 1$).[8]

Then, the formula of the remainder takes the following form.

$$R_n(x) = \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}[a + \theta(x-a)]$$

The formula is *called Taylor's formula of the (arbitrary) function f(x).*

$$\begin{aligned} f(x) = f(a) &+ \frac{(x-a)}{1!} f'(a) + \frac{(x-a)^2}{2!} f''(a) + \frac{(x-a)^3}{3!} f^{(3)}(a) \\ &+ \ldots + \frac{(x-a)^n}{n!} f^{(n)}(a) + \frac{(x-a)^{n+1}}{(n+1)!} [a + \theta(x-a)] \end{aligned} \tag{35}$$

[7] Note that $F(x) = f(x) -$ [expansion for $f(x)$ on the right-hand side of (31)] $= 0$ and similarly $F(a) = 0$.
[8] See Chapter 20, alternate form of MVT.

If in Taylor's formula, we put $a = 0$, we will have,

$$f(x) = f(0) + \frac{x}{1!}f'(0) + \frac{x^2}{2!}f''(0) + \ldots + \frac{x^n}{n!}f^{(n)}(0) + \frac{x^{n+1}}{(n+1)!}f^{(n+1)}(\theta x) \qquad (36)$$

where θ lies between 0 and 1. This special case of *Taylor's formula is called Maclaurin's formula.*

Note (9): For applications of the above formulas, refer to standard books.

It can be easily shown that,

$$R_n(x) = \frac{(x-a)^{n+1}}{(n+1)}f^{(n+1)}(C_{n+1}) \qquad (37)$$

where, C_{n+1} [$= t$ (say)] *is a point lying between a and x.*[9]

Note (10): We can use Taylor's formula to achieve approximations with a prescribed accuracy.

Note (11): Note that Lagrange's form of remainder $R_n(x)$ cannot be used for the exact computation of the value of $R_n(x)$ since the exact location of the point C_{n+1} (between a and x) at which the $(n+1)^{\text{th}}$ derivative is taken, is unknown.

Remark: In approximating e and $\log_e 2$, it is seen that the error introduced could be made as small as we wished, by picking n sufficiently large. By *taking larger values of n means adding up more numbers in the n^{th} Taylor polynomial.*[10]

This suggests the possibility of *attaching a meaning to an infinite series (which is the sum of an infinite number of numbers)*. In fact, this can be done, and we will find that e and $\log_e 2$ can be not only approximated, but also represented by a sum of an infinite collection of numbers. It is even possible to create entirely new functions through the process of summing infinite collections of numbers. This demands study of convergence of sequence and series of numbers and that of functions.

In general, we are interested in the possibility of expressing a function $f(x)$ as a power series

$$\sum_{n=0}^{\infty} C_n(x-a)^n \qquad (38)$$

in powers of $(x - a)$, where "a" can be any fixed number.

[9] For details, refer to *Differential and Integral Calculus* (Vol. I, Second Edition) by N. Piskunov, Mir Publishers, Moscow, English translation 1974 (pp. 145–148).

[10] *Calculus with Analytic Geometry* (Alternate Edition) by Robert Ellis and Denny Gulick, HBJ Publication, 1988 [Examples (4) and (5) (pp. 502–503) approximating e with an error less than 0.001 and $\log_e 2$ with an error less than 0.1, respectively.]

In particular, *if f has derivatives of all orders at a*, we call

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n \tag{39}$$

the *Taylor series of f about the number a*.[11]

The n^{th} *Taylor polynomial* $P_n(x)$ *of f about a* is defined by

$$P_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \ldots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

Note: Key condition for expanding a function into Taylor's series is discussed in Mathematical Analysis by A.F. Bermant and I.G. Aramanovich. (By Mir Publishers, Moscow), Page 676.

[11] If $a = 0$, the Taylor series becomes $\sum_{n=0}^{\infty} f^{(n)}(0)\frac{x^n}{n!}$, which we have already discussed in detail and which is frequently called a Maclaurin series, after the Scottish mathematician Colin Maclaurin (1698–1746).

23 Hyperbolic Functions and Their Properties

23.1 INTRODUCTION

Certain *special combinations of* e^x *and* e^{-x} appear so often in both mathematics and science that they are given *special names*.

Definitions: The functions

$$\sinh x = \frac{e^x - e^{-x}}{2} \quad (1)$$

$$\cosh x = \frac{e^x + e^{-x}}{2} \quad (2)$$

are respectively, called the *hyperbolic sine* and *hyperbolic cosine*.

The terminology suggests that hyperbolic functions must have some connection with trigonometric (circular) functions. In fact, there is. It may not be clear at the moment why these names are appropriate, but it will become apparent as we proceed further. Recall that, the trigonometric (circular) functions are intimately related to the unit circle, $x^2 + y^2 = 1$ (Figure 23.1a) on which any point (x, y) is represented by the parametric equations, $x = \cos t$, $y = \sin t$. In parallel fashion, the parametric equations $x = \cosh t$, $y = \sinh t$ describe the right branch of the unit hyperbola $x^2 - y^2 = 1$ [which is the special case of the hyperbola $((x^2/a^2) - (y^2/b^2)) = 1$] (Figure 23.1b). Moreover, in both cases, the parameter t is related to the shaded area S by $t = 2S$, though this is not obvious in the second case.

Certain Similarities in Formulae

1. There are *six basic hyperbolic functions*, just as there are *six basic trigonometric functions*. The other *four hyperbolic functions are defined in the terms of the hyperbolic sine and hyperbolic cosine*.

   Definitions: The functions

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3)$$

23-Hyperbolic and inverse hyperbolic functions: their properties, derivatives, and integrals.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition. Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.

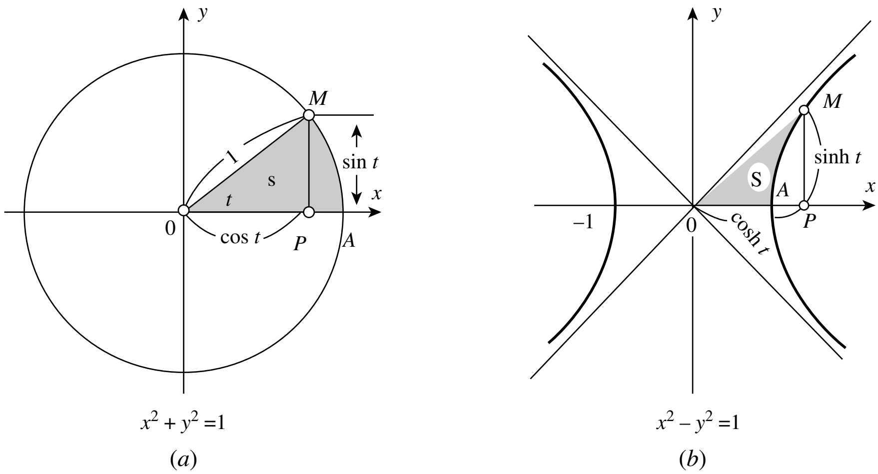


FIGURE 23.1

$$\coth x = \frac{\cosh\ x}{\sinh x} = \frac{e^x + e^{-x}}{e^x - e^{-x}} \tag{4}$$

$$\operatorname{sech} x = \frac{1}{\cosh x} \tag{5}$$

$$\operatorname{cosech} x = \frac{1}{\sinh x} \tag{6}$$

are respectively called the hyperbolic tangent, the hyperbolic cotangent, the hyperbolic secant, and the hyperbolic cosecant.

> **Remark:** It is because of the definitions of sinh x and cosh x that we use the *terminology of circular functions in defining hyperbolic functions*.

2. Hyperbolic functions are connected by a number of algebraic relations similar to those connecting trigonometric functions. In particular, the *fundamental identity* for the hyperbolic functions is

$$\cosh 2x - \sinh 2x = 1 \tag{7}$$

To verify it, we write

$$\begin{aligned}\cosh^2 x - \sinh^2 x &= \left(\frac{e^x + e^{-x}}{2}\right)^2 - \left(\frac{e^x - e^{-x}}{2}\right)^2 \\ &= \frac{1}{4}\left[\left(e^{2x} + e^{-2x} + 2\right) - \left(e^{2x} + e^{-2x} - 2\right)\right] \\ &= 1\end{aligned}$$

(This should be compared with the trigonometric result, $\cos^2 x + \sin^2 x = 1$.)

Similarly, it can be shown that,

$$\cosh^2 x + \sinh^2 x = \cosh 2x \tag{8}$$

(This is analogous to $\cos^2 x - \sin^2 x = \cos 2x$.)

In fact, *any formula for circular functions has its counterpart in hyperbolic functions*. It will be noticed that in the above two cases there is a difference in the signs used, and this applies only to sinh $2x$. *In any formula connecting circular functions of general angles, the corresponding formulae for hyperbolic functions can be obtained* by applying the following rule.

If in any formula connecting $\cos x$, $\sin x$, and $\tan x$, the term $\sin^2 x$ appears (or is implied, as in the case of $\tan^2 x$), then we replace $\sin x$ by $i \sinh x$, $\tan x$ by $i \tanh x$ (where $i=\sqrt{-1}$); $\cos x$ by $\cosh x$, and simplify the expression to obtain the corresponding hyperbolic formula.

Remember that, the product of two sines such as $\sin x \cdot \sin y$ in a formula, will be replaced by $(i \sinh x)\cdot(i \sinh y) = -\sinh x \cdot \sinh y$ to obtain the corresponding formula of hyperbolic functions. Thus, $\sec^2 x = 1 + \tan^2 x$, becomes

$$\text{sech}^2 x = 1 + (i \tanh x)^2 = 1 - \tanh^2 x \tag{9}$$

and $\cos(x \pm y) = \cos x \cos y \pm \sin x \sin y$ becomes

$$\cosh(x \pm y) = \cosh x \cosh y \pm \sinh x \sinh y \tag{10}$$

Further, $\sin(x \pm y) = \sin x \cos y \pm \cos x \sin y$ becomes

$$\sinh(x \pm y) = \sinh x \cosh y \pm \cosh x \sinh y \tag{11}$$

If y is replaced by x in these identities we obtain,

$$\cosh 2x = \cosh^2 x + \sinh^2 x \tag{12}$$

$$\sinh 2x = 2 \sinh x \cdot \cosh x \tag{13}$$

Note (1): From the definitions (1) and (2), we can obtain

$$\cosh x + \sinh x = e^x \tag{14}$$

$$\cosh x - \sinh x = e^{-x} \tag{15}$$

It is, therefore, apparent that *any combination of the exponentials* e^x *and* e^{-x} *can be replaced by a combination of sinh x and cosh x and conversely*.

Let us verify the formula,

$$\cosh(x + y) = \cosh x \cdot \cosh y + \sinh x \cdot \sinh y$$

By definition, we have the left-hand side as,

$$\cos(x + y) = \left(\frac{e^{x+y} + e^{-x-y}}{2}\right)$$

Now, consider the right-hand side,

$$\cosh x \cdot \cosh y + \sinh x \cdot \sinh y = \frac{e^x + e^{-x}}{2} \cdot \frac{e^y + e^{-y}}{2} + \frac{e^x - e^{-x}}{2} \cdot \frac{e^y - e^{-y}}{2}$$

$$= \frac{1}{4}[e^{x+y} + e^{-x+y} + e^{x-y} + e^{-x-y} + e^{x+y} - e^{-x+y} - e^{x-y} + e^{-x-y}]$$

$$= \frac{2[e^{x+y} + e^{-x-y}]}{4} = \frac{e^{x+y} + e^{-x-y}}{2} = \cosh(x+y) = \text{L.H.S.}$$

The important *hyperbolic* and the corresponding *trigonometric formulae* are listed below.

Hyperbolic Functions	Circular Functions
$\cosh 2x - \sinh 2x = 1$	$\cos^2 x + \sin^2 x = 1$
$\sinh 2x = 2 \sinh x + \cosh x$	$\sin 2x = 2 \sin x \cdot \cos x$
$\cosh 2x = \cosh 2x + \sinh 2x$	$\cos 2x = \cos^2 x - \sin^2 x$
$\operatorname{sech}^2 x = 1 - \tanh^2 x$	$\sec^2 x = 1 + \tan^2 x$
$\operatorname{cosech}^2 x = \coth^2 x - 1$	$\operatorname{cosec}^2 x = \cot^2 x + 1$
$\sinh(x \pm y) = \sinh x \cdot \cosh y \pm \cosh x \cdot \sinh y$	$\sin(x \pm y) = \sin x \cdot \cos y \pm \cos x \cdot \sin y$
$\cosh(x \pm y) = \cosh x \cdot \cosh y \pm \sinh x \cdot \sinh y$	$\cos(x \pm y) = \cos x \cdot \cos y \mp \sin x \cdot \sin y$

Note (2): Hyperbolic functions are defined in terms of exponential functions. This is very different from the way we defined trigonometric functions. *However, if you study complex analysis, you will discover that trigonometric functions can also be defined in terms of exponential functions of a complex variable*. Now, we shall discuss the striking connections between the two sets of functions.[1]

23.2 RELATION BETWEEN EXPONENTIAL AND TRIGONOMETRIC FUNCTIONS

The following expansions were obtained in the chapter(s) shown against each:

(i) $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots\ldots\ldots\ldots + \frac{x^n}{n!} + \ldots\ldots$ (see Chapter 13)

(ii) $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots\ldots\ldots\ldots + (-1)^{n-1} \frac{x^{2n-1}}{(2n-1)!} + \ldots\ldots$ (see Chapter 22)

(iii) $\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots\ldots\ldots\ldots + (-1)^{n-1} \frac{x^{2n-2}}{(2n-2)!} + \ldots\ldots$ (see Chapter 22)

It can be shown that these series converge for all values of x, real or complex. Indeed when $x = \alpha + i\beta$, these series will serve as definitions of $e^{\alpha+i\beta}$, $\sin(\alpha + i\beta)$, and $\cos(\alpha + i\beta)$, respectively.

[1] For details, any book on advance trigonometry (i.e., *Trigonometry of Complex Variables*) should be consulted.

(iv) For $x = i\theta$, *a purely imaginary number* (i) becomes

$$e^{i\theta} = 1 + i\theta - \frac{\theta^2}{2!} - i\frac{\theta^3}{3!} + \frac{\theta^4}{4!} + \ldots\ldots\ldots,$$

since $i = \sqrt{-1}$, $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, and so on.

(v) Multiplying (ii) by i and writing θ for x, we get,

$$i\sin\theta = i\theta - i\frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \ldots\ldots\ldots,$$

(vi) For $x = \theta$, (iii) becomes

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \ldots\ldots\ldots,$$

(vii) By adding (v) and (vi), we get (iv). Thus, we have,

$$e^{i\theta} = \cos\theta + i\sin\theta$$

This is a remarkable relation and is generally known as *Euler's Identity*. It exhibits a very simple connection between $\sin\theta$, $\cos\theta$, and $e^{i\theta}$. Evidently,

$$e^{-i\theta} = \cos(-\theta) + i\sin(-\theta), \text{ or}$$

(viii) $e^{-i\theta} = \cos\theta - i\sin\theta$

Solving (vii) and (viii) simultaneously for $\sin\theta$ and $\cos\theta$, we get

(ix) $\sin\theta = \dfrac{e^{i\theta} - e^{-i\theta}}{2i}$ $\left[\text{or } \sin x = \dfrac{e^{ix} - e^{-ix}}{2i}\right]$

(x) $\cos\theta = \dfrac{e^{i\theta} + e^{-i\theta}}{2}$ $\left[\text{or } \cos x = \dfrac{e^{ix} + e^{-ix}}{2}\right]$

These relations are *very important in advanced mathematics*. Also, (ix) and (x) could be used as *definitions of* $\sin\theta$ *and* $\cos\theta$.

Note (3): In many branches of applied mathematics, there are functions very similar to the right-hand side of (ix) and (x) above, which are of definite importance. These are $((e^x - e^{-x})/2)$ and $((e^x + e^{-x})/2)$, *where the exponents are real.*

These simple combinations of the exponential functions are called the "*hyperbolic sine of the variable x*", and the "*hyperbolic cosine of the variable x*" and denoted by $\sinh x$ and $\cosh x$, respectively. That is, by definition,

$$\sinh x = \frac{e^x - e^{-x}}{2} \quad \text{and} \quad \cosh x = \frac{e^x + e^{-x}}{2}$$

Recall that we started this chapter with these definitions. In order to make clear the reference to a hyperbola in these definitions, it may be mentioned that *a trigonometry of hyperbolic functions, comparable to that of the circular functions has been developed.*

It is easy to show that,

$$\sinh x = \frac{1}{i}\sin ix \text{ and } \cosh x = \cos ix, \textit{ where } i = \sqrt{-1}.$$

In the relation

$\sin x = \dfrac{e^{ix} - e^{-ix}}{2i}$, replacing x by ix, we get $\sin ix = \dfrac{e^{-x} - e^{x}}{2i} = -\dfrac{e^{x} - e^{-x}}{2i} = i\dfrac{e^{x} - e^{-x}}{2}$

$$\therefore \sin ix = i \sinh x \quad \text{or} \quad \sinh x = \frac{1}{i}\sin ix$$

Note (4): We know that e^{-x} is positive, therefore the equation, $\cosh x - \sinh x = e^{-x}$ [i.e., equation (ix)] shows that $\cosh x$ is always greater than $\sinh x$. But, for large values of x, e^{-x} is small and $\cosh x \approx \sinh x$.

23.3 SIMILARITIES AND DIFFERENCES IN THE BEHAVIOR OF HYPERBOLIC AND CIRCULAR FUNCTIONS

The graphs of *hyperbolic cosine* and *hyperbolic sine* are shown in Figure 23.2. At $x = 0$, $\cosh x = 1$, and $\sinh x = 0$. Note that these value are same as in the case of corresponding trigonometric functions, at $x = 0$. Therefore, all *the hyperbolic functions have the same values at 0 that the corresponding trigonometric functions have.*

Further, note that,

$$\sinh(-x) = \frac{e^{-x} - e^{x}}{2} = -\frac{e^{x} - e^{-x}}{2} = -\sinh x$$

$$\cosh(-x) = \frac{e^{-x} + e^{x}}{2} = \frac{e^{x} + e^{-x}}{2} = \cosh x$$

Thus, *hyperbolic sine* is an *odd* function and the *hyperbolic cosine* is an *even* function. So the graph of *sinh x is symmetric with respect to the origin and that of cosh x is symmetric about the y-axis*. Here again the hyperbolic functions behave like the ordinary trigonometric (or circular) functions (see Figure 23.2).

Of course, there are *major differences between hyperbolic and circular functions*. For example,

(i) The functions $\sinh x$, $\cosh x$, and $\tanh x$ are obviously *defined for all values of x*. But, the function $\coth x$ is defined everywhere, except at the point $x = 0$ (Figure 23.3).

On the *other hand, the circular function tan x is defined everywhere except at the points* $x = (2k + 1)(\pi/2)$, $(k = 0, \pm 1, \pm 2, \ldots)$. Similarly the function $\cot x$ is defined everywhere except at the points $x = k\pi$, $(k = 0, \pm 1, \pm 2, \ldots)$.

(ii) The *circular functions are periodic*, $\sin(x+2\pi) = \sin x$, $\tan(x+\pi) = \tan x$, and so on. But, *hyperbolic functions are not periodic.*

(iii) Both differ in the range of values they assume.

$\sin x$ varies between -1 and $+1$, i.e., it *oscillates.*

$\sinh x$ varies from $-\infty$ to $+\infty$, i.e., it *steadily increases.*

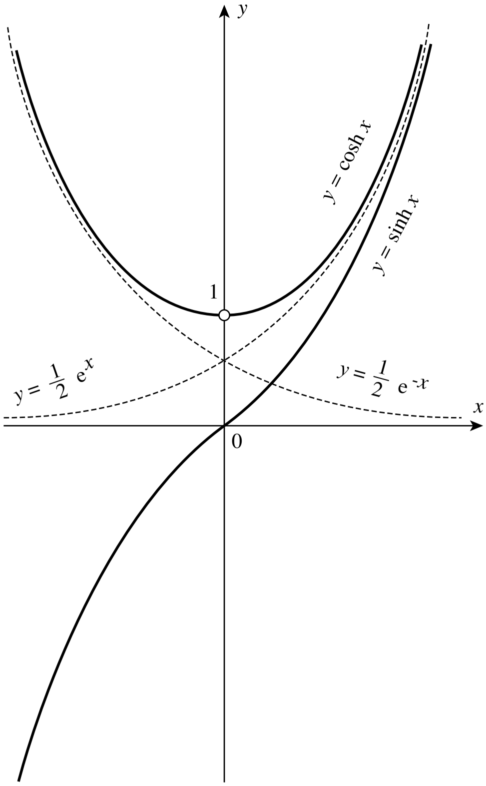


FIGURE 23.2

cos x varies from $-\infty$ to $+\infty$, i.e., it *oscillates*.

cosh x varies from $+\infty$ to 1 to $+\infty$.

tan x varies from $-\infty$ to $+\infty$.

tanh x varies from -1 to $+1$ (Figure 23.4). Also see Note (6), given later.

$|\sec x|$ is never less than 1 $[\because |\cos x| \leq 1]$

sech x is *never greater than 1*, and is always positive (see Figure 23.5).

(iv) Another difference *lies in the behavior of the functions as* $x \to \pm\infty$. *We can say nothing very specific about the behavior of the circular functions* sin x, cos x, tan x, and so on for large values of x. But the hyperbolic functions behave very much like $(e^x)/2$, $(e^{-x})/2$, *unity or zero*, as explained below.

For x large and positive:

$$\cosh x \approx \sinh x \approx \frac{1}{2}e^x$$

$$\tanh x \approx \coth x \approx 1 \, \text{sech}\, x \approx \text{cosech}\, x \approx 2e^{-x} \approx 0$$

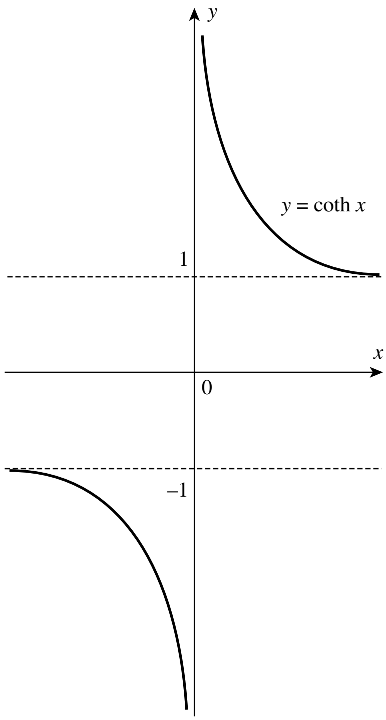


FIGURE 23.3

For x negative and |x| large:

$$\cosh x \approx \sinh x \approx -\frac{e^x - e^{-x}}{2} \approx \frac{1}{2}e^{-x}$$

$$\tanh x \approx \coth x \approx \frac{e^x + e^{-x}}{e^x - e^{-x}} \approx -1$$

$$\operatorname{sech} x \approx -\operatorname{cosech} x \approx -\frac{2}{e^x - e^{-x}} \approx 2e^x \approx 0$$

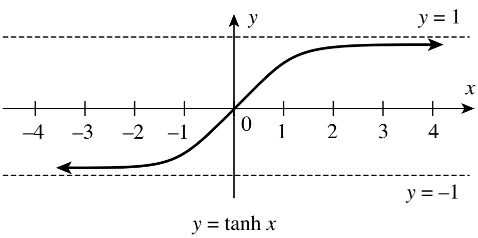
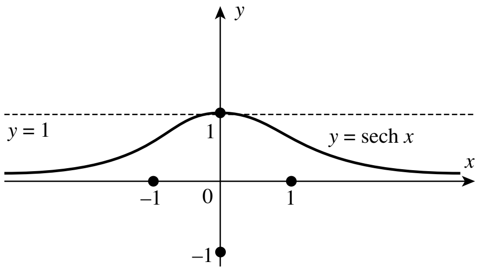


FIGURE 23.4



$y = \text{sech}\, x$

FIGURE 23.5

Note (5): The hyperbolic functions are *not included in the class of basic elementary functions*, but we discuss them here since they are important for applications.

23.4 DERIVATIVES OF HYPERBOLIC FUNCTIONS

The *formulas for the derivatives* of the hyperbolic sine and hyperbolic cosine functions are obtained by considering their definitions (i) and (ii), and differentiating the expressions involving exponential functions. Thus,

$$\frac{\mathrm{d}}{\mathrm{d}x}(\sinh x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{e}^{x} - \mathrm{e}^{-x}}{2}\right) = \frac{\mathrm{e}^{x} + \mathrm{e}^{-x}}{2} = \cosh x$$

$$\text{and } \frac{\mathrm{d}}{\mathrm{d}x}(\cosh x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{e^{x} + e^{-x}}{2}\right) = \frac{e^{x} - e^{-x}}{2} = \sinh x$$

From these formulas and the chain rule we have the following theorem.

Theorem (A): If u is a *differentiable function of x,*

$$\frac{\mathrm{d}}{\mathrm{d}x}(\sinh u) = \cosh u \cdot \frac{\mathrm{d}u}{\mathrm{d}x}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(\cosh u) = \sinh u \cdot \frac{\mathrm{d}u}{\mathrm{d}x}$$

The derivative of tanh x may be found from the exponential definition or we may use the above result(s) (i.e., the derivatives of sinh x and cosh x).

$$\text{Let } y = \tanh\, x = \frac{\sinh x}{\cosh x}$$

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}(\tanh x) &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\sinh x}{\cosh x}\right) \\
&= \frac{\cosh x \cdot \cosh x - \sinh x \cdot \sinh x}{\cosh^2 x} \\
&= \frac{\cosh^2 x \cdot \ - \sinh^2 x}{\cosh^2 x} \\
&= \frac{1}{\cosh^2 x} = \text{sech}^2\, x \left[\because\ \cosh^2 x - \sinh^2 x = 1\right]
\end{aligned}$$

The formulas for the derivatives of the remaining three hyperbolic functions are

$$\frac{\mathrm{d}}{\mathrm{d}x}(\coth x) = -\operatorname{cosec} \mathrm{h}^2 x$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(\operatorname{sech} x) = -\operatorname{sech} x \cdot \tanh x$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(\operatorname{cosech} x) = -\operatorname{cosech} x \cdot \coth x$$

From these formulas and the chain rule, we have the following theorem.

Theorem (B): If u is a differentiable function of x,

$$\frac{\mathrm{d}}{\mathrm{d}x}(\tanh u) = \operatorname{sech}^2 u \frac{\mathrm{d}u}{\mathrm{d}x}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(\coth u) = -\operatorname{cosec} \mathrm{h}^2 u \frac{\mathrm{d}u}{\mathrm{d}x}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(\operatorname{sech} u) = -\operatorname{sech} u \cdot \tanh u \cdot \frac{\mathrm{d}u}{\mathrm{d}x}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(\operatorname{cosec} \mathrm{h}\, u) = -\operatorname{cosec} \mathrm{h}\, u \cdot \coth u \cdot \frac{\mathrm{d}u}{\mathrm{d}x}$$

23.5 CURVES OF HYPERBOLIC FUNCTIONS

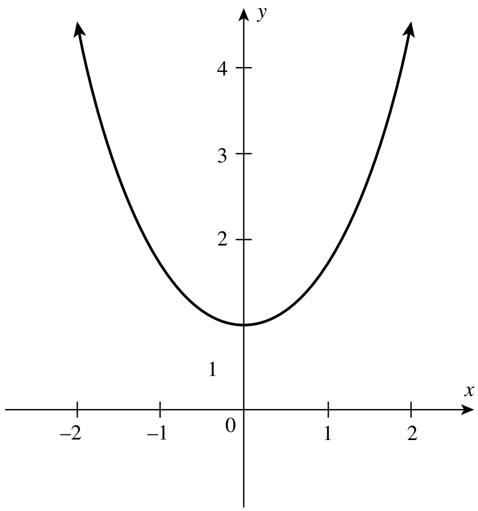
The curves of cosh x and sinh x in Figure 23.2 should be examined again with the assistance of their differential coefficients.

1. $y = \cosh x = \dfrac{e^x + e^{-x}}{2}$

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{e^x - e^{-x}}{2} = \sinh x, \text{ and}$$

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{e^x + e^{-x}}{2} = \cosh x$$

Note that dy/dx vanishes only when x=0. There is, therefore, a turning point on the curve (see Figure 23.6). Also, since dy/dx (=sinh x) is *negative before this point and positive after*, while $(\mathrm{d}^2y)/(\mathrm{d}x^2)$ is positive, the point x=0 is a minimum. There is *no other turning point and no point of inflexion.* The curve of cosh x is an important one. It is called the *catenary*, and is the curve formed by a uniform flexible chain which hangs freely with its ends fixed.



$y = \cosh x$

FIGURE 23.6

2. $y = \sinh x$.

$$\frac{dy}{dx} = \cosh x, \quad \frac{d^2y}{dx^2} = \sinh x$$

Note that dy/dx is *always positive and does not vanish*. Consequently, $\sinh x$ is always increasing and has no turning point. When $x=0$, $(d^2y)/(dx^2) = 0$, and *is negative before and positive after*. Therefore, there is a point of inflexion when $x = 0$; since dy/dx (i.e., $\cosh x) = 1$.

When $x = 0$, the gradient at 0 is unity and the slope is $\pi/4$ (Figure 23.7).

3. $y = \tanh x \cdot \dfrac{dy}{dx} = \text{sech}^2 x$.

Since $\text{sech}^2 x$ is *always positive*, $\tanh x$ is always increasing between $-\infty$ and $+\infty$. Also since $\sinh x$ and $\cosh x$ are *always continuous* and *cosh x never vanishes, tanh x must be a continuous function.*

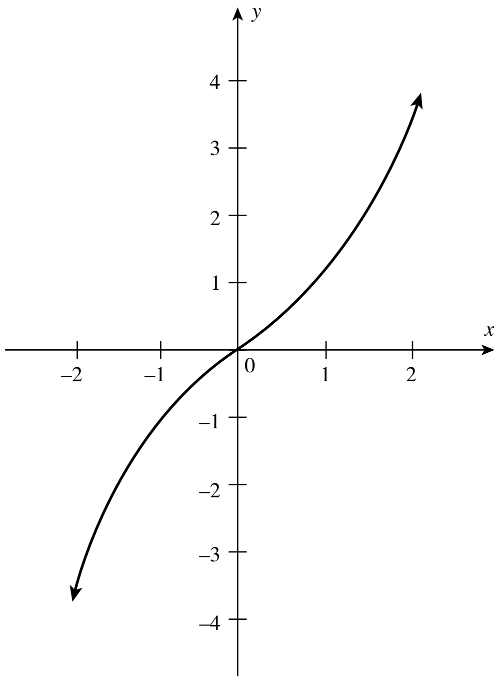
Note (6): The expression for $\tanh x$ can be written in the form:

$$\tanh x = \frac{e^{2x} - 1}{e^{2x} + 1} = 1 - \frac{2}{e^{2x} + 1}$$

From this form, it is evident that while x increases from $-\infty$ to 0, e^{2x} increases from 0 to 1.

$$\therefore 1 - \frac{2}{e^{2x} + 1} \text{ or } \tanh x \textit{ increases from } -1 \textit{ to } 0.$$

Similarly, while x increases from 0 to $+\infty$, $\tanh x$ increases from 0 to 1. The curve therefore has the lines $y = \pm 1$ as its asymptotes which are shown in Figure 23.8.



$y = \sinh x$

FIGURE 23.7

Note (7): Observe that the *derivatives* of the *hyperbolic sine, cosine, and tangent* all have a *plus sign*, whereas those for the *derivatives* of the *hyperbolic cotangent, secant, and cosecant* all have a *minus sign*. Otherwise, the formulas are similar to the corresponding ones for the derivatives of the trigonometric functions.[2]

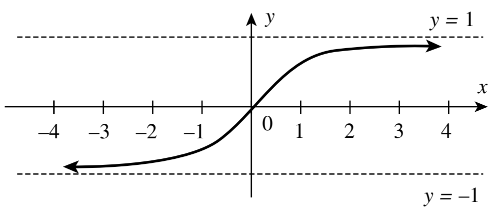


FIGURE 23.8 $y = \tanh x$

[2] Recall that in the case of circular functions, the derivatives of cofunctions (i.e., cos x, cot x, and cosec x) are with negative sign.

23.6 THE INDEFINITE INTEGRAL FORMULAS FOR HYPERBOLIC FUNCTIONS[3]

The indefinite integration formulas for hyperbolic functions from the corresponding differentiation formulas.

$$\int \sinh u \, \mathrm{d}u = \cosh u + c$$

$$\int \cosh u \, \mathrm{d}u = \sinh u + c$$

$$\int \mathrm{sech}^2 u \, \mathrm{d}u = \tanh u + c$$

$$\int \mathrm{cosec\,h}^2\, u \, \mathrm{d}u = -\coth u + c$$

$$\int \sec u \tanh u \, \mathrm{d}u = -\mathrm{sech}\, u + c$$

$$\int \mathrm{cosech}\, u \cdot \coth u \, \mathrm{d}u = -\mathrm{cosech}\, u + c$$

23.7 INVERSE HYPERBOLIC FUNCTIONS

(i) *Inverse Hyperbolic Sine Function.* From the graph of the hyperbolic sine in Figure 23.7, observe that a horizontal line intersects the graph in at most one point. The *hyperbolic sine is, therefore one-to-one*. Furthermore, the hyperbolic sine is continuous and increasing on its domain. Thus, this function has an inverse that we now define.

Definition (A): The *inverse hyperbolic sine function* denoted by $\sinh^{-1} x$, is defined as follows:

$y = \sinh^{-1} x$, if and only if, $x = \sinh y$, where y is any real number (Figure 23.9).

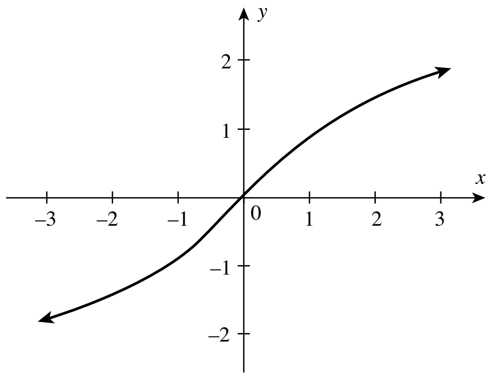
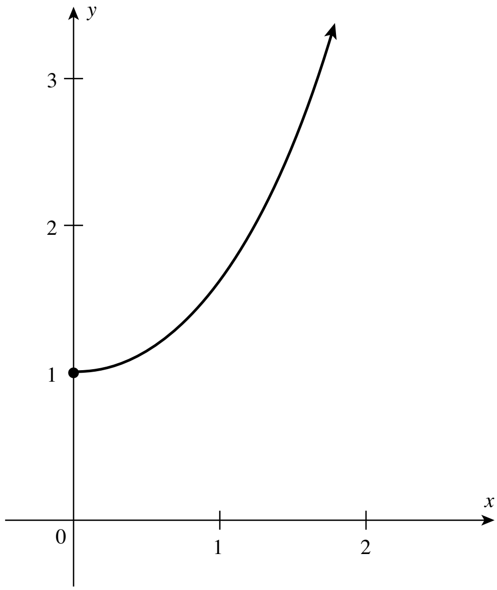


FIGURE 23.9

[3] In fact, this material belongs to Part II of the book. However, it is included here to convey that the techniques applied to integrate hyperbolic functions are similar to those used for trigonometric (circular) functions.



$y = \cosh x, x > 0$

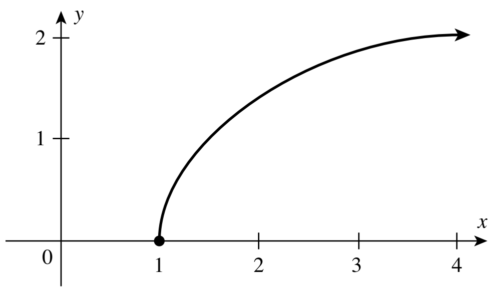
FIGURE 23.10

Both, the *domain and range of* $\sinh^{-1}x$*, are the set R of real numbers*. From the definition (A),

$$\sinh(\sinh^{-1}x) = x \text{ and } \sinh^{-1}(\sinh y) = y.$$

(ii) *Inverse Hyperbolic Cosine Function*. From the graph of the hyperbolic cosine in Figure 23.6, notice that a horizontal line, $y = k$ where $k > 1$, will intersect the graph in two points. Thus, cosh is not one-to-one and *does not have an inverse*. However, as in the case of inverse trigonometric functions, *we restrict the domain and define a new function F* as follows:

$$F(x) = \cosh x,\ x \geq 0 \quad \text{(Figures 23.10 and 23.11)}.$$



$y = \cosh^{-1} x$

FIGURE 23.11

The domain of this function is the interval $[0, +\infty)$ and the range is the interval $[1, +\infty)$. Because F is continuous and increasing on its domain, it has an inverse, called the *inverse hyperbolic cosine function*.

Definition (B): The *inverse hyperbolic cosine function* denoted by $\cosh^{-1}x$, is defined as follows:

$$y = \cosh^{-1}x, \text{ if and only if}, \; x = \cosh y, \text{ where } y \geq 0.$$

The domain of $\cosh^{-1}x$ is in the interval $[1, +\infty)$ and the range is in the interval $[0, +\infty)$. From the definition (B),

$$\cosh(\cosh^{-1}x) = x \text{ if } x \geq 1 \text{ and } \cosh^{-1}(\cosh y) \text{ if } y \geq 0.$$

(iii) *Inverse Hyperbolic Tangent Function.* As with the hyperbolic sine, a horizontal line intersects the graph of the hyperbolic tangent (Figure 23.8) in at most one point. Therefore, the hyperbolic tangent function is one-to-one and has an inverse.

Definition (C): The *inverse hyperbolic tangent function* denoted by $\tanh^{-1}x$ is defined as follows:

$$y = \tanh^{-1}x, \text{ if and only if, } x = \tanh y, \text{ where } y \text{ is any real number.}$$

The domain of the inverse hyperbolic tangent function is the interval $(-1, 1)$ and the range is the set R of real numbers. The graph of $\tanh^{-1}x$ appears in Figure 23.12.

(iv) *Inverse Hyperbolic Cotangent Function.* In this case, a horizontal line intersects the graph of the hyperbolic cotangent function in at most one point. Hence, this function is one-to-one and has an inverse. For convenience, the graphs of both $y = \coth x$ and $y = \coth^{-1}x$ are given in Figures 23.13a and 23.13b. The domain of the inverse hyperbolic cotangent function is $(-\infty, 1) \cup (1, +\infty)$ and the range is $(-\infty, 0) \cup (0, +\infty)$.

Note (8): The *inverse hyperbolic secant and inverse hyperbolic cosecant* functions are not discussed here, since they are seldom used.

23.7.1 Logarithm Equivalents of the Inverse Hyperbolic Functions

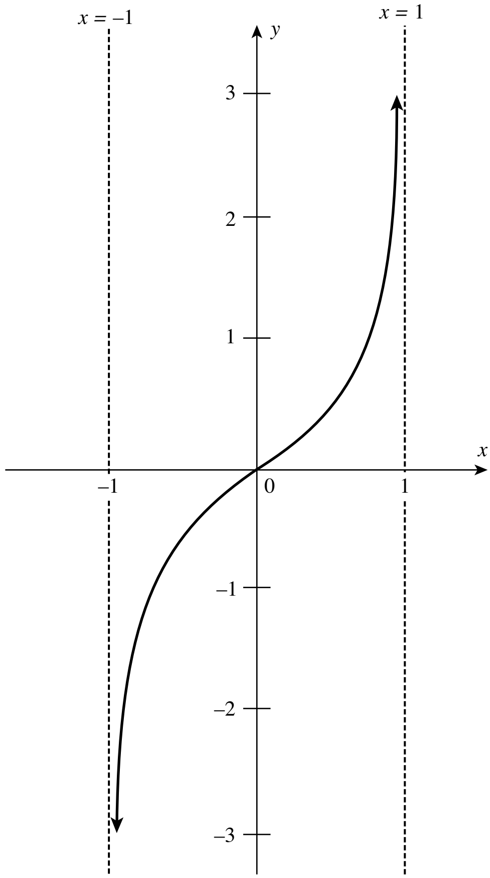
Since the hyperbolic functions are defined in terms of e^x and e^{-x}, it is not too surprising that the *inverse hyperbolic functions can be expressed in terms of the natural logarithm.* Following are these expressions for the four inverse hyperbolic functions we have discussed.

$$\sinh^{-1}x = \log_e(x + \sqrt{x^2+1}), \; x \in R \tag{16}$$

$$\cosh^{-1} = \log_e(x + \sqrt{x^2-1}), \; x \geq 1 \tag{17}$$

$$\tanh^{-1}x = \frac{1}{2}\log_e\frac{1+x}{1-x}, \; |x| < 1 \tag{18}$$

$$\coth^{-1}x = \frac{1}{2}\log_e\frac{x+1}{x-1}, \; |x| > 1 \tag{19}$$



$y = \tanh^{-1} x$

FIGURE 23.12

To prove $\sinh^{-1}x = \log_e(x + \sqrt{x^2+1}),\ x \in R$

Let $y = \sinh^{-1}x$

$\therefore$ From definition (A)

$$x = \sinh y$$

$$\text{or } x = \frac{e^y - e^{-y}}{2}$$

$$2x = e^y - \frac{1}{e^y}$$

$$\therefore\ e^{2y} - 2x \cdot e^y - 1 = 0 \text{ or } (e^y)^2 - 2xe^y - 1 = 0$$

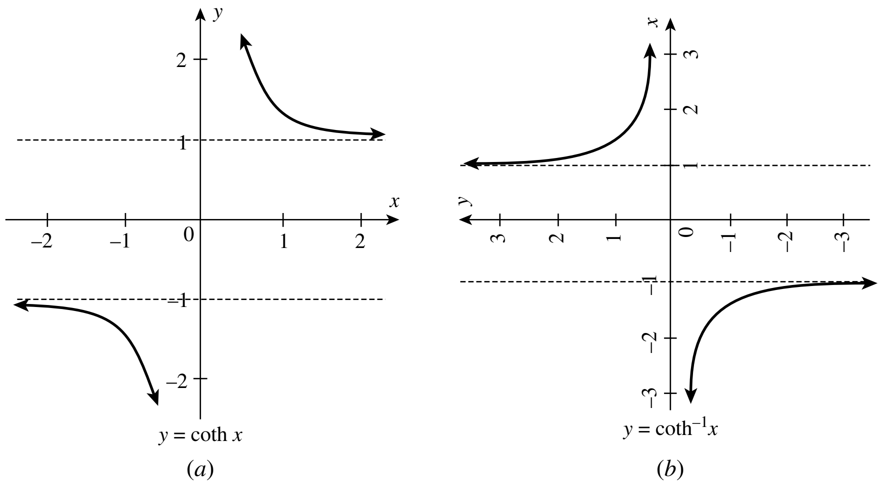


FIGURE 23.13

Solving this equation for e^y by using the quadratic formula, we get,

$$e^y = 2x \pm \frac{\sqrt{4x^2+4}}{2}$$

$$\therefore \; e^y = x \pm \sqrt{x^2+1}$$

We can reject the minus sign in this equation because $e^y > 0$ for all y, while $x - \sqrt{x^2+1}$ is less than zero for all x. Therefore,

$$y = \log_e(x + \sqrt{x^2+1}),$$

But, $y = \sinh^{-1}x$, which means that,

$$y = \sinh^{-1}x = \log_e(x + \sqrt{x^2+1})$$

Other formulas can be proved similarly.

Example (1): Express each of the following in terms of a natural logarithm

(a) $\sinh^{-1}2$

(b) $\tanh^{-1}\left(-\frac{4}{5}\right)$

Solution: (a) We have

$$\sinh^{-1}x = \log_e(x + \sqrt{x^2+1})$$

$$\therefore \sinh^{-1}2 = \log_e(2 + \sqrt{5}) \quad \textbf{Ans.}$$

(b) $\tanh^{-1}\left(-\frac{4}{5}\right)$

We have, $\tanh^{-1}x = \frac{1}{2}\log_e \frac{1+x}{1-x}$, $|x| < 1$

Note that, $x = -\frac{4}{5}$ and $\left|-\frac{4}{5}\right| < 1$

$$\therefore \tan h^{-1}x = \frac{1}{2}\log_e \frac{1-(4/5)}{1+(4/5)}$$

$$= \frac{1}{2}\log_e\left(\frac{1/5}{9/5}\right) = \frac{1}{2}\log_e\left(\frac{1}{9}\right) = \frac{1}{2}\log_e\left(\frac{1}{9}\right)$$

$$= \frac{1}{2}\log_e 3^{-2} = -\log_e 3 \quad \textbf{Ans.}$$

To prove $\tanh^{-1}x = \frac{1}{2}\log_e \frac{1+x}{1-x}$, $|x| < 1$

Let $y = \tanh^{-1}x$

$$x = \tanh y, \quad \text{where } |x| < 1 (\text{i.e., } x \text{ lies between } -1 \text{ and } +1)$$

$$\therefore \quad x = \frac{e^y - e^{-y}}{e^y + e^{-y}} = \frac{e^{2y} - 1}{e^{2y} + 1}$$

$$\therefore \quad x(e^{2y} + 1) = e^{2y} - 1$$

$$\text{or} \quad xe^{2y} + x = e^{2y} - 1$$

$$\text{or} \quad e^{2y}(x - 1) = -(x + 1)$$

$$\text{or} \quad e^{2y} = -\frac{(x+1)}{(x-1)} = \frac{1+x}{1-x}$$

$$\therefore \quad 2y = \log_e \frac{1+x}{1-x}$$

$$\text{or } y = \frac{1}{2}\log_e \frac{1+x}{1-x}, \ |x| < 1$$

23.7.2 Differentiation of Inverse Hyperbolic Functions

Inverse hyperbolic functions correspond to inverse circular functions, and their derivatives are found by similar methods.

(i) Derivative of $\sinh^{-1}x$

Method (1): Let $y = \sinh^{-1}x$

Then $x = \sinh y$

$$\therefore \quad \frac{dx}{dy} = \cosh y$$

$$\therefore \quad \frac{dy}{dx} = \frac{1}{\cosh y} = \frac{1}{\sqrt{1+\sinh^2 y}}$$

$$= \frac{1}{\sqrt{1+x^2}} \quad \text{or} \quad \frac{1}{\sqrt{x^2+1}} \qquad \textbf{Ans.}$$

Method (2): By using the logarithm equivalents, we can compute the derivative of $\sinh^{-1} x$ as follows:

$$\therefore \quad \frac{d}{dx}\left(\sinh^{-1} x\right) = \frac{d}{dx}\log_e\left(x+\sqrt{x^2+1}\right)$$

$$= \frac{1}{x+\sqrt{x^2+1}} \cdot \frac{d}{dx}\left(x+\sqrt{x^2+1}\right)$$

$$= \frac{1}{x+\sqrt{x^2+1}} \cdot \left(1+\frac{1}{2}\cdot\frac{1}{\sqrt{x^2+1}}\cdot 2x\right)$$

$$= \frac{\sqrt{x^2+1}+x}{\sqrt{x^2+1\left(x+\sqrt{x^2+1}\right)}} = \frac{1}{\sqrt{x^2+1}} \qquad \textbf{Ans.}$$

(ii) Derivative of $\cosh^{-1} x$

Let $y = \cosh^{-1} x$. Using the same method as above, we get,

$$\frac{dy}{dx} = \frac{1}{\sqrt{x^2-1}}$$

(iii) Derivative of $\tanh^{-1} x$

$$\text{If } y = \tanh^{-1} x$$

$$x = \tanh y$$

$$\therefore \quad \frac{dx}{dy} = \sec^2 y$$

$$\text{and} \quad \frac{dy}{dx} = \frac{1}{\sec^2 y} = \frac{1}{1-\tanh^2 y} = \frac{1}{1-x^2}$$

The differential coefficient of the reciprocals of the above can be found by the same methods. They are,

$$y = \operatorname{sech}^{-1} x, \qquad \frac{dy}{dx} = -\frac{1}{x\sqrt{1-x^2}}$$

$$y = \operatorname{cosech}^{-1} x, \qquad \frac{dy}{dx} = -\frac{1}{x\sqrt{1+x^2}}$$

$$y = \coth^{-1} x, \qquad \frac{dy}{dx} = -\frac{1}{x^2-1}$$

TABLE 23.1

Functions	Derivatives	Functions	Derivatives
$\sinh x$	$\cosh x$	$\sinh^{-1}x$	$\frac{1}{\sqrt{1+x^2}}$
$\cosh x$	$\sinh x$	$\cosh^{-1}x$	$\frac{1}{\sqrt{x^2-1}}, x > 1$
$\tanh x$	$\text{sech}^2 x$	$\tanh^{-1}x$	$\frac{1}{1-x^2}, \lvert x\rvert < 1$
$\coth x$	$-\text{cosech}^2 x$	$\coth^{-1}x$	$-\frac{1}{x^2-1}, \lvert x\rvert > 1$
$\text{sech}\, x$	$-\text{sech}\, x \cdot \tanh x$	$\text{sech}^{-1}x$	$-\frac{1}{x\sqrt{1-x^2}}, 0 < x < 1$
$\text{cosech}\, x$	$-\text{cosech}\, x \cdot \coth x$	$\text{cosech}^{-1}x$	$-\frac{1}{x\sqrt{1+x^2}}$

The derivatives of all the hyperbolic functions and their corresponding inverse functions are given in Table 23.1.

From these formulas and the chain rule, we can obtain the following results.

If u is a differentiable function of x

- $$\frac{d}{dx}\left(\sinh^{-1}u\right) = \frac{1}{\sqrt{u^2+1}} \cdot \frac{du}{dx} \tag{20}$$
- $$\frac{d}{dx}\left(\cosh^{-1}u\right) = \frac{1}{\sqrt{u^2-1}} \cdot \frac{du}{dx}, \quad u > 1 \tag{21}$$
- $$\frac{d}{dx}\left(\tanh^{-1}u\right) = \frac{1}{1-u^2} \cdot \frac{du}{dx}, \quad |u| < 1 \tag{22}$$
- $$\frac{d}{dx}\left(\coth^{-1}u\right) = \frac{1}{1-u^2} \cdot \frac{du}{dx}, \quad |u| > 1 \tag{23}$$

Later on, the following forms will be found to be important.

1. $y = \sinh^{-1}\frac{x}{a}$

$$\frac{dy}{dx} = \frac{1}{\sqrt{1+(x^2/a^2)}} \cdot \frac{1}{a} = \frac{1}{\sqrt{a^2+x^2}} \quad \text{or} \quad \frac{1}{\sqrt{x^2+a^2}}$$

2. If $y = \cosh^{-1}\frac{x}{a}$

$$\frac{dy}{dx} = \frac{1}{\sqrt{x^2-a^2}}$$

3. If $y = \tanh^{-1}\frac{x}{a}$

$$\frac{dy}{dx} = \frac{a}{a^2-x^2}$$

Logarithm equivalents

$$\sinh^{-1}\frac{x}{a} = \log\left\{\frac{x+\sqrt{x^2+a^2}}{a}\right\}$$

$$\cosh^{-1}\frac{x}{a} = \log\left\{\frac{x+\sqrt{x^2-a^2}}{a}\right\}$$

$$\tanh^{-1}\frac{x}{a} = \frac{1}{2}\log\left\{\frac{a+x}{a-x}\right\}$$

Example (2): Find dy/dx if $y = \tanh^{-1}(\cos 2x)$.

Solution: We have,

$$\frac{d}{dx}\left(\tanh^{-1}u\right) = \frac{1}{1-u^2}\cdot\frac{du}{dx}, \quad \text{where} \quad u = \cos 2x$$

$$\therefore \quad \frac{dy}{dx} = \frac{1}{1-\cos^2 2x}\cdot(-2\sin 2x)$$

$$= \frac{(-2\sin 2x)}{\sin^2 2x}$$

$$= \frac{-2}{\sin 2x} = 2\,\mathrm{cosec}2x \qquad \textbf{Ans.}$$

Example (3): Find dy/dx, if $y = \sinh^{-1}(\tan x)$.

Solution:

$$\frac{d}{dx}(\sinh^{-1}(\tan x)) = \frac{1}{\sqrt{\tan^2 x+1}}\cdot\frac{d}{dx}(\tan x)$$

$$= \frac{1}{\sqrt{\tan^2 x+1}}\cdot\sec^2 x = \frac{\sec^2 x}{|\sec x|} = |\sec x| \qquad \textbf{Ans.}$$

From the formulas for the derivatives of the inverse hyperbolic functions given in Table 23.1, we obtain integration formulas, as follows:

$$\int\frac{du}{\sqrt{u^2+1}} = \sinh^{-1}u + c \tag{24}$$

$$\int\frac{du}{\sqrt{u^2-1}} = \cosh^{-1}u + c \tag{25}$$

$$\int\frac{du}{1-u^2} = \left\{\begin{array}{l}\tanh^{-1}u + c \text{ for } |u| < 1\\ \coth^{-1}u + c \text{ for } |u| > 1\end{array}\right\} \tag{26}$$

$$\int\frac{du}{u\sqrt{1-u^2}} = -\mathrm{sech}^{-1}|u| + c \tag{27}$$

$$\int\frac{du}{u\sqrt{1+u^2}} = -\mathrm{cosec\,h}^{-1}|u| + c \tag{28}$$

Note (9): From Table 23.1, observe that $\tanh^{-1}x$ and $\coth^{-1}x$ have algebraically identical derivatives, but the domain of $\tanh^{-1}x$ is $|x| < 1$ while the domain of $\coth^{-1}x$ is $|x| > 1$.

Hence, *there are two expressions in the formula (26)*. Further note that

$$\frac{1}{1-u^2} = \frac{1/2}{1-u} + \frac{1/2}{1+u}$$

$\therefore$ We can write,

$$\int \frac{du}{1-u^2} = -\frac{1}{2}\log_e|1-u| + \frac{1}{2}\log_e|1+u| + c$$

$$\therefore \int \frac{du}{1-u^2} = \frac{1}{2}\log_e\left|\frac{1+u}{1-u}\right| + c$$

This is an alternative to (26).

Note (10): The main application of inverse hyperbolic functions is in connection with integration, where the following formulas are used.

(I) $$\int \frac{du}{\sqrt{u^2+a^2}} = \sinh^{-1}\frac{u}{a} + c = \log_e(u + \sqrt{u^2+a^2} + c \quad \text{if} \quad a > 0$$

(II) $$\int \frac{du}{\sqrt{u^2-a^2}} = \cosh^{-1}\frac{u}{a} + c = \log_e(u + \sqrt{u^2-a^2} + c \quad \text{if} \quad u > a > 0$$

(III) $$\int \frac{du}{\sqrt{a^2-u^2}} = \begin{cases} \frac{1}{a}\tanh^{-1}\frac{u}{a} + c & \text{for } |u| < a \\ \frac{1}{a}\coth^{-1}\frac{u}{a} + c & \text{for } |u| > a \end{cases}$$

$$= \frac{1}{2a}\log_e\left|\frac{a+u}{a-u}\right| + c \quad \text{if } u \neq a \text{ and } a \neq 0$$

These formulas can be proved by computing the derivatives of the right-hand side and obtaining the integral. We demonstrate the procedure by proving (I) from note 10.

Proof of (I):

$$\frac{d}{dx}\left(\sinh^{-1}\frac{u}{a}\right) = \frac{1}{\sqrt{(u/a)^2+1}} \cdot \frac{1}{a}$$

$$= \frac{\sqrt{a^2}}{\sqrt{u^2+a^2}} \cdot \frac{1}{a} \quad \text{and because } a > 0, \sqrt{a^2} = a; \text{ thus}$$

$$\frac{d}{dx}\left(\sinh^{-1}\frac{u}{a}\right) = \frac{1}{\sqrt{u^2+a^2}}$$

To obtain the natural logarithm representation, we use formula (1) page 696.
We have,

$$\sinh^{-1}x = \log_e(x + \sqrt{x^2+1}),\ x \in R$$

$$\therefore \sinh^{-1}\frac{u}{a} = \log_e\left(\frac{u}{a} + \sqrt{\left(\frac{u}{a}\right)^2 + 1}\right)$$

$$= \log_e\left(\frac{u}{a} + \frac{\sqrt{u^2 + a^2}}{a}\right)$$

$$= \log_e\left(u + \sqrt{u^2 + a^2}\right) - \log_e a$$

Therefore,

$$\sinh^{-1}\frac{u}{a} + c = \log_e\left(u + \sqrt{u^2 + a^2}\right) - \log_e a + c$$

$$= \log_e\left(u + \sqrt{u^2 + a^2}\right) + c_1$$

where $c_1 = c - log_e a$.

In Part II of this book, you will learn various techniques to evaluate the integrals. The formulas (I), (II), and (III) above, give alternate representations of the integral in question. *When evaluating an integral in which one of these forms occurs, the inverse hyperbolic representation may be easier to use and is sometimes less cumbersome to write*. However, in the case of definite integrals, wherein numerical value(s) are obtained, the logarithmic form of the integral may be found more useful.

23.8 JUSTIFICATION FOR CALLING sinh AND cosh AS HYPERBOLIC FUNCTIONS JUST AS sine AND cosine ARE CALLED TRIGONOMETRIC CIRCULAR FUNCTIONS

Recall from trigonometry course (or see Chapter 5) that if t is the angle formed by the x-axis and a line from the origin to the point P(x, y) on the unit circle, then

$$\sin t = y \text{ and } \cos t = x \text{(see Figure 23.14)}.$$

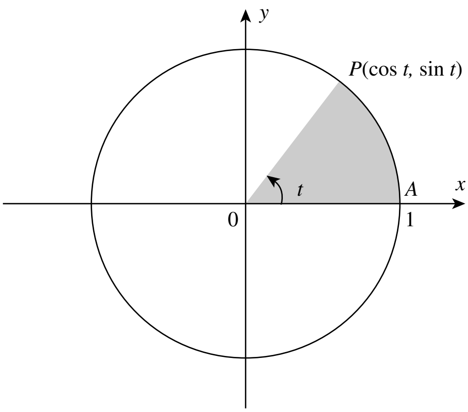


FIGURE 23.14

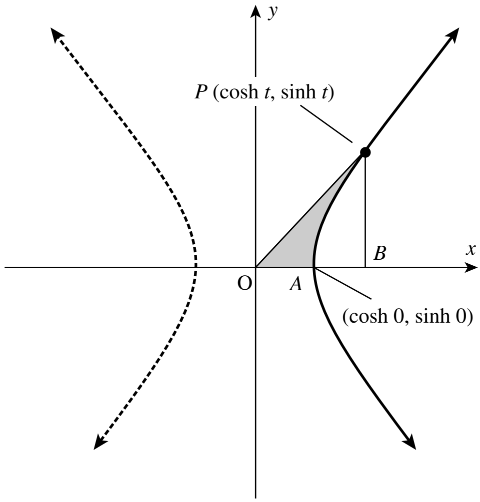


FIGURE 23.15

Now refer to Figure 23.15, where t is any real number. The point $P(\cosh t, \sinh t)$ is on the unit hyperbola because

$$\cosh^2 t - \sinh^2 t = 1.$$

Observe that, *because cosh t is never less than 1, all points (cosh t, sinh t) are on the right branch of the hyperbola.* We now show how the areas of the shaded regions in Figures 23.14 and 23.15 are related. We know that *the area of a circular sector of radius r units and a central angle of radian measure t is given by*$(1/2)r^2t$ *square units*. Therefore, the area of the circular sector in Figure 23.14 is $(1/2)t$ square units, since $r = 1$. The sector AOP in Figure 23.15 is the region bounded by the x-axis, the line OP and the arc AP of the unit hyperbola.

Let the area of sector $AOP = A_1$ square units, the area of sector $OBP = A_2$ square units, and the area of sector $ABP = A_3$ square units.

Then,
$$A_1 = A_2 - A_3 \tag{29}$$

From the formula for determining the area of a triangle

$$A_2 = \frac{1}{2}\cosh t \cdot \sinh t \tag{30}$$

We find A_3 by integration

$$A_3 = \int_0^t \sinh u \, d(\cosh u)$$

$$= \int_0^t \sinh^2 u \, du$$

$$= \frac{1}{2}\int_0^t (\cosh 2u - 1)\mathrm{d}u$$

$$= \frac{1}{2}\left(\frac{\sin 2u}{2} - u\right)\Bigg|_0^t$$

$$= \left(\frac{1}{4}\sinh 2u - \frac{1}{2}u\right)\Bigg|_0^t$$

Therefore,

$$A_3 = \frac{1}{2}\cosh t \cdot \sinh t - \frac{1}{2}t$$

$$A_3 = A_2 - \frac{1}{2}t \quad [\text{using (ii)}]$$

$$\therefore \quad A_3 - A_2 = \frac{1}{2}t \quad \text{or} \quad A_1 = \frac{1}{2}t[\because \quad A_3 - A_2 = A_1]$$

Thus, the measure of the area of circular sector *AOP* of Figure 23.14 and the measure of the area of the sector *AOP* of Figure 23.15 is in each case, one-half of the value of the parameter associated with the point *P*. For the unit circle, the parameter *t* is the radian measure of the angle *AOP*. The parameter *t* for the unit hyperbola is not interpreted as the measure of an angle; the term *hyperbolic radian*, however, is sometimes used in connection with *t*.

Note (11): For exercises, refer to standard books.

APPENDIX A (Related To Chapter-2) Elementary Set Theory

A.1 INTRODUCTION

Set theory is the *basis of modern mathematics*. The great German mathematician George Cantor (1845–1918) is regarded as the father of set theory. He developed, utilized, and stressed the concept of sets in the study of mathematics.

The dictionary meanings of the word *set* are *collection*, *class*, *family*, *aggregate*, *group*, and so on. But there can be collections (or sets) that cannot be identified uniquely. For example, consider a set of *rich people*. From a dictionary point of view, it may be acceptable to use the statement *set of rich people* for a group of people who appear to be rich, but from a mathematical point of view we must define a *rich person* so that one can be identified *(without confusion)* whether he (or she) is rich. From this point of view, the word set is not a well-defined term. It is for this reason that *set* is considered to be an *undefined term* in mathematics.

To have a meaningful discussion about sets, it is necessary to be able to identify the collection (or the set) without any confusion. This demands that there must be a rule that should guide us in identifying the elements of the set under consideration so that one can decide whether a given object belongs to the set under consideration. (Such a rule may specify a property which a single object does or does not have).

Thus, from the point of view of mathematics, we agree to say that *a set is a well-defined collection of objects.* (*Note that, we have not attempted to define the word set*).

A few examples of sets are given below:

(i) Set of natural numbers.
(ii) Set of roots of the equation, $x^2 - 7x + 6 = 0$.
(iii) A set consisting of prime minister of India, capital of the United States, natural numbers 1–10, Taj Mahal and alphabets a–c.

In all these collections, we can identify *each object precisely* and hence they represent sets. On the other hand, *honest people, clever students, handsome boys and beautiful girls,* and so on are *relative terms* and it is not possible to identify them for want of their proper definitions. Hence, *they do not form sets in language of mathematics*. Now, we introduce the following terminology to understand the elementary set theory.

Appendix A Elementary Set Theory: (The language of sets as the back-bone of modern mathematics)

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and A. K. Ghosh.

A.2 ELEMENTS OF A SET

The objects that belong to a set are called elements or members of the set. *The elements of a set need not be related to one another in any obvious way, except that they happen to be put together* (see example (iii) above).

A.3 SET NOTATIONS

The sets are usually denoted by the capital letters $A, B, C, D, \ldots, X, Y, Z$ and their elements are denoted by small letters $a, b, c, d, \ldots, x, y, z$.

If a particular element x belongs to set A, we write $x \in A$. If two elements x and y belong to set A, we shall write $x, y \in A$. However, if an element x *does not belong to set B,* we write $x \notin B$. We use *curly brackets* to enclose the elements of a set. For example, consider the set C given below:

$$\begin{aligned} C &= \{\text{All positive even numbers}\} \\ &= \{2, 4, 6, 8, 10, \ldots\}, \text{ here } 8 \in B,\ 5 \notin B \\ &= \{x | x \text{ is a positive even number}\} \end{aligned}$$

The symbol "|" is used to read "such that".

Each element in a set is separated from the other by a comma.

A.4 SPECIFYING SETS

If the elements of a set do not have any property in common, then it becomes necessary to list all the elements of the set. On the other hand, *if the elements of the set have some property in common*, then it is up to our requirement whether to list the elements of the set or else use the other method to identify correctly the elements of the set. Thus, there are two methods of specifying sets:

(i) *Roster Method or Listing Method or Tabulation Method*: In this method, a set is represented by listing all its elements within braces { }, as shown above. Again, set C of vowels will be written as

$$C = \{a, e, i, o, u\}$$

(ii) *Rule Method or Set Builder Method*: In this method, we state one or more characteristic properties of the elements so that one is able to decide whether a given object is an element of the set. Thus, if D is a set such that its elements x satisfy the property $P(x)$, then we write $D = \{x \mid x \text{ satisfies } P(x)\}$.

Examples:

$$\begin{aligned} A &= \{x|x \text{ is even integer}\} = \{\ldots, -4, -2, 0, 2, 4, 6, \ldots\} \\ B &= \{x|x \text{ is odd number less than } 11\} = \{1,2,3,5,7,9\} \\ C &= \text{The set of roots of the equation } x^3 - 6x^2 + 11x - 6 = 0 \\ &= \{x|x^3 - 6x^2 + 11x - 6 = 0\} \end{aligned}$$

Now, since the roots of the equation $x^3 - 6x^2 + 11x - 6 = 0$ are 1, 2, and 3, we may write $C = \{1, 2, 3\}$.

Note: Sets that we use in mathematics are usually collections of numbers, *points, planes, lines*, and so on, and we shall be concerned with such sets only.

A.5 SINGLETON SET (OR UNIT SET)

A set that contains *only one element* is called *singleton*. Thus, $\{a\}$, $\{5\}$ are singleton sets.

Let $A = \{x|x+5=5\}$. Here, from $x+5=5$, we get $x=0$. Thus, *the set A contains only the element 0*. Hence, $A=\{0\}$, which is *singleton set.*

A.6 THE NULL SET OR THE EMPTY SET

It is possible to characterize a set by a property that would permit no objects to be in the set. For example, the set of *all real roots of the polynomial equation,* $x^2+1=0$. *The set that has no members is called the empty set or the null set and it is denoted by the symbol* ϕ or $\{\}$.

Remark: Here, it may be noted that in different contexts, there can be different null sets. Hence, all null sets are not the same. However, all null sets are denoted by the same symbol. Since there is only one null set (in each context), we call it *the null set* instead of a null set.

Note: We must distinguish between ϕ and $\{\phi\}$. Although set ϕ is a *null set* (i.e., it contains no element), the set $\{\phi\}$ is a *singleton* whose one element is the empty set ϕ. Similarly, $\{0\}$ is a *singleton*.

There are *many relations among sets*, as given below.

A.7 THE CARDINAL NUMBER OF A SET

If a set *A* contains finite number of elements *n*, we denote the *cardinal number* of set *A* by $n(A)$. In other words, $n(A)$ stands for the number of elements in a finite set.

Examples: Consider the following sets:

(i) $A=\{1, 3, 5, 7, 9, 11, 13\}$
(ii) $B=\{a, e, i, o, u\}$
(iii) $C=\{2, 3, 5, 7, 11, 13, 17, 19\}$

Thus, $n(A)=7$, $n(B)=5$, and $n(C)=8$ represent the cardinal numbers of the above sets.

In Chapter 2, we have already introduced the concept of infinite sets (both *countable* and *uncountable*). (*A set is infinite if it is not finite*.) Some examples of infinite sets are as follows:

(i) $N=\{1, 2, 3, 4, \ldots\}$.
(ii) $D=\{1, 4, 9, 16, \ldots\}$.
(iii) $R=$ Set of all real numbers.
(iv) $S=$ Set of all points in a plane.

The cardinal number of *countable infinite set*, $N=\{1, 2, 3, 4, \ldots\}$ (or any other set that is equivalent to *N*), is denoted by the symbol $\aleph_0$ (read aleph-null). The symbol "c" is used to denote the cardinal number of an uncountable infinite set, like the set of all real numbers or the set of points in an open interval or the set of points in a plane.

Remark: The cardinal number of the empty set is zero. We write $n(\phi) = 0$.

A.8 SUBSET OF A SET

If two sets A and B are such that *every* element of A is also an element of set B, then A is called a subset of B. Thus, set A is a subset of B if

$$x \in A \Rightarrow x \in B$$

(Here, the symbol $\Rightarrow$ stands for *implies that*.)

Symbolically, we write this relationship as $A \subseteq B$ and read as A is a subset of B or A is contained in B.[1]

Example: Let $A = \{1, 2, 3\}$ and $B = \{1, 2, 3, 4\}$. Then clearly $A \subseteq B$. Again, let $D = \{2, 4, 6, 8\}$ and $E = \{2, 8, 4, 6\}$. Then, $D \subseteq E$. Also note that $E \subseteq D$.

Remark: Observe that in the term subset, the possibility that both the sets may be equal is included. Thus, every set is a subset of itself.

If set A is *not* a subset of set B, we write $A \nsubseteq B$.

A.9 EQUALITY OF SETS

Definition (1): We say that *two sets are equal if they contain precisely the same elements.*

Again in view of the definition of *subset of a set* (which includes the possibility of their equality), we give the following definition.

Definition (2): Two sets A and B are equal iff $A \subseteq B$ and $B \subseteq A$.

In other words, we say that $A = B$, if every element of set A is an element of set B and every element of set B is an element of set A. The equality of sets A and B may be written in the following symbolic form:

$$x \in A \Leftrightarrow x \in B.$$

The symbol $\Leftrightarrow$ stands for "implies and is implied by" or "if and only if" or "iff".

Note: Importance of Definition (2)

We draw the following two important conclusions from Definition (2):

(i) A set does not change if we change the order in which its elements are tabulated.

(ii) A set does not change if one or more of its elements are repeated.

[1] If $A \subseteq B$, then B is called the superset of A and symbolically we write it as $B \supseteq A$ and read it as B is a superset of A or B contains A.

For example, consider the following sets:

$$A = \{1, 7, 3, 2\} \quad B = \{7, 3, 1, 2, 3\}$$

According to Definition (2), we have $A = B$.

(In general, we never write a set in which its elements are repeated.)

A.10 PROPER SUBSET

Consider a set A that is a subset of set B (i.e., $A \subseteq B$).

If there is at least one element of B *that* is not in A, then A is called a *proper subset* of B, and we write $A \subset B$.

Example: If $A = \{1, 2, 3\}$ and $B = \{1, 2, 3, 5, 6\}$, then $A \subset B$. If A is not a proper subset of B, we write $A \not\subset B$. This will be the situation when there is at least one element $x \in A$, but $x \notin B$. For example, consider $A = \{1, 2, 3\}$ and

$B = \{1, 2, 4, 5, 6\}$. Here, $3 \in A$ but $3 \notin B$. Hence, $A \not\subset B$.

Remark: The null set ϕ is taken as a subset of every set. Thus, every set has at least two subsets: the set itself and the null set.

A.11 COMPARABILITY OF SETS

Two sets can be compared if one of them is subset of the other. Two sets A and B are said to be comparable if $A \subseteq B$ or $B \subseteq A$. If $A \not\subseteq B$ or $B \not\subseteq A$, then A and B are said to be *noncomparable* or *incomparable*.

Example: The sets $A = \{1, 2\}$ and $B = \{1, 2, 3\}$ are *comparable* as $A \subseteq B$. On the other hand, the sets $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$ are *incomparable*.

(The symbols $\subseteq$ and $\supseteq$ in set theory may be compared with the order relations $\leq$ and $\geq$ in arithmetic.)

A.12 SET OF SETS

A set may itself be sometimes an element of another set. A set whose elements are set(s) is called *set of sets*. For example, $A = \{\phi, \{1, 2\}, \{3\}\}$ is a set of sets. An important set of sets is the *power set* defined below.

A.13 POWER SET

If S is any set, then *the family of all the subsets of S* is called the *power set of S* and denoted by $P(S)$.

Example: Let $S = \{1, 2, 3\}$. Then, $P(S) = \{\phi, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$.

It can be shown that if a set S has n elements, then $P(S)$ has 2^n elements.

Remark: The power set of ϕ is $\{\phi\}$ and it has $2^0 = 1$ element. The power set of $\{0\}$ is $\{\{0\}, \phi\}$ and it has $2^1 = 2$ elements.

A.14 UNIVERSAL SET U

In any discussion about sets, all the sets under consideration are to be the subsets of a particular set. Such a set is called the *universal set* or *universe of discourse*. It is denoted by U.

Examples:

(i) In the discussion concerning the set of odd numbers, the set of even numbers, the set of prime numbers, the set of composite numbers, the set of factors, and so on, *the universal set is the set of natural numbers*.

(ii) The set of all real numbers is the *universal set* in the discussion of *subsets of rational and irrational numbers*.

Explanation: The statement *there is no number whose square is 8* is valid if the discussion is limited to the *set of integers or the set of rational numbers*, but it is *invalid if the universal set is the set of all real numbers*.

A.15 OPERATIONS ON SETS

In arithmetic, the *elementary operations* of *addition*, *subtraction*, *multiplication*, and *division* are used *to make new numbers out of old numbers*, that is, to combine two numbers to create third. Similarly, in elementary set theory, there are *binary operations on sets* that generate new sets.(2)

Remark: It will be noted that these operations have many of the algebraic properties of *ordinary addition* and *multiplication of numbers*, although conceptually these operations are quite different from those of numbers.

A.16 THE UNION (LOGICAL SUM) OF TWO SETS *A* AND *B*

It is the set consisting of precisely those elements that belong to either A or B or both A and B. In symbols, $A \cup B = \{x|x \in A \text{ or } B \text{ or both}\}$.

Example: Let $A = \{1, 2, 3\}$ and $B = \{2, 3, 5, 6\}$. Then $A \cup B = \{1, 2, 3, 5, 6\}$. Obviously, then $A \cup \phi = A$. Similarly, if $C = \{7, 8, 9\}$ and $D = \{7, 8, 9, 10\}$, then $C \cup D = D$.

(2) A binary operation defined on sets is a rule, affecting every two sets A and B, that states the manner in which a third set C is to be derived from A and B. The set C is usually (but not always) different from A and B.

The binary operations (on sets) are *union, intersection, and complementation*, which correspond, more or less, to the arithmetic operations of *addition, multiplication, and subtraction*, respectively.

A.17 THE INTERSECTION (LOGICAL PRODUCT) OF TWO SETS *A* AND *B*

It is the set of all those elements *that belong to both* A and B.

In symbols, $A \cap B = \{x|x \in A \text{ and } x \in B\}$.

Examples: Let $A = \{1, 2, 3\}$, $B = \{2, 3, 5, 6\}$, and $C = \{5, 7, 8\}$. Then, $A \cap B = \{2, 3\}$, $B \cap C = \{5\}$, and $A \cap C = \phi$.

A.18 DISJOINT SETS

Two sets A and B are said to be *disjoint sets* if they do not have any element in common.

Let $A = \{1, 2, 3\}$ and $B = \{5, 8, 13\}$. Then, A and B are disjoint sets.

Note: If sets A and B are *disjoint*, then $A \cap B = \phi$, and conversely.

A.19 DIFFERENCE OF TWO SETS *A* AND *B*

The difference of two sets A and B in that order is the set of elements that belong to A but that do not belong to B. We denote the difference of A and B by the set A – B, and read it as "A difference B."

Thus, $A - B = \{x|x \in A, \text{ but } x \notin B\}$.

Example: Let $A = \{1, 2, 3, 4\}$ and $B = \{1, 3, 5, 6,\}$. Then, $A - B = \{2, 4\}$. Also note that $B - A = \{5, 6\}$. Thus, $A - B \neq B - A$.

Remark: In the definition of (A – B), it is not necessary that B should be a subset of A. Thus, A – B is the set of those elements of A that are not in B.

A.20 COMPLEMENT OF A SET

If we consider the difference of sets U and A (where U is the universal set), then this difference is denoted by A' or A^c and it is called the complement of A in U.

Thus, the complement of a given set A (with respect to the universal set U) is the difference of the universal set U and A, in that order, and is denoted by A' or A^c. We write

$$A' = U - A = \{x|x \in U, \text{ but } x \notin A\}$$

$$\text{Clearly, } (A')' = U - A' = U - (U - A) = A$$

$$\phi' = U - \phi = U$$

$$\text{Acordingly, } U' = U - U = \phi$$

Example: Let $N = \{1, 2, 3, 4, \ldots\} = U$ and $A = \{1, 3, 5, 7, \ldots\}$, then $A' = \{2, 4, 6, 8, \ldots\}$.

Remark: Complement of a set is basically the difference of the two sets, of which first set is the universal set U and the other set is a proper subset of U.

APPENDIX B (Related To Chapter-4)

B.1 INTRODUCTION

We know that coordinate geometry (in two variables) deals with the *study of geometric objects* (i.e., points, lines, curves, and areas) *in a plane using algebra*. We are familiar with the representation of real numbers on the *number line*. It was the French philosopher and mathematician Rene Descartes (1596–1650) who introduced the analytic approach in the study of geometry by using algebra. This was achieved by representing points in the plane by ordered pairs of real numbers, called Cartesian coordinates, named after Rene Descartes.[(1)]

B.1.1

It is important to understand how the introduction of Cartesian coordinates allows us to *use numbers and their arithmetic as a tool in studying geometry*. It is also important to remember that *this coordinate system allows us to draw the geometric pictures of algebraic equations that illustrate a great deal of numerical work.*[(2)]

In Chapter 4, we have studied that the *algebraic equations represent lines and curves*. There we also introduced the concept of inclination of a line and its relation with the slope of the line. (We know that inclination of a line relates trigonometric functions with the slope of the line).

In fact, the concept of *slope* is one of the *central concepts* in calculus. In our study of calculus, an important concept to be learnt is the slope of a curve at any point on it. For this purpose, we extend the concept of slope of a line and use it to define the slope of a curve at a point by applying the concept of limit. It will be observed that the subject of calculus is dominated all throughout by the concept of *slope of a curve at a point.*

Of course, at this stage, it is difficult to visualize: *How the slope of the curve can be defined*? It is reasonable to think of a tangent line at a point of the curve and take the slope of this tangent line as the slope of the curve at that point. This is exactly what is done. But, to give the definition of a *tangent to a curve at a point is not simple*. It demands the knowledge of limit concept (which is introduced in Chapter 7a), and subsequently applied in defining a tangent to a curve in Chapter 9 that deals with the concept of derivatives.

Curves represented by second-degree algebraic equations in two variables and their identification: translation of axes

(1) Another French mathematician Pierre de Fermat (1601–1665) is also credited with the invention of coordinate geometry. His work was known after his death. Both Descartes and Fermat introduced two perpendicular lines called axes and agreed to represent any point P in the plane by an ordered pair (x, y), and denoted the point P as $P(x, y)$. In this notation, x and y represent the directed distances (or signed distances) from the y- and x-axes, respectively.

(2) Here, it may be mentioned that, every equation need not represent a curve. For example, the equation $x^2 + y^2 = -5$ does not represent a curve.

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and A. K. Ghosh.

Now, it must be clear that the graph of an algebraic equation is the geometric picture of the equation. Thus, calculus can help us in studying the properties of curves (represented by equations) by operating on the given equations. This indicates how coordinate geometry plays an important role in the foundation and development of calculus.[(3)]

B.1.2 From Lines to the Curves - The Conic Sections

When we speak of an equation of a line l, we mean an equation in the form

$$y = mx + b$$

where m is the slope of the line and b stands for *its intercept with the y-axis.*

This is an equation of degree one in two variables. Also, we have studied other useful forms of the equation of a line. The general equation of first degree in x and y is given by

$$Ax + By + C = 0$$

where A and B are *not* zero, *simultaneously*. (This equation covers all the lines, including vertical lines.)

B.2

Now, we shall study the *curves represented by second degree equations in two variables.* Our interest lies in identifying those second degree equations that represent conic sections. The most general equation of second degree in two variables is

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \text{ (where } A \neq 0,\ B \neq 0) \tag{1}$$

(Note that, irrespective of whether $C = 0$ or $C \neq 0$, Equation (1) will remain *second degree* equation *in two variables.*)

Our interest lies in the special case of the Equation (1), *which does not contain the term Bxy* (i.e., $B = 0$). Thus, it remains to consider only those second degree equations that have *just one second degree term* and those that contain *two second degree terms.*

B.2.1 Equations with Just One Second Degree Term

Such equations represent curves known as *parabolas*. It will also be seen that there are certain limiting forms (of such equations) that do not represent parabola(s), but something different.[(4)]

[(3)] The subject of coordinate geometry is very vast in itself and must be studied separately. We will discuss here only the necessary parts of the subject needed for our purpose.
[(4)] Note that, the following equations represent parabolas:
(i) $y^2 = 8x$ (ii) $x^2 = 12y$ (iii) $x^2 = 4ay,\ a \neq 0$ (iv) $y^2 + 2x - 4y + 3 = 0$ (v) $y^2 + 2x + 3 = 0$
(vi) $y = x^2 - 2x + 3$ (vii) $x^2 = y + 3$
However, the following equations do not represent parabolas:
(a) $y^2 = 4$ (It represents a pair of parallel lines.)
(b) $y^2 = 0$ (It represents a pair of coincident lines, that is, a single line.)
(c) $y^2 = -1$ (It represents an empty set.)

B.2.2 Curves Represented by Equations That Have Two Second Degree Terms

Circles, ellipses, and hyperbolas are the curves whose equations have two second degree terms. Thus, we will be considering the algebraic equations of the form

$$Ax^2 + Cy^2 + Dx + Ey + F = 0 \tag{2}$$

where A and C both are not zero. (Observe that if $A = 0$ or $C = 0$, then Equation (2) will reduce to the equation of a parabola.) Again, it will be noted that there are certain limiting forms of Equation (2), which represent something different, other than the curves mentioned above.

Note: Parabola(s), ellipse(s), and hyperbola(s) are known as conic sections (or more commonly conics), because they can be obtained by the intersection of a double napped right circular cone by a plane.

B.3 THE IDEA OF A DOUBLE NAPPED RIGHT CIRCULAR CONE AND CONICS

Let l be a fixed *vertical line* and m be another line intersecting it at a fixed point V and inclined to it at an angle α (Figure B.1).

Suppose we rotate the line m around the line l in such a way that the *angle α remains constant*, then the surface so generated is a double napped right circular hollow cone. From now on, it will be referred to as the *cone, extending indefinitely far* in both directions (Figure B.2).

Definitions:

- The point V is called the *vertex*.
- The line l is the *axis of the cone*.

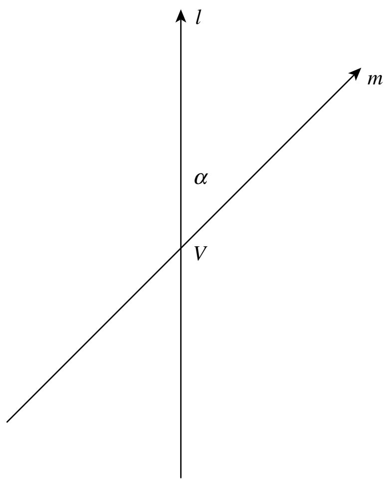


FIGURE B.1

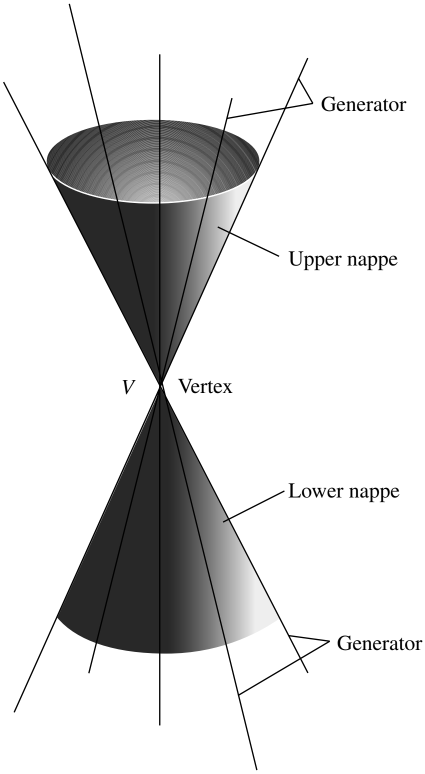


FIGURE B.2

- The rotating line m is called a *generator of the cone*.
- The *vertex* separates the cone into two parts called *nappes*.

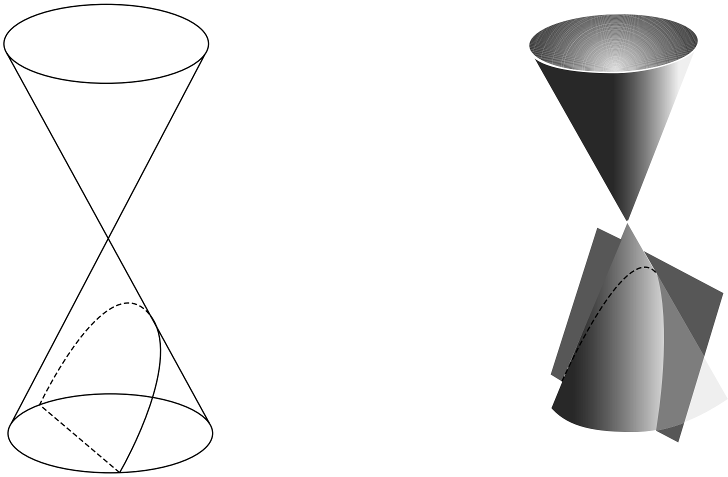
Remark: Note that a *generator* of the cone is a *line lying in the cone and that all generators of a cone contain the point V*.

B.4 CONIC SECTION: DEFINITIONS

If we take the section of a cone by a plane, then the points common to the plane and the cone form the *conic section* (or the conic). The conic sections are classified according to the different positions of the plane with respect to the cone. It will be seen that though a point, a pair of coincident lines, a pair of intersecting lines, and a circle represent conic sections, they are treated separately. A conic section generally refers to a parabola, an ellipse, or a hyperbola. We start with the parabola.

(a) **Parabola:** If the *cutting plane is parallel to the generator*, the section is a parabola (Figure B.3).

(b) **Ellipse:** An *ellipse* is obtained as a conic section if the *cutting plane is parallel to no generator*, in which case the cutting plane intersects each generator (Figure B.4).



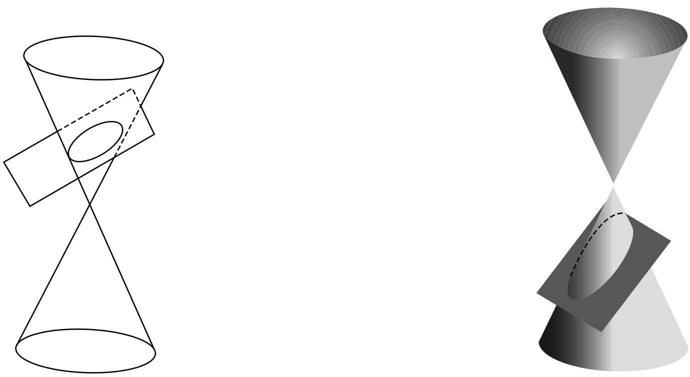
(*a*) (*b*)

FIGURE B.3 Parabola.

Note: A circle is a special case of the ellipse. A circle is formed *if the cutting plane intersecting each generator is also perpendicular to the axis of the cone* (Figure B.5).

(Though a circle represents a conic section, it is not studied under conics. It is treated separately.)

Note: The intersection of a cone with the cutting plane may take place either at the vertex of the cone or at any other part of the nappe, below or above the vertex. If the intersection is *a circle*, *an ellipse*, or *a parabola, the plane cuts entirely across one nappe of the cone.* What happens if *a plane intersects both the nappes?*



(*a*) (*b*)

FIGURE B.4 Ellipse.

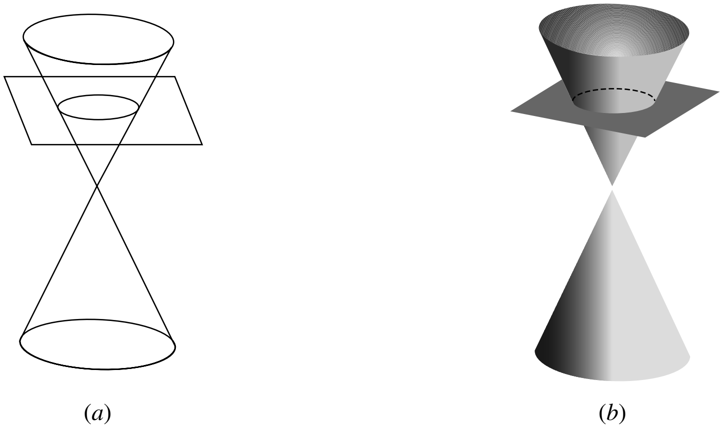


FIGURE B.5 Circle.

(c) **Hyperbola:** When the cutting plane intersects *both nappes of a cone, the conic section obtained is a hyperbola.*

In this case, the cutting plane is parallel to the axis of the cone (Figure B.6).

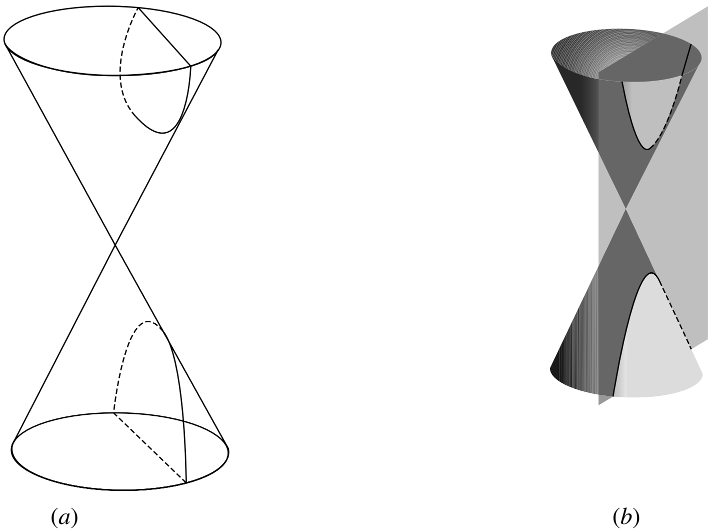
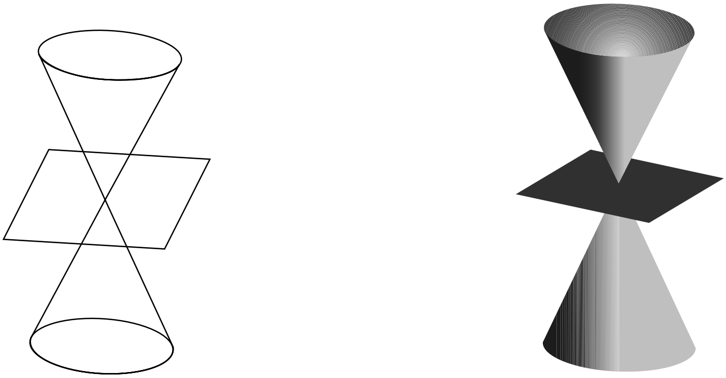


FIGURE B.6 Hyperbola.

B.4.1 Degenerated Conic Sections

When the plane cuts at the vertex of the cone, we have the following degenerate cases of conic sections:

(i) The degenerate case of an ellipse, *a point*, is obtained as a conic section if the *cutting plane contains the vertex but does not contain a generator* (Figure B.7).



(*a*) (*b*)

FIGURE B.7 Point.

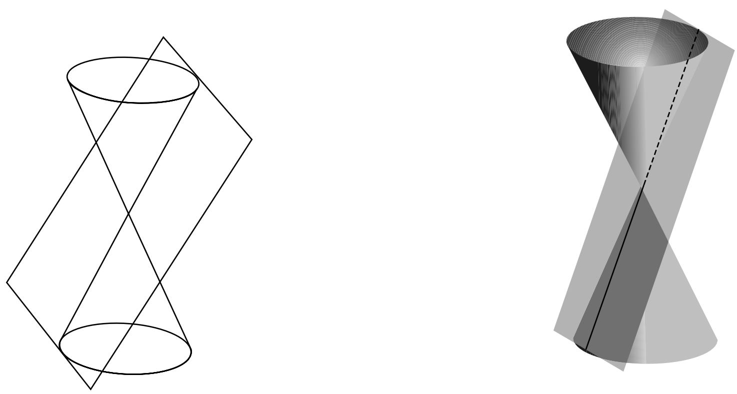
(ii) *A pair of coincident lines*, if the plane *passes through the vertex and contains a generator* (i.e., the plane touches the cone). It is the *degenerated case of a parabola* (Figure B.8).

(iii) If the cutting plane contains the *vertex* of the cone and *two generators*, we obtain the degenerate case of a hyperbola in the form of two intersecting lines (Figure B.9).

Remark: Though a point, a pair of coincident lines, a pair of intersecting lines, and a circle represent conic sections, they are treated separately. A conic section generally refers to *a parabola*, *an ellipse*, or *a hyperbola*.

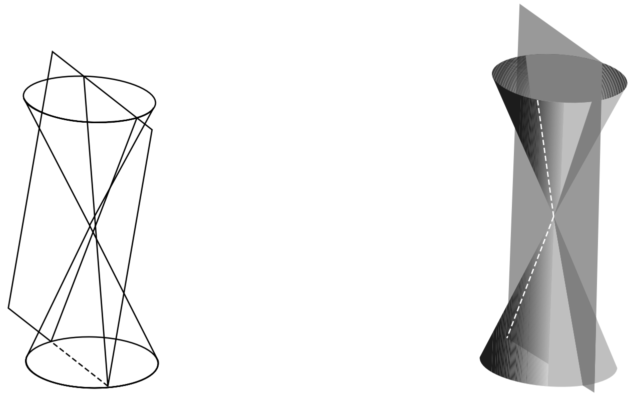
B.4.2 Importance of Conic Sections

The study of properties of conics is very important in geometry, mechanics, physics, and astronomy, such as design of telescopes and antennas, reflectors in flash lights, and automobile headlights.



(*a*) (*b*)

FIGURE B.8 Pair of coincident lines.



(*a*) (*b*)

FIGURE B.9 Two intersecting lines.

The path of a projectile is a *parabola*, if motion is considered to be in a plane and air resistance is neglected. All the planets, namely, Mercury, Venus, Earth, Mars, and others, move around the sun in *elliptical orbits* with the sun at a focus. Indeed, these curves are important tools for the present-day *exploration of outer space* and also for research into the behavior of atomic particles.

We know that the conic sections are plane curves. (Why?) Therefore, it is desirable to use *equivalent definitions* that refer only to the plane in which the curve lies and refer to *special points* and *lines* in this plane, called *foci* (plural of focus), and *directrix*.

B.5 CONICS

Now we define a conic.

Definition: Suppose l is a *fixed line* and F (or S) is *a fixed point* not on the line. Then, the locus of a point P (*in the plane of* l *and* $\underline{F}$) such that the distance of P from the fixed point F (or S) has a *fixed ratio* e to its *distance from the fixed line* l *is called a conic* (or a conic section). By such a distance of P from l, we mean the *length of the perpendicular line segment from* P *to the line* l (Figure B.10).The fixed point is called the *focus*, the fixed line is called the *directrix*, and the fixed ratio is called the *eccentricity* of the conic.

If $e < 1$, the conic is called an *ellipse.*

If $e = 1$, the conic is called a *parabola.*

If $e > 1$, the conic is called a *hyperbola.*

The property defining the conic is called the *focus–directrix property*. It should also be remembered that *para* means equality, *ellipsis* means deficiency, and hyper means *excess*.

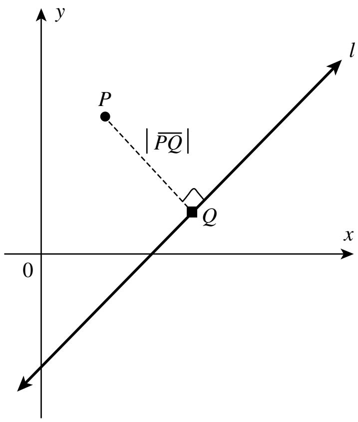


FIGURE B.10

B.6

Now, we shall follow a (*special*) *method* in defining the *standard equations* of conics (i.e., *a parabola*, an *ellipse*, and a *hyperbola*) to get their equations in the simplest form. In this method, we choose the *axes* (not the curves) *in a special way* so that the equation of each conic section is as simple as possible. (Even if we do not choose the origin and the axes conveniently, we would still get the equation(s) of the curves, but they would not be as simple.)

B.6.1 Parabola

Definition: A parabola is the set of *all points in a plane that are equidistant from a fixed line and a fixed point* (not on the line) *in the plane*.

The fixed line is called the *directrix of the parabola* and the fixed point F is called the *focus* (Figures B.11 and B.12).

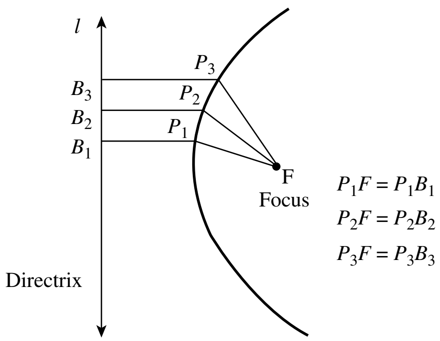


FIGURE B.11

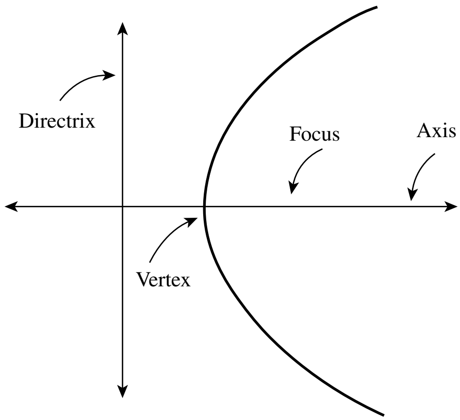


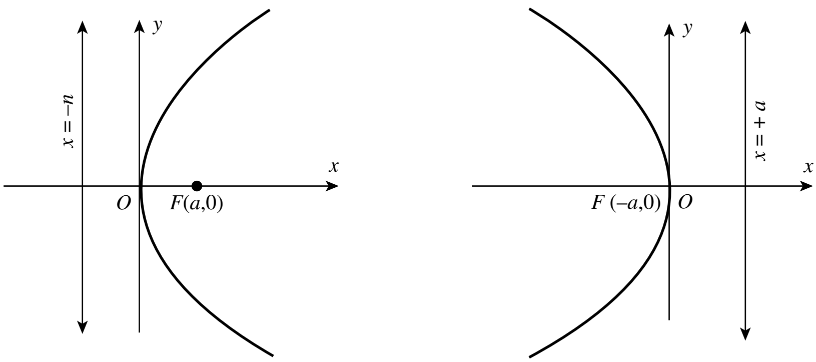
FIGURE B.12

- A line though the focus and perpendicular to the *directrix* is called *the axis of the parabola*.
- The point of intersection of parabola with the axis is called the *vertex of the parabola*.

B.6.1.1 Standard Equations of Parabola The equation of a *parabola* is simplest if the *vertex is at the origin* and the axis of symmetry *is along the x-axis* or *y-axis*. The four possible such orientations of parabola are shown in Figure B.13.

In each figure, F stands for the *focus*.

Note: Here, we do not give the proof of the equations of any conic. For this purpose, any standard book on coordinate geometry may be referred to.



$y^2 = 4ax$ $(a > 0)$ $y^2 = -4ax$ $(a > 0)$

(a) (b)

FIGURE B.13 (a) $y^2 = 4ax$, $(a > 0)$. (b) $y^2 = -4ax$, $(a > 0)$. (c) $x^2 = 4ay$, $(a > 0)$. (d) $x^2 = -4ay$, $(a > 0)$.

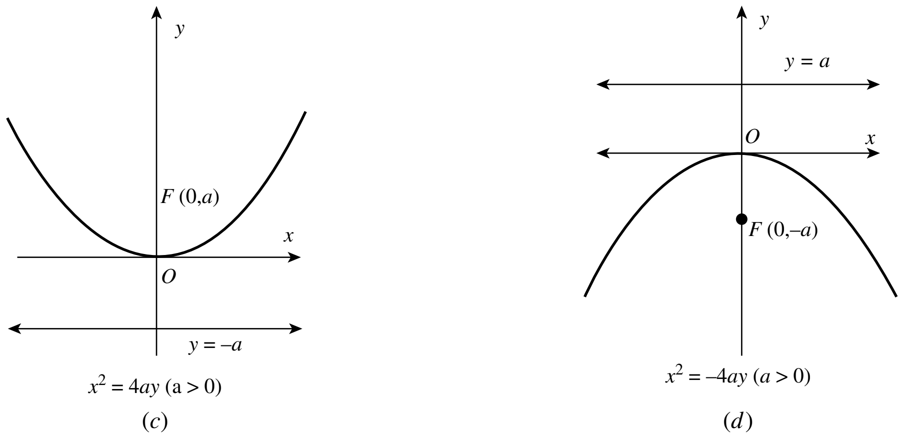


FIGURE B.13 Continued

Note:

(1) *The standard equations of parabolas* have focus on one of the coordinate axes, vertex at the origin, and the directrix is parallel to the other coordinate axis.

(2) From the standard equations of parabolas, we have the following observations:

- A parabola is symmetric with respect to the axis of the parabola. If the equation has a y^2 term, then the axis of symmetry is along the x-axis and if the equation has a x^2 term, then the axis of symmetry is along the y-axis.
- When the axis of symmetry is along the x-axis, the parabola opens
  (a) to the right, if the coefficient of x is positive; and
  (b) to the left, if the coefficient of x is negative.
- When the axis of symmetry is along the y-axis, the parabola opens
  (c) upward, if the coefficient of y is positive, and
  (d) downward, if the coefficient of y is negative.

B.6.1.2 Latus Rectum

Definition: Latus rectum of parabola is a line segment perpendicular to the axis of the parabola through the focus and whose end points lie on the parabola.(Figure B.14).

(It can be easily checked that the length of the latus rectum of the parabola, $y^2 = 4ax$, is $4a$.)

B.6.2 Ellipse

Definition: An ellipse is the set of all points in a plane, the *sum of whose distances* from the *two fixed points* in the plane is a constant.

The *two fixed points* are called the *foci* (plural of focus) *of the ellipse* (Figure B.15).

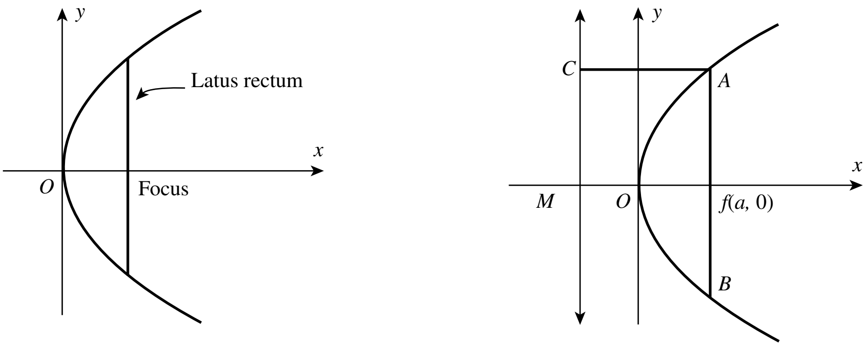


FIGURE B.14

Note: *The constant that* is the *sum of the distances of a point* on the ellipse from the two fixed points is *always greater than the distance between the two fixed points.*

- The *midpoint* of the line segment joining the *foci* is called the *center of the ellipse*.
- The *line segment through the foci* of the ellipse is called the *major axis* and the *line segment through the center and perpendicular to the major axis* is called the *minor axis* (Figure B.16a).
- The *end points* of the major axis are called the *vertices of the ellipse*. (Figure B.16a).
- We denote the length of the *major axis* by $2a$, the length of the *minor axis* by $2b$, and the distance between the foci by $2c$. Thus, the length of the semimajor axis is a and semiminor axis is b (Figure B.16b).

B.6.2.1 Relationship Between Semimajor and Semiminor Axes The relationship between semimajor axis a and semiminor axis b ($a > b$) and the distance of the focus from the center of the ellipse c is shown in Figure B.17.

From this figure, it is easy to show that

$$a = \sqrt{b^2 + c^2} \quad \text{and} \quad c = \sqrt{a^2 + b^2}$$

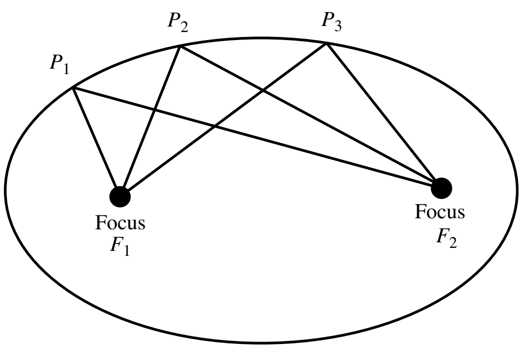


FIGURE B.15 $P_1F_1 + P_1F_2 = P_2F_1 + P_2F_2 = P_3F_1 + P_3F_2$

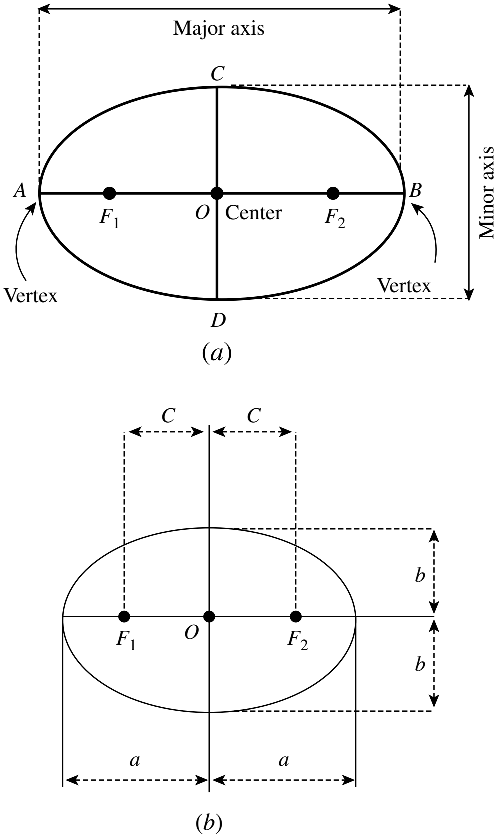


FIGURE B.16

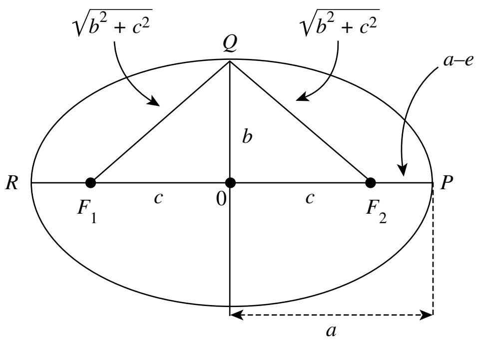


FIGURE B.17

B.6.2.2 Special Cases of an Ellipse From the relation connecting a, b, and c, we have the equation $c^2 = a^2 - b^2$.

If we keep a fixed and vary c from O to a, the resulting ellipse will vary in shape (see Figure B.17). Two cases arise:

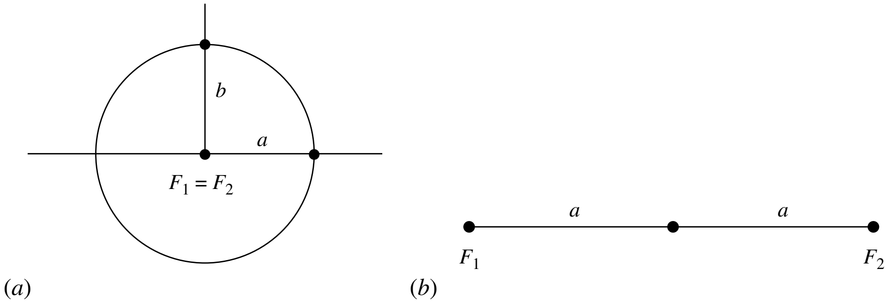


FIGURE B.18

Case (1): When $c = 0$, both foci merge together with the center of the ellipse and $a^2 = b^2$ (i.e., $a = b$) and so the ellipse becomes the *circle*.

Thus, a circle is a special case of an ellipse (Figure B.18a).

Case (2): When $c = a$, then $b = 0$.

The ellipse reduces to *line segment* F_1F_2 joining the two *foci* (Figure B.18b)

B.6.2.3 Standard Equations of an Ellipse The equation of an ellipse is simplest if the center of the ellipse is the origin.

The two such possible orientations are shown in Figure B.19.

Note: The *standard equations* of ellipses have center at the origin and *the major and minor axes are coordinate axes.*

From the standard equations of ellipses (Figure B.19a and b), we have the following observations:

(i) Ellipse is *symmetric with respect to both the coordinate axes* (and origin) since if (x, y) is a point on the ellipse, then $(-x, y)$, $(x, -y)$, and $(-x, -y)$ are also points on the

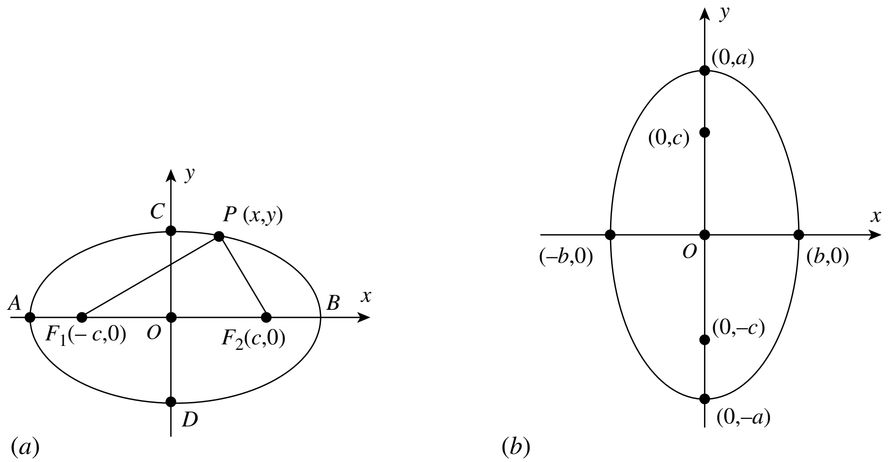


FIGURE B.19 (a) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$. (b) $\dfrac{x^2}{b^2} + \dfrac{y^2}{a^2} = 1$

ellipse. (We have used this property of symmetry in Chapter 8a of Part II, where we compute the area enclosed by an ellipse.)

(ii) The foci always lie on the major axes. The major axis can be determined by finding the intercepts on the axes of symmetry. That is, major axis is along the *x-axis* if the coefficient of x^2 has the larger denominator and is along the *y-axis* if the coefficient of y^2 has the larger denominator.

B.6.2.4 Latus Rectum

Definition: Latus rectum of an ellipse is a *line segment perpendicular to the major axis through any of the foci and whose end points lie on the ellipse* (Figure B.20).

It is easy to show that the *length of the latus rectum* of the ellipse
$(x^2/a^2) + (y^2/b^2) = 1$ is (b^2/a). (Recall that $b =$ semiminor axis and $a =$ semimajor axis.)

B.6.3 Hyperbola

Definition: A hyperbola is the set of all points in a plane, the difference of whose distances from *two fixed points* in the plane is a constant.

- The term *difference* that is used in the definition means the *distance to the farther point minus the distance to the closer point.*
- *The fixed points* are called the *foci* of the hyperbola.
- The *midpoint* of the line segment joining the foci is called the *center of the hyperbola*.
- The line through the foci is called the *transverse axis* and the line through the center and perpendicular to the transverse axis is called the *conjugate axis* (Figure B.21).
- The points at which the hyperbola intersects the transverse axis are called the *vertices of the hyperbola.*

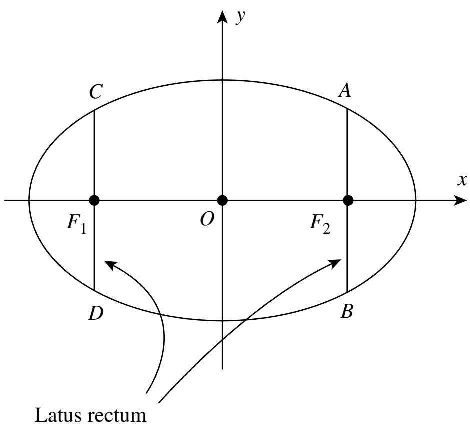


FIGURE B.20

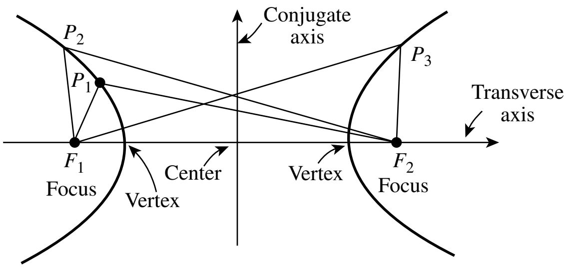


FIGURE B.21 $P_1F_2 - P_1F_1 = P_2F_2 - P_2F_1 = P_3F_1 - P_3F_2$

- We denote the distance between the two foci by $2c$, the distance between two vertices (i.e., length of the transverse axis) by $2a$, and we define the quantity b as $b = \sqrt{c^2 - a^2}$. Also, $2b$ is the length of the conjugate axis. (Figure B.22).

B.6.3.1 Standard Equations of Hyperbola The equation of a hyperbola is *simplest* if the *center* of the hyperbola is at the *origin* and the *foci are on the x-axis or y-axis.*

The two such possible orientations are shown in Figure B.23.

Note: The standard equations of hyperbolas have transverse and conjugate axes as the coordinate axes and the center at the origin.

From the standard equations of hyperbolas (Figure B.23a and b), we have the following observations:

(i) Hyperbola is *symmetric* with respect to both the axes (and origin), since if (x, y) is a point on the hyperbola, then $(-x, y)$, $(x, -y)$, and $(-x, -y)$ are also points on the hyperbola.

(ii) The foci are always on the transverse axis. It is the *positive term* whose *denominator* gives the transverse axis.

For instance, $(x^2/9) - (y^2/16) = 1$ has *transverse axis* along *x-axis of length* 6 units (since $a = 3$ units), while $(y^2/9) - (x^2/16) = 1$ has *transverse axis* along *y-axis* of length 10 units.

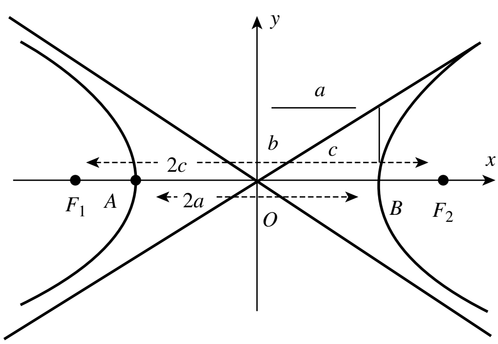


FIGURE B.22

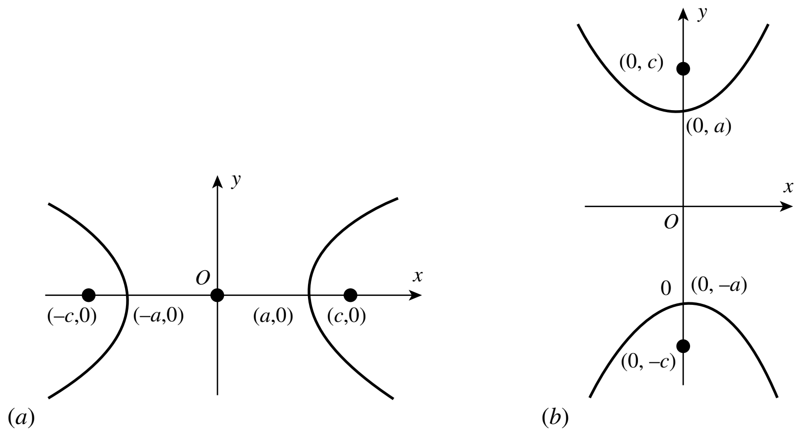


FIGURE B.23 (a) $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$. (b) $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$.

B.6.3.2 Latus Rectum (of Hyperbola) *Latus rectum* of hyperbola is a line segment *perpendicular* to the transverse axis *through any of the foci* and whose end points *lie on the hyperbola.*

It is easy to show that the length of the latus rectum in a hyperbola is $2b^2/a$.

B.7 TRANSLATION OF AXES (OR SHIFT OF ORIGIN)

The *shape of a graph is not changed* by the position of the coordinate axes, but its *equation is changed.* Graphing an equation is frequently made easier by changing from one set of axes to another. Since we may *select* the coordinate axes *as* we please, we generally do so in such a way that the equations will be as simple as possible.

Consider an equation that is given with reference to a set of axes. We may wish to find a simpler equation of its graphs (or we may wish to find if the given equation represents a known curve. In particular, we will be interested to identify the given equation as a conic or its degenerate form). If these different axes are chosen parallel to the given ones, we say that there has been a translation of axes.

Definition: In the Cartesian coordinate system, if we shift the origin to a new point, in the same plane, and take the new axes parallel to the original axes, through this new point, then we say that the new axes are obtained from the old axes, by *translation.*

When we choose new axes in the plane by translation, every point will have *two sets of coordinates*, the *old ones* (x, y) relative to the x- and y-axes and the *new ones* (u, v), relative to the *new axes,* say the u- and v-axes. It is proposed to obtain the relation between the coordinates of a point in two systems of coordinate axes. (It is logical and convenient to assume that in both the sets of axes, the positive numbers lie on the same side of the origin.)

Let (h, k) be the *old coordinates* of the new origin (Figure B.24). By inspection, we see that,

$$x = u + h \quad \text{and} \quad y = v + k \tag{i}$$

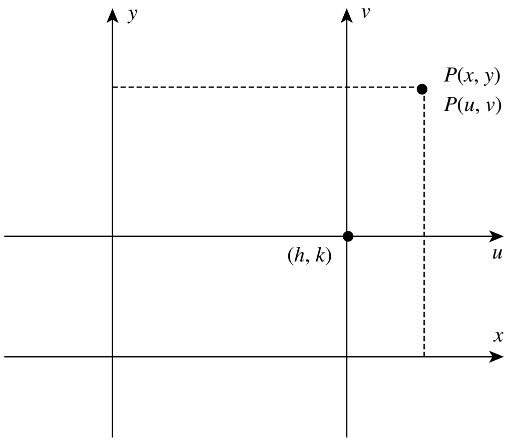


FIGURE B.24

or equivalently

$$u = x - h \quad \text{and} \quad v = y - k \tag{ii}$$

Equations (i) and (ii) are called the equations of transformation. So, if the origin is shifted to (h, k) from (0, 0) (in the old set of axes), the *new coordinates of a point* $P(x, y)$ *will be* $(x - h, y - k)$.

Example (1): Let us find the new coordinates of $P(-6, 5)$ after a translation of axes to a new origin at (2, −4).

Solution: Here $h = 2$ and $k = -4$. It follows that

$$u = x - h = -6 - 2 = -8$$
$$v = y - k = 5 - (-4) = 9$$

Hence, the new coordinates are (−8, 9).

Example (2): The origin is shifted to the points (2, 1). Obtain the equation of the curve in the new frame whose equation in the original form is given by

$$x^2 + y^2 - 4x - 2y - 20 = 0.$$

Solution: Let the new coordinates of a point be (u, v). The equations of transformation are $x = u + h$, $y = v + k$. Here, $(h, k) = (2, 1)$

$$\therefore\ x = u + 2 \quad \text{and} \quad y = v + 1$$

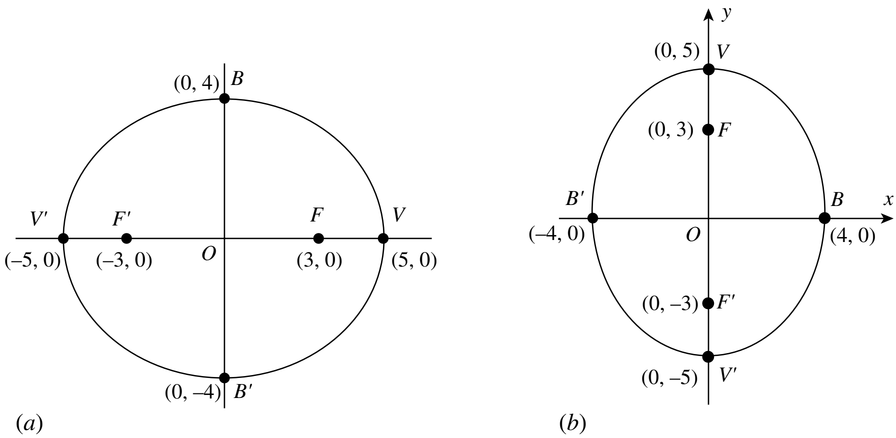


FIGURE B.25 (a) $\dfrac{x^2}{25} + \dfrac{y^2}{16} = 1$. (b) $\dfrac{x^2}{16} + \dfrac{y^2}{25} = 1$.

Substituting for x and y in the given equation, we get

$$(u+2)^2 + (v+1)^2 - 4(u+2) - 2(v+1) - 20 = 0$$
$$\therefore u^2 + 4u + 4 + v^2 + 2v + 1 - 4u - 8 - 2v - 2 - 20 = 0$$
$$\therefore u^2 + v^2 + 4u - 4u + 2v - 2v - 25 = 0$$
$$\therefore u^2 + v^2 = 25$$

It is customary to write (x, y) in place of (u, v) when the transformed equation is obtained.

$\therefore$ The transformed equation is $x^2 + y^2 = 25$.[5]

(The applications and usefulness of the process of translation of axes will be discussed at length later.)

Note (1): An ellipse is called a *central conic* in contrast to a parabola, which has *no center* because it has only one vertex.

Note (2): For an ellipse $a > b$, it follows that for the ellipse having the equation $(x^2/25) + (y^2/16) = 1$, the principal axis is the x-axis. (Note that $a^2 = 25$ and $b^2 = 16$, so $a = 5$ and $b = 4$ (Figure B.25a)). Next, the ellipse having the equation $(x^2/16) + (y^2/25) = 1$ has its principal axis on the y-axis (Figure B.25b).

Note (3): Suppose the center of an ellipse is at the point (h, k) rather than at the origin, and the principal axis is parallel to one of the coordinate axes. Then, by *translation of axes*, we have the following standard forms of the equations of an ellipse:

$$\frac{(x-h)^2}{a^2} + \frac{(y-h)^2}{b^2} = 1 \quad (a > b) \tag{I}$$

if the principal axis is horizontal.

[5] This equation represents a circle with center (2, 1) and radius 5. Of course, the coordinates of the center are with reference to the original axes.

Note: In this particular case, we know the name of the curve, but this may not be the situation, always.

$$\frac{(y-k)^2}{a^2} + \frac{(x-h)^2}{b^2} = 1 \quad (a > b) \tag{II}$$

if the principal axis is vertical.

(Recall that the foci always lie on the major axis. Furthermore, the major axis is along the x-axis if the coefficient of x^2 has the larger denominator and it is along the y-axis if the coefficient of y^2 has the larger denominator.)

Note (4): In the standard equation of an ellipse, we know that $a > b$. On the other hand, for a hyperbola there is no general inequality involving a and b. For instance, in the hyperbola$(x^2/9) - (y^2/16) = 1$, $a = 3$ and $b = 4$, so $a < b$. But in the hyperbola, $(x^2/21) - (y^2/4) = 1$, $a = \sqrt{21}$ and $b = 2$ so that $a > b$.

Note (5): The graphs of the hyperbolas $(x^2/9) - (y^2/16) = 1$ and $(y^2/9) - (x^2/16) = 1$ are shown in Figure B.26.

In the equation of hyperbola, *a may equal b*, in which case the hyperbola is equilateral.

Definition: A hyperbola in which $a = b$ is called an equilateral hyperbola.
(It is of the form $x^2 - y^2 = a^2$).

Definition: An equilateral hyperbola having the equation $x^2 - y^2 = 1$ is called the *unit hyperbola*. A convenient device can be used to obtain equations of the asymptotes of the hyperbola. For instance, for the hyperbola $(x^2/a^2) - (y^2/b^2) = 1$, we replace the right side by zero and obtain $(x^2/a^2) - (y^2/b^2) = 0$. Upon factorizing, this equation becomes $[(x/a) - (y/b)][(x/a) + (y/b)] = 0$, which is equivalent to the two equations $(x/a) - (y/b) = 0$ and $(x/a) + (y/b) = 0 \Leftrightarrow y = (b/a)x$ and $y = -(b/a)x$

These are the *equations of the asymptotes of the given hyperbola.*

(It can be proved that if the equation of a hyperbola is $(x^2/a^2) - (y^2/b^2) = k$, then the equations of the asymptotes are also given by $(x^2/a^2) - (y^2/b^2) = 0$).

Definition: A hyperbola whose asymptotes are at right angles to each other is called a *rectangular hyperbola.*

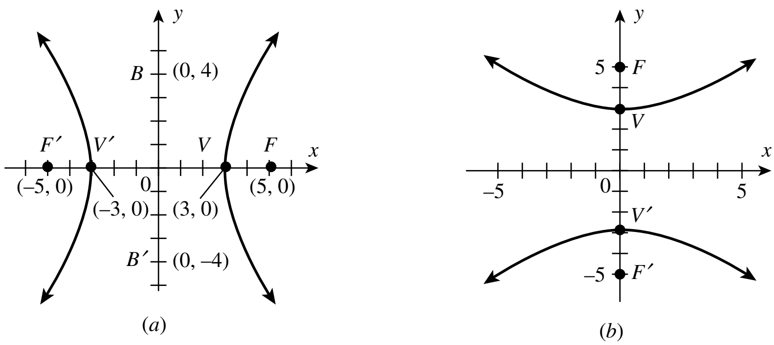


FIGURE B.26 (a) $\frac{x^2}{9} - \frac{y^2}{16} = 1$. (b) $\frac{y^2}{9} - \frac{x^2}{16} = 1$.

Remark: An equilateral hyperbola (i.e., $x^2 - y^2 = a^2$) is a rectangular hyperbola. (Note that the asymptotes of this equilateral hyperbola are $y = \pm x$. Obviously, the asymptotes are at right angles to each other and they are equally inclined to the axes.)

Note (6): If the center of the hyperbola is at (h, k), then an equation of the hyperbola is of the following form:

$$\frac{(x-h)^2}{a^2} - \frac{(y-k)^2}{b^2} = 1 \tag{1}$$

if the principal axis is horizontal.

$$\frac{(y-k)^2}{a^2} - \frac{(x-h)^2}{b^2} = 1 \tag{2}$$

if the principal axis is vertical.

B.7.1 Applications and Usefulness of the Process of Translation of Axes

We know that a given second degree equation in x and y of the form

$$Ax^2 + Cy^2 + Dx + Ey + F = 0\ (A \neq 0,\ C \neq 0)$$

may represent a circle, an ellipse, a hyperbola, or their degenerate forms. Our interest lies in identifying the curve represented by the given equation. This demands that we should be able to express the given equation to a recognizable form. *But how can we do this?* Let us discuss.

We know that the equation of a circle with center at the point (h, k) and the radius r is given by $(x-h)^2 + (y-k)^2 = r^2$. Similarly, we can write down the equations of ellipse(s) and those of hyperbola(s) with their centers at a point (h, k), other than the origin. Let us consider particular equations of these curves for any given (h, k).

(i) The equation $(x-3)^2 + (y-1)^2 = 5^2$ represents a circle with center at the point (3, 1) and radius 5 units. On opening the brackets, the above equation becomes

$$x^2 + y^2 - 6x - 2y - 15 = 0 \tag{A}$$

(ii) The equation $(x+3)^2/64 + (y-4)^2/100 = 1$ represents an ellipse whose center is at $(-3, 4)$. On opening the brackets, the above equation becomes

$$25x^2 + 16y^2 + 150x - 128y - 1119 = 0 \tag{B}$$

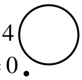
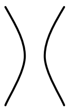
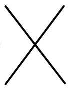
(iii) The equation $(y+2)^2/9 - (x-1)^2/4 = 1$ represents a hyperbola whose center is at $(-2, 1)$. On opening the brackets, the above equation becomes

$$9x^2 - 4y^2 - 18x - 16y + 29 = 0 \tag{C}$$

It is important to remember that a conic section generally refers to a parabola, an ellipse, or a hyperbola. A conic section may be represented by an algebraic equation of second degree. We ask an important question:

Is the graph of the equation of the form $Ax^2 + Cy^2 + Dx + Ey + F = 0$, always represent a conic?

The answer is *no*, unless we admit certain limiting forms. The table below indicated the possibilities with a sample equation of each.

Conics	Limiting Forms
1. $(AC = 0)$ Parabola: $y^2 = 4x$ 	A pair of parallel lines: $y^2 = 4$ Single line: $y^2 = 0$ Empty set: $y^2 = -1$
2. $(AC > 0)$ Ellipse: $\frac{x^2}{9} + \frac{y^2}{4} = 1$ 	Circle: $x^2 + y^2 = 4$  Point: $2x^2 + y^2 = 0$ Empty set: $2x^2 + y^2 = -1$
3. $(AC < 0)$ Hyperbola: $\frac{x^2}{9} - \frac{y^2}{4} = 1$ 	A pair of intersecting lines: $x^2 - y^2 = 0$ 

Note: It must be clear that in case of an ellipse the coefficients A and C have the *same sign*, so that $AC > 0$. (We have also seen this in Equation (B) above. Also, for a circle (which is a special case of an ellipse), A and C are *same* so that $AC > 0$). In case of a hyperbola, A and C have opposite sign so that $AC < 0$. As regards parabola, its equation must have only *one second degree term*. It follows that we have to drop either Ax^2 or Cy^2. In other words, we have to choose either $A = 0$ or $C = 0$, which means $AC = 0$.

Thus, a point, a pair of intersecting lines, a pair of coincident lines, and a circle represent limiting form of conic sections. (Recall that we have already discussed earlier about these possibilities.) Our interest lies in being able to express the given second degree equations in a recognizable form (of a conic) so that its graph can be sketched conveniently.

Now, we ask the question:

Given a complicated second degree equation (in x and y), how do we know what translation will simplify the equation and bring it to a recognizable form?

A familiar process called completing the square provides the answer. In particular, we use this process to eliminate the first degree terms of any expression of the following form:

$$Ax^2 + Cy^2 + Dx + Ey + F = 0 \ (A \neq 0,\ C \neq 0)$$

Example (3): Make a translation that will eliminate the first degree terms of $4x^2 + 9y^2 + 8x - 90y + 193 = 0$ and use this information to identify the curve.

Solution: Consider the given equation:

$$4x^2 + 9y^2 + 8x - 90y + 193 = 0$$
$$\therefore 4(x^2 + 2x + \_) + 9(y^2 - 10y + \_) = -193$$
$$\therefore 4(x^2 + 2x + 1) + 9(y^2 - 10y + 25) = -193 + 4 + 225$$
$$\therefore 4(x+1)^2 + 9(y-5)^2 = 36$$
$$\therefore \frac{(x+1)^2}{9} + \frac{(y-5)^2}{4} = 1$$

The translation $u = x + 1$ and $v = y - 5$ transforms this to $(u^2/9) + (v^2/4) = 1$, which is the *standard form of a horizontal ellipse*.

Example (4): Find the new equation of the curve $3x^2 + 2y^2 - 12x + 4y + 8 = 0$, when the origin is shifted to the points $(2, -1)$.

Solution: We have $(h, k) = (2, -1)$

$$\therefore x = u + h = u + 2, \quad \text{and} \quad y = v + k = v - 1$$

$\therefore$ The given equation transforms into

$$3(u+2)^2 + 2(v-1)^2 - 12(u+2) + 4(v-1) + 8 = 0$$

that is,

$$3(u^2 + 4u + 4) + 2(v^2 - 2v + 1) - 12(u+2) + 4(v-1) + 8 = 0$$

that is,

$$3u^2 + 12u + 12 + 2v^2 - 4v + 2 - 12u - 24 + 4v - 4 + 8$$

that is,

$$3u^2 + 2v^2 - 6 = 0 \text{ or } 3u^2 + 2v^2 = 6$$

The transformed equation is $3x^2 + 2y^2 = 6$ or $(x^2/2) + (y^2/3) = 1$
This equation is recognized as the equation of an ellipse.

Example (5): Show that the graph of the equation $4x^2 + 9y^2 + 8x - 90y + 193 = 0$ is an ellipse.

Solution: Consider the given equation

$$4x^2 + 9y^2 + 8x - 90y + 193 = 0$$
$$\therefore 4(x^2 + 2x + \_) + 9(y^2 - 10y + \_) = -193$$
$$\therefore 4(x^2 + 2x + 1) + 9(y^2 - 10y + 25) = -193 + 4 + 225$$
$$\therefore 4(x+1)^2 + 9(y-5)^2 = 36$$
$$\therefore \frac{(x+1)^2}{9} + \frac{(y-5)^2}{4} = 1$$

This equation represents an ellipse.

Example (6): Consider the equation $6x^2 + 9y^2 - 24x - 54y + 115 = 0$.

We write this equation as $6(x^2 - 4x) + 9(y^2 - 6y) = -115$.
Completing the squares in x and y, we get

$$6(x^2 - 4x + 4) + 9(y^2 - 6y + 9) = -115 + 24 + 81$$

or

$$6(x-2)^2+9(y-3)^2=-10$$

Because the right-hand side of this equation is negative and the left-hand side is nonnegative for all points (x, y), the *graph is the empty set*.

Example (7): Consider the equation $6x^2 + 9y^2 - 24x - 54y + 105 = 0$.

We write this equation as $6(x-2)^2 + 9(y-3)^2 = 0$.
Its graph are the points (2, 3). We can prove in general that the graph of any equation of the form (Q) is either *an ellipse, a point, or the empty set.*
(When the graph is a *point* or the *empty set*, it is said to be *degenerate*.)

Example (8): The equation $4x^2 - 12y^2 + 24x + 96y - 156 = 0$ can be written as

$4(x^2 + 6x) - 12(y^2 - 8y) = 156$ and upon completing the square in x and y, we have

$$\begin{aligned}
&4(x^2+6x+9)-12(y^2-8y+16)=156+36-192\\
&4(x+3)^2-12(y-4)^2=0\\
&(x+3)^2-3(y-4)^2=0\\
&\left[(x+3)-\sqrt{3}\,(y-4)\right]\left[(x+3)+\sqrt{3}\,(y-4)\right]=0\\
&\therefore\ x+3-\sqrt{3}\,(y-4)=0 \quad \text{and} \quad x+3+\sqrt{3}\,(y-4)=0
\end{aligned}$$

These are the equations of two lines through the points $(-3, 4)$

APPENDIX C (Related To Chapter-20)

EXERCISE

Verify the conditions of Rolle's theorem for the following functions on respective intervals and find c, if any, for which $f'(c) = 0$.

Q.(1): The polynomial function

$$y = f(x) = x^3 - 4x$$

is *continuous* and *differentiable* for all x, $-\infty < x < +\infty$.

We have, $f(x) = x^3 - 4x = x\,(x^2 - 4)$.

So if we take $a = -2$ and $b = +2$, then the conditions of Rolle's theorem are satisfied, since $f(-2) = 0$ and $f(+2) = 0$. Thus, the derivative $f'(x) = 3x^2 - 4$, must be zero at least once between -2 and 2. In fact, we can find this by solving $f'(x) = 0$, that is, $3x^2 - 4 = 0$.

We get,

$$x = c_1 = \frac{2}{\sqrt{3}} = \frac{2\sqrt{3}}{3} \quad \text{and} \quad x = c_2 = -\frac{2}{\sqrt{3}} = -\frac{2\sqrt{3}}{3} \text{ Ans.}$$

Q.(2): Verify Rolle's theorem for $f(x) = x^2(1 - x)^2$ in [0,1].

Solution: Here, $f(0) = f(1) = 0$ and $f(x)$ satisfies the conditions of Rolle's theorem.

$$\begin{aligned} f'(x) &= x^2\,[2\,(1-x)\,(-1)] + 2x(1-x)^2 \\ &= -2x^2\,(1-x) + 2x(1-x)^2 \\ &= 2x(1-x)\,[-x + (1-x)] \\ &= 2x(1-x)\,(1-2x) \end{aligned}$$

$$\therefore \quad f'(x) = 0, \text{ when } x = 0,\ x = 1,\ x = \frac{1}{2}$$

Here, $x = c = 1/2$ lies in (0, 1), for which $f'(c) = 0$. Ans.

Appendix C: Solutions to the Exercise on Rolle's Theorem and Mean Value Theorem (Chapter-20)

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and A. K. Ghosh.

Q.(3): Consider the function

$$y = f(x) = 1 - \sqrt[3]{x^2} = 1 - x^{2/3}$$

This function is continuous on the interval $[-1, 1]$, and vanishes at the end points of the interval $[f(-1) = 0$ and $f(1) = 0]$.

Q.(4): It is given that for the function $f(x) = x^3 - 6x^2 + ax + b$ on [1,3]. Rolle's theorem holds with $c = 2 + (1/\sqrt{3})$. Find the values of a and b.

Solution:

$$\begin{aligned} f(x) &= x^3 - 6x^2 + ax + b \\ f'(x) &= 3x^2 - 12x + a \end{aligned}$$

Since Rolle's theorem holds for $f(x)$, we have $f'(c) = 3c^2 - 12c + a$, where $c = (1, 3)$. Putting $f'(c) = 0$, we have

$$\begin{aligned} &3c^2 - 12c + a = 0 \\ &\Rightarrow c = \frac{12 \pm \sqrt{144 - 12a}}{6} \\ &\Rightarrow = \frac{12 \pm 2\sqrt{36 - 3a}}{6} \\ &\Rightarrow = 2 \pm \frac{\sqrt{36 - 3a}}{3} \end{aligned}$$

Here, $c = 2 - ((\sqrt{36 - 3a})/3)$ is not applicable.

$$\therefore \quad 2 + \frac{\sqrt{36 - 3a}}{3} = 2 + \frac{1}{\sqrt{3}}$$

$$\therefore \quad \frac{\sqrt{36 - 3a}}{3} = \frac{1}{\sqrt{3}}$$

$$\therefore \quad \sqrt{36 - 3a} = \sqrt{3}$$

$$\therefore \quad 36 - 3a = 3 \text{ or } a = 11$$

Hence, the function becomes

$$f(x) = x^3 - 6x^2 + 11x + b$$

Now $f(1) = 1 - 6 + 11 + b = 6 + b$. (Similarly, $f(3) = 6 + b$.)

But Rolle's theorem holds. So, $f(1) = 0$.

$$\begin{aligned} &\therefore \quad 6 + b = 0 \\ &\therefore \quad b = -6. \quad \text{Ans.} \end{aligned}$$

Q.(5): On the curve $y = x^2$, find a point at which the tangent is parallel to the chord joining (0, 0) and (1, 1).

Solution: The slope of the chord is

$$\frac{f(b) - f(a)}{b - a} = \frac{1 - 0}{1 - 0} = 1$$

The derivative is $dy/dx = 2x$.

We want x such that $2x = 1$.

Thus, $x = 1/2$. We note that 1/2 is in the open interval (0, 1), as required in the MVT. The corresponding point on the curve is (1/2, 1/4). Ans.

Q.(6): Verify LMVT for the function

$$f(x) = (x - 1)(x - 2)(x - 3) \text{ in } [0, 4]$$

Solution:

$$\begin{aligned} f(x) &= (x - 1)(x - 2)(x - 3) \\ &= x^3 - 6x^2 + 11x - 6 \end{aligned}$$

Since $f(x)$ is a polynomial, it is continuous on [0,4] and differentiable in (0, 4).

Also, $f'(x) = 3x^2 - 12x + 11$.

Now, $f(4) = (4 - 1)(4 - 2)(4 - 3) = 3 \times 2 \times 1 = 6$

And $f(0) = (-1)(-2)(-3) = -6$

By Lagrange's MVT, we have

$$\frac{f(b) - f(a)}{b - a} = f'(c)$$

$$\frac{f(4) - f(0)}{4 - 0} = 3c^2 - 12c + 11$$

$$\text{or} \quad \frac{6 - (-6)}{4} = 3c^2 - 12c + 11$$

$$\text{or} \quad 3c^2 - 12c + 8 = 0$$

$$\begin{aligned} c &= \frac{12 \pm \sqrt{144 - 4(3)\cdot(8)}}{6} \\ &= \frac{12 \pm \sqrt{48}}{6} = \frac{12 \pm 4\sqrt{3}}{6} \\ &= 2 \pm \frac{2\sqrt{3}}{3} \end{aligned}$$

Both these values lie in (0, 4). Hence, LMVT is verified. Ans.

Q.(7): Find a point on the graph of $y = x^3$, where the tangent is parallel to the chord joining (1, 1) and (3, 27).

Solution: $f(x) = x^3$

This function is continuous on [1, 3] and differentiable in (1, 3).[1]

Also, $f'(x) = 3x^2$

Slope of the chord is given by

$$\frac{f(b) - f(a)}{b - a} = \frac{f(4) - f(0)}{4 - 0}$$

$$= \frac{4^3 - 0}{4 - 0} = \frac{64}{4} = 16$$

$\therefore$ By LMVT, we have,

$$16 = 3c^2$$

$$\therefore \quad c = \pm\sqrt{\frac{16}{3}} = \pm 4\frac{\sqrt{3}}{3}$$

Note that, the value

$$c = \frac{4\sqrt{3}}{3} \in (1, 3) \quad \text{Ans.}$$

[1] Observe that the two points on the curve are (1, 1) and (3, 27). Hence, we are concerned only with the closed interval [1, 3], though the function $f(x) = x^3$ is defined for all x.

INDEX

Introduction to Differential Calculus: Systematic Studies with Engineering Applications for Beginners, First Edition.
Ulrich L. Rohde, G. C. Jain, Ajay K. Poddar, and Ajoy K. Ghosh.