Operator Theory: Advances and Applications
Vol. 178

Factorization of Matrix and Operator Functions: The State Space Method

Harm Bart
Israel Gohberg
Marinus A. Kaashoek
André C.M. Ran

Linear Operators & Linear Systems

Birkhäuser
Basel · Boston · Berlin

Authors:

Harm Bart
Econometrisch Instituut
Erasmus Universiteit Rotterdam
Postbus 1738
3000 DR Rotterdam
The Netherlands
e-mail: bart@few.eur.nl

Israel Gohberg
School of Mathematical Sciences
Raymond and Beverly Sackler
Faculty of Exact Sciences
Tel Aviv University
Ramat Aviv 69978
Israel
e-mail: gohberg@math.tau.ac.il

André C.M. Ran
Department of Mathematics, FEW
Vrije Universiteit Amsterdam
De Boelelaan 1081a
1081 HV Amsterdam
The Netherlands
e-mail: ACM.Ran@few.vu.nl

Marinus A. Kaashoek
Department of Mathematics, FEW
Vrije Universiteit
De Boelelaan 1081A
NL-1081 HV Amsterdam
The Netherlands
e-mail: m.a.kaashoek@few.vu.nl

2000 Mathematics Subject Classification: primary 47A48, 47A68, 47B35, 26C15;
secondary 15A21, 15A24, 30G30, 47A62, 90B35, 93B28

Library of Congress Control Number: 2007933911

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

ISBN 978-3-7643-8267-4 Birkhäuser Verlag AG, Basel - Boston - Berlin

ISBN 978-3-7643-8267-4 e-ISBN 978-3-7643-8268-1

9 8 7 6 5 4 3 2 1 www.birkhauser.ch

Dedicated to the memory of

Moshe Livsic

the founding father of the
characteristic function
in operator theory

Contents

Preface

The present book deals with various types of factorization problems for matrix and operator functions. The problems appear in different areas of mathematics and its applications. A unified approach to treat them is developed. The main theorems yield explicit necessary and sufficient conditions for the factorizations to exist and explicit formulas for the corresponding factors. Stability of the factors relative to a small perturbation of the original function is also studied in this book.

The unifying theory developed in the book is based on a geometric approach which has its origins in different fields. A number of initial steps can be found in:

(1) the theory of non-selfadjoint operators, where the study of invariant subspaces of an operator is related to factorization of the characteristic matrix or operator function of the operator involved,

(2) mathematical systems theory and electrical network theory, where a cascade decomposition of an input-output system or a network is related to a factorization of the associated transfer function, and

(3) the factorization theory of matrix polynomials in terms of invariant subspaces of a corresponding linearization.

In all three cases a state space representation of the function to be factored is used, and the factors are expressed in state space form too. We call this approach the *state space method*. It has a large number of applications. For instance, besides the areas referred to above, Wiener-Hopf factorizations of some classes of symbols can also be treated by the state space method.

The present book is the second book which is devoted to the state space factorization theory. The first was published in 1979 as the monograph by H. Bart, I. Gohberg and M.A. Kaashoek, "Minimal factorization of matrix and operator functions," Operator Theory: Advances and Applications **1**, Birkhäuser Verlag. This 1979 book appeared very soon after the first main results were obtained. In fact, some of these results where published in the 1979 book for the first time.

This second book, which is written by the authors of the first book jointly with A.C.M. Ran, consists of four parts. Parts I, II and IV contain a substantial selection from the first book, in a reorganized and updated form. Part III, which covers more than a quarter of the book, is entirely new. This third part is devoted

to the theory of factorization into degree one factors and its connection to the combinatorial problem of job scheduling in operations research. It also contains Maple procedures to calculate degree one factorizations. In contrast to the other parts, this third part is completely finite-dimensional and can be considered as a new advanced chapter of Linear Algebra and its Applications. Almost each chapter in this book offers new elements and in many cases new sections, taking into account a number of new results in state space factorization theory and its applications that have appeared in the period of 25 years after publication of the first book. On the other hand in the present book there is less emphasis on Wiener-Hopf integral equation and its applications than in the first book. However these topics are not entirely absent but, for instance, the applications to transport do not appear in this book. The text is largely self-contained, and will be of interest to experts and students in Mathematics, Sciences and Engineering.

The authors are in the process of writing another book, also devoted to the state space approach to factorization. There the emphasis will be on canonical factorization and symmetric factorization with applications to different classes of convolution equations. For the latter we have in mind the transport equation, singular integral equations, equations with symbols analytic in a strip, and equations involving factorization of non-proper rational matrix functions. Furthermore, a large part of this third book will deal with factorization of matrix functions satisfying various symmetries. A main theme will be the effect on factorization of these symmetries and how the symmetries can be used in effective way to get state space formulas for the factors. Applications to H-infinity control theory, which have been developed in the eighties and nineties, will also be included.

The authors gratefully acknowledge a visitor fellowship for the second author from the Netherlands Organization for Scientific Research (NWO), and the financial support from the School of Economics of the Erasmus University at Rotterdam, from the School of Mathematical Sciences of Tel-Aviv University and the Nathan and Lily Silver Family Foundation, and from the Mathematics Department of the Vrije Universiteit at Amsterdam. These funds allowed us to meet and to work together on the book for different extended periods of time in Amsterdam and Tel-Aviv.

In conclusion, we would like to express our gratitude to Johan Kaashoek who wrote for this book two new sections with Maple procedures for computing certain degree one factorizations. Without his help these sections would not have been. He also read Part III of the book in detail and provided us with several useful comments. We thank our friends and colleagues who made comments on earlier drafts of this book. In particular, we would like to mention Sanne ter Horst for his corrections to the first part of the book, Leonia Lerer for his comments on the first two parts, and Rob Zuidwijk for his remarks about the third part.

The authors *Amsterdam – Rotterdam – Tel-Aviv*
Spring 2007

Chapter 0

Introduction

This monograph is devoted to theory and applications of various types of factorizations for matrix and operator functions belonging to different classes. The types of factorizations described in the book appear in several branches of algebra, analysis and applications. Let us mention a few examples.

In the theory of non-selfadjoint operators [30, 108] there exists the notion of regular factorization of the characteristic matrix or operator function of a given operator. This type of factorization leads to the description of an invariant subspace of the operator involved and, what is more important, to triangular representation of this operator [61]. In systems theory and electrical network theory [27, 84] one encounters the notion of minimal factorization of the transfer function of a system or a network. Such a factorization allows one to represent a system or a network as a cascade connection of systems or networks with simpler synthesis. Sometimes, the situation allows for so-called complete factorizations. These are minimal factorizations where the factors are of the simplest possible type, namely of (McMillan) degree one. Dropping the minimality requirement, factorizations into degree one factors are always possible. Those that have the least possible number of factors are called quasicomplete. Via this notion a connection is made with the two machine flow shop problem from the theory of combinatorial job scheduling. Another type of factorization that we shall consider is that of canonical Wiener-Hopf factorization [45, 57] of some classes of symbols. This factorization, when it exists, allows one to invert Wiener-Hopf, Toeplitz and singular integral operators, and when the factors are known one can also build explicitly the inverses of these operators. Factorization of matrix or operator polynomials into polynomials of lower degree [69, 101] is also a type of factorization we shall discuss.

The matrix and operator functions that are considered have in common that they appear in a natural way as functions of the form

$$W(\lambda) = D + C(\lambda I - A)^{-1}B \tag{1}$$

or (after some transformations) can be represented in this form. In the above formula λ is a complex variable, and A, B, C, and D are matrices or (bounded) linear operators acting between appropriate Banach or Hilbert spaces. When A, B, C, and D are matrices or the underlying spaces are all finite-dimensional, the function W is a rational matrix function which is analytic at infinity. From mathematical systems theory it is known that conversely any rational matrix function which is analytic at infinity admits a representation of the above form. In systems theory the right-hand side of (1) is called a *state space realization* of the function W, and one refers to the space in which A is acting as the *state space*. For this reason we call the method that we are using in this book the *state space method*.

The state space approach has been used very successfully in mathematical systems theory and network theory. In this book we use the method to deal with various classes of factorization problems. The method has also been used in other branches of analysis, for instance, in interpolation theory [8, 43].

Realizations allow us to deal with factorization from a geometric point of view. Special attention is paid to various types of factorizations, for example, to canonical factorization, minimal and non-minimal factorizations, pseudo-canonical factorization, degree one factorizations and others. The problem of numerical computations of the factors of a given matrix or operator function leads in a natural way to questions of stability of divisors under small perturbations. It turns out that in general the factors are unstable. In this book the stable cases are described and estimates are given for the measures of stability.

Not only motivations but also applications play an important role in the book. We shall deal with applications to problems in mathematical systems theory and control, to problems in the theory of algebraic Riccati equations, and to inversion problems for convolution operators. Another special feature is the connection between (generally non-minimal) factorizations into elementary factors and a problem of job scheduling from combinatorial operations research. Applications to the theory of matrix and operator polynomials and rational matrix functions are included too.

Our intention was to make this monograph accessible for readers working in different areas of mathematics. We have in mind Linear Algebra, Linear Operator Theory, Integral Equations, Mathematical Systems Theory and Applied Mathematics. This forced us to make the exposition reasonably self-contained. In particular, we included some known material about characteristic operator functions, angular operators, minimal factorizations of rational matrix functions, the gap between subspaces et cetera.

We shall now give a short description of the contents of the book. Not counting the present introduction, the book consists of four parts.

Part I. The first part has a preparatory character. The motivating problems are described, and the underlying concepts are developed. In this part also the notions of nodes and characteristic function, and of systems and transfer functions are

introduced. The main operations on nodes and systems are studied, and the effect of these operations on the characteristic or transfer functions are described. The basic factorization principle used throughout the book already appears in this part, including its version in terms of angular subspaces and Riccati equations. The problem of realization is also addressed, and the connection with linearization of operator functions is clarified.

Part II. The second part deals with the notions of minimality of realizations and minimality of factorizations. For finite-dimensional systems minimality is equivalent to controllability and observability. For rational matrix functions minimal realizations are constructed in terms of the pole-zero structure of the given function, and minimal factorizations are described in terms of pole-zero cancellation. This part contains also a study of the notion of minimality for various classes of finite- and infinite-dimensional systems. Using the notion of local minimality, the concept of a pseudo-canonical factorization relative to a curve is introduced and analyzed for rational matrix functions with singularities on the given curve.

Part III. The third part is devoted to the problem of factorization into elementary functions, that is, into factors that have a minimal realization with a state space of dimension one, the so-called degree one factors. A new feature is the connection to a job scheduling issue, namely to the two machine flow shop problem from operations research. The latter involves quasicomplete factorizations, that is, generally non-minimal factorizations into degree one factors with the smallest number of factors. Maple procedures are provided to calculate degree one factorizations, complete as well as quasicomplete, of companion based 2×2 rational matrix functions. This part is completely finite-dimensional and can be considered as a new advanced chapter of Linear Algebra and its Applications.

Part IV. The fourth part deals with the behavior of the factors in a factorization under small perturbations of the original function. Canonical factorization is stable in the sense that a rational matrix function which has a canonical factorization keeps this property under small perturbation. In this part we analyze the dependence of the factors on the perturbations using state space realizations. For minimal factorization the situation is different. It can happen that a rational matrix function admits a non-trivial minimal factorization while after a small perturbation the perturbed function has no such factorization. Using the realization theory the minimal factorizations that do not have this kind of instability are identified. The notions of stable, Lipschitz stable and isolated invariant subspaces turn out to play an important role in the analysis. Applications to Riccati equations are included. The case of factorization of real matrix functions is also treated in this part and yields results that differ from the case when the underlying field is complex.

Part I

Motivating Problems, Systems and Realizations

An important notion in this book is that of a time-invariant linear, discrete or continuous, input-output system. This notion is taken from mathematical systems theory. A related notion is that of an operator node. The latter originates from the theory of non-selfadjoint operators. Nodes can be considered as finite- or infinite-dimensional systems with some additional restrictions on the system coefficients. In the two theories different terminologies have been developed for objects that are essentially the same. For instance, the transfer function of a system is the same as the characteristic function of a node. On the other hand the type of problems considered in the two theories are quite different.

This first part, which consists of six chapters, is of a preparatory character. It presents in a unified way various aspects of the two theories. In Chapters 1 and 2 systems and nodes are introduced. The notions of transfer function and characteristic function are defined and discussed. The main operations on systems and nodes – inversion, product, factorization – are introduced and studied in detail. The effects of these operations on the transfer function or characteristic function are analyzed. The main principle of state space factorization theory, used throughout this book, already appears in Chapter 2 (see Section 2.4). Chapter 1 contains also a number of motivating problems. These problems and their variations reappear in different parts of the book. Chapter 3 contains the classification of systems and nodes. Chapter 4 is dedicated to the problem of linearization of analytic operator functions and of transfer functions of systems. In Chapter 5 the state space factorization theorem from Chapter 2 is reformulated in terms of angular subspaces and solutions of algebraic Riccati equations. The final chapter (Chapter 6) presents a first analysis of canonical factorization in terms of the state space method. Included are also applications to convolution equations.

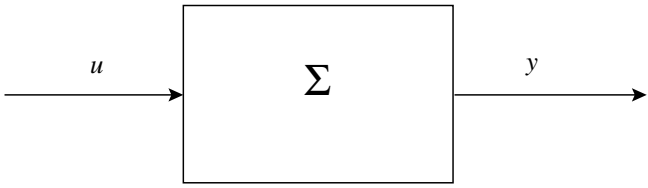
Chapter 1

Motivating Problems

This chapter has an introductory character. It presents a number of problems involving factorization of matrix- and operator-valued functions of different types. The functions considered appear as transfer functions of input output systems (Section 1.1), as characteristic functions of Hilbert space operators (Sections 1.2 and 1.3), as monic matrix polynomials (Section 1.4) or as symbols of Wiener-Hopf and singular integral equations of various types (Sections 1.5 and 1.6). For each of these classes the corresponding factorization is described. This chapter also motivates the state space setting for solving factorization problems.

1.1 Linear time invariant systems and cascade connection

A system Σ can be considered as a physical object which produces an output in response to an input. Schematically,



where u denotes the input and y denotes the output. Mathematically, the input u and the output y are vector-valued functions of a parameter t. The input can be chosen freely (at least in principle), but the output is uniquely determined by the choice of the input. Hence the map $u \mapsto y$ is a well-defined transformation, which is called the *input output operator* of the system.

The way in which the output is generated by the input can be quite complicated. In this section we consider the simplest model and assume that the relation

between input and output is described by a system of differential equations of the following type:

$$\Sigma \left\{ \begin{array}{lcll} x'(t) & = & Ax(t) + Bu(t), & \\ y(t) & = & Cx(t) + Du(t), & t \geq 0, \\ x(0) & = & 0. & \end{array} \right. \tag{1.1}$$

Here the coefficients are linear operators acting between Euclidean spaces,

$$A : \mathbb{C}^m \to \mathbb{C}^m, \quad B : \mathbb{C}^p \to \mathbb{C}^m, \quad C : \mathbb{C}^m \to \mathbb{C}^q, \quad D : \mathbb{C}^p \to \mathbb{C}^q.$$

Whenever convenient, we identify these operators with the corresponding matrices (using the standard bases in the Euclidean spaces).

The space $\mathbb{C}^m$ is called the *state space*, and its elements (vectors) are called *states*. The spaces $\mathbb{C}^p$ and $\mathbb{C}^q$ will be referred to as the *input space* and *output space*, respectively. The operator A is the so-called *state operator* or *main operator* of (1.1), B is the *input operator*, C is the *output operator*, and D is the *external operator*, which is also called the *feed through coefficient*. In what follows we shall call (1.1) a *finite-dimensional linear time-invariant system* or just a *system*. The qualification "finite-dimensional" refers to the finite dimensionality of the underlying spaces, and the word "time-invariant" is reflected by the fact that the coefficients A, B, C and D do not depend on the variable t.

We shall assume that the inputs u of (1.1) are taken from the space $PCE(\mathbb{C}^p)$ which consists of all piecewise continuous $\mathbb{C}^p$-valued functions on $[0, \infty)$ that are exponentially bounded. The latter means that for each $u \in PCE(\mathbb{C}^p)$ there exists real constants M and γ (depending on u), $M \geq 0$, such that $\|u(t)\| \leq Me^{\gamma t}$, $t \geq 0$. Then the output y belongs to the space $PCE(\mathbb{C}^q)$ which consists of all piecewise continuous exponentially bounded $\mathbb{C}^q$-valued functions. In fact, the input output operator of (1.1) is the operator $T : PCE(\mathbb{C}^p) \to PCE(\mathbb{C}^q)$ given by

$$y(t) = (Tu)(t) = Du(t) + \int_0^t Ce^{(t-s)A}Bu(s)\,ds, \qquad t \geq 0, \tag{1.2}$$

To see this, note that

$$x(t) = \int_0^t e^{(t-s)A}Bu(s)\,ds, \qquad t \geq 0, \tag{1.3}$$

is the unique solution of the first equation in (1.1) satisfying the initial condition $x(0) = 0$. Inserting (1.3) into the second equation in (1.1) yields formula (1.2) for the input output operator.

From (1.2) it follows that the input output operator is linear. This explains the use of the term "linear" in connection with (1.1). Furthermore, one sees that (1.1) is *causal*. This means that future inputs do not affect past outputs, i.e., for each $\tau > 0$ the output $y(t)$ on $[0, \tau]$ does not depend on the input $u(t)$, $t > \tau$.

Taking Laplace transforms in (1.1) we arrive at the equivalent form in *frequency domain*:

$$\begin{cases} \lambda\widehat{x}(s) &= A\widehat{x}(\lambda) + B\widehat{u}(\lambda), \\ \widehat{y}(\lambda) &= C\widehat{x}(\lambda) + D\widehat{u}(\lambda). \end{cases} \tag{1.4}$$

Here, for any exponentially bounded vector-valued function v the symbol $\widehat{v}$ denotes its Laplace transform

$$\widehat{v}(\lambda) = \int_0^\infty e^{-\lambda t} v(t)\, dt, \qquad \Re\lambda \geq c,$$

where c is some constant depending on v. From (1.4) one can solve $\widehat{y}(\lambda)$ in terms of $\widehat{u}(\lambda)$. Indeed, on some open open right half-plane of $\mathbb{C}$ we have

$$\widehat{y}(\lambda) = \big(D + C(\lambda I_m - A)^{-1}B\big)\widehat{u}(\lambda),$$

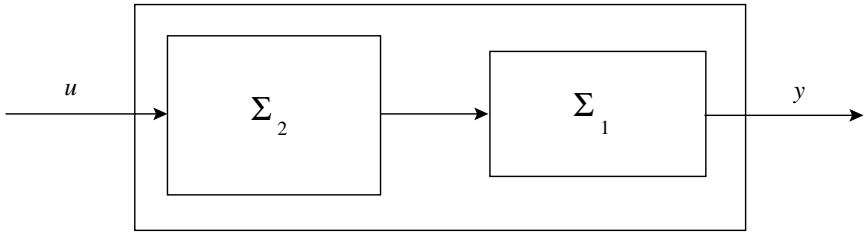
where I_m is the $m \times m$ identity matrix (or, if one prefers, the identity operator on $\mathbb{C}^m$). So in the frequency domain the input output behavior of (1.1) is determined by the function

$$W(\lambda) = D + C(\lambda I_m - A)^{-1}B, \tag{1.5}$$

which is called the *transfer function* of the system (1.1). Since the system is finite-dimensional, the transfer function is a $q \times p$ matrix function all of whose entries are rational functions. Such a function will be referred to as a *rational matrix function*. Notice that the rational matrix function W in (1.5) is analytic at infinity. A rational matrix function with this additional property is said to be *proper*.

We shall see later (in Chapter 4) that any proper rational matrix function is the transfer function of a finite-dimensional time-invariant linear system. That is, given a proper rational matrix function W, one can find matrices A, B, C, D such that (1.5) holds. In this case we call the right-hand side of (1.5) or the corresponding system (1.1) a *realization* of W. This connection allows one to study problems involving a rational matrix function in terms of the four matrices appearing in its realization. We refer to this approach as the *state space method*. In particular, we shall use the state space method to solve factorization problems.

The problem to factorize a rational matrix function into factors of simpler type appears naturally in system theory when one considers cascade connections. By definition the *cascade connection* of two systems is the system which one obtains when the output of the first system is taken to be the input of the second system. Schematically:



Let W_1 and W_2 be the transfer functions of the systems Σ_1 and Σ_2, respectively. Then, given the input u, the output y_2 of Σ_2 is given by $\widehat{y}_2(\lambda) = W_2(\lambda)\widehat{u}(\lambda)$. Since the input of Σ_1 is the output of Σ_2, it follows that the output y is given $\widehat{y}(\lambda) = W_1(\lambda)\widehat{y}_2(\lambda)$. Clearly then, the transfer function W of the cascade connection of these two systems is given by the product $W = W_1W_2$ of W_1 and W_2, that is,

$$\widehat{y}(\lambda) = W_1(\lambda)W_2(\lambda)\widehat{u}(\lambda) = W(\lambda)\widehat{u}(\lambda).$$

Let us analyze this in terms of the operators appearing in the representation (1.1). For $j = 1, 2$, let W_j be the transfer function of the system

$$\Sigma_j \left\{ \begin{array}{rcll} x_j'(t) & = & A_jx_j(t) + B_ju_j(t), & \\ y_j(t) & = & C_jx_j(t) + D_ju_j(t), & t \geq 0, \\ x(0) & = & 0. & \end{array} \right.$$

We take $y_2 = u_1$, in other words, we form the cascade connection. Taking

$$x = \left[\begin{array}{c} x_1 \\ x_2 \end{array} \right]$$

as the state vector for the system thus obtained, we have

$$\begin{aligned} x'(t) &= \left[\begin{array}{c} A_1x_1(t) + B_1u_1(t) \\ A_2x_2(t) + B_2u_2(t) \end{array} \right] = \left[\begin{array}{c} A_1x_1(t) + B_1C_2x_2(t) + B_1D_2u_2(t) \\ A_2x_2(t) + B_2u_2(t) \end{array} \right] \\ &= \left[\begin{array}{cc} A_1 & B_1C_2 \\ 0 & A_2 \end{array} \right] x(t) + \left[\begin{array}{c} B_1D_2 \\ B_2 \end{array} \right] u_2(t) \end{aligned}$$

and

$$\begin{aligned} y(t) &= y_1(t) = C_1x_1(t) + D_1u_1(t) \\ &= C_1x_1(t) + D_1C_2x_2(t) + D_1D_2u_2(t) \\ &= \left[\begin{array}{cc} C_1 & D_1C_2 \end{array} \right] x(t) + D_1D_2u_2(t). \end{aligned}$$

Thus the transfer function $W = W_1W_2$ is also given by

$$W(\lambda) = D_1D_2 + \left[\begin{array}{cc} C_1 & D_1C_2 \end{array} \right] \left(\lambda - \left[\begin{array}{cc} A_1 & B_1C_2 \\ 0 & A_2 \end{array} \right] \right)^{-1} \left[\begin{array}{c} B_1D_2 \\ B_2 \end{array} \right]. \tag{1.6}$$

The fact that the transfer function of the cascade connection is the product of the transfer functions of the corresponding systems is the basis for the state space approach to factorization used in this monograph. We shall develop the state

space factorization method for various types of factorization, including canonical factorization (Chapter 6), minimal factorization (Chapter 9), and factorization of rational matrix functions into a product of elementary ones (see Chapter 10 and also Theorems 2.7 and 8.15). Factorization of the latter type corresponds to cascade synthesis (of systems) involving components of simplest possible type (cf., [39] and the references therein)

As can be expected from (1.6) the problem of finding a factorization of W is related to presence of invariant subspaces of the main operator in a realization of W. This relation is one of the leading principles of this monograph. It also turns up in the theory of characteristic operator functions which we shall discuss in the next two sections.

1.2 Characteristic operator functions and invariant subspaces (1)

In the theory of characteristic functions the main object is a bounded linear operator acting on a Hilbert space, and the characteristic function serves as a unitary invariant for the operator. In this section we consider operators close to selfadjoint ones.

Let A be a bounded linear operator acting on a Hilbert space H. The adjoint of A will be denoted by A^*. The imaginary part of A, given by $\frac{1}{2i}(A - A^*)$, is a selfadjoint operator on H, and hence there exists a Hilbert space G and operators $K : G \to H$ and $J : G \to G$ such that

$$KJK^* = \frac{1}{2i}(A - A^*)$$

and J is a *signature operator*. By definition, the latter means that J is invertible and $J^{-1} = J = J^*$. From A, K and J we construct the following operator-valued function:

$$W(\lambda) = I + 2iK^*(\lambda - A)^{-1}KJ, \qquad \lambda \in \rho(A). \tag{1.7}$$

Here $\rho(A)$ is the *resolvent set* of A, that is, the set of $\lambda \in \mathbb{C}$ such that $\lambda - A$ is (boundedly) invertible.

The operator-valued function W defined by (1.7) is called the Livsic-Brodskii *characteristic operator function* of A or, more precisely, of the *operator node* $(A, KJ, 2iK^*, I; H, G)$. A Hilbert space operator J satisfying $J = J^* = J^{-1}$ is called a *signature operator*. This function has special symmetry properties. Indeed, using $KJK^* = \frac{1}{2i}(A - A^*)$ one easily checks that

$$W(\lambda)^* JW(\lambda) = J - 2i(\lambda - \bar{\lambda})JK^*(\bar{\lambda} - A^*)^{-1}(\lambda - A)^{-1}KJ.$$

It follows that

$$\begin{aligned} W(\lambda)^* JW(\lambda) &= J, \qquad \lambda \in \rho(A) \cap \mathbb{R}, \\ W(\lambda)^* JW(\lambda) &\leq J, \qquad \lambda \in \rho(A),\ \Im\lambda \leq 0. \end{aligned}$$

These formulas remain true if the positions of $W(\lambda)$ and $W(\lambda)^*$ are interchanged. It is possible, using the above formulas, to give an intrinsic characterization of the class of functions that appear as Livsic-Brodskii characteristic functions (see [30]).

The characteristic operator function can be viewed as the transfer function of the following system

$$\begin{cases} x'(t) &= Ax(t) + KJu(t), \\ y(t) &= 2iK^*x(t) + u(t), \qquad t \geq 0, \\ x(0) &= 0. \end{cases}$$

The above system will be called a *Brodskii J-system*; this term will also be used for the corresponding operator node $(A, KJ, 2iK^*, I; H, G)$.

Suppose that A is unitary equivalent to an operator B, i.e., $A = UBU^*$, where $U : H_1 \to H$ is unitary. Then

$$KJK^* = \frac{1}{2i}(A - A^*) = \frac{1}{2i}(UBU^* - UB^*U^*) = \frac{1}{2i}U(B - B^*)U^*.$$

Taking $L = U^*K$, we see that the system $(B, LJ, 2iL^*, I; H_1, G)$ is also a Brodskii J-system, and that this system has the same transfer function W as the system $(A, KJ, 2iK^*, I; H, G)$. So the characteristic operator function W does not change under unitary equivalence. Under a certain additional minimality condition the converse is also true. Indeed, if two characteristic operator functions W_1 and W_2 given by

$$\begin{aligned} W_1(\lambda) &= I + 2iK_1^*(\lambda - A_1)^{-1}K_1J, \qquad \lambda \in \rho(A_1), \\ W_2(\lambda) &= I + 2iK_2^*(\lambda - A_2)^{-1}K_2J, \qquad \lambda \in \rho(A_2), \end{aligned}$$

coincide in some neighborhood of infinity and the corresponding systems are simple (cf., Subsection 7.5.1) then the operators A_1 and A_2 are unitary equivalent. Actually there exists a unitary operator U such that $UA_1 = A_2U$ and $UK_1 = K_2$ (see [30], Theorem I.3.2). This fact is of particular interest when the imaginary part of A is small. For instance, when A has rank one, then W reduces to a scalar function, and hence the infinite-dimensional operator A is determined up to unitary equivalence by a scalar function.

The product of two Brodskii characteristic operator functions W_1 and W_2 is again a Brodskii characteristic operator function. To see this, write

$$\begin{aligned} W_1(\lambda) &= I + 2iK_1^*(\lambda - A_1)^{-1}K_1J, \qquad \lambda \in \rho(A_1), \\ W_2(\lambda) &= I + 2iK_2^*(\lambda - A_2)^{-1}K_2J, \qquad \lambda \in \rho(A_2), \end{aligned}$$

Here $A_1 : H_1 \to H_1$ and $A_2 : H_2 \to H_2$. As in the previous section it straightforward to check that the function $W = W_1W_2$ is the transfer function of the

system

$$\left(\begin{bmatrix} A_1 & 2iK_1JK_2^* \\ 0 & A_2 \end{bmatrix}, \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} J, 2i \begin{bmatrix} K_1^* & K_2^* \end{bmatrix}, I; H_1 \oplus H_2, G \right).$$

Here $H_1 \oplus H_2$ is the Hilbert space direct sum of H_1 and H_2. Put

$$A = \begin{bmatrix} A_1 & 2iK_1JK_2^* \\ 0 & A_2 \end{bmatrix}, \qquad K = \begin{bmatrix} K_1 \\ K_2 \end{bmatrix}.$$

Then $\frac{1}{2i}(A - A^*) = KJK^*$. So the function W is the characteristic operator function of the operator A.

Notice that the operator A constructed in the previous paragraph has the space H_1 as an invariant subspace. This fact contains a hint for constructing factorizations within the class of characteristic operator functions.

To be more precise, let $\Theta = (A, KJ, 2iK^*, I; H, G)$ be a Brodskii system, and assume that H_0 is an invariant subspace of A. Let Π be the orthogonal projection onto H_0. Put

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \qquad K = \begin{bmatrix} K_1 \\ K_2 \end{bmatrix}$$

with respect to the decomposition $H = H_0 \oplus H_0^\perp$. Then

$$\Theta_1 = (A_{11}, K_1J, 2iK_1^*, I; H_0, G), \qquad \Theta_2 = (A_{22}, K_2J, 2iK_2^*, I; H_0^\perp, G)$$

are Brodskii J-systems. Indeed, the imaginary part of A is given by

$$\frac{1}{2i}\left(\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} - \begin{bmatrix} A_{11}^* & 0 \\ A_{12}^* & A_{22}^* \end{bmatrix} \right) = \begin{bmatrix} K_1JK_1^* & K_1JK_2^* \\ K_2JK_1^* & K_2JK_2^* \end{bmatrix},$$

so in particular $\frac{1}{2i}(A_{11} - A_{11}^*) = K_1JK_1^*$ and $\frac{1}{2i}(A_{22} - A_{22}^*) = K_2JK_2^*$. Moreover, $\frac{1}{2i}A_{12} = K_1JK_2^*$. This implies that the product of the characteristic operator function of A_{11} and the characteristic operator function of A_{22} (i.e., the product of the transfer function of the systems Θ_1 and Θ_2) is the characteristic operator function of A.

Under appropriate minimality conditions there is a one-one correspondence between invariant subspaces of A and factorizations of the characteristic operator function W of A as the product of two characteristic operator functions. Thus, in certain cases, the problem of finding invariant subspaces of an operator A may be solved by factorization of its characteristic operator function. For an example of the application of this technique involving the unicellularity of a Volterra operator, see Section XXVIII.11 in [47].

1.3 Characteristic operator functions and invariant subspaces (2)

The Livsic-Brodskii characteristic operator function has been designed to study operators that are not far from being selfadjoint. There are also characteristic operator functions that have been introduced in order to deal with operators that are close to unitary operators. Among them are the characteristic operator function of Sz.-Nagy and Foias and the one of M.G. Kreĭn (see [33] and [108] for references). Here we shall only discuss the characteristic operator function of Kreĭn.

The Kreĭn *characteristic operator function* has the form

$$V(\lambda) = J(K^*)^{-1}\big(J - R^*(I - \lambda A)^{-1}R\big). \tag{1.8}$$

Here $A : H \to H$, $R : G \to H$, $J : G \to G$, $K : G \to G$ are operators, the underlying spaces G and H are complex Hilbert spaces,

$$J = J^* = J^{-1}, \quad I - AA^* = RJR^*, \quad J - R^*R = K^*JK, \tag{1.9}$$

and the operators A and K are invertible. Instead of $(K^*)^{-1}$ we also write K^{-*}. With this (1.8) becomes $V(\lambda) = JK^{-*}\big(J - R^*(I - \lambda A)^{-1}R\big)$.

Obviously, (1.8) does not directly fit into the framework developed in Section 1.1. However, replacing λ by λ^{-1} and using (1.9), one can transform (1.8) into

$$U(\lambda) = K - JK^{-*}R^*A(\lambda - A)^{-1}R.$$

This is the transfer function of the system

$$\begin{cases} x'(t) &= Ax(t) + Ru(t), \\ y(t) &= -JK^{-*}R^*Ax(t) + Ku(t), \quad t \geq 0, \\ x(0) &= 0. \end{cases}$$

The above system will be called a *Kreĭn J-system*; this term will also be used for the corresponding operator node

$$\Theta = (A, R, -JK^{-*}R^*A, K; H, G). \tag{1.10}$$

Observe that the external operator of a Kreĭn J-system is invertible.

The product of two Kreĭn characteristic operator functions is again a Kreĭn characteristic operator function. To see this, suppose

$$U_j(\lambda) = K_i - JK_j^{-*}R_j^*A_i(\lambda - A_j)^{-1}R_j, \qquad j = 1, 2,$$

where

$$I - A_jA_j^* = R_jJR_j^*, \qquad J - R_j^*R_j = K_j^*JK_j. \tag{1.11}$$

Then $U = U_1U_2$ is the transfer function of the operator node

$$\left(\begin{bmatrix} A_1 & -R_1JK_2^{-*}R_2^*A_2 \\ 0 & A_2 \end{bmatrix}, \begin{bmatrix} R_1K_2 \\ R_2 \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} -JK_1^{-*}R_1^*A_1 & -K_1JK_2^{-*}R_2^*A_2 \end{bmatrix}, K_1K_2 \right).$$

Moreover, this operator node is a Kreĭn J-system. Indeed, using (1.11) we have

$$I - \begin{bmatrix} A_1 & -R_1JK_2^{-*}R_2^*A_2 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} A_1^* & 0 \\ -A_2^*R_2K_2^{-1}JR_1^* & A_2^* \end{bmatrix}$$

$$= \begin{bmatrix} R_1K_2 \\ R_2 \end{bmatrix} J \begin{bmatrix} K_2^*R_1^* & R_2^* \end{bmatrix},$$

and

$$J - \begin{bmatrix} K_2^*R_1^* & R_2^* \end{bmatrix} \begin{bmatrix} R_1K_2 \\ R_2 \end{bmatrix} = K_2^*K_1^*JK_1K_2,$$

while finally

$$\begin{bmatrix} -JK_1^{-*}R_1^*A_1 & -K_1JK_2^{-*}R_2^*A_2 \end{bmatrix}$$

$$= -JK_1^{-*}K_2^{-*} \begin{bmatrix} K_2^*R_1^* & R_2^* \end{bmatrix} \begin{bmatrix} A_1 & -R_1JK_2^{-*}R_2^*A_2 \\ 0 & A_2 \end{bmatrix}.$$

This proves that $U = U_1U_2$ is the transfer function of a Kreĭn J-system. Notice that the main operator A is given by

$$A = \begin{bmatrix} A_1 & -R_1JK_2^{-*}R_2^*A_2 \\ 0 & A_2 \end{bmatrix},$$

and hence the space on which A_1 acts is an invariant subspace for A.

Let us consider the reverse implication. Our starting point is a Kreĭn J-system as in (1.10), with A acting on the Hilbert space H and A being invertible. Assume that H_0 is an invariant subspace of A. With respect to the decomposition $H_0 \oplus H_0^\perp$ of the state space H, write

$$A = \begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix}, \qquad R = \begin{bmatrix} B_1 \\ R_2 \end{bmatrix}.$$

Suppose that A_1 or, equivalently, A_2 is invertible. From $RJR^* = I - AA^*$ one easily deduces that $R_2JR_2^* = I - A_2A_2^*$. Since A_2 is assumed to be invertible, this

shows that $I - R_2 J R_2^*$ is invertible. But then $I - J R_2^* R_2$ is invertible, and hence the same holds true for $J - R_2^* R_2$. The invertibility of $J - R_2^* R_2$ implies (see [33]) the existence of an invertible operator K_2 such that $J - R_2^* R_2 = K_2^* J K_2$. Put $K_1 = K K_2^{-1}$ and $R_1 = B_1 K_2^{-1}$. Then K_1 is also invertible. We claim that $A_{12} = -R_1 J K_2^{-*} R_2^* A_2$. In order to prove this, we first note that $A_{12} = -B_1 J R_2^* A_2^{-*} = -R_1 K_2 J R_2^* A_2^{-*}$. Furthermore we have

$$\begin{aligned} K_2 J R_2^* A_2^{-*} &= J K_2^{-*}(K_2^* J K_2) J R_2^* A_2^{-*} \\ &= J K_2^{-*}(J - R_2^* R_2) J R_2^* A_2^{-*} \\ &= J K_2^{-*}(R_2^* - R_2^* R_2 J R_2^*) A_2^{-*} \\ &= J K_2^{-*} R_2^*(I - R_2 J R_2^*) A_2^{-*} = J K_2^{-*} R_2^* A_2. \end{aligned}$$

Thus $A_{12} = -R_1 J K_2^{-*} R_2^* A_2$. It follows that one can decompose the function $U(\lambda) = K - J K^{-*} R^* A(\lambda - A)^{-1} R$ as a product of two functions corresponding to Kreĭn J-systems. In fact, $U = U_1 U_2$, where

$$U_j(\lambda) = K_j - J K_j^{-*} R_j^* A_j (\lambda - A_j)^{-1} R_j, \qquad j = 1, 2$$

with the coefficients satisfying (1.11).

We conclude this section with an interesting characterization of Kreĭn J-systems. Let G and H be complex Hilbert spaces, and let J be a signature operator on G, that is, $J = J^* = J^{-1}$. A bounded linear operator T on G is called *J-unitary* if T is invertible and $T^{-1} = J T^* J$. By definition, a node $\Theta = (A, B, C, D; H, G)$ is a Kreĭn J-system if A and D are invertible and

$$I - A A^* = B J B^*, \quad J - B^* B = D^* J D, \quad C = -J D^{-*} B^* A.$$

A straightforward calculation shows that these conditions are equivalent to the requirement that the external operator D of Θ is invertible and the operator

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{1.12}$$

on $H \oplus G$ is $\widetilde{J}$-unitary. Here $\widetilde{J} = I \oplus J$. Notice that $\widetilde{J} = \widetilde{J}^* = \widetilde{J}^{-1}$. The class of nodes $\Theta = (A, B, C, D; H, G)$ for which the operator (1.12) is $\widetilde{J}$-unitary (but D not necessarily invertible) is closed under multiplication. Characteristic operator functions of the form $D + \lambda C(I - \lambda A)^{-1} B = D + C(\lambda^{-1} - A)^{-1} B$, where A, B, C and D are such that (1.12) is unitary, have been investigated (cf., [31]; see also Section 3.3 below).

Observe that if (1.12) is $\widetilde{J}$-unitary and $U(\lambda) = D + C(\lambda - A)^{-1} B$, then

$$U(\lambda)^*(-J) U(\lambda) = -J - (1 - |\lambda|^2) B^* (\bar{\lambda} - A^{-*}(\lambda - A)^{-1} B.$$

So we have

$$\begin{aligned} U(\lambda)^*(-J)U(\lambda) &= -J, \qquad |\lambda| = 1 \\ U(\lambda)^*(-J)U(\lambda) &\leq -J, \qquad |\lambda| < 1. \end{aligned}$$

It is possible to give an intrinsic characterization of the class of functions that appear as transfer functions of Kreĭn J-systems (cf., [33]; see also [1] for the case of matrix functions).

1.4 Factorization of monic matrix polynomials

By definition a *monic $m \times m$ matrix polynomial* of *degree ℓ* is a function of the form

$$L(\lambda) = \lambda^\ell I_m + \lambda^{\ell-1} A_{\ell-1} + \cdots + \lambda A_1 + A_0,$$

where $A_0, \ldots, A_{\ell-1}$ are $m \times m$ matrices. Given such a function, introduce

$$A = \begin{bmatrix} 0 & I_m & 0 & \cdots & 0 \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & I_m \\ -A_0 & -A_1 & \cdots & \cdots & -A_{\ell-1} \end{bmatrix}, \tag{1.13}$$

the *first companion operator matrix* associated with L, and

$$C = \begin{bmatrix} 0 & \cdots & 0 & I_m \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{bmatrix}. \tag{1.14}$$

Then the (*pointwise*) *inverse* L^{-1} of L, given by $L^{-1} = L(\lambda)^{-1}$, has the realization

$$L^{-1}(\lambda) = C(\lambda I_m - A)^{-1} B \tag{1.15}$$

(see [66]).

To prove this identity, consider the set of differential equations in $\mathbb{C}^n$-valued vector functions y given by

$$\begin{cases} y^{(\ell)}(t) + A_{\ell-1} y^{(\ell-1)}(t) + \cdots + A_1 y'(t) + A_0 y(t) = u(t), \\ y(0) = 0,\ y'(0) = 0\ ,\ \ldots,\ y^{(\ell-1)}(0) = 0. \end{cases} \tag{1.16}$$

Let us transform this to a higher-dimensional first-order system in the usual way, by introducing

$$x(t) = \begin{bmatrix} y(t)^\top & y'(t)^\top & \dots & y^{(\ell-1)}(t)^\top \end{bmatrix}^\top.$$

Then, because of (1.16), we have

$$\begin{cases} x'(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t), \qquad\qquad t \geq 0, \\ x(0) &= 0, \end{cases} \tag{1.17}$$

where A, B, and C are defined by (1.13) and (1.14). Taking Laplace transform in (1.17) and eliminating $\widehat{x}(s)$ we obtain that

$$\widehat{y}(s) = C(s - A)^{-1}B\widehat{u}(s).$$

On the other hand, taking Laplace transform in (1.16) we get $L(s)\widehat{y}(s) = \widehat{u}(s)$. Thus (1.15) has been established.

We shall consider the problem of finding and describing factorizations of $L(\lambda)$ of the form $L(\lambda) = L_2(\lambda)L_1(\lambda)$, where L_1 and L_2 are again monic matrix polynomials. Certain invariant subspaces of the operator A play an important role in solving this problem (see Section 3.4).

1.5 Wiener-Hopf integral operators and factorization

In this section we outline the factorization method of [59] to solve systems of Wiener-Hopf integral equations. Such a system may be written as a single *Wiener-Hopf equation*

$$\phi(t) - \int_0^\infty k(t-s)\phi(s)\,ds = f(t), \qquad 0 \leq t < \infty, \tag{1.18}$$

where ϕ and f are m-dimensional vector functions and $k \in L_1^{m\times m}(-\infty,\infty)$, that is, the kernel function k is an $m \times m$ matrix function whose entries are in $L_1(-\infty,\infty)$. We assume that the given vector function f has its component functions in $L_p[0,\infty)$, and we express this property by writing $f \in L_p^m[0,\infty)$. Throughout this section p will be fixed and $1 \leq p < \infty$. The problem we shall consider is to find a solution ϕ of equation (1.18) that also belongs to the space $L_p^m[0,\infty)$.

Equation (1.18) has a unique solution $\phi \in L_p^m[0,\infty)$ for any right-hand side $f \in L_p^m[0,\infty)$ if and only if the *Wiener-Hopf integral operator* $I - \mathbf{K} : L_p^m[0,\infty) \to L_p^m[0,\infty)$ is invertible, where

$$\big(\mathbf{K}\phi\big)(t) = \int_0^\infty k(t-s)\phi(s)\,ds, \qquad t \geq 0.$$

The usual method (see [59]) to solve equation (1.18) is as follows. First assume that (1.18) has a solution ϕ in $L_p^m[0,\infty)$. Extend ϕ and f to the full real line by putting

$$\phi(t) = 0, \qquad f(t) = -\int_0^\infty k(t-s)\phi(s)\,ds, \quad t < 0.$$

Then $\phi, f \in L_p^m(-\infty,\infty)$ and the full line convolution equation

$$\phi(t) - \int_{-\infty}^\infty k(t-s)\phi(s)\,ds = f(t), \qquad -\infty < t < \infty$$

is satisfied. By applying the Fourier transformation and leaving the part of f that is given in the right-hand side, one gets

$$\big(I_m - K(\lambda)\big)\Phi_+(\lambda) - F_-(\lambda) = F_+(\lambda), \qquad \lambda \in \mathbb{R}, \tag{1.19}$$

where

$$K(\lambda) = \int_{-\infty}^\infty e^{i\lambda t}k(t)\,dt, \qquad F_+(\lambda) = \int_0^\infty e^{i\lambda t}f(t)\,dt,$$

$$\Phi_+(\lambda) = \int_0^\infty e^{i\lambda t}\phi(t)\,dt, \qquad F_-(\lambda) = \int_{-\infty}^0 e^{i\lambda t}f(t)\,dt.$$

Note that the functions K and F_+ are given, but the functions Φ_+ and F_- have to be found. In fact in this way the problem to solve (1.18) is reduced to that of finding two functions Φ_+ and F_- such that (1.19) holds, while furthermore Φ_+ and F_- must be as above with $\phi \in L_p^m[0,\infty)$ and $f \in L_p^m(-\infty,0]$.

To find Φ_+ and F_- of the desired form such that (1.19) holds, one factorizes the $m \times m$ matrix function $I_m - K(\lambda)$. This function is called the *symbol* of the integral equation (1.18). Assume that the symbol admits a factorization of the form

$$I_m - K(\lambda) = \big(I_m + G_-(\lambda)\big)\big(I_m + G_+(\lambda)\big), \qquad \lambda \in \mathbb{R}, \tag{1.20}$$

where

$$G_+(\lambda) = \int_0^\infty e^{i\lambda t}g_+(t)\,dt, \qquad G_-(\lambda) = \int_{-\infty}^0 e^{i\lambda t}g_-(t)\,dt,$$

with $g_+ \in L_1^{m\times m}[0,\infty)$ and $g_- \in L_1^{m\times m}(-\infty,0]$ while, in addition, the determinants $\det\big(I_m + G_+(\lambda)\big)$ and $\det\big(I_m + G_-(\lambda)\big)$ do not vanish in the closed upper and lower half-plane, respectively. We shall refer to the factorization (1.20) as a *right canonical factorization of $I_m - K(\lambda)$ with respect to the real line*. Under the conditions stated above the functions $\big(I_m + G_+(\lambda)\big)^{-1}$ and $\big(I_m + G_-(\lambda)\big)^{-1}$ admit representations as Fourier transforms:

$$\big(I_m + G_+(\lambda)\big)^{-1} = I_m + \int_0^\infty e^{i\lambda t}\gamma_+(t)\,dt, \tag{1.21}$$

$$\big(I_m + G_-(\lambda)\big)^{-1} = I_m + \int_{-\infty}^0 e^{i\lambda t}\gamma_-(t)\,dt, \tag{1.22}$$

with $\gamma_+ \in L_1^{m\times m}[0,\infty)$ and $\gamma_- \in L_1^{m\times m}(-\infty,0]$. Using the factorization (1.20) and suppressing the variable λ, equation (1.19) can be rewritten as

$$(I_m + G_+)\Phi_+ - (I_m + G_-)^{-1}F_- = (I_m + G_-)^{-1}F_+. \tag{1.23}$$

Let $\mathcal{P}$ be the projection acting on the Fourier transforms of $L_p^m(-\infty,\infty)$-functions according to the following rule

$$\mathcal{P}\left(\int_{-\infty}^{\infty} e^{i\lambda t}h(t)\,dt\right) = \int_0^{\infty} e^{i\lambda t}h(t)\,dt.$$

Applying $\mathcal{P}$ to (1.23), one gets

$$(I_m + G_+)\Phi_+ = \mathcal{P}\big((I_m + G_-)^{-1}F_+\big),$$

and hence

$$\Phi_+ = (I_m + G_+)^{-1}\mathcal{P}\big((I_m + G_-)^{-1}F_+\big),$$

which is the formula for the solution of equation (1.19). To obtain the solution ϕ of the original equation (1.18), i.e., to obtain the inverse Fourier transform of Φ_+, one can employ the formulas (1.21) and (1.22). In fact,

$$\phi(t) = f(t) + \int_0^{\infty} \gamma(t,s)f(s)\,ds, \qquad t \geq 0,$$

where the kernel $\gamma(t,s)$ is given by

$$\gamma(t,s) = \begin{cases} \gamma_+(t-s) + \displaystyle\int_0^s \gamma_+(t-r)\gamma_-(r-s)\,dr, & 0 \leq s < t, \\ \gamma_-(t-s) + \displaystyle\int_0^t \gamma_+(t-r)\gamma_-(r-s)\,dr, & 0 \leq t < s. \end{cases} \tag{1.24}$$

We conclude the description of this factorization method by mentioning that the equation (1.18) has a unique solution in $L_p^m[0,\infty)$ for each f in $L_p^m[0,\infty)$ if and only if its symbol admits a factorization as in (1.20). For details, see [45], [59].

To illustrate the method, let us consider a special choice for the right-hand side f (cf., [59]). Take

$$f(t) = e^{-iqt}x_0, \tag{1.25}$$

where x_0 is a fixed vector in $\mathbb{C}^m$ and q is a complex number with $\Im q < 0$. Then

$$F_+(\lambda) = \int_0^{\infty} e^{i(\lambda-q)t}x_0\,dt = \frac{i}{\lambda - q}x_0, \qquad \Im\lambda \geq 0.$$

Now observe that

$$\frac{i}{\lambda - q}\Big(\big(I_m + G_-(\lambda)\big)^{-1} - \big(I_m + G_-(q)\big)^{-1}\Big)x_0$$

is the Fourier transform of an $L_p^m(-\infty, 0]$-function and hence it vanishes when applying the projection $\mathcal{P}$. It follows that in this case the formula for Φ_+ may be written as

$$\Phi_+(\lambda) = \frac{i}{\lambda - q}\big(I_m + G_+(\lambda)\big)^{-1}\big(I_m + G_-(q)\big)^{-1}x_0.$$

Recall that the solution ϕ is the inverse Fourier transform of Φ_+. So we have

$$\phi(t) = e^{-iqt}\left(I_m + \int_0^t e^{iqs}\gamma_+(s)\,ds\right)\big(I_m + G_-(q)\big)^{-1}x_0. \tag{1.26}$$

1.6 Block Toeplitz equations and factorization

In this section we consider the discrete analogue of a Wiener-Hopf integral equation, that is, a block *Toeplitz equation*. So we consider an equation of the type

$$\sum_{k=0}^{\infty} a_{j-k}\xi_k = \eta_j, \qquad j = 0, 1, 2, \ldots . \tag{1.27}$$

Throughout we assume that the coefficients a_j are given complex $m \times m$ matrices satisfying

$$\sum_{j=-\infty}^{\infty} \|a_j\| < \infty,$$

and $\eta = (\eta_j)_{j=0}^{\infty}$ is a given vector from $\ell_p^m = \ell_p(\mathbb{C}^m)$. The problem is to find $\xi = (\xi_k)_{k=0}^{\infty} \in \ell_p^m$ such that (1.27) is satisfied. We shall restrict ourselves to the case when $1 \le p \le 2$; the final results however are valid for $2 < p \le \infty$ as well.

Assume $\xi \in \ell_p^m$ is a solution of (1.27). Then one can write (1.27) in the form

$$\sum_{k=-\infty}^{\infty} a_{j-k}\xi_k = \eta_j, \qquad j = 0, \pm 1, \pm 2, \ldots , \tag{1.28}$$

where $\xi_k = 0$ for $k < 0$ and η_j is defined by (1.28) for $j < 0$. Multiplying both sides of (1.28) by λ^j with $|\lambda| = 1$ and summing over j, one gets

$$a(\lambda)\xi_+(\lambda) - \eta_-(\lambda) = \eta_+(\lambda), \qquad |\lambda| = 1, \tag{1.29}$$

where the functions a, η_+, η_-, ξ_+ and ξ_+ are given by

$$a(\lambda) = \textstyle\sum_{j=-\infty}^{\infty} \lambda^j a_j, \qquad \eta_+(\lambda) = \sum_{j=0}^{\infty} \lambda^j \eta_j,$$

$$\xi_+(\lambda) = \textstyle\sum_{j=0}^{\infty} \lambda^j \xi_j, \qquad \eta_-(\lambda) = \sum_{j=-\infty}^{-1} \lambda^j \eta_j.$$

In this way the problem to solve (1.27) is reduced to that of finding two functions ξ_+ and η_- such that (1.29) holds, while moreover, ξ_+ and η_- must be as above with $(\xi_j)_{j=0}^{\infty}$ and $(\eta_{-j-1})_{j=0}^{\infty}$ from ℓ_p^m.

The usual way (cf., [59] or the book [42]) of solving (1.29) is again by factorizing the *symbol* a of the given block Toeplitz equation. Assume that a admits a *right canonical factorization with respect to the unit circle*. By definition this means that a can be written as

$$a(\lambda) = h_-(\lambda)h_+(\lambda), \qquad |\lambda| = 1, \tag{1.30}$$

$$h_+(\lambda) = \sum_{j=0}^{\infty} \lambda^j h_j^+, \qquad h_-(\lambda) = \sum_{j=-\infty}^{0} \lambda^j h_j^-,$$

where $(h_j^+)_{j=0}^{\infty}$ and $(h_{-j}^-)_{j=0}^{\infty}$ belong to the space $\ell_1^{m\times m}$ of all absolutely convergent sequences of complex $m \times m$ matrices such that $\det h_+(\lambda) \neq 0$ for $|\lambda| \leq 1$ and $\det h_-(\lambda) \neq 0$ for $|\lambda| \geq 1$ (including $\lambda = \infty$). Then h_+^{-1} and h_-^{-1} also admit a representation of the form

$$h_+^{-1}(\lambda) = \sum_{j=0}^{\infty} \lambda^j \gamma_j^+, \qquad h_-^{-1}(\lambda) = \sum_{j=-\infty}^{0} \lambda^j \gamma_j^-,$$

with $(\gamma_j^+)_{j=0}^{\infty}$ and $(\gamma_{-j}^-)_{j=0}^{\infty}$ from $\ell_1^{m\times m}$. Defining the projection $\mathcal{P}$ by

$$\mathcal{P}\left(\sum_{j=-\infty}^{\infty} \lambda^j b_j\right) = \sum_{j=0}^{\infty} \lambda^j b_j,$$

one gets from (1.29) and (1.30) the identity $\xi_+ = h_+^{-1}\mathcal{P}\big(h_-^{-1}\eta_+\big)$. Here, for convenience, the variable λ is suppressed. The solution of the original equation (1.27) can now be written as

$$\xi_k = \sum_{s=0}^{\infty} \gamma_{ks}\eta_s, \qquad k = 0, 1, \dots, \tag{1.31}$$

where

$$\gamma_{ks} = \begin{cases} \sum_{r=0}^{s} \gamma_{k-r}^+ \gamma_{r-s}^-, & s < k, \\ \sum_{r=0}^{s=k} \gamma_{s-r}^+ \gamma_{r-s}^-, & s = k, \\ \sum_{r=0}^{k} \gamma_{k-r}^+ \gamma_{r-s}^-, & s > k. \end{cases} \tag{1.32}$$

The assumption that a admits a right canonical factorization as in (1.30) is equivalent to the requirement that for each $\eta = (\eta_j)_{j=0}^{\infty}$ in ℓ_p^m the equation (1.27) has a unique solution $\xi = (\xi_k)_{k=0}^{\infty}$ in ℓ_p^m. For details we refer to [59], [42].

By way of illustration, we consider the special case when

$$\eta_j = q^j \eta_0, \qquad j = 0, 1, \dots .$$

Here η_0 is a fixed vector in $\mathbb{C}^m$ and q is a complex number with $|q| < 1$. Then clearly

$$\eta_+(\lambda) = \frac{1}{1-\lambda q}\eta_0, \quad |\lambda| \leq 1,$$

and one checks without difficulty that formula (1.31) becomes

$$\xi_k = q^k \sum_{s=0}^{k} q^{-s}\gamma_s^+ h_-^{-1}(q^{-1})\eta_0, \qquad k = 0, 1, \ldots .$$

This is the analogue of formula (1.26) in the previous section.

Notes

The material in this chapter is standard, and can be found in much more detail in various monographs, books, and papers. For the theory of time invariant systems we refer to the books [84], [114], and the more recent [36]. Further information on the theory of characteristic operator functions can be found in the books [30] and [108]; see also the survey paper [7] and the references therein. For the general theory of matrix polynomials, including the monic case, we refer to the book [69]. For the corresponding theory of operator polynomials, see [101]. The idea to think of the inverse of a monic matrix or operator polynomial as a characteristic function appears and has been developed in [11]. Sections 5–7 contain standard material about Wiener-Hopf integral equations and block Toeplitz equations. For more information on these equations and the corresponding operators see the monographs [45], [62], [63], [64] and [29]. A first introduction to the theory of Wiener-Hopf integral equations and the theory of (block) Toeplitz operators can be found in Chapters XII and XIII of [46] and Chapters XXIII–XXV of [47], respectively. For an extensive review (with many additional references) of the factorization theory of matrix functions relative to a curve and its applications to inversion of singular integral operators of different types, including Wiener-Hopf and block Toeplitz operators, the reader is referred to the recent survey paper [57].

Chapter 2

Operator Nodes, Systems, and Operations on Systems

In this chapter the concepts of an operator node (abstract system) and its transfer function are introduced and developed systematically. Important operations on operator nodes (abstract systems) and the corresponding operations on the associated transfer functions are studied in detail: inversion (Section 2.2), products (Section 2.3) and factorization (Section 2.4). With an eye on future applications, a detailed analysis of the relationships between the various results is given in the final section.

2.1 Operator nodes, systems and transfer functions

An *operator node* is a collection of three complex Banach spaces X, U, Y, and four bounded linear operators

$$A : X \to X, \qquad B : U \to X, \qquad C : X \to Y, \qquad D : U \to Y.$$

We shall denote such a node by $\Theta = (A, B, C, D; X, U, Y)$. Whenever convenient, we shall think about the operators in an operator node as the coefficients of a (possibly infinite-dimensional) time invariant system, either in continuous time, that is

$$\begin{cases} x'(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t), \qquad t \geq 0, \\ x(0) &= 0, \end{cases}$$

or in a discrete time setting, i.e.,

$$\begin{cases} x(n+1) &= Ax(n)+Bu(n), \\ y(n) &= Cx(n)+Du(n), \qquad n=0,1,2,\ldots . \\ x(0) &= 0. \end{cases}$$

Therefore, in the sequel, we shall use the word *system* also to denote an operator node. Furthermore, we shall freely use the terminology of system theory in the operator node setting.

Let $\Theta = (A, B, C, D; X, U, Y)$ be a system (or operator node). The spaces U, X and Y are called the *input space*, *state space* and *output space* of the system, respectively. The operator A is referred to as the *state space operator* or *main operator* of the system Θ. When A is given or can be viewed as a matrix, the terms *main matrix* and *state matrix* will be used too. We call D the *external operator* of Θ.

In the situation where $U = Y$, we denote Θ by $(A, B, C, D; X, Y)$. When, in addition, D is the identity operator $I = I_Y$ on Y, we simply write $(A, B, C; X, Y)$ instead of $(A, B, C, I; X, Y)$, and in that case we refer to Θ as a *unital system*. When no confusion can arise, the spaces X, U and Y will sometimes be dropped altogether, resulting in the notation $\Theta = (A, B, C, D)$.

By definition, the *transfer function* of $\Theta = (A, B, C, D; X, U, Y)$ is the function W_Θ given by

$$W_\Theta(\lambda) = D + C(\lambda - A)^{-1}B, \qquad \lambda \in \rho(A).$$

Here $\rho(A)$ is the resolvent set of A. Note that the transfer function is *proper* in the sense that

$$\lim_{\lambda\to\infty} W_\Theta(\lambda) = D \tag{2.1}$$

exists. The transfer function W_Θ has to be considered as an analytic operator function, defined on an open neighborhood of ∞ on the Riemann sphere $\mathbb{C}\cup\{\infty\}$. Instead of (2.1), we often write $W_\Theta(\infty) = D$.

When the external operator D is invertible, we say that the system is *biproper*; when $D = 0$, the system is called *strictly proper*. Mutatis mutandis, these terms are also used for the transfer function .

Two systems

$$\Theta_1 = (A_1, B_1, C_1, D_1; X_1, U, Y), \qquad \Theta_2 = (A_2, B_2, C_2, D_2; X_2, U, Y),$$

having the same input and output space, are said to be *similar*, written $\Theta_1 \simeq \Theta_2$, if $D_1 = D_2$ and there exists an invertible operator $S : X_1 \to X_2$ such that

$$A_1 = S^{-1}A_2S, \qquad B_1 = S^{-1}B_2, \qquad C_1 = C_2S.$$

In this case we say that S is a *system similarity* from Θ_1 to Θ_2. Notice that $\simeq$ is reflexive, symmetric and transitive. Obviously, similar systems have the same transfer function.

Let W be an operator function, analytic on an open subset Ω of $\mathbb{C} \cup \{\infty\}$. We say that the system $\Theta = (A, B, C, D; X, U, Y)$ is a *realization for W on Ω* if $\Omega \subset \rho(A) \cup \{\infty\}$ and $W(\lambda) = W_\Theta(\lambda)$ for each $\lambda \in \Omega$. If there is no danger of confusion (e.g., when W is a rational matrix function), we shall simply use the term "realization" and omit the additional qualifiers. The term *realization* will also be used to denote any expression of the form $W(\lambda) = D + C(\lambda - A)^{-1}B$.

2.2 Inversion

Let $\Theta = (A, B, C, D; X, U, Y)$ be a system which is biproper, that is, the external operator D is invertible. Consider the corresponding linear time invariant system

$$\begin{cases} x'(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t), \qquad t \geq 0, \\ x(0) &= 0. \end{cases} \tag{2.2}$$

As D is invertible, we can solve u in terms of y from the second equation in (2.2). Inserting the solution into the first equation yields

$$\begin{cases} x'(t) &= (A - BD^{-1}C)x(t) + BD^{-1}y(t), \\ u(t) &= -D^{-1}Cx(t) + D^{-1}y(t), \qquad t \geq 0, \\ x(0) &= 0. \end{cases}$$

The corresponding node will be denoted by $\Theta^\times$, i.e.,

$$\Theta^\times = (A - BD^{-1}C, BD^{-1}, -D^{-1}C, D^{-1}; X, Y, U),$$

and $\Theta^\times$ will be called the *associate or inverse system* of Θ. By slight abuse of notation we write $A^\times$ for $A - BD^{-1}C$, and we call $A^\times$ the *associate state space operator* or *associate main operator* of Θ. Whenever this is feasible, the terms *associate main matrix* and *associate state matrix* will also be employed.

The slight abuse of notation we mentioned lies in the fact that $A^\times$ does not depend on A only, but also on the other operators appearing in the system Θ. A direct computation gives $(\Theta^\times)^\times = \Theta$.

Theorem 2.1. *Let $\Theta = (A, B, C, D; X, U, Y)$ be a biproper system, and let $W = W_\Theta$ be its transfer function. Put $A^\times = A - BD^{-1}C$, and take $\lambda \in \rho(A)$. Then $W(\lambda)$ is invertible if and only if λ belongs to $\rho(A^\times)$. In that case, for $\lambda \in \rho(A) \cap \rho(A^\times)$, the following identities hold*

$$\begin{aligned} W(\lambda)^{-1} &= D^{-1} - D^{-1}C(\lambda - A^\times)^{-1}BD^{-1}, \\ (\lambda - A^\times)^{-1} &= (\lambda - A)^{-1} - (\lambda - A)^{-1}BW(\lambda)^{-1}C(\lambda - A)^{-1}. \end{aligned}$$

Moreover,

$$\begin{aligned} W(\lambda)D^{-1}C(\lambda - A^\times)^{-1} &= C(\lambda - A)^{-1}, \\ (\lambda - A^\times)^{-1}BD^{-1}W(\lambda) &= (\lambda - A)^{-1}B, \end{aligned}$$

where again $\lambda \in \rho(A) \cap \rho(A^\times)$.

The first expression says that on $\rho(A) \cap \rho(A^\times)$, the (*pointwise*) *inverse* W^{-1} of W, given by $W^{-1}(\lambda) = W(\lambda)^{-1}$, coincides with the transfer function $W_{\Theta^\times}$ of the system $\Theta^\times$. The last two identities can be written in different forms, for instance as

$$\begin{aligned} W(\lambda)^{-1}C(\lambda - A)^{-1} &= D^{-1}C(\lambda - A^\times)^{-1}, \\ (\lambda - A)^{-1}BW(\lambda)^{-1} &= (\lambda - A^\times)^{-1}BD^{-1}. \end{aligned}$$

We shall give two proofs of the theorem.

First proof of Theorem 2.1. Put $W^\times = W_{\Theta^\times}$. For $\lambda \in \rho(A) \cap \rho(A^\times)$, one has

$$\begin{aligned} W(\lambda)W^\times(\lambda) &= \big(D + C(\lambda - A)^{-1}B\big)\big(D^{-1} - D^{-1}C(\lambda - A^\times)^{-1}BD^{-1}\big) \\ &= I_Y + C(\lambda - A)^{-1}BD^{-1} - C(\lambda - A^\times)^{-1}BD^{-1} + \\ &\qquad -C(\lambda - A)^{-1}BD^{-1}C(\lambda - A^\times)^{-1}BD^{-1}. \end{aligned}$$

Now use that

$$BD^{-1}C = A - A^\times = \big(\lambda - A^\times\big) - (\lambda - A)\,.$$

It follows that $W(\lambda)W^\times(\lambda) = I_Y$. Analogously one has $W^\times(\lambda)W(\lambda) = I_U$. The expression for $(\lambda - A^\times)^{-1}$ as well as the last two identities in the theorem are obtained in a similar way. □

For the second proof of Theorem 2.1 we use Schur complements. First we define this notion. Consider the 2×2 operator matrix

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} : Z_1 \dot{+} U \to Z_2 \dot{+} Y.$$

Here M_{ij}, $i, j = 1, 2$, are bounded linear operators acting between complex Banach spaces, M_{11} maps Z_1 into Z_2, M_{12} maps U into Z_2, and so on. Assume that M_{22} is invertible. Then by Gauss elimination, M admits the following factorization

$$M = \begin{bmatrix} I_{Z_2} & M_{12}M_{22}^{-1} \\ 0 & I_Y \end{bmatrix} \begin{bmatrix} \Delta & 0 \\ 0 & M_{22} \end{bmatrix} \begin{bmatrix} I_{Z_1} & 0 \\ M_{22}^{-1}M_{21} & I_U \end{bmatrix}, \tag{2.3}$$

where $\Delta = M_{11} - M_{12}M_{22}^{-1}M_{21}$. The operator Δ is called the *Schur complement of* M_{22} *in* M. Since the first and third factor in the right-hand side of (2.3) are

invertible operators, the operator M is invertible if and only if Δ is invertible, and in that case

$$M^{-1} = \begin{bmatrix} \Delta^{-1} & -\Delta^{-1}M_{12}M_{22}^{-1} \\ -M_{22}^{-1}M_{21}\Delta^{-1} & M_{22}^{-1} + M_{22}^{-1}M_{21}\Delta^{-1}M_{12}M_{22}^{-1} \end{bmatrix}.$$

Similarly, if M_{11} is invertible, then the operator $\Lambda = M_{22} - M_{21}M_{11}^{-1}M_{12}$ is called the *Schur complement of* M_{11} *in* M. Since M_{11} is invertible, we have

$$M = \begin{bmatrix} I_{Z_2} & 0 \\ M_{21}M_{11}^{-1} & I_Y \end{bmatrix} \begin{bmatrix} M_{11} & 0 \\ 0 & \Lambda \end{bmatrix} \begin{bmatrix} I_{Z_1} & M_{11}^{-1}M_{12} \\ 0 & I_U \end{bmatrix}. \tag{2.4}$$

Hence, when M_{11} is invertible, we see that M is invertible if and only if the Schur complement Λ is invertible, and in this case

$$M^{-1} = \begin{bmatrix} M_{11}^{-1} + M_{11}^{-1}M_{12}\Lambda^{-1}M_{21}M_{11}^{-1} & -M_{11}^{-1}M_{12}\Lambda^{-1} \\ -\Lambda^{-1}M_{21}M_{11}^{-1} & \Lambda^{-1} \end{bmatrix}.$$

Thus if both M_{11} and M_{22} are invertible, then Δ is invertible if and only if the same holds true for Λ. Moreover, by comparing the two formulas for M^{-1}, we see that

$$\Delta^{-1} = M_{11}^{-1} + M_{11}^{-1}M_{12}\Lambda^{-1}M_{21}M_{11}^{-1}, \tag{2.5}$$

$$\Lambda^{-1} = M_{22}^{-1} + M_{22}^{-1}M_{21}\Delta^{-1}M_{12}M_{22}^{-1}. \tag{2.6}$$

Besides these inversion formulas, we also have the intertwining relations

$$\Lambda M_{22}^{-1}M_{21} = M_{21}M_{11}^{-1}\Delta, \qquad M_{12}M_{22}^{-1}\Lambda = \Delta M_{11}^{-1}M_{12}. \tag{2.7}$$

Second proof of Theorem 2.1. We apply the results about Schur complements mentioned above to the operator matrix

$$M(\lambda) = \begin{bmatrix} A - \lambda I_X & B \\ C & D \end{bmatrix} : X \dotplus U \to X \dotplus Y.$$

According to our hypothesis, D is invertible. Thus the Schur complement of D in $M(\lambda)$ is well defined and is given by

$$\Delta(\lambda) = A - \lambda I_X - BD^{-1}C = A^{\times} - \lambda I_X.$$

Now take $\lambda \in \rho(A)$. Then $A - \lambda I_X$ is invertible, and the Schur complement of $A - \lambda I_X$ in $M(\lambda)$ exists and is equal to $W(\lambda)$. It follows that $W(\lambda)$ is invertible if and only if $\Delta(\lambda)$ is invertible, that is, $W(\lambda)$ is invertible if and only if $\lambda \in \rho(A^{\times})$. Next assume that $\lambda \in \rho(A) \cap \rho(A^{\times})$. Then (2.5) and (2.6), specified for $M = M(\lambda)$, yield the inversion formulas in Theorem 2.1. The last two identities in the theorem are immediate from the intertwining relations (2.7). □

By applying the Schur complement results mentioned above to

$$M = \begin{bmatrix} -A & B \\ C & D \end{bmatrix} : X \dot{+} U \to X \dot{+} Y,$$

we obtain another useful identity. Indeed, assume that A and D are invertible. Then $D + CA^{-1}B$ is invertible if and only if $A + BD^{-1}C$ is invertible, and in this case (2.6) yields

$$\big(D + CA^{-1}B\big)^{-1} = D^{-1} - D^{-1}C\big(A + BD^{-1}C\big)^{-1}BD^{-1}.$$

2.3 Products

Let $\Theta_1 = (A_1, B_1, C_1, D_1; X_1, U_1, Y)$ and $\Theta_2 = (A_2, B_2, C_2, D_2; X_2, U, Y_2)$ be two systems such that the output space Y_2 of Θ_2 coincides with the input space U_1 of Θ_1. Let W_1 and W_2 be the transfer functions of Θ_1 and Θ_2, respectively. Because of the assumption $Y_2 = U_1$, the product $W(\lambda) = W_1(\lambda)W_2(\lambda)$ is well defined whenever $\lambda \in \rho(A_1) \cap \rho(A_2)$.

The next theorem shows how to obtain the product function $W = W_1W_2$ from Θ_1 and Θ_2. Let $\Theta = (A, B, C, D; X, U, Y)$ be the system built from Θ_1 and Θ_2 by putting $X = X_1 \dot{+} X_2$ and

$$\begin{aligned} A &= \begin{bmatrix} A_1 & B_1C_2 \\ 0 & A_2 \end{bmatrix} : X_1 \dot{+} X_2 \to X_1 \dot{+} X_2, \\ B &= \begin{bmatrix} B_1D_2 \\ B_2 \end{bmatrix} : U \to X_1 \dot{+} X_2, \\ C &= \begin{bmatrix} C_1 & D_1C_2 \end{bmatrix} : X_1 \dot{+} X_2 \to Y, \\ D &= D_1D_2 : U \to Y. \end{aligned}$$

The system Θ is called the *product* of Θ_1 and Θ_2 and is denoted by $\Theta = \Theta_1\Theta_2$.

Theorem 2.2. *Let*

$$\Theta_1 = (A_1, B_1, C_1, D_1; X_1, U_1, Y), \quad \Theta_2 = (A_2, B_2, C_2, D_2; X_2, U, Y_2)$$

be two systems and assume that the output space Y_2 of Θ_2 coincides with the input space U_1 of Θ_1. Write W_1, W_2 and W for the transfer function of Θ_1, Θ_2 and $\Theta = \Theta_1\Theta_2$, respectively. Then

$$W(\lambda) = W_1(\lambda)W_2(\lambda), \qquad \lambda \in \rho(A_1) \cap \rho(A_2) \subset \rho(A),$$

where A is the main operator of Θ.

Proof. Take $\lambda \in \rho(A_1) \cap \rho(A_2)$. Then, as can be verified by direct computation,

$$(\lambda - A)^{-1} = \begin{bmatrix} (\lambda - A_1)^{-1} & H(\lambda) \\ 0 & (\lambda - A_2)^{-1} \end{bmatrix} : X_1 \dot{+} X_2 \to X_1 \dot{+} X_2,$$

where $H(\lambda) = -(\lambda - A_1)^{-1} B_1 C_2 (\lambda - A_2)^{-1}$. Using this and the expressions for B, C and D given prior to the theorem, we have $W(\lambda) = D + C(\lambda - A)^{-1}B$. The right-hand side of the latter identity transforms into

$$\begin{aligned} & D_1 D_2 + \begin{bmatrix} C_1 & D_1 C_2 \end{bmatrix} \begin{bmatrix} (\lambda - A_1)^{-1} & H(\lambda) \\ 0 & (\lambda - A_2)^{-1} \end{bmatrix} \begin{bmatrix} B_1 D_2 \\ B_2 \end{bmatrix} \\ &= D_1 D_2 + \begin{bmatrix} C_1 (\lambda - A_1)^{-1} & C_1 H(\lambda) + D_1 C_2 (\lambda - A_2)^{-1} \end{bmatrix} \begin{bmatrix} B_1 D_2 \\ B_2 \end{bmatrix} \\ &= \big(D_1 + C_1(\lambda - A_1)^{-1} B_1\big)\big(D_2 + C_2(\lambda - A_2)^{-1} B_2\big). \end{aligned}$$

Thus $W(\lambda) = W_1(\lambda) W_2(\lambda)$, as desired. □

Note that the product $W_1 W_2$ is defined on the intersection $\rho(A_1) \cap \rho(A_2)$ of the resolvent sets of A_1 and A_2, whereas W is defined for $\lambda \in \rho(A)$. We have $\rho(A_1) \cap \rho(A_2) \subset \rho(A)$, and in general this inclusion is strict. Equality occurs when, for instance, $\rho(A)$ is connected or $\sigma(A_1) \cap \sigma(A_2) = \emptyset$. Note that $\rho(A_1) \cap \rho(A_2)$ is a neighborhood of infinity.

One verifies easily that $(\Theta_1 \Theta_2)^\times \simeq \Theta_2^\times \Theta_1^\times$, the natural identification of $X_1 \dot{+} X_2$ and $X_2 \dot{+} X_1$ being a system similarity between $(\Theta_1 \Theta_2)^\times$ and $\Theta_2^\times \Theta_1^\times$.

Modulo standard identifications of direct sums of Banach spaces, the product of systems is associative. So, when systems $\Theta_1, \ldots, \Theta_k$ are given, one can unambiguously define the product $\Theta_1 \cdots \Theta_k$. In general, one has to assume that the appropriate output and input spaces are coinciding, so that (in particular) the product of the external operators in question is well defined. We give details for the situation where all the input and output spaces are one and the same, while all given systems $\Theta_1, \ldots, \Theta_k$ are unital.

For $j = 1, \ldots, k$ write $\Theta_j = (A_j, B_j, C_j; X_j, Y)$, and introduce

$$A = \begin{bmatrix} A_1 & B_1 C_2 & \cdots & B_1 C_n \\ 0 & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & B_{n-1} C_n \\ 0 & \cdots & 0 & A_n \end{bmatrix} : X_1 \dot{+} \cdots \dot{+} X_k \to X_1 \dot{+} \cdots \dot{+} X_k,$$

$$B = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix} : Y \to X_1 \dot{+} \cdots \dot{+} X_k \to X_1 \dot{+} \cdots \dot{+} X_k,$$

$$C = \begin{bmatrix} C_1 \, C_2 \cdots C_n \end{bmatrix} : X_1 \dot{+} \cdots \dot{+} X_k \to Y.$$

The product of the (unital) systems $\Theta_1, \ldots, \Theta_k$, in that order, is now the (unital) system

$$\Theta_1 \cdots \Theta_k = (A, B, C; X_1 \dot{+} \cdots \dot{+} X_k, Y).$$

Again the transfer function of the product is the product of the transfer functions:

$$W_{\Theta_1 \cdots \Theta_k}(\lambda) = W_{\Theta_1}(\lambda) \cdots W_{\Theta_k}(\lambda), \qquad \lambda \in \rho(A) \subset \bigcap_{j=1}^{k} \rho(A_j).$$

This follows by a repeated application of Theorem 2.2.

2.4 Factorization and matching of invariant subspaces

In this section we study factorization of biproper systems and their transfer functions. The main theorem will serve as the basis for the more involved factorization results to be given in the sequel. Subspaces of Banach spaces are always assumed to be closed , otherwise we use the term linear manifold.

Theorem 2.3. *Let $\Theta = (A, B, C, D; X, Y)$ be a biproper system, let M and $M^\times$ be subspaces of X, and assume*

$$X = M \dot{+} M^\times. \tag{2.8}$$

Write

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \qquad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \qquad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \tag{2.9}$$

for the operator matrix representations of A, B and C with respect to the decomposition $X = M \dot{+} M^\times$. Assume $D = D_1 D_2$, where D_1 and D_2 are invertible operators on Y and introduce

$$\Theta_1 = (A_{11}, B_1 D_2^{-1}, C_1, D_1; M, Y), \tag{2.10}$$

$$\Theta_2 = (A_{22}, B_2, D_1^{-1} C_2, D_2; M^\times, Y). \tag{2.11}$$

Then $\Theta = \Theta_1 \Theta_2$ if and only if

$$A[M] \subset M, \qquad A^\times[M^\times] \subset M^\times, \tag{2.12}$$

where, as before, $A^\times = A - BD^{-1}C$. In that case the transfer function W_Θ admits the factorization

$$W_\Theta(\lambda) = W_{\Theta_1}(\lambda)W_{\Theta_2}(\lambda), \qquad \lambda \in \rho(A_{11}) \cap \rho(A_{22}) \subset \rho(A).$$

The above identity, holding on $\rho(A_{11}) \cap \rho(A_{22})$, can be rewritten as

$$D + C(\lambda - A)^{-1}B = \big(D_1 + C_1(\lambda - A_{11})^{-1}B_1\big)\big(D_2 + C_2(\lambda - A_{22})^{-1}B_2\big).$$

The left-hand side of this expression is defined and analytic on $\rho(A)$, while the two factors in the right-hand side are defined and analytic on the sets $\rho(A_{11})$ and $\rho(A_{22})$, respectively. In particular, the factors may be defined and analytic on domains where the left-hand side is not. This will turn out to be relevant in applications. For a more detailed discussion, see Section 2.5 (cf., the remark made after Theorem 2.5 below).

We shall refer to (2.8) as the *matching condition*, and when this condition is satisfied we refer to $M, M^\times$ as a pair of matching subspaces. A pair of matching subspaces $M, M^\times$ satisfying (2.12) will be called a *supporting pair of subspaces* for Θ.

Proof. The first part of this theorem is an immediate consequence of the definition of the product of two systems. The details are as follows.

Assume $\Theta = \Theta_1\Theta_2$. Then we know from the definition of the product that M is invariant under A. Identifying $M \dot{+} M^\times$ and $M^\times \dot{+} M$, we have $\Theta^\times = \Theta_2^\times\Theta_1^\times$, and hence we conclude that $M^\times$ is invariant under $A^\times$. This proves the only if part of the theorem.

To prove the if part, we argue as follows. The fact that M is invariant under A implies that $A_{21} = 0$. As

$$A^\times = A - BD^{-1}C = \begin{bmatrix} A_{11} - B_1D_2^{-1}D_1^{-1}C_1 & A_{12} - B_1D_2^{-1}D_1^{-1}C_2 \\ -B_2D_2^{-1}D_1^{-1}C_1 & A_{22} - B_2D_2^{-1}D_1^{-1}C_1 \end{bmatrix}$$

leaves the space $M^\times$ invariant, we have $A_{12} = B_1D_2^{-1}D_1^{-1}C_2$. But then the conclusion $\Theta = \Theta_1\Theta_2$ follows directly from the definition of the product of two systems.

The second statement in the theorem follows immediately from the first and Theorem 2.2. □

Elaborating on Theorem 2.3, we consider the case when the input/output space Y is finite-dimensional. In that case the second part of (2.12) is equivalent to a rank condition, an observation that will be used in an essential way in Section 9.3. Here are the details.

Proposition 2.4. *Let $\Theta = (A, B, C, D; X, Y)$ be a biproper system, let M and $M^\times$ be subspaces of X, and assume $X = M \dot{+} M^\times$. Suppose, in addition, that the*

dimension $\dim Y$ *of the input/output space* Y *is finite. Then* $A^\times M^\times \subset M^\times$ *if and only if*

$$\operatorname{rank}\begin{bmatrix} A_{12} & B_1 \\ C_2 & D \end{bmatrix} = \dim Y. \tag{2.13}$$

Here A_{12}, B_1 *and* C_2 *are as in* (2.9).

Proof. To see this we use Schur complements (cf., Section 2.2). Since D is invertible, we can use formula (2.3) to show that

$$\operatorname{rank}\begin{bmatrix} A_{12} & B_1 \\ C_2 & D \end{bmatrix} = \operatorname{rank} D + \operatorname{rank}\big(A_{12} - B_1 D^{-1} C_2\big).$$

Now rank D is equal to the dimension of Y which is assumed to be finite. Thus (2.13) amounts to $A_{12} - B_1 D^{-1} C_2 = 0$ which, in turn, is equivalent to the second part of (2.12). □

In a certain sense Theorem 2.3 gives a complete description of all possible factorizations of a system Θ. Indeed, if $\Theta \simeq \Theta'_1 \Theta'_2$ for some systems Θ'_1 and Θ'_2 having invertible external operators, then there exists a supporting pair of subspaces $M, M^\times$ for Θ such that $\Theta_1 \simeq \Theta'_1$ and $\Theta_2 \simeq \Theta'_2$, where Θ_1 and Θ_2 are as in Theorem 2.3. In this sense Theorem 2.3 gives a complete description of all possible factorizations of Θ.

Matching pairs of subspaces correspond to projections. So Theorem 2.3 can also be formulated in terms of projections. In fact, we have the following result.

Theorem 2.5. *Let* $\Theta = (A, B, C, D; X, Y)$ *be a biproper system, and let* W *be its transfer function. Put* $A^\times = A - BD^{-1}C$, *and let* Π *be a projection on* X *such that*

$$A[\operatorname{Ker}\Pi] \subset \operatorname{Ker}\Pi, \qquad A^\times[\operatorname{Im}\Pi] \subset \operatorname{Im}\Pi. \tag{2.14}$$

Assume $D = D_1 D_2$, *where* D_1 *and* D_2 *are invertible operators on* Y *and introduce, for* $\lambda \in \rho(A)$,

$$\begin{aligned} W_1(\lambda) &= D_1 + C(\lambda - A)^{-1}(I - \Pi)BD_2^{-1}, \\ W_2(\lambda) &= D_2 + D_1^{-1}C\Pi(\lambda - A)^{-1}B. \end{aligned}$$

Then $W(\lambda) = W_1(\lambda)W_2(\lambda)$ *for all* $\lambda \in \rho(A)$.

This factorization can be rewritten as

$$\begin{aligned} D + C(\lambda - A)^{-1}B = \big(D_1 + C(\lambda - A)^{-1}(I - \Pi)BD_2^{-1}\big)\times \\ \times\big(D_2 + D_1^{-1}C\Pi(\lambda - A)^{-1}B\big). \end{aligned}$$

It holds on $\rho(A)$, the resolvent set of A. However, in many cases (relevant for applications), the factors in the right-hand side have an analytic extension to

larger domain. This is already suggested by Theorem 2.3. In fact, one may wonder how exactly the two factorization results Theorems 2.3 and 2.5 relate to each other. We shall discuss this point in detail in Section 2.5 below.

A projection Π satisfying (2.14) will be called a *supporting projection* for the system Θ.

Proof. For $\lambda \in \rho(A)$, we have

$$\begin{aligned} W_1(\lambda)W_2(\lambda) &= D + C(\lambda - A)^{-1}(I - \Pi)B + C\Pi(\lambda - A)^{-1}B + \\ &\quad + C(\lambda - A)^{-1}(I - \Pi)BD^{-1}C\Pi(\lambda - A)^{-1}B \\ &= D + C(\lambda - A)^{-1}(I - \Pi)B + C\Pi(\lambda - A)^{-1}B + \\ &\quad + C(\lambda - A)^{-1}(I - \Pi)(A - A^\times)\Pi(\lambda - A)^{-1}B. \end{aligned}$$

Now $\Pi A = \Pi A \Pi$ and $A^\times \Pi = \Pi A^\times \Pi$, hence $(I - \Pi)A^\times \Pi = 0$ and

$$(I - \Pi)(A - A^\times)\Pi = A\Pi - \Pi A = \Pi(\lambda - A) - (\lambda - A)\Pi.$$

From this, the desired identity is immediate. □

Suppose $\Theta = (A, B, C; X, Y)$ is a unital system, so the external operator of Θ is $I = I_Y$. Let Π be a supporting projection for Θ. With respect to the decomposition $X = \operatorname{Ker}\Pi \dot{+} \operatorname{Im}\Pi$, write A, B, C as in (2.9). The system

$$\operatorname{pr}_\Pi(\Theta) = (A_{22}, B_2, C_2; \operatorname{Im}\Pi, Y) \tag{2.15}$$

will be called the *projection* of Θ associated with Π (the terminology is taken from [30]). Observe that

$$\operatorname{pr}_{I-\Pi}(\Theta) = (A_{11}, B_1, C_1; \operatorname{Ker}\Pi, Y). \tag{2.16}$$

One easily verifies that $\operatorname{pr}_\Pi(\Theta^\times) = \operatorname{pr}_\Pi(\Theta)^\times$. Note that (2.15) and (2.16) are defined for any projection Π of the state space X. By Theorem 2.3, the projection Π is a supporting projection for the system Θ if and only if $\Theta = \operatorname{pr}_{I-\Pi}(\Theta)\operatorname{pr}_\Pi(\Theta)$. In fact, the following slightly more general theorem, involving a product of possibly more than two factors (see the end of Section 2.3), holds true.

Theorem 2.6. *Let $\Theta = (A, B, C; X, Y)$ be a unital system (i.e., the external operator is the identity on Y), and let $\Pi_1, \dots, \Pi_n$ be mutually disjoint projections of X such that $\Pi_1 + \cdots + \Pi_n$ is the identity on X. Then*

$$\Theta = \operatorname{pr}_{\Pi_1}(\Theta)\operatorname{pr}_{\Pi_2}(\Theta)\cdots\operatorname{pr}_{\Pi_n}(\Theta)$$

if and only if for $j = 1, \dots, n-1$, the projection $\Pi_{j+1} + \cdots + \Pi_n$ is a supporting projection for Θ.

Proof. To prove the theorem one can employ the same arguments as in the prove of Theorem 2.3. Of course the decomposition $X = \operatorname{Ker}\Pi \dot{+} \operatorname{Im}\Pi$ has to be replaced by the decomposition $X = X_1 \dot{+} \cdots \dot{+} X_n$, where $X_j = \operatorname{Im}\Pi_j$, and with respect to the latter decomposition one writes A, B and C in block matrix form. □

Theorem 2.6 is formulated as a factorization result for systems, this in line with the first part of Theorem 2.3. We could as well have stated it as a factorization result for transfer functions, thereby generalizing the second part of Theorem 2.3 or Theorem 2.5.

As an application of Theorem 2.6 we prove the following result.

Theorem 2.7. *Let $\Theta = (A, B, C; X, \mathbb{C}^m)$ be a unital system with a finite-dimensional state space X, and let W be the transfer function of Θ. Assume that A is diagonalizable. Then W admits a factorization of the following form*

$$W(\lambda) = (I_m + \frac{1}{\lambda - \lambda_1} R_1) \cdots (I_m + \frac{1}{\lambda - \lambda_n} R_n),$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A counted according to algebraic multiplicity, and $R_1, \ldots, R_n$ are $m \times m$ matrices of rank at most one.

Recall that A is called *diagonalizable* if A is similar to a diagonal matrix. In other words, A is diagonalizable if and only if its Jordan matrix is diagonal.

Proof. Since A is diagonalizable, we can find a basis $e_1, \ldots, e_n$ of the finite-dimensional space X such that the matrix of A with respect to this basis is diagonal, say

$$A = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

Here $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A counted according to algebraic multiplicity. Next, we consider the associate main operator $A^\times$. We can choose a basis $f_1, \ldots, f_n$ of X such that the matrix of $A^\times$ has lower triangular form. Then clearly f_n is an eigenvector of $A^\times$. We may assume that the vectors $e_1, \ldots, e_n$ are ordered in such a way that

$$X = \operatorname{span}\{e_1, \ldots, e_{n-1}\} \dot{+} \operatorname{span}\{f_n\}.$$

Here $\operatorname{span}\{V\}$ denotes the linear hull of V. For convenience we put

$$X_0 = \operatorname{span}\{e_1, \ldots, e_{n-1}\}, \qquad X_n = \operatorname{span}\{f_n\}.$$

Clearly, $X = X_0 \dot{+} X_n$, the space X_0 is invariant under A, and the space X_n is invariant under $A^\times$.

Let Π be the projection of X onto X_n along X_0. Then Π is a supporting projection for Θ. Let $W = W_0 W_n$ be the corresponding factorization of W. Then W_n is

the transfer function of the node $\Theta_n = \mathrm{pr}_{\Pi}(\Theta)$. Write $\Theta_n = (A_n, B_n, C_n; X_n, \mathbb{C}^m)$. The matrix of A with respect to the basis $\{e_1, \ldots, e_{n-1}, f_n\}$ of X is given by

$$A = \begin{bmatrix} \lambda_1 & & & * \\ & \ddots & & \vdots \\ & & \lambda_{n-1} & * \\ & & & a_n \end{bmatrix},$$

and it follows that $a_n = \lambda_n$. Now $a_n = \lambda_n$ may be viewed as the matrix of the operator $A_n : X_n \to X_n$ with respect to the basis singleton $\{f_n\}$ of X_n. Hence $\sigma(A_n) = \{\lambda_n\}$ and

$$W_n(\lambda) = I + \frac{1}{\lambda - \lambda_n} R_n,$$

where R_n is an operator on $\mathbb{C}^m$ of rank at most one.

Next consider the factor W_0 which is the transfer function of the system $\Theta_0 = p_{I-\Pi}(\Theta)$. Write $\Theta_0 = (A_0, B_0, C_0; X_0, \mathbb{C}^n)$. Then A_0 is the restriction of A to $X_0 = \mathrm{span}\,\{e_1, \ldots, e_{m-1}\}$. Note that A_0 is again diagonalizable. Therefore we can repeat the above argument with W_0 and Θ_0 in place of W and Θ, respectively. In a finite number of steps, we thus obtain the desired result. □

A factorization of the type appearing in Theorem 2.7 is called a *factorization into elementary factors*. An in depth analysis of such factorizations, including connections with problems of job scheduling, will given in Part III of this book. See also Theorem 8.15 for an alternative version of Theorem 2.7.

2.5 Factorization and inversion revisited

The previous section contains two factorization results: Theorems 2.3 and 2.5. These theorems contain different expressions for the factors, and they also feature different domains on which the factorizations are valid. For systems with a finite-dimensional state space the differences are not substantial. In the infinite-dimensional case, however, the situation is more involved. We shall now analyze the situation in detail by presenting a synthesis of Theorems 2.3 and 2.5. Along the way, we will also clarify the relationship between these factorization results on the one hand and the inversion result Theorem 2.1 on the other. The analysis in question should be kept in mind whenever the results of the previous two sections are applied.

It is convenient to fix some notation. Throughout $\Theta = (A, B, C, D; X, Y)$ stands for a biproper system. Recall that a projection Π on X is a supporting projection for Θ if $\mathrm{Ker}\,\Pi$ is A-invariant and $\mathrm{Im}\,\Pi$ is $A^\times$-invariant. Clearly this is equivalent to the requirement that the complementary projection $I - \Pi$ is a supporting projection for $\Theta^\times = (A^\times, BD^{-1}, -D^{-1}C, D^{-1}; X, Y)$, the inverse or associate system of Θ. Here, as usual, $A^\times = A - BD^{-1}C$.

Theorem 2.8. *Let W and $W^\times$ be the transfer functions of the (biproper) systems Θ and $\Theta^\times$, respectively, i.e.,*

$$\begin{aligned} W(\lambda) &= D + C(\lambda - A)^{-1}B, && \lambda \in \rho(A), \\ W^\times(\lambda) &= D^{-1} - D^{-1}C(\lambda - A^\times)^{-1}BD^{-1}, && \lambda \in \rho(A^\times). \end{aligned}$$

Suppose Π is a supporting projection for Θ or, equivalently, $I - \Pi$ is a supporting projection for $\Theta^\times$. Write

$$A = \begin{bmatrix} A_1 & A_0 \\ 0 & A_2 \end{bmatrix}, \quad A^\times = \begin{bmatrix} A_1^\times & 0 \\ A_0^\times & A_2^\times \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}$$

for the operator matrix representations of $A, A^\times, B$ and C with respect to the decomposition $X = \operatorname{Ker}\Pi \dotplus \operatorname{Im}\Pi$, thus, in particular,

$$A_1^\times = A_1 - B_1D^{-1}C_1, \quad A_2^\times = A_2 - B_2D^{-1}C_2.$$

Assume $D = D_1D_2$, where D_1 and D_2 are invertible operators on Y, and introduce

$$\widetilde{W}_1(\lambda) = \begin{cases} D_1 + C(\lambda - A)^{-1}(I - \Pi)BD_2^{-1}, & \lambda \in \rho(A), \\ D_1 + C_1\big(\lambda - A_1\big)^{-1}B_1D_2^{-1}, & \lambda \in \rho(A_1), \end{cases}$$

$$\widetilde{W}_2(\lambda) = \begin{cases} D_2 + D_1^{-1}C\Pi(\lambda - A)^{-1}B, & \lambda \in \rho(A), \\ D_2 + D_1^{-1}C_2\big(\lambda - A_2\big)^{-1}B_2, & \lambda \in \rho(A_2), \end{cases}$$

$$\widetilde{W}_1^\times(\lambda) = \begin{cases} D_1^{-1} - D_1^{-1}C(I - \Pi)(\lambda - A^\times)^{-1}BD^{-1}, & \lambda \in \rho(A^\times), \\ D_1^{-1} - D_1^{-1}C_1\big(\lambda - A_1^\times\big)^{-1}B_1D^{-1}, & \lambda \in \rho(A_1^\times), \end{cases}$$

$$\widetilde{W}_2^\times(\lambda) = \begin{cases} D_2^{-1} - D^{-1}C(\lambda - A^\times)^{-1}\Pi BD_2^{-1}, & \lambda \in \rho(A^\times), \\ D_2^{-1} - D^{-1}C_2\big(\lambda - A_2^\times\big)^{-1}B_2D_2^{-1}, & \lambda \in \rho(A_2^\times). \end{cases}$$

The following statements hold true:

(i) *The functions $\widetilde{W}_1$, $\widetilde{W}_2$ are well defined and analytic on their domains $\Omega_1 = \rho(A) \cup \rho(A_1)$, $\Omega_2 = \rho(A) \cup \rho(A_2)$, respectively, and*

$$W(\lambda) = \widetilde{W}_1(\lambda)\widetilde{W}_2(\lambda), \quad \lambda \in \Omega_1 \cap \Omega_2 = \rho(A).$$

Similarly, the functions $\widetilde{W}_1^\times$, $\widetilde{W}_2^\times$ are well defined and analytic on their domains $\Omega_1^\times = \rho(A^\times) \cup \rho(A_1^\times)$, $\Omega_2^\times = \rho(A^\times) \cup \rho(A_2^\times)$, respectively, and

$$W^\times(\lambda) = \widetilde{W}_2^\times(\lambda)\widetilde{W}_1^\times(\lambda), \quad \lambda \in \Omega_1^\times \cap \Omega_2^\times = \rho(A^\times).$$

(ii) *The operators* $W(\lambda)$ *and* $W^\times(\lambda)$ *are invertible for the same values of* λ *and for these they are each others inverse. In fact,*

$$\{\lambda \in \rho(A) \mid W(\lambda) \text{ invertible}\} = \{\lambda \in \rho(A^\times) \mid W^\times(\lambda) \text{ invertible}\}$$
$$= \rho(A) \cap \rho(A^\times),$$

and for λ *in these coinciding sets we have* $W(\lambda)^{-1} = W^\times(\lambda)$.

(iii) *The sets* $\rho(A) \setminus \rho(A_1)$, $\rho(A) \setminus \rho(A_2)$ *and* $\rho(A) \setminus \big(\rho(A_1) \cap \rho(A_2)\big)$ *coincide; also, for* λ *in one (and hence all) of these sets, the operators*

$$\begin{aligned} \widetilde{W}_1(\lambda) &= D_1 + C(\lambda - A)^{-1}(I - \Pi)BD_2^{-1}, \\ \widetilde{W}_2(\lambda) &= D_2 + D_1^{-1}C\Pi(\lambda - A)^{-1}B \end{aligned}$$

are not invertible. Similarly, the sets $\rho(A^\times) \setminus \rho(A_1^\times)$, $\rho(A^\times) \setminus \rho(A_2^\times)$ *and* $\rho(A^\times) \setminus \big(\rho(A_1^\times) \cap \rho(A_2^\times)\big)$ *coincide; also, for* λ *in one (and hence all) of these sets, the operators*

$$\begin{aligned} \widetilde{W}_1^\times(\lambda) &= D_1^{-1} - D_1^{-1}C(I - \Pi)(\lambda - A^\times)^{-1}BD^{-1}, \\ \widetilde{W}_2^\times(\lambda) &= D_2^{-1} - D^{-1}C(\lambda - A^\times)^{-1}\Pi BD_2^{-1} \end{aligned}$$

are not invertible.

(iv) *The operators* $\widetilde{W}_1(\lambda)$ *and* $\widetilde{W}_1^\times(\lambda)$ *are invertible for the same values of* λ *and for these they are each others inverse. In fact,*

$$\{\lambda \in \Omega_1 \mid \widetilde{W}_1(\lambda) \text{ invertible}\} = \{\lambda \in \Omega_1^\times \mid \widetilde{W}_1^\times(\lambda) \text{ invertible}\}$$
$$= \rho(A_1) \cap \rho(A_1^\times)$$

and, for λ *in these coinciding sets,* $\widetilde{W}_1(\lambda)^{-1} = \widetilde{W}_1^\times(\lambda)$. *Analogously, the operators* $\widetilde{W}_2(\lambda)$ *and* $\widetilde{W}_2^\times(\lambda)$ *are invertible for the same values of* λ *and for these they are each others inverse. In fact,*

$$\{\lambda \in \Omega_2 \mid \widetilde{W}_2(\lambda) \text{ invertible}\} = \{\lambda \in \Omega_2^\times \mid \widetilde{W}_2^\times(\lambda) \text{ invertible}\}$$
$$= \rho(A_2) \cap \rho(A_2^\times)$$

and, for λ *in these coinciding sets,* $\widetilde{W}_2(\lambda)^{-1} = \widetilde{W}_2^\times(\lambda)$.

Theorem 2.8 contains the earlier factorization results as special cases. Indeed, for Theorem 2.5 restrict in (i) to $\rho(A)$, for Theorem 2.3 (second part), restrict to $\rho(A_1) \cap \rho(A_2)$. In general, the factors $\widetilde{W}_1(\lambda)$ and $\widetilde{W}_2(\lambda)$ appearing in (i) are defined and analytic on domains which are larger than $\rho(A)$. This is of significance for obtaining such special factorizations as those needed for solving Wiener-Hopf, Toeplitz or singular integral equations (cf., Chapter 6 below). In that context, it

is also necessary to have information on the sets where the factors take invertible values and to have expressions for the inverses. These issues are covered by (iii) and (iv). Statement (ii) is added for completeness and is a reformulation of part of Theorem 2.1.

In certain important cases, assertion (iii) is redundant (completely or partly) because the coinciding sets mentioned there are empty. Restricting ourselves to the first part of (iii), the point in question is the relationship between $\rho(A)$, $\rho(A_1)$ and $\rho(A_2)$. It is convenient to clear this issue up first. We begin by recording the following simple lemma (in which X_1 may be read as $\operatorname{Ker}\Pi$ and X_2 as $\operatorname{Im}\Pi$).

Lemma 2.9. *Let X_1 and X_2 be Banach spaces, and let*

$$A = \begin{bmatrix} A_1 & A_0 \\ 0 & A_2 \end{bmatrix} : X_1 \dot{+} X_2 \to X_1 \dot{+} X_2$$

be a bounded linear operator. Suppose two of the operators A_1, A_2 and A are invertible. Then so are all three of them.

The same conclusion is of course valid when the zero in the representation of A is in the upper right instead of in the lower left corner.

Proof. Our hypotheses implies that at least one of the operators A_1 and A_2 is invertible. Suppose A_1 is. Then the Schur complement of A_1 in A exists. In view op the (block) upper triangular form of A, this Schur complement is just the operator A_2. From the intermezzo on Schur complements in Section 2.2 it is now clear that A is invertible if and only if A_2 is. The case when A_2 is invertible, can be dealt with analogously: use that the Schur complement of A_2 in A is A_1. □

Next we pass to resolvent sets. In Theorem 2.3 we already came across the inclusion $\rho(A_1) \cap \rho(A_2) \subset \rho(A)$. Now we see from Lemma 2.9 that this inclusion can be made more precise as follows

$$\rho(A_1) \cap \rho(A_2) = \rho(A) \cap \rho(A_1) = \rho(A) \cap \rho(A_2). \tag{2.17}$$

From this it is clear that the three sets mentioned in the first part of statement (iii) of Theorem 2.8 do indeed coincide, that is,

$$\rho(A) \setminus \big(\rho(A_1) \cap \rho(A_2)\big) = \rho(A) \setminus \rho(A_1) = \rho(A) \setminus \rho(A_2). \tag{2.18}$$

These (coinciding) sets are empty if and only if $\rho(A) \subset \rho(A_1) \cap \rho(A_2)$ which, together with the inclusion mentioned in the beginning of this paragraph, comes down to $\rho(A) = \rho(A_1) \cap \rho(A_2)$ or, if one prefers, $\sigma(A) = \sigma(A_1) \cup \sigma(A_2)$. However, more strikingly, we see from (2.17) that the set determined by (2.17) is already empty under the weaker requirement $\rho(A) \subset \rho(A_1) \cup \rho(A_2)$. In terms of spectra, this condition may be rewritten as

$$\sigma(A_1) \cap \sigma(A_2) \subset \sigma(A).$$

We conclude that one relevant case where Theorem 2.8 (iii) is redundant occurs when the state space of the given system Θ is finite-dimensional (and hence one can work with matrices and determinants). Another such situation occurs in the important case of Wiener-Hopf factorization (see Chapter 6 below). The reason there is that the spectra of A_1 and A_2 are disjoint and likewise those of $A_1^\times$ and $A_2^\times$.

From (2.17) it is also clear that when Ω_1 and Ω_2 are as in Theorem 2.8 (i), then indeed, as is stated there, $\Omega_1 \cap \Omega_2 = \rho(A)$. The analogous identity $\Omega_1^\times \cap \Omega_2^\times = \rho(A^\times)$ comes about in the same way.

Proof of Theorem 2.8. Take λ in $\rho(A) \cap \rho(A_1) = \rho(A) \cap \rho(A_2)$. With

$$H(\lambda) = -\big(\lambda - A_1\big)^{-1} A_0 \big(\lambda - A_2\big)^{-1},$$

we have

$$\begin{aligned}(\lambda - A)^{-1}(I - \Pi) &= \begin{bmatrix} \big(\lambda - A_1\big)^{-1} & H(\lambda) \\ 0 & \big(\lambda - A_2\big)^{-1} \end{bmatrix} \begin{bmatrix} I_{\operatorname{Ker}\Pi} & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} (\lambda - A_1)^{-1} & 0 \\ 0 & 0 \end{bmatrix},\end{aligned}$$

which leads to

$$D_1 + C(\lambda - A)^{-1}(I - \Pi)BD_2^{-1} = D_1 + C_1(\lambda - A_1)^{-1}B_1D_2^{-1}.$$

Thus $\widetilde{W}_1$ is well defined. The same conclusion hold for $\widetilde{W}_2$. Indeed, for $\lambda \in \rho(A) \cap \rho(A_2)$ we have

$$\begin{aligned}\Pi(\lambda - A)^{-1} &= \begin{bmatrix} 0 & 0 \\ 0 & I_{\operatorname{Im}\Pi} \end{bmatrix} \begin{bmatrix} \big(\lambda - A_1\big)^{-1} & H(\lambda) \\ 0 & \big(\lambda - A_2\big)^{-1} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & (\lambda - A_2)^{-1} \end{bmatrix},\end{aligned}$$

and hence

$$D_2 + D_1^{-1}C\Pi(\lambda - A)^{-1}B \,=\, D_2 + D_1^{-1}C_2\big(\lambda - A_2\big)^{-1}B_2.$$

The analyticity of the functions $\widetilde{W}_1$ and $\widetilde{W}_2$ on, respectively, the (open) sets $\Omega_1 = \rho(A) \cup \rho(A_1)$ and $\Omega_2 = \rho(A) \cup \rho(A_2)$ is obvious. With respect to the factorization in the first part of (i), recall that $\Omega_1 \cap \Omega_2 = \rho(A)$, and use the conclusion of Theorem 2.5.

This proves the first part of statement (i). The second part is just the first, reformulated for the inverse system $\Theta^\times$. Further (ii) comes down to part of Theorem 2.1. So we can move on to (iii).

The equality of the sets in (iii) has already been established above. Take $\lambda \in \rho(A) \setminus \rho(A_1)$. Then

$$\widetilde{W}_1(\lambda) = D_1 + C(\lambda - A)^{-1}(I - \Pi)BD_2^{-1}.$$

Suppose $\widetilde{W}_1(\lambda)$ is invertible. By Theorem 2.1, this can only happen when λ belongs to the resolvent set of

$$\begin{aligned} A - (I-\Pi)BD_2^{-1}D_1^{-1}C &= A - (I-\Pi)BD^{-1}C \\ &= \Pi A + (I-\Pi)A^\times \\ &= \begin{bmatrix} 0 & 0 \\ 0 & A_2 \end{bmatrix} + \begin{bmatrix} A_1^\times & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} A_1^\times & 0 \\ 0 & A_2 \end{bmatrix}, \end{aligned}$$

and it follows that $\lambda \in \rho(A_1^\times) \cap \rho(A_2)$. Given the choice of λ, we now have that $\lambda \notin \rho(A_1)$ on the one hand and $\lambda \in \rho(A) \cap \rho(A_2)$ on the other. By (2.17), however, $\rho(A) \cap \rho(A_2) = \rho(A) \cap \rho(A_1)$. So $\lambda \in \rho(A_1)$, and we have arrived at a contradiction. Thus for the values of λ considered here, $\widetilde{W}_1(\lambda)$ is never invertible. A similar reasoning gives the same conclusion for $\widetilde{W}_2(\lambda) = D_2 + D_1^{-1}C\Pi(\lambda - A)^{-1}B$.

This proves the first part of statement (iii). The second part is just the first, reformulated for the inverse system $\Theta^\times$. The arguments for the two parts of (iv) are analogous. We concentrate on the first.

By (iii), the operator $\widetilde{W}_1(\lambda)$ is not invertible whenever λ belongs to the set $\rho(A) \setminus \rho(A_1)$. So we can restrict our attention to the complement of this set in $\Omega_1 = \rho(A) \cup \rho(A_1)$. This complement coincides with $\rho(A_1)$. Apply now Theorem 2.1 to the system $(A_1, B_1D_2^{-1}, C_1, D_1; \operatorname{Ker}\Pi, Y)$ which has the restriction of $\widetilde{W}_1$ to $\rho(A_1)$ as its the transfer function. The conclusion is that, for $\lambda \in \rho(A_1)$, the operator $\widetilde{W}_1(\lambda)$ is invertible if and only if λ belongs to $\rho(A_1) \cap \rho(A_1^\times)$. Also, for these values of λ,

$$\begin{aligned} \widetilde{W}_1(\lambda)^{-1} &= D_1^{-1} - D_1^{-1}C_1\big(\lambda - (A_1 - B_1D_2^{-1}D_1^{-1}C_1)\big)^{-1}B_1D^{-1} \\ &= D_1^{-1} - D_1^{-1}C_1\big(\lambda - A_1^\times\big)^{-1}B_1D^{-1}, \end{aligned}$$

in other words, $\widetilde{W}_1(\lambda)^{-1} = \widetilde{W}_1^\times(\lambda)$. □

In the above considerations, we came across the sets (2.17) and (2.18). Let us, for convenience, denote them by Ω and Ω_0:

$$\Omega = \rho(A_1) \cap \rho(A_2) = \rho(A) \cap \rho(A_1) = \rho(A) \cap \rho(A_2),$$

$$\Omega_0 = \rho(A) \setminus \big(\rho(A_1) \cap \rho(A_2)\big) = \rho(A) \setminus \rho(A_1) = \rho(A) \setminus \rho(A_2).$$

Here

$$A = \begin{bmatrix} A_1 & A_0 \\ 0 & A_2 \end{bmatrix}$$

as before. Without going into the proof, we note that the sets Ω and Ω_0 have a special structure in relation to $\rho(A)$. Indeed, Ω is the union of the connected components of $\rho(A)$ that have a nonempty intersection with both $\rho(A_1)$ and $\rho(A_2)$, and these are the connected components of Ω. Likewise, Ω_0 is the union of the connected components of $\rho(A)$ that do not intersect $\rho(A_1)$ or $\rho(A_2)$, and these are the connected components of Ω_0. As a consequence, the unbounded components of $\rho(A)$ and Ω coincide. In the finite-dimensional case, these unbounded components are the only ones that exist.

We illustrate Theorem 2.8 with an example exhibiting the different aspects of the result.

Example. Write $\mathbb{Z}$, $\mathbb{Z}_-$ and $\mathbb{Z}_+$ for the set of integers, (strictly) negative integers and non-negative integers (including zero) , respectively. The system Θ that we will consider has $\ell_1(\mathbb{Z})$ for its state space, $\mathbb{C}$ for its input/output space and the identity operator on $\mathbb{C}$ as external operator. The other operators in $\Theta = (A, B, C; \ell_1(\mathbb{Z}), \mathbb{C})$ are

$$A : \ell_1(\mathbb{Z}) \to \ell_1(\mathbb{Z}),$$

$$(Ax)_j = x_{j+1}, \qquad x \in \ell_1(\mathbb{Z}),\ j \in \mathbb{Z},$$

$$B : \mathbb{C} \to \ell_1(\mathbb{Z}),$$

$$(Bz)_{-1} = z, \quad (Bz)_1 = -z, \qquad (Bz)_j = 0, \quad j \in \mathbb{Z},\ j \neq -1, 1,$$

$$C : \ell_1(\mathbb{Z}) \to \mathbb{C},$$

$$Cx = x_0 - (x_{-2} + x_{-3} + x_{-4} + \cdots), \qquad x \in \ell_1(\mathbb{Z}).$$

We refrain from giving the analogous expressions for $A^\times = A - BC$ as they can be obtained directly from those for A, B and C.

The spaces $\ell_1(\mathbb{Z}_+)$ and $\ell_1(\mathbb{Z}_-)$ will be viewed in the customary manner as subspaces of $\ell_1(\mathbb{Z})$. Doing this, we have the direct sum decomposition

$$\ell_1(\mathbb{Z}) = \ell_1(\mathbb{Z}_-) \dot{+} \ \ell_1(\mathbb{Z}_+). \tag{2.19}$$

As is easily verified $\ell_1(\mathbb{Z}_-)$ is an invariant subspace for A, and $\ell_1(\mathbb{Z}_+)$ is an invariant subspace for $A^\times$. So the projection Π of $\ell_1(\mathbb{Z})$ along $\ell_1(\mathbb{Z}_-)$ onto $\ell_1(\mathbb{Z}_+)$ is a supporting projection for Θ. Also $I - \Pi$ is a supporting projection for $\Theta^\times = (A^\times, B, -C; \ell_1(\mathbb{Z}), \mathbb{C})$.

We shall now explain what Theorem 2.8 means for the situation specified above, thereby taking for D_1 and D_2 the identity operator on $\mathbb{C}$. In line with the theorem, we write

$$\begin{aligned} A &= \begin{bmatrix} A_1 & A_0 \\ 0 & A_2 \end{bmatrix} : \ell_1(\mathbb{Z}_-) \dot{+} \ell_1(\mathbb{Z}_+) \to \ell_1(\mathbb{Z}_-) \dot{+} \ell_1(\mathbb{Z}_+), \\ A^\times &= \begin{bmatrix} A_1^\times & 0 \\ A_0^\times & A_2^\times \end{bmatrix} : \ell_1(\mathbb{Z}_-) \dot{+} \ell_1(\mathbb{Z}_+) \to \ell_1(\mathbb{Z}_-) \dot{+} \ell_1(\mathbb{Z}_+), \\ B &= \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} : \mathbb{C} \to \ell_1(\mathbb{Z}_-) \dot{+} \ell_1(\mathbb{Z}_+), \\ C &= \begin{bmatrix} C_1 & C_2 \end{bmatrix} : \ell_1(\mathbb{Z}_-) \dot{+} \ell_1(\mathbb{Z}_+) \to \mathbb{C}. \end{aligned}$$

Our first task is to determine the spectra of $A_1, A_2, A, A_1^\times, A_2^\times$ and $A^\times$.

First note that A_1 and A_2 are unilateral shifts. So, as is well known, these operators have the closed unit disc $\mathbb{D}$ as their spectrum. Since A is the (bilateral) backward shift on $\ell_1(\mathbb{Z})$, the spectrum of A is $\mathbb{T}$, the unit circle in the complex plane.

Next consider $A_1^\times : \ell_1(\mathbb{Z}_-) \to \ell_1(\mathbb{Z}_-)$. For $x \in \ell_1(\mathbb{Z}_+)$ we have

$$\begin{aligned} (A_1^\times x)_{-1} &= x_{-2} + x_{-3} + x_{-4} + \cdots, \\ A_1^\times x)_j &= x_{j+1}, \quad j = -2, -3, \ldots, \end{aligned}$$

and so, modulo the standard identification of $\ell_1(\mathbb{Z}_-)$ and $\ell_1(\mathbb{Z}_+)$, it is the "Fibonacci operator" featuring in [110], Section V.4, Problem 11. Thus, as is stated there, its spectrum is $\mathbb{D} \cup \{\phi\}$ with $\phi = \frac{1}{2} + \frac{1}{2}\sqrt{5}$ (golden ratio). To see this, we argue as follows. For $|\lambda| < 1$, the $\ell_1(\mathbb{Z}_-)$-sequence $(\ldots, 0, 0, 0, 1)$ does not belong to the image of $\lambda - A_1^\times$. So $\mathbb{D} \subset \sigma(A_1^\times)$. Clearly, $A_1^\times$ is a rank one perturbation of a unilateral shift. So, for $|\lambda| > 1$, the operator $\lambda - A_1^\times$ is a rank one perturbation of an invertible operator, hence Fredholm of index zero. Therefore the only way for λ, taken outside the closed unit disc, to be in the spectrum of $A_1^\times$ is to be an eigenvalue of $A_1^\times$. It is easily verified that this is the case if and only if $\lambda^2 - \lambda - 1 = 0$, that is $\lambda = \phi$, and in that case, the essentially unique eigenvector associated with ϕ is the $\ell_1(\mathbb{Z}_-)$-sequence $\big(\ldots, \phi^{-3}, \phi^{-2}, \phi^{-1}, 1\big)$.

We now turn to $A_2^\times : \ell_1(\mathbb{Z}_+) \to \ell_1(\mathbb{Z}_+)$. If $x \in \ell_1(\mathbb{Z}_+)$, then

$$\begin{aligned}(A_2^\times x)_0 &= x_1,\\ (A_2^\times x)_1 &= x_0 + x_2,\\ (A_2^\times x)_j &= x_{j+1}, \quad j = 2, 3, 4, \ldots,\end{aligned}$$

and it is clear that $A_2^\times$ is a contraction. Hence $\sigma(A_2^\times) \subset \mathbb{D}$. Also, each λ in the open unit disc is an eigenvalue of $A_2^\times$ with eigenvector

$$\big(1,\ \lambda,\ (\lambda^2 - 1),\ \lambda(\lambda^2 - 1),\ \lambda^2(\lambda^2 - 1),\ \lambda^3(\lambda^2 - 1),\ \ldots\big).$$

As spectra are closed, it follows that $\sigma(A_2^\times) = \mathbb{D}$.

It remains to determine the spectrum of $A^\times$. From what we now know about $A_2^\times$ and the matrix representation of $A^\times$ with respect to the decomposition (2.19), it is clear that each λ in the open unit disc is an eigenvalue of $A^\times$. It follows that $\mathbb{D} \subset \sigma(A^\times)$, which can be rewritten as $\sigma(A_2^\times) \subset \sigma(A^\times)$. But then

$$\sigma(A^\times) = \sigma(A^\times) \cup \sigma(A_2^\times) = \sigma(A_1^\times) \cup \sigma(A_2^\times),$$

with the second identity based on (2.17), and hence $\sigma(A^\times) = \mathbb{D} \cup \{\phi\}$.

As an intermediate step and aid for the reader, we summarize the results obtained about the spectra of the operators $A, A_1, A_2, A^\times, A_1^\times$ and $A_2^\times$. With an eye on the formulation of Theorem 2.8, we do this in terms of their resolvent sets:

$$\begin{aligned}\rho(A) &= \{\lambda \in \mathbb{C} \mid |\lambda| \neq 1\},\\ \rho(A_1) &= \rho(A_2) = \rho(A_2^\times) = \{\lambda \in \mathbb{C} \mid |\lambda| > 1\},\\ \rho(A^\times) &= \rho(A_1^\times) = \{\lambda \in \mathbb{C} \mid |\lambda| > 1,\ \lambda \neq \phi\}\end{aligned}$$

where, as before, $\phi = \frac{1}{2} + \frac{1}{2}\sqrt{5}$. The different sets featuring in the theorem are now easy to determine. For instance, focussing on the first part of Theorem 2.8 (iii), the three coinciding sets $\rho(A) \setminus \rho(A_1)$, $\rho(A) \setminus \rho(A_2)$ and $\rho(A) \setminus \big(\rho(A_1) \cap \rho(A_2)\big)$ are all equal to the open unit disc.

Next we compute the transfer function W of Θ. Identifying operators on $\mathbb{C}$ with complex numbers (via the action of multiplication), W is a scalar function. The resolvent of A is given by

$$(\lambda - A)^{-1} = \begin{cases} \displaystyle\sum_{k=1}^{\infty} \lambda^{-k} A^{k-1}, & |\lambda| > 1,\\ \displaystyle\sum_{k=0}^{\infty} -\lambda^k S^{k+1}, & |\lambda| < 1.\end{cases}$$

Here S is the inverse of the operator A, i.e., $S = A^{-1}$ is the forward shift on $\ell_1(\mathbb{Z})$. For $|\lambda| > 1$, we now get

$$W(\lambda) = 1 + \sum_{k=1}^{\infty} \lambda^{-k} C A^{k-1} B = 1 - \frac{2}{\lambda^2} - \frac{1}{\lambda^3} = \frac{\lambda^3 - 2\lambda - 1}{\lambda^3}.$$

A similar argument, using that $CSB = 1$ and $CS^jB = 0$ for $j > 1$, yields that W vanishes on the interior of $\mathbb{D}$.

For the transfer function $W^\times$ of the system $\Theta^\times = \big(A^\times, B, -C; \ell_1(\mathbb{Z}), \mathbb{C}\big)$, we have

$$W^\times(\lambda) = W(\lambda)^{-1} = \frac{\lambda^3}{\lambda^3 - 2\lambda - 1}, \qquad \lambda \in \rho(A^\times) = \{\lambda \in \mathbb{C} \mid |\lambda| > 1, \lambda \neq \phi\}.$$

Bearing in mind that $\rho(A^\times) \subset \rho(A)$, the quickest way to see this is to use Theorem 2.8 (ii) or Theorem 2.1. If one prefers to avoid the use of these theorems, the statement can also be checked by computing the Laurent expansion of $W^\times$ at infinity from

$$W^\times(\lambda) = 1 - \sum_{k=1}^{\infty} \lambda^{-k} C \left(A^\times\right)^{k-1} B, \qquad |\lambda| > \phi,$$

and applying the uniqueness theorem for analytic functions.

We end the example by considering the factorizations of W and $W^\times$ induced by the decomposition (2.19) and the associated projections Π and $I - \Pi$. In other words, using the notation of Theorem 2.8, we analyze the situation with respect to $\widetilde{W}_1$, $\widetilde{W}_2$, $\widetilde{W}_1^\times$ and $\widetilde{W}_2^\times$. First let us consider $\widetilde{W}_1$ and $\widetilde{W}_2$.

The domain of $\widetilde{W}_1$ is

$$\Omega_1 = \rho(A) \cup \rho(A_1) = \{\lambda \in \mathbb{C} \mid |\lambda| \neq 1\}.$$

For $|\lambda| > 1$, we have

$$\begin{aligned} \widetilde{W}_1(\lambda) &= 1 + C_1\big(\lambda - A_1\big)^{-1} B_1 = 1 + \sum_{k=1}^{\infty} \lambda^{-k} C_1 A_1^{k-1} B_1 \\ &= 1 - \sum_{k=2}^{\infty} \lambda^{-k} = 1 - \frac{1}{\lambda^2}\left(\frac{1}{1 - \frac{1}{\lambda}}\right) = \frac{\lambda^2 - \lambda - 1}{\lambda^2 - \lambda}. \end{aligned}$$

Also, calculating $CS^j(I - \Pi)B$, we see that for $|\lambda| < 1$,

$$\widetilde{W}_1(\lambda) = 1 + C(\lambda - A)^{-1}(I - \Pi)B = 1 - \sum_{k=0}^{\infty} \lambda^k C S^{k+1}(I - \Pi)B = 0.$$

This is in line with Theorem 2.8 (iii). The domain Ω_2 of $\widetilde{W}_2$ is

$$\Omega_2 = \rho(A) \cup \rho(A_2) = \{\lambda \in \mathbb{C} \mid |\lambda| \neq 1\}.$$

For $|\lambda| > 1$, we have

$$\begin{aligned}\widetilde{W}_2(\lambda) &= 1 + C_2\big(\lambda - A_2\big)^{-1}B_2 = 1 + \sum_{k=1}^{\infty} \lambda^{-k} C_2 A_2^{k-1} B_2 \\ &= 1 - \frac{1}{\lambda^2} = \frac{\lambda^2 - 1}{\lambda^2}.\end{aligned}$$

Taking $|\lambda| < 1$ and computing $C\Pi S^j B$, we get

$$\widetilde{W}_2(\lambda) = 1 + C\Pi(\lambda - A)^{-1}B = 1 - \sum_{k=0}^{\infty} \lambda^k C\Pi S^{k+1} B = 0,$$

again in agreement with Theorem 2.8 (iii). In connection with Theorem 2.8 (i), we note that $W(\lambda) = \widetilde{W}_1(\lambda)\widetilde{W}_2(\lambda)$ on $\rho(A) = \{\lambda \in \mathbb{C} \mid |\lambda| \neq 1\}$. For values of λ in $\rho(A_1)$, i.e., for $|\lambda| > 1$, this is corroborated by the simple identity

$$\frac{\lambda^3 - 2\lambda - 1}{\lambda^3} = \left(\frac{\lambda^2 - \lambda - 1}{\lambda^2 - \lambda}\right)\left(\frac{\lambda^2 - 1}{\lambda^2}\right). \tag{2.20}$$

For values of λ in $\rho(A) \setminus \rho(A_1)$, i.e., for $|\lambda| < 1$, the factorization has the trivial form $0 = 0 \times 0$.

Next we turn to $\widetilde{W}_1^{\times}$ and $\widetilde{W}_2^{\times}$. The domain of $\widetilde{W}_1^{\times}$ is

$$\Omega_1^{\times} = \rho(A^{\times}) \cup \rho(A_1^{\times}) = \{\lambda \in \mathbb{C} \mid |\lambda| > 1,\ \lambda \neq \phi\},$$

and for λ in this set

$$\widetilde{W}_1^{\times}(\lambda) = \widetilde{W}_1(\lambda)^{-1} = \frac{\lambda^2 - \lambda}{\lambda^2 - \lambda - 1}.$$

A fast way to see this is via Theorem 2.8 (iv), but (if one wants to avoid the use of the theorem) one can also use the Laurent expansion of the resolvent $\big(\lambda - A_1^{\times}\big)^{-1}$ for $|\lambda| > \phi$ (cf., what was said about the computation of $W^{\times}$). The domain of $\widetilde{W}_2^{\times}$ is

$$\Omega_2^{\times} = \rho(A^{\times}) \cup \rho(A_2^{\times}) = \{\lambda \in \mathbb{C} \mid |\lambda| > 1\},$$

and for λ in this set

$$\widetilde{W}_1^{\times}(\lambda) = \widetilde{W}_1(\lambda)^{-1} = \frac{\lambda^2}{\lambda^2 - 1}.$$

For this, one can rely on Theorem 2.8 (iv), but again an alternative approach can be taken via the Laurent expansion of $\big(\lambda - A_2^{\times}\big)^{-1}$ for $|\lambda| > 1$. The factorization $W^{\times}(\lambda) = \widetilde{W}_2^{\times}(\lambda)\widetilde{W}_1^{\times}(\lambda)$ on

$$\rho(A^{\times}) = \{\lambda \in \mathbb{C} \mid |\lambda| > 1,\ \lambda \neq \phi\}$$

exhibited in Theorem 2.8 (i) is corroborated by taking reciprocals in (2.20). The second part of Theorem 2.8 (iii) is redundant because the three coinciding sets $\rho(A^\times) \setminus \rho(A_1^\times)$, $\rho(A^\times) \setminus \rho(A_2^\times)$ and $\rho(A^\times) \setminus \big(\rho(A_1^\times) \cap \rho(A_2^\times)\big)$ happen to be empty here. This finishes the example.

We conclude this section by comparing the two original forms that we have of the factorization principle – Theorems 2.3 and 2.5 – in light of what we have seen above. When invertibility of the factors plays a role, Theorem 2.3 is the more effective of the two. On the other hand, the representation of the factors in Theorem 2.5 is somewhat more straightforward than that in Theorem 2.3 and will be often used, tacitly having in mind the above considerations and Theorem 2.8. The latter is concerned with representations of the type $D + C(\lambda I - A)^{-1}B$ but, via the necessary modifications, it can be made to hold also for the more general realizations of the form $D + C(\lambda G - A)^{-1}B$ which are appropriate for handling non-proper functions (cf., Section 9.3). We will refrain from giving further details later on. To keep things in perspective: in dealing with rational matrix functions and finite-dimensional realizations, the finer details that are involved do not play a role.

Notes

This chapter is based on the text of the first chapter of [14]. Here the presentation of the material has been made more systematic. Some of the ideas are inspired by the theory of characteristic operator functions; see the references in the notes to the previous chapter. The final section is new. For linear fractional decompositions in state space form we refer to [81]. For a brief description of the history of the factorization principle presented in this chapter, we refer to the book I. Gohberg, M.A. Kaashoek (Eds), *Constructive methods of Wiener-Hopf factorization*, OT **21**, Birkhäuser Verlag, Basel, 1986.

Chapter 3

Various Classes of Systems

In this chapter we review the notions and results from the previous chapter for various classes of systems. Included are Brodskii systems (Section 3.1), Kreĭn systems (Section 3.2), unitary systems (Section 3.3), monic systems (Section 3.4) and polynomial systems (Section 3.5). The final section (Section 3.6) concerns a change of variable in the transfer function defined by a Möbius transform.

3.1 Brodskii systems

In this section we shall see how the results on inversion, products, and factorization obtained in the previous chapter apply to the Brodskii systems introduced in Section 1.2. By definition, a system $\Theta = (A, B, C; H, G)$ is a Brodskii J-system if H and G are Hilbert spaces, $J = J^* = J^{-1}$ and

$$A - A^* = BC, \qquad C = 2iJB^*.$$

A system which is similar to a Brodskii J-system need not be of this type, but it is a Brodskii J-system provided that the system similarity is a unitary operator (cf., [30], page 11). On the other hand, if two Brodskii J-systems are similar, say with system similarity S, then one can prove that there exists a unitary operator U that provides the similarity too. In fact for U one may take the unitary operator appearing in the polar decomposition $S = U\sqrt{S^*S}$ of S.

Let $\Theta = (A, B, C; H, G)$ be a Brodskii J-system. As the external operator is equal to the identity operator on G, the associate system $\Theta^\times = (A^\times, B, -C; H, G)$ is well defined. Note that $A^\times = A - BC = A^*$. So in this case the associate main operator of Θ depends exclusively on A and coincides with the adjoint of A. From the relationships between the operator A, B and C in Θ it follows that the associate system $\Theta^\times$ is a Brodskii $(-J)$-system.

Suppose now that Π is an orthogonal projection of H and $A[\operatorname{Ker}\Pi] \subset \operatorname{Ker}\Pi$. Then automatically $A^*[\operatorname{Im}\Pi] \subset \operatorname{Im}\Pi$, and hence Π is a supporting projection for

Θ. So we can apply Theorem 2.6 to show that

$$\Theta = \mathrm{pr}_{I-\Pi}(\Theta)\mathrm{pr}_{\Pi}(\Theta).$$

The systems $\mathrm{pr}_{\Pi}(\Theta)$ and $\mathrm{pr}_{I-\Pi}(\Theta)$ are Brodskii J-systems again (cf., [30], page 6). This result leads to an important multiplicative representation of the Livsic-Brodskii characteristic operator function (cf., [30], page 143).

3.2 Kreĭn systems

Next we consider Kreĭn J-systems introduced in Section 1.3. By definition, a system $\Theta = (A, R, -JK^{-*}R^*A, K; H, G)$ is a Kreĭn J-system if J is a signature operator,

$$I - AA^* = RJR^*, \qquad J - R^*R = K^*JK,$$

and the operators A and K are invertible. Since A is invertible, $I - RJR^*$ is invertible. But then we can apply the operator identity (2.5) and $J = K^*JK + R^*R$ to obtain that

$$I + RK^{-1}JK^{-*}R^* = (I - RJR^*)^{-1}.$$

Hence

$$(AA^*)^{-1} = (I - RJR^*)^{-1} = I + RK^{-1}JK^{-*}R^*.$$

It follows that $\Theta^{\times} = \big(A^{-*}, RK^{-1}, JR^*A^{-*}, K^{-1}; H, G\big)$. From this we see that $\Theta^{\times}$ is a Kreĭn $(-J)$-system. Observe that in this case the associate main operator $A^{\times}$ depends again exclusively on A and coincides with A^{-*}.

Let Π be an orthogonal projection of H. With respect to the decomposition $H = \operatorname{Ker}\Pi \oplus \operatorname{Im}\Pi$, we write

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \qquad R = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}.$$

Suppose now that $A_{21} = 0$ (i.e., $\operatorname{Ker}\Pi$ is an invariant subspace for A) and A_{11} is invertible. Then A_{22} is invertible too and $\operatorname{Im}\Pi$ is an invariant subspace for $A^{\times} = A^{-*}$. From $RJR^* = I - AA^*$ it follows that $R_2JR_2^* = I - A_{22}A_{22}^*$. But this implies (see [33]) the existence of an invertible operator K_2 on G such that $J - R_2^*R_2 = K_2^*JK_2$. Put $K_1 = KK_2^{-1}$. Then K_1 is also invertible and $K = K_1K_2$. We are now in a position to apply Theorem 2.6 (with $n = 2$). The result is a factorization $\Theta = \Theta_1\Theta_2$, where Θ_1 and Θ_2 can be described explicitly with the help of formulas (2.10) and (2.11). It can be shown that Θ_1 and Θ_2 are Kreĭn J-systems (cf., [32]).

3.3 Unitary systems

A system or operator node $(A, B, C, D; X, U, Y)$ with X, U and Y Hilbert spaces is said to be *unitary* if the operator

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} : X \oplus U \to X \oplus Y$$

is unitary. The operator matrix in this formula is usually referred to as the *system matrix*.

Let $\Theta = (A, B, C, D; X, U, Y)$ be a unitary system. Then its main operator A is a contraction, that is, $\|A\| \leq 1$. It follows that the corresponding transfer function $W_\Theta(\lambda) = D + C(\lambda - A)^{-1}B$ is analytic on the exterior $|\lambda| > 1$ of the closed unit disc. It can be shown (see, e.g., Theorem XXVIII.2.1 in [47]) that

$$\|W_\Theta(\lambda)\| \leq 1, \qquad |\lambda| > 1.$$

The converse statement is also true, that is, if W is analytic on the exterior of the closed unit disc (including the point ∞), and its values are contractions from the Hilbert space U to the Hilbert space Y, then $W = W_\Theta$ for some unitary system Θ. Moreover, under some additional minimality conditions, the system Θ is unique up to unitary equivalence.

As we mentioned above the main operator of a unitary system is a contraction. Conversely, any contraction appears as the main operator of a unitary system. To see this, let A on X be a contraction. Put $\mathcal{D}_{A^*} = \overline{D_{A^*}X}$ and $\mathcal{D}_A = \overline{D_A X}$. Here, for a contraction T, the operator D_T is the defect operator $D_T = (I - T^*T)^{1/2}$. Since $A^*D_{A^*} = D_A A^*$, the operator A^* maps $\mathcal{D}_{A^*}$ into $\mathcal{D}_A$. Now define $B : \mathcal{D}_{A^*} \to X$, $C : X \to \mathcal{D}_A$ and $D : \mathcal{D}_{A^*} \to \mathcal{D}_A$ by $Bu = D_{A^*}u$, $Cx = D_A x$ and $Du = -A^*u$. Then the system $(A, B, C, D; X, \mathcal{D}_{A^*}, \mathcal{D}_A)$ is unitary and has A as its main operator. For this system the transfer function is given by

$$W(\lambda) = -A^* + D_A(\lambda - A)^{-1}D_{A^*} : \mathcal{D}_{A^*} \to \mathcal{D}_A,$$

that is, up to the change of variable $\lambda \mapsto \lambda^{-1}$ it coincides with the Sz-Nagy-Foias characteristic operator function for A; see [108].

Notice that the external operator D of a unitary system does not have to be invertible, and hence Theorem 2.1 need not apply to unitary systems. However (see Proposition XXVIII.2.7 in [47]), if $\Theta = (A, B, C, D; X, U, Y)$ is a unitary system and $|\lambda| > 1$, then $W_\Theta(\lambda)$ is invertible if and only if $\bar{\lambda}^{-1} \in \rho(A)$, and in that case

$$W_\Theta(\lambda)^{-1} = D^* + \lambda B^*(I - \lambda A^*)^{-1}C^*.$$

The product of two unitary systems is again a unitary system (Theorem XXVIII.6.1 in [47]). Also, invariant subspaces of the main operator of a unitary system induce factorizations but to get the factors another method than the one

of Section 2.4 has to be used because the external operator may not be invertible. To get an analogue of Theorem 2.3 for unitary systems one proceed as follows.

Let $\Theta = (A, B, C, D; X, U, Y)$ be a unitary system, and let X_1 be an invariant subspace of A. Let X_2 be the orthogonal complement of X_1 in X. Put

$$U_2 = U, \qquad Y_1 = Y,$$

and consider the following block operator matrix representations

$$\begin{aligned} A &= \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} : X_1 \oplus X_2 \to X_1 \oplus X_2, \\ B &= \begin{bmatrix} B_{12} \\ B_{22} \end{bmatrix} : U_2 \to X_1 \oplus X_2, \\ C &= \begin{bmatrix} C_{11} & C_{12} \end{bmatrix} : X_1 \oplus X_2 \to Y_1. \end{aligned}$$

Thus

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & B_{12} \\ 0 & A_{22} & B_{22} \\ C_{11} & C_{12} & D \end{bmatrix}. \tag{3.1}$$

Since this operator is unitary, $A_{11}^* A_{11} + C_{11}^* C_{11} = I_{X_1}$. Now put

$$U_1 = \left\{ \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \in X_1 \oplus Y_1 \,\Big|\, A_{11}^* x_1 + C_{11}^* y_1 = 0 \right\},$$

and define $B_1 : U_1 \to X_1$ and $D_1 : U_1 \to Y_1$ by

$$B_1 \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = x_1, \qquad D_1 \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = y_1.$$

Then the system $\Theta_1 = (A_{11}, B_1, C_{11}, D_1; X_1, U_1, Y_1)$ is a unitary system, which is called the *left projection* of Θ associated to invariant subspace X_1.

Next, consider the product

$$\begin{bmatrix} A_{11}^* & C_{11}^* & 0 \\ B_1^* & D_1^* & 0 \\ 0 & 0 & I_{X_2} \end{bmatrix} \begin{bmatrix} A_{11} & B_{12} & A_{12} \\ C_{11} & D & C_{12} \\ 0 & B_{22} & A_{22} \end{bmatrix}, \tag{3.2}$$

which acts as an operator from $X_1 \oplus U_2 \oplus X_2$ to $X_1 \oplus U_1 \oplus X_2$. Since (3.1) is unitary, the $(1,1)$-entry in the 3×3 operator matrix defined by the product in (3.2) is equal to I_{X_1}. On the other hand, the product in (3.2) defines a unitary

operator, because both its factors are unitary. Hence the product in (3.2) is of the form

$$\begin{bmatrix} I_{X_1} & 0 & 0 \\ 0 & D_2 & C_2 \\ 0 & B_{22} & A_{22} \end{bmatrix}$$

for certain operators $C_2 \colon X_2 \to Y_1$ and $D_2 \colon U_2 \to U_1$. Now put $Y_2 = U_1$. Then the operators A_{22}, B_{22}, C_2 and D_2 form a unitary system Θ_2,

$$\Theta_2 = (A_{22}, B_{22}, C_2, D_2; X_2, U_2, Y_2),$$

which is called the *right projection* of Θ associated to the invariant subspace H_1. The following theorem is the analogue of Theorem 2.3 for unitary systems.

Theorem 3.1. *Let Θ be a unitary system, and let X_1 be an invariant subspace for the main operator of Θ. Then the left projection Θ_1 and the right projection Θ_2 of Θ associated with X_1 are unitary systems, and $\Theta = \Theta_1\Theta_2$.*

Proof. We continue to use the notation introduced in the two paragraphs preceding the theorem. We already know that Θ_1 and Θ_2 are unitary systems. From (3.2) and the fact that Θ_1 is unitary it follows that

$$\begin{bmatrix} A_{11} & B_{12} & A_{12} \\ C_{11} & D & C_{12} \\ 0 & B_{22} & A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & B_1 & 0 \\ C_{11} & D_1 & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & D_2 & C_2 \\ 0 & B_{22} & A_{22} \end{bmatrix}.$$

This identity is equivalent to the statement that $\Theta = \Theta_1\Theta_2$. □

3.4 Monic systems

Let $T : X \to X$, $R : Y \to X$ and $Q : X \to Y$ be operators, the underlying spaces X and Y being complex Banach spaces, and let ℓ be a positive integer. The system $\Theta = (T, R, Q, 0; X, Y)$ is called a *monic system* of *degree* ℓ if the operator

$$\operatorname{col}\left(QT^{j-1}\right)_{j=1}^{\ell} = \begin{bmatrix} Q \\ QT \\ \vdots \\ QT^{\ell-1} \end{bmatrix} : X \to Y^{\ell}$$

is invertible and its inverse is of the form

$$\operatorname{row}\left(U_{j-1}\right)_{j=1}^{\ell} = [U_0, \ldots, U_{\ell-1}] : Y^{\ell} \to X$$

with $U_{\ell-1} = R$. The integer ℓ is uniquely determined by these properties. Monic systems have been introduced and studied in [11], [12]. Its external operator being zero, a monic system is strictly proper.

To justify our terminology we make the following remark. Suppose $\Theta = (T, R, Q, 0; X, Y)$ is a monic system of degree ℓ, and let $U_0, \ldots, U_{\ell-1}$ be as above. Then the transfer function W_Θ of Θ,

$$W_\Theta(\lambda) = Q(\lambda - T)^{-1}R, \qquad \lambda \in \rho(A),$$

coincides with the inverse L^{-1} of the monic operator polynomial L defined by

$$L(\lambda) = \lambda^\ell I - \sum_{j=0}^{\ell-1} \lambda^j Q T^\ell U_i.$$

Furthermore, it can be shown that L can also be written as

$$L(\lambda) = \lambda^\ell I - \sum_{j=0}^{\ell-1} \lambda^j V_i T^\ell R,$$

where $\mathrm{col}(V_{\ell-j})_{j=1}^{\ell}$ is the inverse of the invertible operator $\mathrm{row}(T^{\ell-j}R)_{j=1}^{\ell}$. For the proofs of these statements we refer to [11] and [12].

Suppose now, conversely, that L is a given monic operator polynomial the coefficients of which are operators on Y. Then one can construct a monic system Θ for which $W_\Theta = L^{-1}$. Indeed, if

$$L(\lambda) = \lambda^\ell I + \sum_{j=0}^{\ell-1} \lambda^j A_j$$

and

$$C_{1,L} = \begin{bmatrix} 0 & I & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & I \\ -A_0 & -A_1 & \ldots & -A_{\ell-1} \end{bmatrix},$$

then

$$\Theta = \big(C_{1,L},\, \mathrm{col}(\delta_{j\ell} I)_{j=1}^{\ell},\, \mathrm{row}(\delta_{j\ell} I)_{j=1}^{\ell},\, 0\, ;\, Y^\ell, Y\big) \tag{3.3}$$

has the desired properties. Here δ_{ij} is the Kronecker delta. The operator $C_{1,L}$ is known as the *first companion operator* associated with L, and for that reason (3.3) will be called the *first companion system* corresponding to L. When Y is finite-dimensional it is possible to construct a system Θ with $W_\Theta = L^{-1}$ from the spectral data (eigenvalues, eigenvectors and associated eigenvectors) of L. The construction may be found in [66], [69] (see also Chapter 8).

As was mentioned before the external operator of a monic system is equal to the zero operator, and hence, as for the unitary systems in the previous section, the system $\Theta^\times$ is not defined. Nevertheless, as we have seen in the preceding paragraphs, there still is a construction of the inverse of the transfer function in terms of the system.

If Θ_1 and Θ_2 are monic systems of degree ℓ_1 and ℓ_2, respectively, then $\Theta_1\Theta_2$ is a monic system of degree $\ell_1 + \ell_2$. A system which is similar to a monic system of degree ℓ is again a monic system of degree ℓ.

In Section 2.4 we introduced the notion of a supporting projection for systems having invertible external operator. This notion does not apply to monic systems because its external operator, being the zero operator, is not invertible. Still, a similar concept has been introduced in [11], [12]. We shall review some of the material presented there.

Let $\Theta = (T, R, Q, 0; X, Y)$ be a monic system of degree ℓ, and let Π be a projection of X. We say that Π is a *monic supporting projection* for Θ if $\operatorname{Ker}\Pi$ is a non-trivial invariant subspace for T and there exists a positive integer m (necessarily unique and less than ℓ) such that

$$\operatorname{col}(QT^{j-1})_{j=1}^m\big|_{\operatorname{Ker}\Pi} : \operatorname{Ker}\Pi \to Y^m \tag{3.4}$$

is invertible. We call m the *degree* of the monic supporting projection. The operator (3.4) is invertible if and only if this is the case for the operator

$$\Pi \operatorname{row}(T^{k-1}R)_{j=1}^k : Y^k \to \operatorname{Im}\Pi. \tag{3.5}$$

Here $k = \ell - m$.

Let $\Theta = (T, R, Q, 0; X, Y)$ be a monic system, and let Π be a monic supporting projection for Θ. Let m be the degree of Π. Put $X_1 = \operatorname{Ker}\Pi$, and define $T_1 : X_1 \to X_1$ and $Q_1 : X_1 \to Y$ by $T_1x = Tx$ and $Q_1x = Qx$. The invertibility of the operator in (3.4) now implies that $\operatorname{col}(Q_1T_1^{j-1})_{j=1}^m$ is invertible. Hence there exists a unique $R_1 : Y \to X_1$ such that $\Theta_1 = (T_1, R_1, Q_1, 0; X_1, Y)$ is a monic system. This system, which has degree m, is called the *left projection* of Θ associated with Π.

Put $X_2 = \operatorname{Im}\Pi$, $k = \ell - m$, and define $T_2 : X_2 \to X_2$ and $R_2 : Y \to X_2$ by $T_2x = \Pi Tx$ and $R_2y = \Pi Ry$. Since $\operatorname{Ker}\Pi$ is invariant under T, we have $\Pi T\Pi = \Pi T$. This, together with the invertibility of the operator in (3.5), implies that $\operatorname{row}(T_2^{k-j}R_2)_{j=1}^k$ is invertible. Therefore there exists a unique $Q_2 : X_2 \to Y$ such that $\Theta_2 = (T_2, R_2, Q_2, 0; X_2, Y)$ is a monic system. This system, which has degree $k = \ell - m$, is called the *right projection* of Θ associated with Π.

Let $\Theta = (T, R, Q, 0; X, Y)$ be a monic system, and let Π be a monic supporting projection for Θ. Let Θ_1 and Θ_2 be the associated left and right projections. Then Θ and $\Theta_1\Theta_2$ are similar (see [12], Theorem 2.2). Conversely, if $\Theta_1 = (T_1, R_1, Q_1, 0; X_1, Y)$ and $\Theta_2 = (T_2, R_2, Q_2, 0; X_2, Y)$ are monic systems such that Θ and $\Theta_1\Theta_2$ are similar, then there exists a monic supporting projection for Θ such that the associated left and right projections of Θ are similar to

Θ_1 and Θ_2, respectively. To prove this we may assume that $\Theta = \Theta_1\Theta_2$. But then one can take Π to be the canonical projection of $X_1 \dot{+} X_2$ along X_1 onto X_2.

By Theorem 2.2, the factorization of a monic system implies a factorization of the corresponding transfer function. Since in this case the transfer functions are inverses of monic operator polynomials, we can employ the theory explained above to derive factorizations for monic operator polynomials with the factors being monic operator polynomials too. In fact, the following factorization result holds true (cf., Theorem 8 in [65], Theorem 13 in [67]; see also the book [69]). Let $L(\lambda) = A_0 + \lambda A_1 + \cdots + \lambda^{\ell-1}A_{\ell-1} + \lambda^\ell I$ be a monic operator polynomial, and let $\Theta = (T, R, Q, 0; X, Y)$ be a fixed monic system such that $W_\Theta = L^{-1}$. Let Π be a monic supporting projection for Θ of degree m, and let $\Theta_1 = (T_1, R_1, Q_1, 0; X_1, Y)$ and $\Theta_2 = (T_2, R_2, Q_2, 0; X_2, Y)$ be the associated left and right projections of Θ. Put

$$L_1(\lambda) = \lambda^m I - \sum_{j=0}^{m-1} \lambda^j Q_1 T_1^m W_j,$$

where $\operatorname{row}(W_i)_{j=0}^{m-1} = \left(\operatorname{col}(Q_1 T_1^{j-1})_{j=1}^{m}\right)^{-1}$, and

$$L_2(\lambda) = \lambda^{\ell-m} I - \sum_{j=0}^{\ell-m-1} \lambda^i Q_2 T_2^{\ell-m} V_j,$$

with $\operatorname{row}(V_j)_{j=0}^{\ell-m-1} = \left(\operatorname{col}(Q_2 T_2^{j-1})_{j=1}^{\ell-m}\right)^{-1}$. Note that $W_{\Theta_1} = L_1^{-1}$ and $W_{\Theta_2} = L_2^{-1}$. As $\Theta \simeq \Theta_1\Theta_2$, we may conclude from Theorem 2.2 that $L = L_2L_1$.

Conversely, given a factorization $L = L_2L_1$ where L_1 and L_2 are monic operator polynomials, there exists a monic supporting projection Π of Θ such that for the associated left and right projections Θ_1 and Θ_2 we have $L_1^{-1} = W_{\Theta_1}$ and $L_2^{-1} = W_{\Theta_2}$. In fact, for Π we may take the projection $\Pi = I - P$, the projection P being defined by

$$P = \left(\operatorname{col}(QT^j)_{j=0}^{\ell-m}\right)^{-1} \operatorname{col}\left(Q_1 T_1^j\right)_{j=0}^{\ell-1} \left(\operatorname{col}\left(Q_1 T_1^j\right)_{j=0}^{m-1}\right)^{-1} \operatorname{col}(QT^j)_{j=0}^{m-1},$$

where m is the degree of L_1 and (T_1, Q_1, R_1) is a monic system with transfer function L_1^{-1} (see [65], Section 5 and [69]). By using an appropriate Möbius transformation (see the final section of this chapter), this factorization theorem for monic operator polynomials can also be deduced from Theorem 2.3; see Theorem 3.5 in Section 3.6 below.

In the previous paragraph the correspondence between the monic supporting projections Π of Θ and the right divisors L_1 of L is not one-one. The reason may be explained as follows. Suppose Π is a monic supporting projection for Θ of degree m, and let Π' be another projection of X such that $\operatorname{Ker}\Pi = \operatorname{Ker}\Pi'$. Then it is immediately clear from the definition that Π' is also a monic supporting projection for Θ of degree m. Furthermore, the left projections of Θ associated

with Π and Π' coincide. So what really matters in the definition of the monic supporting projection Π is the existence of a T-invariant complemented subspace X_1 of X such that the operator

$$\operatorname{col}(QT^{j-1})_{j=1}^{m}\big|_{X_1} : X_1 \to Y^m$$

is invertible. Such a subspace X_1 is called a *supporting subspace* for Θ (cf., [65], Section 5). The correspondence between the supporting subspaces of Θ and the right divisors of L, $L^{-1} = W_\Theta$, is one to one (cf., [67] or [12]; see also the discussion in the paragraph after the proof of Theorem 9.3.

For the companion type system (3.3) the supporting subspaces may be characterized in a simple way. Indeed, a closed subspace X_1 of Y^ℓ is a supporting subspace for (3.3) if and only if X_1 is invariant under the first companion operator $C_{1,L}$ and X_1 is an algebraic complement in Y^ℓ of the subspace of Y^ℓ consisting of all vectors for which the first m coordinates are zero. This is the contents of the first part of Theorem 1.6 in [67]. Note that condition (iii) in this theorem is superfluous (see also [69]).

3.5 Polynomial systems

In the previous section we have seen that the inverses of monic operator polynomials can be seen as transfer functions of certain systems. Now we shall deal with arbitrary polynomials. A system $\Theta = (A, B, C, D; X, Y)$ is called a *polynomial system* if its main operator A is nilpotent. If, in addition, $D = I$, we say that Θ is a *comonic polynomial system*. The transfer function of a (comonic) polynomial system is obviously a (comonic) polynomial in λ^{-1}. An operator polynomial is said to be *comonic* if its constant term is the identity operator on the underlying space.,

Now conversely. Let P be a regular operator polynomial of degree ℓ whose coefficients are operators on Y. Here *regular* means that $P(\lambda)$ is invertible for at least one λ. In order to show that $P(\lambda^{-1})$ is the transfer function of a polynomial system, it suffices to consider the comonic case, where $P(0)$ is the identity operator on Y. Put $L(\lambda) = \lambda^\ell P(\lambda^{-1})$. Then L is a monic operator polynomial of degree ℓ. Let $\Delta = (T, R, Q, 0; X, Y)$ be a monic system such that the transfer function W_Δ is equal to L^{-1}. Then

$$L(\lambda) = \lambda^\ell I - \sum_{j=0}^{\ell-1} \lambda^j Q T^\ell U_j,$$

where

$$\operatorname{row}(U_{j-1})_{j=1}^{\ell} = \left(\operatorname{col}(QT^{j-1})_{j=1}^{\ell}\right)^{-1}. \tag{3.6}$$

Further, $U_{\ell-1} = R$ (see the previous section). So

$$U_0 Q + \cdots + U_{\ell-2} Q T^{\ell-2} + R Q T^{\ell-1} = I,$$

and hence

$$T - RQT^{\ell} = \big(\operatorname{col}(QT^{j-1})_{j=1}^{\ell}\big)^{-1}\big[\delta_{i,j-1}I\big]_{i,j=1}^{l}\big(\operatorname{col}(QT^{j-1})_{j=1}^{\ell}\big).$$

It follows that $T - RQT^{\ell}$ is nilpotent of order ℓ. Also, using (3.6), one sees that $(T - RQT^{\ell})U_j = U_{j-1}$ for $j = 1, \ldots, \ell - 1$. But then

$$\begin{aligned} I - QT^{\ell}(\lambda - T + RQT)^{-1}R &= I - \sum_{j=1}^{\ell} \lambda^{-j} QT^{\ell} U_{\ell-j} \\ &= \lambda^{-\ell} L(\lambda) = P(\lambda^{-1}). \end{aligned}$$

So $P(\lambda^{-1})$ is the transfer function of the comonic polynomial system

$$\Theta = \big(T - RQT^{\ell}, R, -QT^{\ell}; X, Y\big). \tag{3.7}$$

Summarizing we obtain: the class of transfer functions of (comonic) polynomial systems coincides with that of the (comonic) polynomials in λ^{-1}. We observe that the system $\Theta = (T - T^{\ell}RQ, T^{\ell}R, -Q; X, Y)$ is also a comonic polynomial system and its transfer function is also equal to $P(\lambda^{-1})$.

Consider the system (3.7). Note that $\Theta^{\times} = (T, R, -QT^{\ell}; X, Y)$ is the corresponding associate system. On $\rho(T) \setminus \{0\}$ the transfer function of $\Theta^{\times}$ coincides with $P(\lambda^{-1})^{-1}$. From this fact one easily infers that $P(\lambda^{-1}) = QT^{\ell-1}(I - \lambda T)^{-1}R$.

The product of two (comonic) polynomial systems is again a (comonic) polynomial system. A system that is similar to a (comonic) polynomial system is also a (comonic) polynomial system. If $\Theta = (A, B, C; X, Y)$ is a comonic polynomial system, then for any supporting projection Π of Θ the systems $\operatorname{pr}_{\Pi}(\Theta)$ and $\operatorname{pr}_{I-\Pi}(\Theta)$ are comonic polynomial systems.

3.6 Möbius transformation of systems

From the definition of the transfer function of a system it is clear that such a function is analytic at infinity. Therefore, in general, the theory developed in the previous sections can be applied to an arbitrary analytic operator function after a suitable transformation of the independent variable. For this reason we study in this section the effect of a Möbius transformation on complex variable λ.

Throughout this section φ will be the Möbius transformation

$$\varphi(\lambda) = \frac{p\lambda + q}{r\lambda + s}. \tag{3.8}$$

Here p, q, r and s are complex numbers and $ps - qr \neq 0$. We consider φ as a map from the Riemann sphere $\mathbb{C} \cup \infty$ onto itself. The inverse map φ^{-1} is given by

$$\varphi^{-1}(\lambda) = \frac{-s\lambda + q}{r\lambda - p}.$$

Theorem 3.2. *Let $W(\lambda) = D + C(\lambda - A)^{-1}B$ be the transfer function of the system $\Theta = (A, B, C, D; X, Y)$, and let φ be the Möbius transformation* (3.8)*. Assume that $T = p - rA$ $(= pI_X - rA)$ is invertible. Then $W_\varphi(\lambda) = W\big(\varphi(\lambda)\big)$ is the transfer function of the system*

$$\Theta_\varphi = \Big((-(q - sA)T^{-1},\, T^{-1}B,\, (ps - qr)CT^{-1},\, D + rCT^{-1}B\,;X,\, Y\Big). \tag{3.9}$$

Proof. As $T = p - rA$ is invertible, the inverse map φ^{-1} is analytic on the spectrum $\sigma(A)$ of (A). So $\varphi^{-1}(A)$ is well defined. In fact, the operator $\varphi^{-1}(A) = -(q - sA)T^{-1}$ is equal to the main operator of Θ_φ. By the spectral mapping theorem, the resolvent set of $\varphi^{-1}(A)$ is given by

$$\rho\big(\varphi^{-1}(A)\big) = \big\{\lambda \in \mathbb{C} \mid \varphi(\lambda) \in \rho(A) \cup \{\infty\}\big\}.$$

It follows that the function W_φ as well as the transfer function of Θ_φ are defined on the same open set, namely on $\rho(\varphi^{-1}(A))$.

To prove that the two functions coincide there, take λ in $\rho(\varphi^{-1}(A))$. Assume first that $r\lambda + s \neq 0$. Then

$$\begin{aligned}
W(\lambda) &= D + C\left(\frac{\lambda p + q}{r\lambda + s} - A\right)^{-1} B \\
&= D + (r\lambda + s)C\big(\lambda(p - rA) + q - sA\big)^{-1}B \\
&= D + (r\lambda + s)C\big(\lambda - \varphi^{-1}(A)\big)^{-1}T^{-1}B \\
&= D + C\big(r\big(\lambda - \varphi^{-1}(A)\big) + r\varphi^{-1}(A) + s\big)\big(\lambda - \varphi^{-1}(A)\big)^{-1}T^{-1}B \\
&= D + rCT^{-1}B + (ps - qr)CT^{-1}\big(\lambda - \varphi^{-1}(A)\big)^{-1}T^{-1}B,
\end{aligned}$$

where we use that $r\varphi^{-1}(A) + s = (ps - qr)(p - rA)^{-1}$.

Next, assume that $\lambda \in \rho(\varphi^{-1}(A))$ and $r\lambda + s = 0$. In this case, $r = 0$ implies $s = 0$, because $\lambda \in \mathbb{C}$. Since $ps - qr \neq 0$, we cannot have $r = s = 0$. So $r \neq 0$. Notice that $\varphi(\lambda) = \infty$ and $W_\varphi(\lambda) = D$. On the other hand, it is not difficult to check that the value of the transfer function of Θ_φ in $\lambda = -sr^{-1}$ is equal to D too. This completes the proof. □

The system Θ_φ introduced in formula (3.9) has several interesting properties; some of them will be discussed below.

Proposition 3.3. *For $j = 1, 2$, let $\Theta_j = (A_j, B_j, C_j, D_j; X_j, Y)$ be a system such that both $(\Theta_1)_\varphi$ and $(\Theta_2)_\varphi$ exist. Then $(\Theta_1\Theta_2)_\varphi$ exists too and*

$$(\Theta_1\Theta_2)_\varphi = (\Theta_1)_\varphi(\Theta_2)_\varphi. \tag{3.10}$$

Proof. Recall that $\Theta_1\Theta_2 = (A, B, C, D; X_1 \dotplus X_2, Y)$, where

$$A = \begin{bmatrix} A_1 & B_1C_2 \\ 0 & A_2 \end{bmatrix}, \qquad B = \begin{bmatrix} B_1D_2 \\ B_2 \end{bmatrix},$$

$$C = \begin{bmatrix} C_1 & D_1C_2 \end{bmatrix}, \qquad D = D_1D_2.$$

The fact that $(\Theta_1)_\varphi$ and $(\Theta_2)_\varphi$ exist comes down to the invertibility of $p - rA_1$ and $p - rA_2$. Since $\sigma(A)$ is a subset of $\sigma(A_1) \cup \sigma(A_2)$, it follows that $p - rA$ is invertible too. Thus $(\Theta_1\Theta_2)_\varphi$ is well defined. Equality (3.10) follows now by a direct computation. □

Let $\mathcal{C}$ be a class of systems such that for each $\Theta \in \mathcal{C}$ the system Θ_φ is well defined. Assume that $\mathcal{C}$ is closed under multiplication. For instance, we could take for $\mathcal{C}$ the class of Brodskii systems for which Θ_φ exists. Then we can form the class $\mathcal{C}_\varphi = \{\Theta_\varphi \mid \Theta \in \mathcal{C}\}$, and by Proposition 3.3 the new class $\mathcal{C}_\varphi$ is again closed under multiplication. In this way one can also establish certain relationships between different classes of systems.

For example (cf., [33]), let $\Theta = (A, R, C, K; H, G)$ be a Kreĭn J-system, and let Ψ be the Möbius transformation

$$\Psi(\lambda) = \alpha\left(\frac{\lambda + i}{\lambda - i}\right).$$

Here $|\alpha| = 1$, and we assume that $\alpha \in \rho(A)$. It follows that Θ_Ψ is well defined. Put

$$\begin{aligned} A_\Psi &= -i(\alpha + A)(\alpha - A)^{-1}, \\ B_\Psi &= (\alpha - A)^{-1}R, \\ C_\Psi &= -2i\alpha C(\alpha - A)^{-1}, \\ D_\Psi &= K + C(\alpha - A)^{-1}R. \end{aligned}$$

Using the properties of Kreĭn J-systems, one sees that

$$B_\Psi J B_\Psi^* = (\alpha - A)^{-1}(I - AA^*)(\bar{\alpha} - A^*)^{-1}. \tag{3.11}$$

It follows that

$$A_\Psi - A_\Psi^* = -2iB_\Psi J B_\Psi^*. \tag{3.12}$$

This last identity is one of the defining properties of a Brodskii $(-J)$-system. However, note that in general Θ_Ψ is not a Brodskii system, because its external operator D_Ψ may not be equal to the identity operator. As $A^\times = (A^*)^{-1}$ for Kreĭn systems, we have $\alpha \in \rho(A^*)$, and hence D_Ψ is invertible. We shall prove that the system

$$\Delta = (A_\Psi, B_\Psi, D_\Psi^{-1}C_\Psi; H, G) \tag{3.13}$$

is a Brodskii$(-J)$-system. In view of (3.12) it suffices to prove that $D_\Psi^{-1}C_\Psi = -2iJB_\Psi^*$. In other words we have to show that $C_\Psi = -2iD_\Psi JB_\Psi^*$. To do this, observe that

$$B_\Psi JB_\Psi^* = A^*(\bar\alpha - A^*)^{-1} + \alpha(\alpha - A)^{-1}$$

(cf., (3.11)). It follows that

$$C_\Psi = -2i\big(-CA^*(\bar\alpha - A^*)^{-1} + CB_\Psi JB_\Psi^*\big). \tag{3.14}$$

From the definition of Θ_Ψ it is clear that $D_\Psi = K + CB_\Psi$. So $-2iD_\Psi JB_\Psi^* = -2i(KJB_\Psi^* + CB_\Psi JB_\Psi^*)$. Employing the properties of Kreĭn J-systems, we have

$$\begin{aligned} KJR^* &= J(K^*)^{-1}(J - R^*R)JR^* \\ &= J(K^*)^{-1}R^* - J(K^*)^{-1}R^*(I - AA^*) \\ &= J(K^*)^{-1}R^*AA^* = -CA^*. \end{aligned}$$

Combining this with (3.14) yields $C_\Psi = -2iD_\Psi JB_\Psi^*$, and hence the system (3.13) is a Brodskii $(-J)$-system. The relationship between the systems Δ and Θ can also be expressed in terms of the corresponding characteristic operator functions W_Δ and W_Θ. We have

$$W_\Delta(\lambda) = JW_\Theta(\alpha)^*JW_\Theta\left(\bar\alpha\,\frac{\lambda + i}{\lambda - i}\right).$$

This is clear from the definition of Δ and the fact that $D_\Psi^{-1} = W_\Theta(\alpha)^{-1} = JW_\Theta(\alpha)^*J$.

In the next couple of paragraphs we will consider the effect of the Möbius transformation on inversion and factorization.

Proposition 3.4. *Let $\Theta = (A, B, C, D; X, Y)$ be a system such that Θ_φ exists. Assume that the external operators of Θ and Θ_φ are invertible. Then $(\Theta^\times)_\varphi$ and $(\Theta_\varphi)^\times$ exist while, moreover,*

$$(\Theta^\times)_\varphi = (\Theta_\varphi)^\times. \tag{3.15}$$

Proof. Let W and W_φ be the transfer functions of Θ and Θ_φ, respectively. By assumption the operators $W(\infty) = D$ and $W_\varphi(\infty) = D + rC(p - rA)^{-1}B$ are invertible. If $r \neq 0$, then $W(\mathrm{pr}^{-1}) = W_\varphi(\infty)$, and so $\mathrm{pr}^{-1} \in \rho(A^\times)$, where $A^\times$ is the main operator of $\Theta^\times$. Hence $p - rA^\times$ is invertible. This conclusion is also correct if $r = 0$, because then $p \neq 0$. It follows that $(\Theta^\times)_\varphi$ exists. Also, as $D + rC(p - rA)^{-1}B$ is invertible, $(\Theta_\varphi)^\times$ exists too. Finally, using

$$\big(D + rC(p - rA)^{-1}B\big)^{-1} = D^{-1} - rD^{-1}C(p - rA^\times)^{-1}BD^{-1},$$

formula (3.15) is proved by a direct computation. □

Let $\Theta = (A, B, C, D; X, Y)$ be a system such that Θ_φ exists. Assume that the external operators of Θ and Θ_φ are invertible. Let Π be a supporting projection for Θ, i.e.,

$$A[\operatorname{Ker}\Pi] \subset \operatorname{Ker}\Pi, \qquad A^\times[\operatorname{Im}\Pi] \subset \operatorname{Im}\Pi.$$

In general one may not conclude that Π is a supporting projection for Θ_φ too. But if the state space of X is finite-dimensional, then the conclusion is correct. So let us assume that $\dim X$ is finite. Let Θ_1 and Θ_2 be the factors of Θ corresponding to Π, and let $D = D_1 D_2$ be a factorization of D with D_1 and D_2 invertible (i.e., Θ_1 and Θ_2 are given by formulas (2.10) and (2.11), respectively). As X is finite-dimensional, the systems $(\Theta_1)_\varphi$ and $(\Theta_2)_\varphi$ are well defined, and we have (cf., Proposition 3.3)

$$\Theta_\varphi = (\Theta_1)_\varphi (\Theta_2)_\varphi. \tag{3.16}$$

This factorization corresponds to Π (as a supporting projection for Θ_φ) and a special factorization of $D + rC(p - rA)^{-1}B$ into invertible factors induced by $D = D_1 D_2$. In the particular case that φ is a translation and the external operator of Θ is the identity operator we may replace (3.16) by

$$\operatorname{pr}_\Pi(\Theta_\varphi) = (\operatorname{pr}_\Pi(\Theta))_\varphi, \qquad \operatorname{pr}_{I-\Pi}(\Theta_\varphi) = (\operatorname{pr}_{I-\Pi}(\Theta))_\varphi.$$

Möbius transformations may be employed to derive from Theorem 2.3 factorization theorems for transfer functions that do not take an invertible value at infinity. To illustrate this we shall give a new proof of the division theorem for monic operator polynomials (cf., [67], Theorem 13, see also [69]) based on Theorem 2.3.

Theorem 3.5. *Suppose L is a monic operator polynomial of degree ℓ, and let $\Delta = (T, R, Q, O; X, Y)$ be a monic system such that the transfer function of Δ is equal to L^{-1}. Let $X_1 \subset X$ be a supporting subspace, i.e., the space X_1 is a non-trivial (complemented) invariant subspace for T such that for some positive integer m (necessarily unique and less than ℓ)*

$$\operatorname{col}(QT^{j-1})_{j=1}^m\big|_{X_1} : X_1 \to Y^m$$

is invertible. Define $T_1 : X_1 \to X_1$ and $Q_1 : X_1 \to Y$ by $T_1 x = Tx$ and $Q_1 x = Qx$. Then $\operatorname{col}(Q_1 T_1^{j-1})_{j=1}^m$ is invertible, say with inverse $\operatorname{row}(W_{j-1})_{j=1}^m$. Introduce

$$L_1(\lambda) = \lambda^m I - \sum_{j=0}^{m-1} \lambda^j Q_1 T_1^m W_1.$$

Then L_1 is a right divisor of L.

Proof. Consider the Möbius transformation $\Psi(\lambda) = (\alpha\lambda + 1)\lambda^{-1}$. Here α is a fixed complex number in the unbounded component of the resolvent set of T. As $T - \alpha$ is invertible, the system $\Theta = \Delta_\varphi$ is well defined. Put $A = -(\alpha - T)^{-1}$, $B = (\alpha - T)^{-1}R$, $C = -Q(\alpha - T)^{-1}$ and $D = Q(\alpha - T)^{-1}R$. Then $\Theta = (A, B, C, D; X, Y)$, and the external operator D of Θ is equal to $L(\alpha)^{-1}$. Further, the associate main

operator of Θ is equal to

$$A^\times = A - BD^{-1}C = -(\alpha - T)^{-1} + (\alpha - T)^{-1}RL(\alpha)Q(\alpha - T)^{-1}. \tag{3.17}$$

With the supporting subspace X_1 we associate the projection P defined by

$$Px = \big(\operatorname{col}(Q_1T_1^{j-1})_{j=1}^m\big)^{-1}\big(\operatorname{col}(QT^{j-1})_{j=1}^m\big)x.$$

We know that $\operatorname{Im} P = X_1$ is invariant under T. As α is in the unbounded component of T, it follows that X_1 is also invariant under $A = -(\alpha - T)^{-1}$. Furthermore, we see that

$$A\big|_{X_1} = -(\alpha - T_1)^{-1}. \tag{3.18}$$

Next we consider $X_2 = \operatorname{Ker} P$. From the theory of monic systems we know that $X_2 = \operatorname{Im}\big(\operatorname{row}(T^{j-1}R)_{j=1}^{\ell-m}\big)$. We shall prove that X_2 is invariant under $A^\times$. In order to do this, recall that $QT^jR = 0,\ \ j = 0, \dots, \ell - 2$ and $QT^{\ell-1}R = I$. It follows that, for $s = 0, \dots, \ell - 1$,

$$Q(\alpha - T)^{-1}T^sR = \alpha^sQ(\alpha - T)^{-1}R = \alpha^sL(\alpha)^{-1}.$$

Using (3.17) we obtain $A^\times R = 0$. Also, for $1 \le s \le \ell - 1$, we have

$$\begin{aligned} A^\times T^sR &= -(\alpha - T)^{-1}T^sR + (\alpha - T)^{-1}RL(\alpha)Q(\alpha - T)^{-1}T^sR \\ &= -(\alpha - T)^{-1}T^sR + \alpha^s(\alpha - T)^{-1}R \\ &= T^{s-1}R + \alpha T^{s-2}R + \cdots + \alpha^{s-1}R. \end{aligned}$$

So we know the action $A^\times$ on $T^sRy,\ 0 \le s \le \ell - 1$. It follows that X_2 is invariant under $A^\times$. Furthermore, we see that the restriction of $A^\times$ to X_2 is nilpotent of order $\ell - m$. Observe that by now we have proved that $\Pi = I - P$ is a supporting projection for Θ.

As $\alpha - T_1$ is invertible, the same is true for the operator $L_1(\alpha)$. Put $D_1 = L_1(\alpha)^{-1}$ and $D_2 = L_1(\alpha)L(\alpha)^{-1}$. Then $D = D_1D_2$. Let

$$A = \begin{bmatrix} A_{11} & B_{12} \\ 0 & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}$$

be the operator matrix representations for A, B and C with respect to the decomposition $X = X_1 \dot{+} X_2$. Consider the systems

$$\begin{aligned} \Theta_1 &= (A_{11}, B_1D_2^{-1}, C_1, D_1; X_1, Y), \\ \Theta_2 &= (A_{22}, B_2, D_1^{-1}C_2, D_2; X_2, Y), \end{aligned}$$

and, for $j = 1, 2$, let W_j be the transfer function of Θ_j. As $AX_1 \subset X_1$ and $A^\times X_2 \subset X_2$, we can apply Theorem 2.3 to show that $\Theta = \Theta_1\Theta_2$, and hence

$$L\big(\Psi(\alpha)\big)^{-1} = W_1(\lambda)W_2(\lambda). \tag{3.19}$$

First we shall prove that $L_1(\Psi(\lambda))^{-1} = W_1(\lambda)$. Let Δ_1 be the system $\Delta_1 = (T_1, R_1, Q_1, 0; X_1, Y)$, where T_1 and Q_1 are as before and $R_1 = W_{m-1}$. So Δ_1 is a monic system whose transfer function is equal to L_1^{-1}. Thus in order to prove that $L(\Psi(\lambda))^{-1}$ is equal to $W_1(\lambda)$, it suffices to show that $(\Delta_1)_\Psi = \Theta_1$. Note that

$$(\Delta_1)_\Psi = \big(-(\alpha - T_1)^{-1}, (\alpha - T_1)^{-1}R_1, -Q_1(\alpha - T_1)^{-1}, L_1(\alpha)^{-1}; X_1, Y\big).$$

From (3.18) we know that $A_{11} = -(\alpha - T_1)^{-1}$. By definition $D_1 = L_1(\alpha)^{-1}$. Further

$$C_1 = -Q(\alpha - T)\big|_{X_1} = -Q_1(\alpha - T_1).$$

It remains to prove that $B_1 D_2^{-1} = (\alpha - T_1)^{-1}R_1$. Take $y \in Y$. By applying (3.18), first for T, Q, R and next for T_1, Q_1, R_1, we obtain

$$\begin{aligned}\big(\mathrm{col}(QT^{j-1})_{j=1}^m\big)(\alpha - T)^{-1}RL(\alpha)L_1(\alpha)^{-1}y &= \\ &= \big(\mathrm{col}\big(\alpha^{j-1}L(\alpha)^{-1}\big)_{j=1}^m\big)L(\alpha)L_1(\alpha)^{-1}y \\ &= \big(\mathrm{col}(Q_1T_1^{j-1})_{j=1}^m\big)(\alpha - T_1)^{-1}R_1y.\end{aligned}$$

But then $B_1D_2^{-1}y = PBD_2^{-1}y = (\alpha - T_1)^{-1}R_1y$, and we have proved that $\Theta_1 = (\Delta_1)_\Psi$.

As the restriction of $A^\times$ to X_2 is nilpotent of order $\ell - m$, we know that $W_2(\lambda)^{-1}$ is a polynomial in λ^{-1} of degree at most $\ell - m$. Put

$$L_2(\lambda) = W_2(\Psi^{-1}(\lambda))^{-1},$$

where Ψ^{-1} is the inverse map of the Möbius transformation Ψ. In other words, $\Psi^{-1}(\lambda) = (\lambda - \alpha)^{-1}$. It follows that $L_2(\lambda)$ is a polynomial in λ of degree at most $\ell - m$. From (3.19) we see that $L(\lambda) = L_2(\lambda)L_1(\lambda)$, and hence L_1 is a right divisor of L. □

Notes

The main references for Sections 3.1 and 3.2 are [30] and [32], respectively. Section 3.3 follows [47], Section XXVIII.7. The results in Section 3.4 originate from [11], [12]. These two papers were inspired on the one hand by the theory of characteristic operator functions and on the other hand by the theory of monic operator polynomials developed in [65], [66] (see also the books [69] and [101]). The concept of a monic system does not appear in the latter publications but its role is played by the notion of a standard triple. We note that a triple (Q, T, R) of operators is a standard triple for a monic operator polynomial L if and only if $\Theta = (T, R, Q, 0; X, Y)$ is a monic system and $W_\Theta = L^{-1}$. Sections 3.5 and 3.6 are taken from [14], Chapter 1.

Chapter 4

Realization and Linearization of Operator Functions

The main problem addressed in this chapter is the realization problem for operator-valued functions. Given such a function the problem is to find a system for which the transfer function coincides with the given function. In the first section we consider rational operator functions, and in the second analytic ones. In Section 4.3 it is shown that, in a certain sense, the transfer function of a system with an invertible external operator can be reduced to a linear function, and we use this reduction to describe the singularities of the transfer function. In the final section a connection between Schur complements and linearization is described.

4.1 Realization of rational operator functions

We start our considerations with he following result.

Theorem 4.1. *Given the operator polynomials*

$$H(\lambda) = \sum_{j=0}^{\ell-1} \lambda^j H_j, \qquad L(\lambda) = \lambda^\ell I + \sum_{j=0}^{\ell-1} \lambda^j L_j,$$

with coefficients acting on the complex Banach space Y, let

$$A = \begin{bmatrix} 0 & I & \dots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & I \\ -L_0 & -L_1 & \dots & -L_{\ell-1} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ I \end{bmatrix}, \quad C = \begin{bmatrix} H_0 & \dots & H_{\ell-1} \end{bmatrix}. \tag{4.1}$$

Then $\Theta = (A, B, C, 0; Y^\ell, Y)$ *is a system such that*

$$W_\Theta(\lambda) = H(\lambda)L(\lambda)^{-1}, \qquad \lambda \in \rho(A).$$

Proof. We know already (see Section 3.4) that

$$L^{-1}(\lambda) = Q(\lambda - A)^{-1}B, \qquad \lambda \in \rho(A), \tag{4.2}$$

where $Q = \begin{bmatrix} I & 0 & \cdots & 0 \end{bmatrix}$. For $\lambda \in \rho(A)$, define $C_1(\lambda), \dots, C_\ell(\lambda)$ by

$$\operatorname{col}\big(C_j(\lambda)\big)_{j=1}^{\ell} = (\lambda - A)^{-1}B.$$

From (4.2) we see that $C_1(\lambda) = L^{-1}(\lambda)$. As $(\lambda - A)\big(\operatorname{col}\big(C_j(\lambda)\big)_{j=1}^{\ell}\big) = B$, the special form of A in (4.1) yields

$$C_j(\lambda) = \lambda^{j-1}C_1(\lambda), \qquad j = 1, \dots, \ell.$$

It follows that $C(\lambda - A)^{-1}B = \sum_{j=0}^{\ell-1} H_j C_{j+1}(\lambda) = H(\lambda)L(\lambda)^{-1}$, and the proof is complete. □

Let us employ Theorem 4.1 to obtain a realization for a proper rational operator function W, the values of which act on a complex Banach space Y. By definition, an operator function is *rational* if it can be transformed into an operator polynomial by multiplication with a scalar polynomial. Such a function is meromorphic with a finite set of poles.

Theorem 4.2. *Let* W *be a proper rational operator function whose values act on the complex Banach space* Y*, and put* $D = W(\infty)$*. Then there is a Banach space* X*, and there are bounded linear operators* $A : X \to X$*,* $B : Y \to X$ *and* $C : X \to Y$ *such that* $\rho(A)$ *coincides with the (finite) set of poles of* A *and*

$$W(\lambda) = D + C(\lambda - A)^{-1}B, \qquad \lambda \in \rho(A).$$

Proof. By the definition given above, there exist a scalar polynomial q and an operator polynomial P such that

$$W(\lambda) = \frac{1}{q(\lambda)}P(\lambda), \qquad \lambda \in \mathbb{C},\ q(\lambda) \neq 0.$$

Put $D = W(\infty)$ and introduce

$$H(\lambda) = q(\lambda)\big(W(\lambda) - D\big) = P(\lambda) - q(\lambda)D.$$

Then H is an operator polynomial with coefficients acting on Y. Obviously we may assume q to be monic. Then, clearly, the operator polynomial $L(\lambda) = q(\lambda)I_Y$ is monic and

$$W(\lambda) = D + H(\lambda)L(\lambda)^{-1}, \qquad \lambda \in \mathbb{C},\ q(\lambda) \neq 0.$$

Moreover, as $W(\infty) = D$, we have $\lim_{\lambda\to\infty} H(\lambda)L(\lambda)^{-1} = 0$, and hence the degree of H is strictly less than that of L. The desired conclusion now comes about by applying Theorem 4.1. □

The case when W is a proper rational matrix function corresponds to the situation where Y is finite-dimensional and X can be taken to be finite-dimensional too. A deeper analysis of the relation between the poles of a rational matrix function W and the eigenvalues of the main operator A in a realization of W will be given in Chapter 8.

4.2 Realization of analytic operator functions

By a *Cauchy contour* Γ we shall mean the positively oriented boundary of a bounded Cauchy domain in $\mathbb{C}$. Such a contour consists of a finite number of non-intersecting closed rectifiable Jordan curves.

Let Γ be a Cauchy contour around zero, and let Y be a complex Banach space. With Γ and Y we associate the space $C(\Gamma, Y)$ of all Y-valued continuous functions on Γ endowed with the supremum norm. The canonical embedding of Y into $C(\Gamma, Y)$ will be denoted by τ, i.e., $\tau(y)(z) = y$ for each $y \in Y$ and $z \in \Gamma$. Further we define $\omega : C(\Gamma, Y) \to Y$ by setting

$$\omega(f) = \frac{1}{2\pi i} \int_\Gamma \frac{1}{\zeta} f(\zeta)\, d\zeta.$$

Since 0 is assumed to be in the interior domain of Γ, we may conclude that $\omega\tau$ is the identity operator on Y. This observation will be used in Section 4.3 below.

By $\mathcal{L}(Y)$ we mean the Banach algebra of all bounded linear operators on Y.

Theorem 4.3. *Let Ω be the interior domain of Γ, and let W be an operator function, analytic on Ω, continuous towards the boundary Γ, and with values in $\mathcal{L}(Y)$. Define operators V and M on $C(\Gamma, Y)$ by*

$$\big(Vf\big)(z) = zf(z), \qquad \big(Mf\big)(z) = W(z)f(z).$$

Then

$$W(\lambda) = I + \omega(V - VM)(\lambda - V)^{-1}\tau, \qquad \lambda \in \Omega \subset \rho(V).$$

In other words, the system $\Theta = (V, \tau, \omega(V - VM); C(\Gamma, Y), Y)$ is a realization for W on Ω.

Proof. First observe that $\Omega \subset \rho(V)$. In fact, $\sigma(V) = \Gamma$,

$$\big[(\lambda - V)^{-1} g\big](z) = \frac{1}{\lambda - z} g(z), \qquad \lambda \notin \Gamma;\ z \in \Gamma.$$

It follows that for $\lambda \notin \Gamma$ we have

$$\omega(V - VM)(\lambda - V)^{-1}\tau y = \left(\frac{1}{2\pi i} \int_\Gamma \frac{1}{\lambda - \zeta} \Big(I - W(\zeta)\Big)\, d\zeta\right) y. \tag{4.3}$$

For $\lambda \in \Omega$ the right-hand side of (4.3) is equal to $W(\lambda)y - y$, and the theorem is proved. □

The associate main operator $V^\times$ of the system Θ introduced in the above theorem is $V^\times = V - \tau\omega(V - VM)$. It follows that

$$(V^\times f)(z) = zf(z) - \frac{1}{2\pi i}\int_\Gamma \Big(I - W(\zeta)\Big) f(\zeta)\, d\zeta. \tag{4.4}$$

In the next section (see Theorem 4.6) we shall show that, in a sense to be made precise, $V^\times$ is a linearization of W on Ω.

Theorem 4.3 can be proved for any bounded open set Ω in $\mathbb{C}$, regardless of possible boundary conditions. In that case the space $C(\Gamma, Y)$ must be replaced by an appropriate Banach space, which one has to define in terms of the behavior of W near the boundary (cf., [97]; see also the next theorem).

If Ω is an unbounded open set containing zero, then one cannot expect that W admits a representation of the form $D + C(\lambda - A)^{-1}B$, because the behavior of W near infinity may be irregular. However one can always write W in the alternative form $D + \lambda C(I - \lambda A)^{-1}B$. This follows from the next theorem by changing λ into λ^{-1}.

Theorem 4.4. *Let Ω be an open neighborhood of infinity in the Riemann sphere $\mathbb{C} \cup \{\infty\}$ not containing the origin, and let $W : \Omega \to \mathcal{L}(Y)$ be analytic. Define X to be the space of all Y-valued functions, analytic on Ω, such that*

$$\|f\|_\bullet = \sup_{z\in\Omega} \frac{\|f(z)\|}{\max(1, \|W(z)\|)} < \infty.$$

The space X endowed with norm $\|\cdot\|_\bullet$ is a Banach space. The canonical embedding of Y into X is denoted by τ. Further $\gamma : X \to Y$ is defined by $\gamma(f) = f(\infty)$. Finally, let M and V be the operators on X given by

$$\begin{aligned}
(Mf)(z) &= W(z)f(\infty), \qquad z \in \Omega,\\
(Vf)(z) &= \begin{cases} z\big(f(z) - f(\infty)\big), & z \in \Omega \setminus \{\infty\},\\ \lim\limits_{z\to\infty} z\big(f(z) - f(\infty)\big), & z = \infty. \end{cases}
\end{aligned}$$

Then

$$W(\lambda) = W(\infty) + \gamma(\lambda - V)^{-1}VM\tau, \qquad \lambda \in \Omega \subset \rho(V) \cup \{\infty\}.$$

In other words, the system $(V, VM\tau, \gamma, W(\infty); X, Y)$ is a realization for W on Ω.

Proof. It is straightforward to check that the operators τ, γ, V and M are well defined bounded linear operators. Next we prove that for each $\lambda \in \Omega \setminus \{\infty\}$ the operator $\lambda - V$ is invertible. Indeed, take $\lambda \in \Omega \setminus \{\infty\}$. For $g \in X$, put

$$h(z) = \begin{cases} \dfrac{zg(\lambda) - \lambda g(z)}{z - \lambda}, & z \in \Omega \setminus \{\lambda, \infty\},\\ g(\lambda) - \lambda g'(\lambda), & z = \lambda,\\ g(\lambda), & z = \infty. \end{cases}$$

Then $h \in X$, and by direct computation one sees that

$$\big((\lambda - V)h\big)(z) = \lambda g(z), \qquad z \in \Omega.$$

Now λ is nonzero (since Ω does not contain the origin), and it follows that $\lambda - V$ is surjective. As is easily verified, $\lambda - V$ is injective too. We conclude that $\lambda \in \rho(V)$ and $(\lambda - V)^{-1}g = \lambda^{-1}h$.

Now take $g = VM\tau y$, i.e., $g(z) = z\big(W(z)y - W(\infty)y\big)$ for each $z \in \Omega$. Then

$$\big((\lambda - V)^{-1}g\big)(z) = \begin{cases} \dfrac{z}{\lambda - z}\big(W(z) - W(\lambda)\big)y, & z \in \Omega \setminus \{\lambda, \infty\}, \\ -\lambda W'(\lambda)y, & z = \lambda, \\ \big(W(\lambda) - W(\infty)\big)y, & z = \infty, \end{cases}$$

and so $\gamma(\lambda - V)^{-1}g = \big((\lambda - V)^{-1}g\big)(\infty) = \big(W(\lambda) - W(\infty)\big)y$. Thus

$$\gamma(\lambda - V)^{-1}VM\tau = W(\lambda) - W(\infty), \qquad \lambda \in \Omega \setminus \{\infty\} \subset \rho(V),$$

and the theorem is proved. □

In the previous theorem the condition that Ω does not contain the origin may be replaced by the requirement that $\mathbb{C} \setminus \Omega \neq \emptyset$. In the latter case one takes a point in $\mathbb{C} \setminus \Omega$, and with appropriate changes the theorem remains valid. If $\mathbb{C} \setminus \Omega = \emptyset$, i.e., if Ω is the full Riemann sphere, the theorem does not go through, but on the other hand in that case the function W is constant.

4.3 Linearization

Let W be the transfer function of a system $\Theta = (A, B, C, D; X, Y)$ with invertible external operator D. Then $W^{-1}(\lambda) = D^{-1} - D^{-1}C(\lambda - A^\times)^{-1}BD^{-1}$ for λ in a neighborhood of ∞. Here $A^\times = A - BD^{-1}C$ is the associate main operator. In this section we shall point out another connection between W and $A^\times$. In fact we shall prove that $A^\times$ appears as a linearization of W.

First we define the notion of linearization. Let Ω be an open set in $\mathbb{C}$, and let $W : \Omega \to \mathcal{L}(Y)$ be analytic. An operator $T \in \mathcal{L}(X)$ is called a *linearization* of W on Ω if there exist a Banach space Z and analytic operator functions E and F on Ω such that

$$W(\lambda) \dot{+} I_Z = E(\lambda)\big(\lambda - T\big)F(\lambda), \qquad \lambda \in \Omega, \tag{4.5}$$

while the maps $E(\lambda), F(\lambda) : Y \dot{+} Z \to X$ are bijective for each λ in Ω. Here we follow the convention that for operators $R : Y \to Y$ and $Q : Z \to Z$ the operator $R \dot{+} Q$ stands for the diagonal operator on $Y \dot{+} Z$ build from R and Q, that is,

$$R \dot{+} Q = \begin{bmatrix} R & 0 \\ 0 & Q \end{bmatrix} : Y \dot{+} Z \to Y \dot{+} Z.$$

The operator function $W(\lambda) \dot{+} I_z$ in (4.5) is called the *Z-extension* of W. If two operator functions W_1 and W_2, analytic on Ω, are connected as in (4.5), i.e., if $W_1(\lambda) = E(\lambda)W_2(\lambda)F(\lambda)$ for $\lambda \in \Omega$, with E, F analytic on Ω and $E(\lambda), F(\lambda)$ invertible for each $\lambda \in \Omega$, then W_1 and W_2 are said to be *analytically equivalent* on Ω.

Theorem 4.5. *Let $\Theta = (A, B, C, D; X, Y)$ be a system with an invertible external operator D, and assume that B has a left inverse. Then $A^\times$ is a linearization on $\rho(A)$ of the transfer function W_Θ. In fact, if B^+ is a left inverse of B and $Z = \operatorname{Ker} B^+$, then the Z-extension of W_Θ is analytically equivalent to $(\lambda - A)^{-1}$ on $\rho(A)$.*

Further relevant details can be found in the proof.

Proof. For $\lambda \in \rho(A)$, define $E(\lambda), F(\lambda) : Y \dot{+} Z \to X$ by

$$\begin{aligned} E(\lambda)(y,z) &= BD^{-1}y + z + BD^{-1}C(\lambda - A)^{-1}z, \\ F(\lambda)(y,z) &= (\lambda - A)^{-1}(By + z), \end{aligned}$$

where $y \in Y$ and $z \in Z$. Then $E(\lambda)$ and $F(\lambda)$ are bounded linear operators depending analytically on the parameter $\lambda \in \rho(A)$. Also $E(\lambda)$ and $F(\lambda)$ are invertible with inverses $E(\lambda)^{-1}, F(\lambda)^{-1} : X \to Y \dot{+} Z$ given by

$$\begin{aligned} E(\lambda)^{-1}x &= \big(DB^+x - C(\lambda - A)^{-1}\big(I_X - BB^+\big)x,\ \big(I_X - BB^+\big)x\big), \\ F(\lambda)^{-1}x &= \big(B^+(\lambda - A)x,\ \big(I_X - BB^+\big)(\lambda - A)x\big)\big), \end{aligned}$$

where $x \in X$. Finally,

$$E(\lambda)\big(W_\Theta(\lambda) \dot{+} I_Z\big) = (\lambda - A^\times)F(\lambda), \qquad \lambda \in \rho(A). \tag{4.6}$$

The (straightforward) computations are left to the reader. □

By applying Theorem 4.5 to the associate system $\Theta^\times$ we may conclude that under the conditions of Theorem 4.5 the operator A is a linearization on $\rho(A^\times)$ of the transfer function $W_{\Theta^\times}$. So, roughly speaking, the operator A appears as a linearization of $W_\Theta^{-1}(\lambda)$ and $A^\times$ as a linearization of W_Θ. For finite-dimensional systems this corresponds to the fact that the eigenvalues of A are related to the poles of W_Θ and the eigenvalues of $A^\times$ are related to the zeroes of W_Θ. We shall return to this relation in more detail in Chapter 8.

Theorem 4.5 has a counterpart for the situation where C has a right inverse C^+, say. One then takes $Z = \operatorname{Ker} C$ and proves (4.6) with the equivalence operators $E(\lambda), F(\lambda) : Y \dot{+} Z \to X$ given by

$$\begin{aligned} E(\lambda)(y,z) &= (\lambda - A)\big(C^+y + z\big), \\ F(\lambda)(y,z) &= C^+Dy - \big(I_X - C^+C\big)(\lambda - A)^{-1}By + z, \end{aligned}$$

where $y \in Y$ and $z \in Z$. For the inverses $E(\lambda)^{-1}, F(\lambda)^{-1} : X \to Y \dot{+} Z$ of these operators, we have the expressions

$$\begin{aligned} E(\lambda)^{-1}x &= \big(C(\lambda - A)^{-1}x,\ \big(I_X - C^+C\big)(\lambda - A)^{-1}x\big), \\ F(\lambda)^{-1}x &= \big(D^{-1}Cx,\ \big(I_X - C^+C\big)\big(I_X + (\lambda - A)^{-1}BD^{-1}C\big)x\big), \end{aligned}$$

where $x \in X$.

If B or C has a generalized inverse then one has to allow for extensions on both sides (see [96] for details). Always, irrespective of any invertibility condition on B or C, the functions $W_\Theta(\lambda) \dot{+} I_X$ and $(\lambda - A^\times) \dot{+} I_Y$ are analytically equivalent on $\rho(A)$. In fact (cf., [52], Theorem 4.5),

$$W_\Theta(\lambda) \dot{+} I_X = E(\lambda)\big((\lambda - A^\times) \dot{+} I_Y\big)F(\lambda), \qquad \lambda \in \rho(A), \tag{4.7}$$

where the equivalence operators and their inverses are given by

$$\begin{aligned} E(\lambda) &= \begin{bmatrix} -C(\lambda - A)^{-1} & W_\Theta(\lambda) \\ (\lambda - A)^{-1} & -(\lambda - A)^{-1}B \end{bmatrix} : X \dot{+} Y \to Y \dot{+} X, \\ F(\lambda) &= \begin{bmatrix} (\lambda - A)^{-1}B & I_X \\ D^{-1}W_\Theta(\lambda) & D^{-1}C \end{bmatrix} : Y \dot{+} X \to X \dot{+} Y, \\ E^{-1}(\lambda) &= \begin{bmatrix} BD^{-1} & \lambda - A^\times \\ D^{-1} & D^{-1}C \end{bmatrix} : Y \dot{+} X \to X \dot{+} Y, \\ F^{-1}(\lambda) &= \begin{bmatrix} -D^{-1}C & I_Y \\ (\lambda - A)^{-1}(\lambda - A^\times) & -(\lambda - A)^{-1}B \end{bmatrix} : X \dot{+} Y \to Y \dot{+} X. \end{aligned}$$

In all these expressions, the dependance of the parameter $\lambda \in \rho(A)$ is analytic.

The equivalence after two-sided extension embodied in (4.7) sheds new light on Theorem 2.1. Indeed, it is now clear that $W_\Theta(\lambda)$ and $\lambda - A^\times$ share not only the property of being (non-)invertible, but $W_\Theta(\lambda)$ and $\lambda - A^\times$ have all Fredholm characteristics in common. Further details can be derived from [18], Section 2. We give two examples. The first is about the nullity and says that the operator $(\lambda - A)^{-1}B$ maps $\operatorname{Ker} W_\Theta(\lambda)$ one-to-one onto $\operatorname{Ker}(\lambda - A^\times)$, hence $\operatorname{Ker} W_\Theta(\lambda)$ and $\operatorname{Ker}(\lambda - A^\times)$ have the same (possibly infinite) dimension. The second is concerned with the defect and reads as follows. If M is a (closed) complement of $\operatorname{Im} W_\Theta(\lambda)$ in Y, then $BD^{-1}[M]$ is a (closed) complement of $\operatorname{Im}(\lambda - A^\times)$ in X and BD^{-1} maps M one-to-one onto $BD^{-1}[M]$, hence $\operatorname{Im} W_\Theta(\lambda)$ and $\operatorname{Im}(\lambda - A^\times)$ have the same (possibly infinite) codimension in Y and X, respectively.

We now make the connection with Theorem 4.3 (see reference [52], Theorems 2.2 and 2.3).

Theorem 4.6. *Let Γ be a Cauchy contour around zero in $\mathbb{C}$, and let W be an operator function, analytic on the interior domain of Ω of Γ, continuous towards the boundary Γ and with values in $\mathcal{L}(Y)$. Let T on $C(\Gamma, Y)$ be defined by*

$$\big(Tf\big)(z) = zf(z) - \frac{1}{2\pi i}\int_{\Gamma}\big(I_Y - W(\zeta)\big)f(\zeta)\,d\zeta.$$

Then T is a linearization of W on Ω and

$$\sigma(T) = \Gamma \cup \{\lambda \in \Omega \mid W(\lambda) \text{ not invertible}\}. \tag{4.8}$$

Proof. With the notation of Theorem 4.3, we have

$$W(\lambda) = I + \omega(V - VM)(\lambda - V)^{-1}\tau, \qquad \lambda \in \Omega \subset \rho(V).$$

Since 0 is in the interior domain of Γ, the operator τ has a left inverse. In fact, as noted in the beginning of Section 4.2, we have $\omega\tau = I_Y$. As $\Omega \subset \rho(V)$, Theorem 4.5 shows that $W(\lambda) \dot{+} I_{\operatorname{Ker}\omega}$ is analytically equivalent to $\lambda - V^{\times}$ on Ω, in other words $V^{\times}$ is a linearization of W on Ω. Here $V^{\times}$ is the associate main operator of the system $\Theta = \big(V, \tau, \omega(V - VM); C(\Gamma, Y), Y\big)$. So $V^{\times}$ is given by formula (4.4), and hence $V^{\times} = T$.

To prove (4.8), recall that $\sigma(V) = \Gamma$ and consider the transfer function W_{Θ} of the system $\Theta = \big(V, \tau, \omega(V - VM); C(\Gamma, Y), Y\big)$. We know that $W_{\Theta}(\lambda) = W(\lambda)$ for $\lambda \in \Omega$. For $\lambda \in \mathbb{C} \setminus \overline{\Omega}$ we have $W_{\Theta}(\lambda) = I$. This is clear from (4.3). So $W_{\Theta}(\lambda)$ is invertible for each λ in the exterior domain of Γ. As $\rho(V) = \mathbb{C} \setminus \Gamma$, we may apply Theorem 2.1 to show that

$$\sigma(V^{\times}) \cap \Omega = \{\lambda \in \Omega \mid W(\lambda) \text{ not invertible}\},$$

and $\sigma(V^{\times}) \cap [\mathbb{C} \setminus \overline{\Omega}] = \emptyset$. So to prove(4.8) it remains to show that Γ is contained in $\sigma(V^{\times})$.

Take $\lambda_0 \in \Gamma$, and assume that $\lambda_0 \in \rho(V^{\times})$. So $W_{\Theta^{\times}}(\lambda)$ is defined in some connected open neighborhood of U of λ_0. Observe that on $[\mathbb{C} \setminus \overline{\Omega}] \cap U$ the function $W_{\Theta^{\times}}$ is identically equal to I. By analyticity $W_{\Theta^{\times}}(\lambda) = I$ for each $\lambda \in U$. But then, applying Theorem 2.1 to $\Theta^{\times}$, one obtains $\lambda_0 \in \rho(V)$. This contradicts $\Gamma = \sigma(V)$, and the proof is complete. □

Theorem 4.5 is applicable only to systems with an invertible external operator. To obtain a linearization of the transfer function of a system with a non-invertible external operator one can employ an appropriate Möbius transformation. In some cases a linearization can be given directly. For example, if $\Theta = (T, R, Q, 0; X, Y)$ is a monic system, then T is a linearization on $\mathbb{C}$ of W_{Θ}^{-1}. Recall (cf., Section 3.4) that in this case W_{Θ}^{-1} is a monic operator polynomial. To describe the linearization in more detail, put $Z = Y^{\ell-1}$ where ℓ is the degree of

Θ, and let $E(\lambda), F(\lambda) : X \to Y \dot{+} Z$ be given by

$$\begin{aligned} E(\lambda)x &= \Big(Qx,\ \mathrm{col}\,\big(QT^j(T-\lambda)x\big)_{j=0}^{\ell-2}\Big), \\ F(\lambda)x &= \left(QT^{\ell-1}x + \sum_{j=1}^{\ell-1} L_j(\lambda)QT^{j-1}x,\ -\mathrm{col}\,(QT^jx)_{j=0}^{\ell-2}\right). \end{aligned}$$

Here $L_j(\lambda) = \lambda^{\ell-j}I - \sum_{s=0}^{\ell-1-j} \lambda^s QT^\ell U_{s+j}$, where

$$\begin{bmatrix} U_0 & \cdots & U_{\ell-1} \end{bmatrix} = \big(\mathrm{col}\,(QT^{j-1})_{j=1}^{\ell}\big)^{-1}.$$

Then (cf., [12], Theorem 3.1; see also [11]) the operators $E(\lambda)$ and $F(\lambda)$ are invertible and

$$F(\lambda)(\lambda - T) = [W_\Theta^{-1}(\lambda) \dot{+} I_Z]E(\lambda), \qquad \lambda \in \mathbb{C}.$$

Notice that Theorem 4.5 is also applicable to the Livsic-Brodskii characteristic operator function

$$W(\lambda) = I + 2iK^*(\lambda - A)^{-1}KJ,$$

provided K has a left inverse, and in that case the adjoint operator A^* is a linearization on $\rho(A)$ of W. A similar remark holds for Kreĭn nodes.

4.4 Linearization and Schur complements

Let X_1, X_2, Y_1 and Y_2 be Banach spaces and $K : X_1 \to X_2$ and $L : Y_1 \to Y_2$ be bounded linear operators. The operators K and L are called *equivalent* when there exist invertible operators $G : X_2 \to Y_2$ and $H : Y_1 \to X_1$ such that $L = GKH$. Extending this notion, we say that K and L are *equivalent after extension* if there exist Banach spaces X_0 and Y_0 such that

$$K \dot{+} I_{X_0} = \begin{bmatrix} K & 0 \\ 0 & I_{X_0} \end{bmatrix} : X_1 \dot{+} X_0 \to X_2 \dot{+} X_0$$

and

$$L \dot{+} I_{Y_0} = \begin{bmatrix} L & 0 \\ 0 & I_{Y_0} \end{bmatrix} : Y_1 \dot{+} Y_0 \to Y_2 \dot{+} Y_0$$

are equivalent. Equivalence and equivalence after extension are reflexive, symmetric and transitive properties.

In this section we apply the notions of equivalence and equivalence after extension to Schur complements (see the paragraph after the first proof of Theorem 2.1), and explain the connection with the linearization results from the previous section. Throughout M is the following 2×2 operator matrix

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} : X_1 \dot{+} Y_1 \to X_2 \dot{+} Y_2. \tag{4.9}$$

We shall assume that both M_{11} and M_{22} are invertible. In this case the Schur complements of M_{11} and M_{22} in M are well defined and given by

$$\Lambda = M_{22} - M_{21}M_{11}^{-1}M_{12}, \qquad \Delta = M_{11} - M_{12}M_{22}^{-1}M_{21},$$

respectively.

Theorem 4.7. *Let M be given by* (4.9)*, and assume that M_{11} and M_{22} are invertible. Then the Schur complement Λ of M_{11} in M is equivalent after extension to the Schur complement Δ of M_{22} in M.*

In fact

$$\Lambda \dot{+} I_{X_1} = E(\Delta \dot{+} I_{Y_1})F$$

with the invertible operators E and F and their inverses given by

$$E = \begin{bmatrix} -M_{21}M_{11}^{-1} & \Lambda \\ M_{11}^{-1} & M_{11}^{-1}M_{12} \end{bmatrix} : X_2 \dot{+} Y_1 \to Y_2 \dot{+} X_1,$$

$$F = \begin{bmatrix} -M_{11}^{-1}M_{12} & I_{X_1} \\ I_{Y_1} - M_{22}^{-1}M_{21}M_{11}^{-1}M_{12} & M_{22}^{-1}M_{21} \end{bmatrix} : Y_1 \dot{+} X_1 \to X_1 \dot{+} Y_1,$$

$$E^{-1} = \begin{bmatrix} -M_{12}M_{22}^{-1} & \Delta \\ M_{22}^{-1} & M_{22}^{-1}M_{21} \end{bmatrix} : Y_2 \dot{+} X_1 \to X_2 \dot{+} Y_1,$$

$$F^{-1} = \begin{bmatrix} -M_{22}^{-1}M_{21} & I_{Y_1} \\ I_{X_1} - M_{11}^{-1}M_{12}M_{22}^{-1}M_{21} & M_{11}^{-1}M_{12} \end{bmatrix} : X_1 \dot{+} Y_1 \to Y_1 \dot{+} X_1.$$

Proof. The verification can be done by direct computation. The following reasoning, however, gives more insight (cf. the remark made after the proof).

From the basic identities (2.3) and (2.4), one immediately gets

$$\begin{bmatrix} M_{11} & 0 \\ 0 & \Lambda \end{bmatrix} =$$

$$= \begin{bmatrix} I_{X_2} & M_{12}M_{22}^{-1} \\ -M_{21}M_{11}^{-1} & G \end{bmatrix} \begin{bmatrix} \Delta & 0 \\ 0 & M_{22} \end{bmatrix} \begin{bmatrix} I_{X_1} & -M_{11}^{-1}M_{12} \\ M_{22}^{-1}M_{21} & H \end{bmatrix}$$

with $G = I_{Y_2} - M_{21}M_{11}^{-1}M_{12}M_{22}^{-1}$ and $H = I_{Y_1} - M_{22}^{-1}M_{21}M_{11}^{-1}M_{12}$. We also have the simple equalities

$$\begin{bmatrix} M_{11} & 0 \\ 0 & \Lambda \end{bmatrix} = \begin{bmatrix} 0 & I_{X_2} \\ I_{Y_2} & 0 \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & M_{11} \end{bmatrix} \begin{bmatrix} 0 & I_{Y_1} \\ I_{X_1} & 0 \end{bmatrix},$$

$$\begin{bmatrix} \Lambda & 0 \\ 0 & M_{11} \end{bmatrix} = \begin{bmatrix} I_{Y_2} & 0 \\ 0 & M_{11} \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & I_{X_1} \end{bmatrix} = \begin{bmatrix} \Lambda & 0 \\ 0 & I_{Y_2} \end{bmatrix} \begin{bmatrix} I_{Y_1} & 0 \\ 0 & M_{11} \end{bmatrix},$$

$$\begin{bmatrix} \Delta & 0 \\ 0 & M_{22} \end{bmatrix} = \begin{bmatrix} I_{X_2} & 0 \\ 0 & M_{22} \end{bmatrix} \begin{bmatrix} \Delta & 0 \\ 0 & I_{Y_1} \end{bmatrix} = \begin{bmatrix} \Delta & 0 \\ 0 & I_{Y_1} \end{bmatrix} \begin{bmatrix} I_{X_1} & 0 \\ 0 & M_{22} \end{bmatrix}.$$

The desired result is now easily obtained by appropriately combining parts of these identities. □

Now let $\Theta = (A, B, C, D; X, U, Y)$ be a system with an invertible external operator D, and put

$$M(\lambda) = \begin{bmatrix} A - \lambda & B \\ C & D \end{bmatrix}.$$

Take $\lambda \in \rho(A)$. Then both $A - \lambda$ and D are invertible, and we can apply Theorem 4.7. In this case the Schur complement of $A - \lambda$ in $M(\lambda)$ is equal to $W_\Theta(\lambda)$, where W_Θ is the transfer function of Θ, and the Schur complement of D in $M(\lambda)$ is equal to $A^\times - \lambda$. Thus Theorem 4.7 shows that $W_\Theta(\lambda)$ and $\lambda - A^\times$ are equivalent after extension whenever $\lambda \in \rho(A)$. From the way the equivalence operators were constructed, we see that the (invertible) operators establishing the equivalence depend analytically on the parameter λ. Hence in this way we recover (4.7) as a corollary of Theorem 4.7.

One can also combine other parts of the identities given in the proof of Theorem 4.7. This gives three additional results. The four results thus obtained differ slightly from each other. There is no need to present all details here.

Theorem 4.7 is concerned with equivalence after two-sided extension. In certain situations, as in Theorem 4.5, one can make do with one sided extension.

Theorem 4.8. *Let M be given by* (4.9) *with M_{11} and M_{22} being invertible, and let Λ and Δ be the Schur complements of M_{11} and M_{22} in M, respectively. Assume, in addition, that either $M_{12} : Y_1 \to X_2$ is left invertible or $M_{21} : X_1 \to Y_2$ right invertible. Then there exists a Banach space Z such that $\Lambda \dotplus I_Z$ and Δ are equivalent.*

The proof of the above result is similar to that of Theorem 4.5. The details, which are omitted, can be found in [18].

Notes

The problem of realization is a classical problem in system theory, and has many different faces. The literature on this subject is rich. For references we refer to the text books [84], [36]. The material of the first section is standard, cf., [9], Theorem 4.20. Sections 4.2 and 4.3 are based on the paper [52]. Theorem 4.4 is related to the linearization result proved in [28]. For other versions of the realization theorems in Section 4.2 we refer to [96]. The material in Section 4.4 is taken from [18]. We return to the topic of realization in Chapters 7 and 8.

Chapter 5

Factorization and Riccati Equations

In this chapter the state space factorization theory from Section 2.4 is presented using a different terminology. Here it will be based on the notion of an angular operator and the algebraic Riccati equation.

5.1 Angular subspaces and angular operators

Throughout this chapter, X is a complex Banach space and $\mathbb{P}$ is a given fixed projection of X along X_1 onto X_2, so $\operatorname{Ker}\mathbb{P} = X_1$ and $\operatorname{Im}\mathbb{P} = X_2$. Matrix representations of operators acting on X will always be taken with respect to the decomposition $X = X_1 \dot{+} X_2$.

A closed subspace N of X is called *angular* (with respect to $\mathbb{P}$) if $X = \operatorname{Ker}\mathbb{P} \dot{+} N = X_1 \dot{+} N$. If R is a bounded linear operator from X_2 into X_1, then the space

$$N_R = \{Rx + x \mid x \in X_2\} = \operatorname{Im} \begin{bmatrix} R \\ I \end{bmatrix} \tag{5.1}$$

is angular with respect to $\mathbb{P}$. The next proposition shows that any angular subspace is of this form. The operator R appearing here is uniquely determined. It is called the *angular operator* for N. In the next proposition we shall describe a few different ways to express the angular operator.

Proposition 5.1. *Let N be a closed subspace of X. The following statements are equivalent:*

(i) *N is angular with respect to $\mathbb{P}$,*

(ii) *$N = N_R$ for some bounded linear operator R from X_2 into X_1,*

(iii) *the restriction $\mathbb{P}|_N : N \to X_2$ is bijective.*

In that case the angular operator R for N is given by

$$Rx = (I - \mathbb{P})\big(\mathbb{P}\big|_N\big)^{-1}x = \big(\mathbb{P}\big|_N\big)^{-1}x - x, \qquad x \in X_2. \tag{5.2}$$

Proof. As already observed, if $N = N_R$, then N is angular. To prove the converse, assume that N is angular with respect to $\mathbb{P}$, and let Q be the projection of X onto N along X_1. Put $Rx = (Q - \mathbb{P})x$ for $x \in X_2$. Then $N = N_R$.

Suppose that N is angular with angular operator R. The bijectivity of $\mathbb{P}\big|_N$ is clear from the fact that $\mathbb{P}(Rx + x) = x$ for all $x \in X_2$.

Next assume that $\mathbb{P}\big|_N$ is bijective and define R by (5.2). We shall prove that $N = N_R$. First, take $x \in X_2$. Then $Rx + x = (\mathbb{P}\big|_N)^{-1}x \in N$, and hence $N_R \subset N$. Conversely, if $u \in N$, then $v = \mathbb{P}u \in X_2$ and $Rv + v = u$. It follows that $N \subset N_R$, and the proof is complete. □

The next proposition tells us when the kernel of a given projection is an angular subspace with respect to $\mathbb{P}$.

Proposition 5.2. *Let $\mathbb{Q}$ be a another projection of X. Then* $\operatorname{Ker}\mathbb{Q}$ *is angular with respect to $\mathbb{P}$ if and only if the restriction $\mathbb{Q}\big|_{\operatorname{Ker}\mathbb{P}} : \operatorname{Ker}\mathbb{P} \to \operatorname{Im}\mathbb{Q}$ is bijective, and in that case the angular operator R for* $\operatorname{Ker}\mathbb{Q}$ *is given by*

$$Rx = -\big(\mathbb{Q}\big|_{\operatorname{Ker}\mathbb{P}}\big)^{-1}\mathbb{Q}x, \qquad x \in X_2. \tag{5.3}$$

Proof. Observe that $\operatorname{Ker}\mathbb{Q}$ is angular with respect to $\mathbb{P}$ if and only if $\operatorname{Ker}\mathbb{P}$ is angular with respect to $\mathbb{Q}$. So the first part of the proposition follows by applying Proposition 5.1 to $\operatorname{Ker}\mathbb{P}$ and $\mathbb{Q}$.

Next, assume that $\mathbb{Q}\big|_{\operatorname{Ker}\mathbb{P}}$ is bijective. To determine the angular operator R for $\operatorname{Ker}\mathbb{Q}$, note that

$$0 = \mathbb{Q}(Rx + x) = \big(\mathbb{Q}\big|_{\operatorname{Ker}\mathbb{P}}\big)Rx + \mathbb{Q}x$$

for each $x \in X_2$. From this, formula (5.3) is clear. □

In the next proposition we consider the image of X_2 under a general operator, and give conditions under which it is angular with respect to $\mathbb{P}$.

Proposition 5.3. *Let*

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} : X_1 \dot{+} X_2 \to X_1 \dot{+} X_2$$

be an invertible bounded linear operator on $X = X_1 \dot{+} X_2$. Then $S[X_2]$ is angular with respect to $\mathbb{P}$ if and only if S_{22} is bijective, and in that case $R = S_{12}S_{22}^{-1}$ is its angular operator.

Proof. Put $N = S[X_2]$, and let S_0 be the restriction of S to X_2 considered as an operator from X_2 into N. Then S_0 is bijective. Also, let $\mathbb{P}\big|_N$ be the restriction of $\mathbb{P}$ to N considered as an operator into X_2. Since $\big(\mathbb{P}\big|_N\big)S_0 = S_{22}$, we see that $\mathbb{P}\big|_N$ is bijective if and only if this is the case for S_{22}. Apply now Proposition 5.1 and use that $(I - \mathbb{P})S_0 u = S_{12}u,\ u \in X_2$. □

5.2 Angular subspaces and the algebraic Riccati equation

The following question is of interest in view of Theorem 5.5 below. Given an angular subspace N of X and an operator T on X, when is N invariant under T? The next proposition shows that the answer involves an algebraic (operator) Riccati equation.

Proposition 5.4. *Let N be an angular subspace of X with respect to $\mathbb{P}$, and let*

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} : X_1 \dot{+} X_2 \to X_1 \dot{+} X_2 \tag{5.4}$$

be an operator on $X = X_1 \dot{+} X_2$. Then N is invariant under T if and only if the angular operator R for N satisfies

$$RT_{21}R + RT_{22} - T_{11}R - T_{12} = 0. \tag{5.5}$$

Moreover, in that case the operators $T\big|_N$ and $T_{22} + T_{21}R$ are similar.

Equation (5.5) is usually referred to as an *algebraic Riccati equation*, or more precisely, a nonsymmetric version of it. The 2×2 operator matrix in (5.4) is often referred to as the *Hamiltonian* of (5.5).

Proof. Let R be the angular operator for N, and let E be the operator given by

$$E = \begin{bmatrix} I & R \\ 0 & I \end{bmatrix} : X_1 \dot{+} X_2 \to X_1 \dot{+} X_2. \tag{5.6}$$

Note that E is invertible and maps X_2 in a one to one way onto N. It follows that T leaves N invariant if and only if $E^{-1}TE$ leaves X_2 invariant. A direct computation yields

$$E^{-1}TE = \begin{bmatrix} T_{11} - RT_{21} & -RT_{21}R - RT_{22} + T_{11}R + T_{12} \\ T_{21} & T_{22} + T_{21}R \end{bmatrix}. \tag{5.7}$$

This formula shows that $E^{-1}TE$ leaves X_2 invariant if and only if (5.5) is satisfied. This proves the first part of the proposition.

Next, let E_2 be the restriction of E to X_2 considered as an operator from X_2 into N. Then E_2 is invertible. In fact, E_2^{-1} is the restriction of E^{-1} to N viewed as an operator from N into X_2. Using (5.7) we see that $E_2^{-1}(T\big|_N)E_2 = T_{22} + T_{21}R$, and hence $T\big|_N$ and $T_{22} + T_{21}R$ are similar. □

5.3 Angular operators and factorization

In this section we use the concepts introduced in the previous section to bring the factorization theorem for systems in a somewhat different form. The main point is that throughout we work with a fixed decomposition $X = X_1 \dot{+} X_2$ of the state space X of the system that has to be factorized and the factors are described with respect to this decomposition. In the finite-dimensional case this corresponds to working with a fixed coordinate system.

Theorem 5.5. *Let $W(\lambda) = D + C(\lambda I - A)^{-1}B$ be the transfer function of a biproper system $\Theta = (A, B, C, D; X, Y)$. Let $\mathbb{P}$ be a projection of X along X_1 onto X_2, and let N be an angular subspace of X with respect to $\mathbb{P}$ with angular operator R. So $R : X_2 \to X_1$ and $N = N_R$ as in* (5.1)*. Assume that*

$$A[X_1] \subset X_1, \qquad A^{\times}[N] \subset N, \tag{5.8}$$

and let $D = D_1 D_2$ with D_1 and D_2 invertible operators on Y. Write

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \qquad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \qquad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}$$

for the matrix representations of A, B and C with respect to the decomposition $X = X_1 \dot{+} X_2$. Then R satisfies the algebraic Riccati equation

$$RB_2D^{-1}C_1R - R(A_{22} - B_2D^{-1}C_2) + (A_{11} - B_1D^{-1}C_1)R \tag{5.9}$$

$$+ (A_{12} - B_1D^{-1}C_2) = 0.$$

Furthermore $W = W_1W_2$, where

$$\begin{aligned} W_1(\lambda) &= D_1 + C_1(\lambda - A_{11})^{-1}(B_1 - RB_2)D_2^{-1}, \\ W_2(\lambda) &= D_2 + D_1^{-1}(C_1R + C_2)(\lambda - A_{22})^{-1}B_2, \\ W_1^{-1}(\lambda) &= D_1^{-1} - D_1^{-1}C_1(\lambda - A_1^{\times})^{-1}(B_1 - RB_2)D^{-1}, \\ W_2^{-1}(\lambda) &= D_2^{-1} - D^{-1}(C_1R + C_2)(\lambda - A_2^{\times})^{-1}B_2D_2^{-1}, \end{aligned}$$

with $A_1^{\times} = A_{11} - (B_1 - RB_2)D^{-1}C_1$ and $A_2^{\times} = A_{22} - B_2D^{-1}(C_1R + C_2)$.

Proof. Put

$$\Theta_1 = \big(A_{11}, (B_1 - RB_2)D_2^{-1}, C_1, D_1; X_1, Y\big),$$

$$\Theta_2 = \big(A_{22}, B_2, D_1^{-1}(C_1R + C_2), D_2; X_2, Y\big).$$

Then $\Theta \simeq \Theta_1\Theta_2$. More precisely $\Theta_1\Theta_2 = (E^{-1}AE, E^{-1}B, CE, D; X, Y)$, where E is the invertible operator

$$E = \begin{bmatrix} I & R \\ 0 & I \end{bmatrix}.$$

To see this, for convenience introduce

$$\hat{A} = E^{-1}AE = \begin{bmatrix} A_{11} & A_{12} - RA_{22} + A_{11}R \\ 0 & A_{22} \end{bmatrix},$$

$$\hat{B} = E^{-1}B = \begin{bmatrix} B_1 - RB_2 \\ B_2 \end{bmatrix}, \qquad \hat{C} = CE = \begin{bmatrix} C_1 & C_1R + C_2 \end{bmatrix},$$

and set $\hat{\Theta} = (\hat{A}, \hat{B}, \hat{C}, D; X, Y)$. Observe that

$$\begin{aligned} \hat{A}^\times &= E^{-1}A^\times E \\ &= \begin{bmatrix} A_{11} - (B_1 - RB_2)D^{-1}C_1 & H \\ -B_2D^{-1}C_1 & A_{22} - B_2D^{-1}(C_1R + C_2) \end{bmatrix}, \end{aligned}$$

where H is the left-hand side of (5.9). Now E maps X_1 onto X_1 and X_2 onto N. Thus (5.8) implies that

$$\hat{A}[X_1] \subset X_1, \qquad \hat{A}^\times[X_2] \subset X_2.$$

It follows that (5.9) is satisfied (see also Proposition 5.4).

Apply now Theorem 2.3 to show that $\hat{\Theta} = \Theta_1\Theta_2$. As $\widehat{\hat{\Theta}} \simeq \Theta$, the proof of the first part of the theorem is complete. The formulas for the factors and their inverses are now immediate. □

Suppose that the angular subspace N in Theorem 5.5 is the image of X_2 under the invertible operator

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} : X_1 \dot{+} X_2 \to X_1 \dot{+} X_2.$$

Then we know from Proposition 5.3 that S_{22} is invertible and the angular operator R for N is given by $R = S_{12}S_{22}^{-1}$. So then the formulas for Θ_1 and Θ_2 become

$$\Theta_1 = \big(A_{11}, (B_1 - S_{12}S_{22}^{-1}B_2)D_2^{-1}, C_1, D_1; X_1, Y\big),$$

$$\Theta_2 = \big(A_{22}, B_2, D_1^{-1}(C_1S_{12}S_{22}^{-1} + C_2), D_2; X_2, Y\big).$$

Obviously, corresponding formulas for the factors W_1 and W_2 hold. In fact, for the particular case when $D = D_1 = D_2 = I$, we get

$$\Theta_1 = (A_{11}, B_1 - S_{12}S_{22}^{-1}B_2, C_1; X_1, Y) \tag{5.10}$$

$$\Theta_2 = (A_{22}, B_2, C_1S_{12}S_{22}^{-1} + C_2; X_2, Y). \tag{5.11}$$

We shall use this to prove the following analogue of Theorem 4 in the L. Sakhnovich paper [104]; see also [87].

Corollary 5.6. *Let $\Theta = (A, B, C; X, Y)$ and $\tilde{\Theta} = (\tilde{A}, \tilde{B}, \tilde{C}; X, Y)$ be systems such that*

$$AS - S\tilde{A} = B\tilde{C}, \qquad S\tilde{B} = B, \qquad CS = -\tilde{C} \tag{5.12}$$

for some operator $S : X \to X$. Let $\mathbb{P}$ be a projection of X along X_1 onto X_2, and assume that

$$A[X_1] \subset X_1, \qquad \tilde{A}[X_2] \subset X_2.$$

If the operators $S : X \to X$ and $S_{22} = \mathbb{P}S\mathbb{P}\big|_{X_2} : X_2 \to X_2$ are invertible, then $\Theta \simeq \Theta_1\Theta_2$, where Θ_1 and Θ_2 are given by (5.10) *and* (5.11)*, respectively.*

Proof. Formula (5.12) and the invertibility of S imply that the associate system $\Theta^\times$ and $\tilde{\Theta}$ are similar, the similarity being given by S. As

$$S^{-1}A^\times S = S^{-1}AS - S^{-1}BCS = \tilde{A},$$

the space $N = SX_2$ is invariant under $A^\times$. The fact that S_{22} is invertible implies that N is angular with respect to $\mathbb{P}$. But then the remarks made in the paragraph preceding this corollary yield the desired factorization. □

The next theorem is a two-sided version of Theorem 5.5.

Theorem 5.7. *Let $W(\lambda) = D + C(\lambda I - A)^{-1}B$ be the transfer function of a biproper system $\Theta = (A, B, C, D; X, Y)$, and let $\mathbb{P}$ be a projection of X along X_1 onto X_2. Further, let N_1 and N_2 be closed subspaces of X such that*

$$X = X_1 \dot{+} N_2 = N_1 \dot{+} X_2,$$

i.e., N_2 is angular with respect to $\mathbb{P}$ and N_1 is angular with respect to $I - \mathbb{P}$. Let $R_{12} : X_2 \to X_1$ and $R_{21} : X_1 \to X_2$ be the corresponding angular operators. Assume

$$X = N_1 \dot{+} N_2, \qquad A[N_1] \subset N_1, \qquad A^\times[N_2] \subset N_2, \tag{5.13}$$

and let $D = D_1 D_2$ with D_1 and D_2 invertible operators on Y. Write

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \qquad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \qquad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}$$

for the matrix representations of A, B and C with respect to the decomposition $X = X_1 \dot{+} X_2$. Introduce $R_1 = I - R_{12}R_{21}$ and $R_2 = I - R_{21}R_{12}$. Then $R_1 : X_1 \to X_1$ and $R_2 : X_2 \to X_2$ are invertible. Also put

$$A_1^\times = A_{11} - B_1 D^{-1} C_1 - R_{12} A_{21} + R_{12} B_2 D^{-1} C_1,$$

$$A_2^\times = A_{22} - B_2 D^{-1} C_2 + A_{21} R_{12} - B_2 D^{-1} C_1 R_{12}.$$

Then $W = W_1 W_2$, where

$$\begin{aligned} W_1(\lambda) &= D_1 + (C_1 + C_2 R_{21})\big(\lambda - (A_{11} + A_{12}R_{21})\big)^{-1} \times \\ &\qquad \times R_1^{-1}(B_1 - R_{12}B_2)D_2^{-1}, \\ W_2(\lambda) &= D_2 + D_1^{-1}(C_1 R_{12} + C_2)R_2^{-1}\big(\lambda - (A_{22} - R_{21}A_{12})\big)^{-1} \times \\ &\qquad \times (B_2 - R_{21}B_1), \\ W_1^{-1}(\lambda) &= D_1^{-1} - D_1^{-1}(C_1 + C_2 R_{21})R_1^{-1}(\lambda - A_1^\times)^{-1}(B_1 - R_{12}B_2)D^{-1}, \\ W_2^{-1}(\lambda) &= D_2^{-1} - D^{-1}(C_1 R_{12} + C_2)(\lambda - A_2^\times)^{-1}R_2^{-1}(B_2 - R_{21}B_1)D_2^{-1}. \end{aligned}$$

Before proving the theorem we present a lemma.

Lemma 5.8. *Let N_1 and N_2 be closed subspaces of X such that*

$$X = X_1 \dot{+} N_2 = N_1 \dot{+} X_2,$$

that is, N_2 is angular with respect to $\mathbb{P}$ and N_1 is angular with respect to $I - \mathbb{P}$ where $\mathbb{P}$ be a projection of X along X_1 onto X_2. Let $R_{12} : X_2 \to X_1$ and $R_{21} : X_1 \to X_2$ be the corresponding angular operators. Then the following statements are equivalent.

(i) $X = N_1 \dot{+} N_2$,
(ii) $I - R_{21}R_{12}$ *is invertible,*
(iii) $I - R_{12}R_{21}$ *is invertible,*
(iv) $F = \begin{bmatrix} I & R_{12} \\ R_{21} & I \end{bmatrix} : X_1 \dot{+} X_2 \to X_1 \dot{+} X_2$ *is invertible.*

In case the equivalent conditions (i)–(iv) *hold, the projection $\mathbb{P}_N$ of X along N_1 onto N_2 is given by*

$$\mathbb{P}_N = \begin{bmatrix} R_{12} \\ I \end{bmatrix} \big(I - R_{21}R_{12}\big)^{-1} \begin{bmatrix} -R_{21} & I \end{bmatrix},$$

while the complementary projection $I - \mathbb{P}_N$ can be written as

$$I - \mathbb{P}_N = \begin{bmatrix} I \\ R_{21} \end{bmatrix} \big(I - R_{12}R_{21}\big)^{-1} \begin{bmatrix} I & -R_{12} \end{bmatrix}.$$

Proof. The equivalence of (ii), (iii) and (iv) is straightforward. Observe that F maps X_1 and X_2 in a one-one manner onto N_1 and N_2, respectively. As $X = X_1 \dot{+} X_2$, it is clear that $X = N_1 \dot{+} N_2$ if and only if F is invertible. So (i) and (iv) are equivalent.

To complete the proof it remains to prove the formula for $\mathbb{P}_N$. Observe that the given formula does define a projection. Its image and kernel are given by

$$\operatorname{Im} \begin{bmatrix} R_{12} \\ I \end{bmatrix}, \qquad \operatorname{Im} \begin{bmatrix} I \\ R_{21} \end{bmatrix},$$

respectively, so it is indeed equal to the projection $\mathbb{P}_N$. □

Proof of Theorem 5.7. From Lemma 5.8 we know that the operator

$$F = \begin{bmatrix} I & R_{12} \\ R_{21} & I \end{bmatrix} : X_1 \dot{+} X_2 \to X_1 \dot{+} X_2$$

is invertible. Introduce $\hat{\Theta} = \big(\hat{A}, \hat{B}, \hat{C}, D; X, Y\big)$ with

$$\hat{A} = F^{-1}AF, \qquad \hat{B} = F^{-1}B, \qquad \hat{C} = CF.$$

Then the biproper systems $\hat{\Theta}$ and Θ are similar, and so they have the same transfer function, namely W. Note that

$$\hat{A}[X_1] \subset X_1, \qquad \hat{A}^{\times} X_2 \subset [X_2]$$

where, following standard convention $\hat{A}^{\times} = \hat{A} - \hat{B}D^{-1}\hat{C}$, and so $\hat{A}^{\times} = F^{-1}A^{\times}F$. Write

$$\hat{A} = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ 0 & \hat{A}_{22} \end{bmatrix}, \qquad \hat{B} = \begin{bmatrix} \hat{B}_1 \\ \hat{B}_2 \end{bmatrix}, \qquad \hat{C} = \begin{bmatrix} \hat{C}_1 & \hat{C}_2 \end{bmatrix},$$

and put

$$\Theta_1 = \big(\hat{A}_{11}, \hat{B}_1 D_2^{-1}, \hat{C}_1, D_1; X_1, Y\big),$$

$$\Theta_2 = \big(\hat{A}_{22}, \hat{B}_2, D_1^{-1}\hat{C}_2, D_2; X_2, Y\big).$$

Then $\Theta = \Theta_1\Theta_2$, and on $\rho(\hat{A}_{11}) \cap \rho(\hat{A}_{22}) \subset \rho(\hat{A}) = \rho(A)$, the function W is the product of the transfer functions of the biproper systems Θ_1 and Θ_2.

The inverse of F is given by

$$F^{-1} = \begin{bmatrix} R_1^{-1} & -R_1^{-1}R_{12} \\ -R_{21}R_1^{-1} & I + R_{21}R_1^{-1}R_{12} \end{bmatrix} : X_1 \dot{+} X_2 \to X_1 \dot{+} X_2.$$

Using this and the expression for F, one easily sees that

$$\hat{A}_{11} = R_1^{-1}\big(A_{11} + A_{12}R_{21} - R_{12}A_{21} - R_{12}A_{22}R_{21}\big),$$
$$\hat{B}_1 D_2^{-1} = R_1^{-1}\big(B_1 - R_{12}B_2\big)D_2^{-1}, \qquad \hat{C}_1 = C_1 + C_2R_{21}.$$

Now R_{21} satisfies the algebraic Riccati equation

$$R_{21}A_{12}R_{21} + R_{21}A_{11} - A_{22}R_{21} - A_{21} = 0,$$

and it follows that $\hat{A}_{11} = A_{11} + A_{12}R_{21}$. Thus, for the transfer function of Θ_1, we have

$$D_1 + \hat{C}_1(\lambda - \hat{A}_{11})^{-1}\hat{B}_1 D_2^{-1}$$
$$= D_1 + \big(C_1 + C_2R_{21}\big)\big(\lambda - (A_{11} + A_{12}R_{21})\big)^{-1}R_1^{-1}\big(B_1 - R_{12}B_2\big)D_2^{-1},$$

that is, it is precisely the factor W_1 in the theorem.

Next we compute the transfer function of Θ_2. Using the alternative formula

$$F^{-1} = \begin{bmatrix} I + R_{12}R_2^{-1}R_{21} & -R_{12}R_2^{-1} \\ -R_2^{-1}R_{21} & R_2^{-1} \end{bmatrix} : X_1 \dot{+} X_2 \to X_1 \dot{+} X_2$$

for the inverse of F, we obtain

$$\begin{aligned} \hat{A}_{22} &= R_2^{-1}(-R_{21}A_{11}R_{12} - R_{21}A_{12} + A_{21}R_{12} + A_{22}) \\ &= R_2^{-1}(A_{22} - R_{21}A_{12})R_2^{-1}, \end{aligned}$$
$$\hat{B}_2 = R_2^{-1}(B_2 - R_{21}B_1), \qquad D_1^{-1}\hat{C}_1 = D_1^{-1}(C_1R_{12} + C_2).$$

Hence the transfer function of Θ_2 is given by

$$D_2 + D_1^{-1}\hat{C}_2(\lambda - \hat{A}_{22})^{-1}\hat{B}_2$$
$$= D_2 + D_1^{-1}(C_1R_{12} + C_2)R_2^{-1}\big(\lambda - (A_{22} - R_{21}A_{12})\big)^{-1}(B_1 - R_{12}B_2)D_2^{-1},$$

so it coincides with the factor W_2 in the theorem.

This proves that the factorization claimed in the theorem holds on

$$\rho(A_{11} + A_{12}R_{21}) \cap \rho(A_{22} - R_{21}A_{12}) \subset \rho(A).$$

What remains to be done is deducing the formulas for the inverses. But this amounts to repeating the work with Θ replaced by its associate system $\Theta^\times$, thereby employing the Riccati equation

$$R_{12}(A_{21} - B_2D^{-1}C_1)R_{12} + R_{12}(A_{22} - B_2D^{-1}C_2)$$
$$-(A_{11} - B_1D^{-1}C_1)R_{12} - (A_{12} - B_1D^{-1}C_2) = 0,$$

for R_{12} instead of the one for R_{21} used above. The details are omitted. □

5.4 Angular spectral subspaces and the algebraic Riccati equation

In this section Proposition 5.4 is specified further for the case when the angular subspace N is a spectral subspace of T. We begin with some preliminaries that will be useful in the next chapter too.

Let Γ be a Cauchy contour (see Section 4.2). We say that Γ *splits the spectrum* $\sigma(S)$ of a bounded linear operator S if Γ and $\sigma(S)$ have empty intersection. In that case $\sigma(S)$ decomposes into two disjoint compact sets σ_+ and σ_- such that σ_+ is in the inner domain of Γ and σ_- is in the outer domain of Γ. If Γ splits the spectrum of S, then we have a *Riesz projection*, also called *spectral projection*, associated with S and Γ, namely

$$P(S;\Gamma) = \frac{1}{2\pi i}\int_\Gamma (\lambda - S)^{-1}\, d\lambda.$$

The subspace $N = \operatorname{Im} P(S;\Gamma)$ will be called the *spectral subspace* for S corresponding to the contour Γ (or to the spectral set σ_+).

Lemma 5.9. *Let Y_1 and Y_2 be complex Banach spaces, and consider the operator*

$$S = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} : Y_1 \dot{+} Y_2 \to Y_1 \dot{+} Y_2. \tag{5.14}$$

Let Π be any projection of $Y = Y_1 \dot{+} Y_2$ such that $\operatorname{Ker}\Pi = Y_1$. Then the compression $\Pi S\big|_{\operatorname{Im}\Pi} : \operatorname{Im}\Pi \to \operatorname{Im}\Pi$ and $S_{22} : Y_2 \to Y_2$ are similar. Furthermore, Y_1 is a spectral subspace for S if and only if $\sigma(S_{11}) \cap \sigma(S_{22}) = \emptyset$, and in that case $\sigma(S) = \sigma(S_{11}) \cup \sigma(S_{22})$ while, in addition,

$$Y_1 = \operatorname{Im} P(S;\Gamma) = \operatorname{Im}\left(\frac{1}{2\pi i}\int_\Gamma (\lambda - S)^{-1}\, d\lambda\right), \tag{5.15}$$

where Γ is a Cauchy contour around $\sigma(S_{11})$ separating $\sigma(S_{11})$ from $\sigma(S_{22})$.

Proof. Let P be the projection of $Y = Y_1 \dot{+} Y_2$ along Y_1 onto Y_2. As $\operatorname{Ker} P = \operatorname{Ker}\Pi$, we have $P = P\Pi$ and the map

$$E = P\big|_{\operatorname{Im}\Pi} : \operatorname{Im}\Pi \to Y_2$$

is an invertible operator. Write S_0 for the compression $\Pi S\big|_{\operatorname{Im}\Pi}$ of S to $\operatorname{Im}\Pi$ viewed as an operator from $\operatorname{Im}\Pi$ into $\operatorname{Im}\Pi$, and take $x = \Pi y$. Then

$$ES_0x = P\Pi S\Pi y = PS\Pi y = PSP\Pi y = S_{22}Ex,$$

hence S_0 and S_{22} are similar.

Now suppose $\sigma(S_{11}) \cap \sigma(S_{22}) = \emptyset$. Let λ be an arbitrary complex number. Since $\sigma(S_{11}) \cap \sigma(S_{22}) = \emptyset$, at least one of the two operators $\lambda - S_{11}$ and $\lambda - S_{22}$ is invertible. But then, by applying Lemma 2.9 with $X_1 = Y_1$, $X_2 = Y_2$ and $A = \lambda - S$, we see that $\lambda - S$ is invertible if and only both $\lambda - S_{11}$ and $\lambda - S_{22}$ are invertible. It follows that $\rho(S_{11}) \cap \rho(S_{22}) = \rho(S)$, an identity which can be rewritten as $\sigma(S) = \sigma(S_{11}) \cup \sigma(S_{22})$.

Still under the assumption that $\sigma(S_{11}) \cap \sigma(S_{22}) = \emptyset$, let Γ be a Cauchy contour Γ around $\sigma(S_{11})$ separating $\sigma(S_{11})$ from $\sigma(S_{22})$. Then Γ splits the spectrum of S. In fact, if $\lambda \in \Gamma$, then both $\lambda - S_{11}$ and $\lambda - S_{22}$ are invertible and

$$(\lambda - S)^{-1} = \begin{bmatrix} (\lambda - S_{11})^{-1} & (\lambda - S_{11})^{-1} S_{12} (\lambda - S_{22})^{-1} \\ 0 & (\lambda - S_{22})^{-1} \end{bmatrix},$$

which leads to an expression of the type

$$P(S; \Gamma) = \begin{bmatrix} I & * \\ 0 & 0 \end{bmatrix}$$

for the Riesz projection associated with S and Γ. In particular, it is clear that $Y_1 = \operatorname{Im} P(S; \Gamma)$. So Y_1 is a spectral subspace for S and (5.15) holds.

Finally, assume that $Y_1 = \operatorname{Im} Q$, where Q is a Riesz projection for S. Put $\Pi = I - Q$, and let S_0 be the restriction of S to $\operatorname{Im} \Pi$. Then $\sigma(S_{11}) \cap \sigma(S_0) = \emptyset$. By the first part of the proof, the operators S_0 and S_{22} are similar. So $\sigma(S_0) = \sigma(S_{22})$, and we have shown that $\sigma(S_{11}) \cap \sigma(S_{22}) = \emptyset$. □

Next we present the analogue of Proposition 5.4 for spectral subspaces. Recall that $\mathbb{P}$ is a projection of X along X_1 onto X_2.

Proposition 5.10. *Let N be an angular subspace of X with respect to $\mathbb{P}$, and let T be the operator on X given by* (5.4)*. Then N is a spectral subspace for T if and only if the angular operator R for N satisfies the algebraic Riccati equation* (5.5) *and*

$$\sigma(T_{11} - RT_{21}) \cap \sigma(T_{22} + T_{21}R) = \emptyset.$$

More precisely the following holds. If $N = \operatorname{Im} P(T; \Gamma)$, where Γ is a Cauchy contour that splits $\sigma(T)$, then $\sigma(T_{22} + T_{21}R)$ is inside Γ and $\sigma(T_{11} - RT_{21})$ is outside Γ. Conversely, if Γ is a Cauchy contour such that $\sigma(T_{22} + T_{21}R)$ is inside Γ and $\sigma(T_{11} - RT_{21})$ is outside Γ, then the spectrum of T does not intersect with Γ and $N = \operatorname{Im} P(T; \Gamma)$.

Proof. Let R be the angular operator for N, and let E be the operator given by (5.6). We know that E is invertible and maps X_2 in a one to one way onto N. Since a spectral subspace of T is invariant under T, we may without loss of generality

assume that the angular operator R for N satisfies the Riccati equation (5.5). Then formula (5.7) shows that

$$E^{-1}TE = \begin{bmatrix} T_{11} - RT_{21} & 0 \\ T_{21} & T_{22} + T_{21}R \end{bmatrix}. \tag{5.16}$$

Since E maps X_2 in a one to one way onto N, the space N is a spectral subspace for T if and only if X_2 is a spectral subspace for $E^{-1}TE$. and we can apply Lemma 5.9 to get the desired result. □

Notes

The notion of an angular operator is standard in operator theory and goes back to [90]. The theory of Riccati equations is important in system theory; see, e.g., the text books [84], [36]. For more details on this subject we also refer to the monograph [91] or Section 1.6 in [70]. For the basic facts about Cauchy domains, Riesz projections and spectral subspaces used in Section 5.4 we refer to Sections I.1–I.3 in [46]. This chapter is a rewritten and reorganized version of Chapter 5 in [14].

Chapter 6

Canonical Factorization and Applications

As we have seen in Chapter 1 canonical factorization serves as tool to solve Wiener-Hopf integral equations and their discrete analogues, the block Toeplitz equations. In this chapter the state space factorization method developed in Chapter 2 is used to solve the problem of canonical factorization (necessary and sufficient conditions for its existence) and to derive explicit formulas for its factors. This is done in Section 6.1 for rational matrix functions. The results are applied to invert Wiener-Hopf integral equations (Section 6.2) and block Toeplitz operators (Section 6.3) with a rational matrix symbol.

6.1 Canonical factorization of rational matrix functions

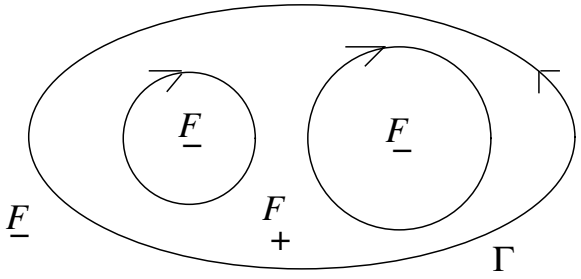
In this section we shall consider the factorization theorems of Section 2.4 (see also Section 2.5) for the special case when the two factors have disjoint spectra. First we introduce some additional terminology.

For a Cauchy contour Γ we let F_+ denote the interior domain of Γ and F_- will be the complement of $\overline{F}_+$ in the Riemann sphere $\mathbb{C} \cup \{\infty\}$. Note that it is assumed that $\infty \in F_-$.

Let W be a rational $m \times m$ matrix function, with $W(\infty) = I$, analytic on an open neighborhood of Γ, whose values on Γ are invertible matrices. By a *right canonical factorization* of W with respect to Γ we mean a factorization

$$W(\lambda) = W_-(\lambda)W_+(\lambda), \qquad \lambda \in \Gamma, \tag{6.1}$$

where W_- and W_+ are rational $m \times m$ matrix functions, analytic and taking invertible values on an open neighborhood of $\overline{F}_-$ and $\overline{F}_+$, respectively.



If in (6.1) the factors W_- and W_+ are interchanged, then we speak of a *left canonical factorization*.

Theorem 6.1. *Let $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ be the transfer function of the unital system $\Theta = (A, B, C; \mathbb{C}^n, \mathbb{C}^m)$, and let Γ be a Cauchy contour. Assume that Γ splits the spectra of A and $A^\times$. Then W admits a right canonical factorization with respect to Γ if and only if*

$$\mathbb{C}^n = \operatorname{Im} P(A;\Gamma) \dotplus \operatorname{Ker} P(A^\times;\Gamma). \tag{6.2}$$

In that case, such a right canonical factorization is given by

$$W(\lambda) = W_-(\lambda)W_+(\lambda), \qquad \lambda \in \rho(A),$$

where the factors and their inverses are given by

$$\begin{aligned} W_-(\lambda) &= I_m + C(\lambda I_n - A)^{-1}(I_n - \Pi)B, \\ W_+(\lambda) &= I_m + C\Pi(\lambda I_n - A)^{-1}B, \\ W_-^{-1}(\lambda) &= I_m - C(I_n - \Pi)(\lambda I_n - A^\times)^{-1}B, \\ W_+^{-1}(\lambda) &= I_m - C(\lambda I_n - A^\times)^{-1}\Pi B, \end{aligned}$$

with Π the projection of $\mathbb{C}^n$ along $\operatorname{Im} P(A;\Gamma)$ onto $\operatorname{Ker} P(A^\times;\Gamma)$.

For left canonical factorizations an analogous theorem holds. In the result in question, the direct sum decomposition (6.2) is replaced by the decomposition $\mathbb{C}^n = \operatorname{Ker} P(A;\Gamma) \dotplus \operatorname{Im} P(A^\times;\Gamma)$.

Proof. Let Θ be as in the first part of the theorem. Assume that (6.2) holds. Note that $X_1 = \operatorname{Im} P(A;\Gamma)$ is invariant for A and $X_2 = \operatorname{Ker} P(A^\times;\Gamma)$ is invariant for $A^\times$. So, by definition, the projection Π is a supporting projection for Θ. Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \qquad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \qquad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}$$

be the matrix representations of A, B and C with respect to the decomposition $\mathbb{C}^n = X_1 \dot{+} X_2$. Then

$$\begin{aligned} \mathrm{pr}_{\Pi}(\Theta) &= (A_{22}, B_2, C_2; X_2, \mathbb{C}^m), \\ \mathrm{pr}_{I-\Pi}(\Theta) &= (A_{11}, B_1, C_1; X_1, \mathbb{C}^m), \end{aligned}$$

and we know that $\Theta = \mathrm{pr}_{I-\Pi}(\Theta)\mathrm{pr}_{\Pi}(\Theta)$. It follows (see Sections 2.4 and 2.5) that

$$W(\lambda) = W_{\Theta}(\lambda) = W_-(\lambda)W_+(\lambda) \tag{6.3}$$

for each $\lambda \in \rho(A_{11}) \cap \rho(A_{22})$.

As X_1 is a spectral subspace for A, we can apply Lemma 5.9 to show that $\sigma(A_{11}) \cap \sigma(A_{22}) = \emptyset$. But then $\rho(A) = \rho(A_{11}) \cap \rho(A_{22})$ and it follows that (6.3) holds for each $\lambda \in \rho(A)$. Also, we see from Lemma 5.9 that

$$\sigma(A_{11}) = \sigma(A) \cap F_+, \qquad \sigma(A_{22}) = \sigma(A) \cap F_-. \tag{6.4}$$

In a similar way one proves that

$$\sigma(A_{11}^{\times}) = \sigma(A^{\times}) \cap F_+, \qquad \sigma(A_{22}^{\times}) = \sigma(A^{\times}) \cap F_-. \tag{6.5}$$

As $W_-(\lambda) = I + B_1(\lambda - A_{11})^{-1}C_1$, we know that W_- is defined and analytic on the complement of $\sigma(A_{11})$ and $W_-(\lambda)$ is invertible for $\lambda \notin \sigma(A_{11}^{\times})$. So using the first parts of (6.4) and (6.5), it follows that W_- is analytic and has invertible values on an open neighborhood of $\overline{F}_-$. In the same way, using the second parts of (6.4) and (6.5), one proves that W_+ is analytic and has invertible values on an open neighborhood of $\overline{F}_+$.

Conversely, let $W = W_-W_+$ be a right canonical factorization with respect to Γ. By a simple modification of the factors we can reach the situation where $W_-(\infty) = W_+(\infty) = W(\infty) = I_m$. It is our task to show that $\mathbb{C}^n = \mathrm{Im}\, P(A;\Gamma) \dot{+} \mathrm{Ker}\, P(A^{\times};\Gamma)$. First the identity

$$\mathrm{Im}\, P(A;\Gamma) \cap \mathrm{Ker}\, P(A^{\times};\Gamma) = \{0\}$$

will be established.

Suppose $x \in \mathrm{Im}\, P(A;\Gamma) \cap \mathrm{Ker}\, P(A^{\times};\Gamma)$, and consider $(\lambda - A)^{-1}x$. This function is analytic on an open neighborhood of $\overline{F}_-$. On the other hand $(\lambda - A^{\times})^{-1}x$ is analytic on an open neighborhood of $\overline{F}_+$. For λ in the intersection of $\rho(A)$ and $\rho(A^{\times})$, we have

$$\begin{aligned} W(\lambda)C(\lambda - A^{\times})^{-1} &= C(\lambda - A^{\times})^{-1} + C(\lambda - A)^{-1}BC(\lambda - A^{\times})^{-1} \\ &= C(\lambda - A^{\times})^{-1} + C(\lambda - A)^{-1}(A - A^{\times})(\lambda - A^{\times})^{-1} \\ &= C(\lambda - A)^{-1}, \end{aligned}$$

and it follows that $W_+(\lambda)C(\lambda - A^{\times})^{-1} = W_-(\lambda)^{-1}C(\lambda - A)^{-1}$. The analyticity properties of the factors W_-, W_+ and their inverses now imply that the function

$W_+(\lambda)C(\lambda-A^\times)^{-1}x = W_-(\lambda)^{-1}C(\lambda-A)^{-1}x$ is analytic on the extended complex plane. By Liouville's theorem it must be constant. As the function in question has the value zero at infinity, it is identically zero. Hence both $C(\lambda - A^\times)^{-1}x$ and $C(\lambda - A)^{-1}x$ vanish. Next use the identity

$$(\lambda - A^\times)^{-1}BC(\lambda - A)^{-1} = (\lambda - A)^{-1} - (\lambda - A^\times)^{-1}$$

to obtain $(\lambda - A^\times)^{-1}x = (\lambda - A)^{-1}x$. But then this function is analytic on the extended complex plane too. Using Liouville's theorem again, we see that it must be identically zero. Thus $x = 0$.

Observe that up to this point in the proof we have not used the finite dimensionality of the state space.

We now finish the proof by a duality argument. Introduce

$$W^*(\lambda) = I_m + B^*(\lambda I_n - A^*)^{-1}C^*,$$

and let Γ^* be the adjoint curve of Γ, i.e., the curve obtained from Γ by complex conjugation. Then $W^*(\lambda) = W_+^*(\lambda)W_-^*(\lambda)$ is a left canonical factorization. On the basis of a similar argument as above, we may conclude that $\operatorname{Ker} P(A^*, \Gamma^*) \cap \operatorname{Im} P\big((A^\times)^*, \Gamma^*\big) = 0$. It follows that

$$\operatorname{Ker} P(A^*, \Gamma^*) + \operatorname{Im} P((A^\times)^*, \Gamma^*) = \mathbb{C}^n.$$

In first instance, this holds for the closure of $\operatorname{Ker} P(A^*, \Gamma^*) + \operatorname{Im} P((A^\times)^*, \Gamma^*)$, but in $\mathbb{C}^n$ all linear manifolds are closed. □

With minor modifications we could have worked in Theorem 6.1 with two curves, one splitting the spectrum of A and the other splitting the spectrum of $A^\times$ (cf., [87]).

Finally, let us mention that Theorem 6.1 remains true if the Cauchy contour Γ is replaced by the closure of the real line on the Riemann sphere $\mathbb{C} \cup \infty$. In that case F_+ is the open upper half-plane and F_- is the open lower half-plane.

6.2 Application to Wiener-Hopf integral equations

In this section the general factorization result proved in the preceding sections is used to provide explicit formulas for solutions of the vector-valued Wiener-Hopf equation

$$\phi(t) - \int_0^\infty k(t-s)\phi(s)\,ds = f(t), \qquad 0 \le t < \infty, \tag{6.6}$$

where ϕ and f are m-dimensional vector functions and $k \in L_1^{m\times m}(-\infty, \infty)$, i.e., the kernel function k is an $m \times m$ matrix function of which the entries are in $L_1(-\infty, \infty)$. We assume that the given vector function f has its component functions in $L_p[0, \infty)$, and we express this property by writing $f \in L_p^m[0, \infty)$. Throughout this section p will be fixed and $1 \le p < \infty$. Given the kernel function k and the

right-hand side f, the problem we shall consider is to find a solution ϕ for equation (6.6) that also belongs to the space $L_p^m[0,\infty)$. As was mentioned in Section 1.5, equation (6.6) has a unique solution in $L_p^m[0,\infty)$ for each f in $L_p^m[0,\infty)$ if and only if its symbol $I_m - K(\lambda)$ admits a factorization as in (1.25).

Our aim is to apply the factorization theory developed in the previous sections to get the canonical factorization (1.25). Therefore, in the sequel we assume that the symbol is a rational $m \times m$ matrix function. As $K(\lambda)$ is the Fourier transform of an $L_1^{m\times m}(-\infty,\infty)$–function, the symbol is continuous on the real line. In particular, $I_m - K(\lambda)$ has no poles on the real line. Furthermore, by the Riemann-Lebesgue lemma,

$$\lim_{\lambda\in\mathbb{R},\ |\lambda|\to\infty} K(\lambda) = 0,$$

which implies that the symbol $I_m - K(\lambda)$ has the value I_m at ∞.

The assumption that $I_m - K(\lambda)$ is rational is equivalent to the requirement that the kernel function k is in the linear space spanned by all functions of the form

$$h(t) = \begin{cases} p(t)e^{i\alpha t}, & t > 0, \\ \\ q(t)e^{i\beta t}, & t < 0, \end{cases}$$

where $p(t)$ and $q(t)$ are matrix polynomials in t with coefficients in $\mathbb{C}^{m\times m}$, and α and β are complex numbers with $\Im\alpha > 0$ and $\Im\beta < 0$.

Since $I_m - K(\lambda)$ is rational, continuous on the real line, and takes the value I_m at ∞, one can construct (see Section 4.1) a system $\Theta = (A, B, C; \mathbb{C}^n, \mathbb{C}^m)$ such that A has no real eigenvalues and

$$I_m - K(\lambda) = I_m + C(\lambda I_n - A)^{-1}B.$$

In the next theorem we express the solvability of equation (6.6) in terms of such a realization and give explicit formulas for its solutions in the same terms.

Theorem 6.2. *Let $I_m - K(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ be a realization for the symbol of equation* (6.6)*, and assume A has no real eigenvalues. In order that* (6.6) *has a unique solution ϕ in $L_p^m[0,\infty)$ for each f in $L_p^m[0,\infty)$ the following two conditions are necessary and sufficient:*

(i) *$A^\times = A - BC$ has no real eigenvalues;*

(ii) *$\mathbb{C}^n = M \dot{+} M^\times$, where M is the spectral subspace of A corresponding to the eigenvalues of A in the upper half-plane, and $M^\times$ is the spectral subspace of $A^\times$ corresponding to the eigenvalues of $A^\times$ in the lower half-plane.*

Assume conditions (i) *and* (ii) *hold true, and let Π be the projection of $\mathbb{C}^n$ along M onto $M^\times$. Then the symbol $I_m - K(\lambda)$ admits a right canonical factorization with respect to the real line that has the form*

$$I_m - K(\lambda) = \big(I_m + G_-(\lambda)\big)\big(I_m + G_+(\lambda)\big), \qquad \lambda \in \mathbb{R},$$

where the factors and their inverses are given by

$$\begin{aligned} I_m + G_+(\lambda) &= I_m + C\Pi(\lambda I_n - A)^{-1}B, \\ I_m + G_-(\lambda) &= I_m + C(\lambda I_n - A)^{-1}(I_n - \Pi)B, \\ \big(I_m + G_+(\lambda)\big)^{-1} &= I_m - C(\lambda I_n - A^\times)^{-1}\Pi B, \\ \big(I_m + G_-(\lambda)\big)^{-1} &= I_m - C(I_n - \Pi)(\lambda I_n - A^\times)^{-1}B. \end{aligned}$$

The functions γ_+ and γ_- in (1.26) *and* (1.27) *have the representation*

$$\begin{aligned} \gamma_+(t) &= +iCe^{-itA^\times}\Pi B, \qquad t > 0, \\ \gamma_-(t) &= -iC(I_n - \Pi)e^{-itA^\times t}B, \qquad t < 0. \end{aligned}$$

Finally, the solution ϕ to (6.6) *can be written as*

$$\phi(t) = f(t) + \int_0^\infty \gamma(t,s)f(s)\,ds,$$

where

$$\gamma(t,s) = \begin{cases} +iCe^{-itA^\times}\Pi e^{isA^\times}B, & s < t, \\ -iCe^{-itA^\times}(I_n - \Pi)e^{isA^\times}B, & s > t. \end{cases}$$

Proof. We have already mentioned that equation (6.6) has a unique solution in $L_p^m[0,\infty)$ for each f in $L_p^m[0,\infty)$ if and only if the symbol $I_m - K(\lambda)$ admits a right canonical factorization as in (1.25). So to prove the necessity and sufficiency of the conditions (i) and (ii), it suffices to show that the conditions (i) and (ii) together are equivalent to the statement that $I_m - K(\lambda)$ admits a right canonical factorization as in (1.25). We first observe that condition (i) is equivalent to the requirement that $I_m - K(\lambda)$ is invertible for all $\lambda \in \mathbb{R}$ (see Theorem 2.1). But then we can apply the version of Theorem 6.1 referred to in the remark made at the end of Section 6.1 to prove the first part of the theorem.

Next assume that conditions (i) and (ii) hold true. Applying Theorem 6.1 once again, we get the desired formulas for $I_m + G_+(\lambda)$, $I_m + G_-(\lambda)$ and their inverses. The formulas for γ_+ and γ_- are now obtained by noticing that for $\lambda \in \rho(A^\times)$, $\Im\lambda \geq 0$,

$$\int_0^\infty e^{i\lambda t}e^{-itA^\times}\Pi\,dt = i(\lambda I_n - A^\times)^{-1}\Pi,$$

while for $\lambda \in \rho(A^\times)$, $\Im\lambda \leq 0$,

$$\int_{-\infty}^0 e^{i\lambda t}(I_n - \Pi)e^{-itA^\times}\,dt = -i(I_n - \Pi)(\lambda I_n - A^\times)^{-1}.$$

The proof of the latter identity uses (the first conclusion in) Lemma 5.9.

It remains to prove the final formula for $\gamma(t,s)$. We use (1.24), and compute first that

$$\gamma_+(t-r)\gamma_-(r-s) = Ce^{-i(t-r)A^\times}\Pi BC(I-\Pi.)e^{-i(r-s)A^\times}B$$

Here and below $I = I_n$. Now $\operatorname{Ker}\Pi = M$ is A-invariant and $\operatorname{Im}\Pi = M^\times$ is $A^\times$-invariant. Thus $\Pi A(I-\Pi) = 0$ and $(I-\Pi)A^\times\Pi = 0$, from which it follows that

$$\Pi BC(I-\Pi) = \Pi(A-A^\times)(I-\Pi) = \Pi A^\times - A^\times\Pi.$$

But then

$$\begin{aligned}\gamma_+(t-r)\gamma_-(r-s) &= Ce^{-i(t-r)A^\times}(A^\times\Pi - \Pi A^\times)e^{-i(r-s)A^\times}B \\ &= -i\frac{d}{dr}Ce^{-i(t-r)A^\times}\Pi e^{-i(r-s)A^\times}B.\end{aligned}$$

Inserting this in (1.24) we obtain for $s < t$ that

$$\begin{aligned}\gamma(t,s) &= iCe^{-i(t-s)A^\times}\Pi B - \int_0^s i\frac{d}{dr}Ce^{-i(t-r)A^\times}\Pi e^{-i(r-s)A^\times}B\,dr \\ &= iCe^{-i(t-s)A^\times}\Pi B - Ce^{-i(t-r)A^\times}\Pi e^{-i(r-s)A^\times}B|_{r=0}^s \\ &= iCe^{-itA^\times}\Pi e^{isA^\times}B,\end{aligned}$$

while for $s > t$ we get

$$\begin{aligned}\gamma(t,s) &= -iC(I-\Pi)e^{-i(t-s)A^\times}B + \int_0^t i\frac{d}{dr}Ce^{-i(t-r)A^\times}\Pi e^{-i(r-s)A^\times}B\,dr \\ &= -iC(I-\Pi)e^{-i(t-s)A^\times}B - Ce^{-i(t-r)A^\times}\Pi e^{-i(r-s)A^\times}B|_{r=0}^t \\ &= -iCe^{-itA^\times}(I-\Pi)e^{isA^\times}B.\end{aligned}$$

This completes the proof. □

Corollary 6.3. *Let $I_m - K(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ be a realization for the symbol of equation (6.6). Assume that A and $A^\times = A - BC$ have no spectrum on the real line, and that*

$$\mathbb{C}^n = \operatorname{Im}P \dotplus \operatorname{Ker}P^\times,$$

where P and $P^\times$ are the Riesz projections of A and $A^\times$, respectively, corresponding to the spectra in the upper half-plane. Fix $x \in \operatorname{Ker}P$, and let $f(t) = Ce^{-itA}x$, $t \geq 0$. Then f belongs to $L_p^m[0,\infty)$ and the unique solution ϕ in $L_p^m[0,\infty)$ of equation (6.6) with this particular f in the right-hand side is given by

$$\phi(t) = Ce^{-itA^\times}\Pi x, \qquad 0 \leq t < \infty.$$

Here Π is the projection of $\mathbb{C}^n$ onto $\operatorname{Ker}P^\times$ along $\operatorname{Im}P$.

Proof. Since $x \in \operatorname{Ker} P$, the vector $e^{-itA}x$ is exponentially decaying in norm when $t \to \infty$, and thus the function f belongs to $L_p^m[0,\infty)$. Furthermore, the conditions (i) and (ii) in Theorem 6.2 are fulfilled, and hence for this f equation (6.6) has a unique solution $\phi \in L_p^m[0,\infty)$. Moreover from Theorem 6.2 we know that ϕ is given by

$$\phi(t) = f(t) + iCe^{-itA^\times}\left(\int_0^t \Pi e^{isA^\times} BCe^{-isA}x\,ds\right) \\ -iCe^{-itA^\times}\left(\int_t^\infty (I-\Pi)e^{isA^\times} BCe^{-isA}x\,ds\right).$$

Now use that

$$e^{isA^\times} BCe^{-isA} = ie^{isA^\times}(iA^\times - iA)e^{-isA} = i\frac{d}{ds}e^{isA^\times}e^{-isA}.$$

It follows that

$$\phi(t) = f(t) - Ce^{-itA^\times}\left(\Pi e^{isA^\times}e^{-isA}x\Big|_0^t\right) \\ +Ce^{-itA^\times}\left((I-\Pi)e^{isA^\times}e^{-isA}x\Big|_t^\infty\right).$$

Since $(I-\Pi) = (I-\Pi)P^\times$, the function $(I-\Pi)e^{isA^\times} = (I-\Pi)P^\times e^{isA^\times}$ is exponentially decaying for $s \to \infty$. As we have seen, the same holds true for $e^{-isA}x$. Thus

$$\phi(t) = f(t) - Ce^{-itA^\times}\Pi e^{itA^\times}e^{-itA}x + Ce^{-itA^\times}\Pi x \\ -Ce^{-itA^\times}(I-\Pi)e^{itA^\times}e^{-itA}x$$

$$= f(t) + Ce^{-itA^\times}\Pi x - Ce^{-itA}x = Ce^{-itA^\times}\Pi x,$$

which completes the proof. □

Finally, let us return to the special case that the known function f is given by formula (1.25), and assume that the conditions (i) and (ii) in the Theorem 6.2 hold true. Then the solution ϕ admits the following representation

$$\phi(t) \quad = \quad e^{-iqt}\left(I_m + i\int_0^t Ce^{i(q-A^\times)s}\Pi B\,ds\right) \\ \times\big(I_m - C(I-\Pi)(q-A^\times)^{-1}B\big)x_0;$$

see the expression (1.32).

6.3 Application to block Toeplitz operators

In the previous section the factorization theory was applied to Wiener-Hopf integral equations. In this section we carry out a similar program for their discrete analogues, block Toeplitz equations (cf., Section 1.6). So we consider an equation of the type

$$\sum_{k=0}^{\infty} a_{j-k}\xi_k = \eta_j, \qquad j = 0, 1, 2, \ldots . \tag{6.7}$$

Throughout we assume that the coefficients a_j are given complex $m \times m$ matrices satisfying

$$\sum_{j=-\infty}^{\infty} \|a_j\| < \infty,$$

and $\eta = (\eta_j)_{j=0}^{\infty}$ is a given vector from $\ell_p^m = \ell_p(\mathbb{C}^m)$. The problem is to find $\xi = (\xi_k)_{k=0}^{\infty} \in \ell_p^m$ such that (6.7) is satisfied.

As before, we shall apply our factorization theory. For that reason we assume that the symbol $a(\lambda) = \sum_{j=-\infty}^{\infty} \lambda^j a_j$ is a rational $m \times m$ matrix function whose value at ∞ is I_m. Note that a has no poles on the unit circle. Therefore the conditions on a are equivalent to the following assumptions:

(j) the sequence $(a_j - \delta_{j0} I_m)_{j=0}^{\infty}$ is a linear combination of sequences of the form

$$\left(\alpha^j j^r D\right)_{j=0}^{\infty},$$

where $|\alpha| < 1$, r is a non-negative integer and D is a complex $m \times m$ matrix;

(jj) the sequence $(a_{-j})_{j=1}^{\infty}$ is a linear combination of sequences of the form

$$(\beta^{-j} j^s E)_{j=1}^{\infty}, \qquad (\delta_{jk} F)_{j=1}^{\infty},$$

where $|\beta| > 1$, s and k are nonnegative integers and E and F are complex $m \times m$ matrices.

Since $a(\lambda)$ is rational and $a(\infty) = I$, one can construct (see Section 4.1) a system $\Theta = (A, B, C, ; \mathbb{C}^n, \mathbb{C}^m)$ such that A has no unimodular eigenvalues and

$$a(\lambda) = I_m + C(\lambda I_n - A)^{-1} B \tag{6.8}$$

is a realization for a. The next theorem is the analogue of Theorem 6.2.

Theorem 6.4. *Let* (6.8) *be a realization for the symbol a of the equation* (6.7)*, and assume A has no unimodular eigenvalues. Then* (6.7) *has a unique solution $\xi = (\xi_k)_{k=0}^{\infty}$ in ℓ_p^m for each $\eta = (\eta_j)_{j=0}^{\infty}$ in ℓ_p^m if and only if the following two conditions are satisfied:*

(i) *$A^\times = A - BC$ has no unimodular eigenvalues;*

(ii) *$\mathbb{C}^n = M \dot{+} M^\times$, where M is the spectral subspace of A corresponding to the eigenvalues of A inside the unit circle, and $M^\times$ is the spectral subspace of $A^\times$ corresponding to the eigenvalues of $A^\times$ outside the unit circle.*

Assume conditions (i) *and* (ii) *are satisfied, and let* Π *be the projection of* $\mathbb{C}^n$ *along* M *onto* $M^\times$*. Then the symbol* a *admits a right canonical factorization with respect to the unit circle that has the form*

$$a(\lambda) = h_-(\lambda)h_+(\lambda), \qquad |\lambda| = 1,$$

where the factors and their inverses are given by

$$\begin{aligned} h_+(\lambda) &= I_m + C\Pi(\lambda - A)^{-1}B, \\ h_-(\lambda) &= I_m + C(\lambda - A)^{-1}(I - \Pi)B, \\ h_+^{-1}(\lambda) &= I_m - C(\lambda - A^\times)^{-1}\Pi B, \\ h_-^{-1}(\lambda) &= I_m - C(I - \Pi)(\lambda - A^\times)^{-1}B. \end{aligned}$$

The sequences $(\gamma_j^+)_{j=0}^\infty$ *and* $(\gamma_{-j}^-)_{j=0}^\infty$ *in* (1.32) *have the representation*

$$\begin{aligned} \gamma_0^+ &= I_m + C(A^\times)^{-1}\Pi B, \\ \gamma_j^+ &= C(A^\times)^{-(j+1)}\Pi B, \qquad j = 1, 2, \dots, \\ \gamma_0^- &= I_m, \\ \gamma_j^- &= -C(I_n - \Pi)(A^\times)^{-(j+1)}B, \qquad j = -1, -2, \dots. \end{aligned}$$

Finally, the solution ξ *to* (6.7) *can be written as* $\xi_k = \sum_{s=0}^\infty \gamma_{ks}\eta_s$ *where*

$$\gamma_{ks} = \begin{cases} C(A^\times)^{-(k+1)}\Pi(A^\times)^s B, & s < k, \\ I_m + C(A^\times)^{-(s+1)}\Pi(A^\times)^s B, & s = k, \\ -C(A^\times)^{-(k+1)}(I_n - \Pi)(A^\times)^s B, & s > k. \end{cases}$$

Proof. The proof of Theorem 6.4 is similar to that of Theorem 6.2. Here we only derive the final formula for γ_{ks}.

With respect to the formulas for γ_j^+, we note that $\operatorname{Im}\Pi$ is $A^\times$-invariant and the restriction of $A^\times$ to $\operatorname{Im}\Pi$ is invertible. So, with slight abuse of notation as far as inverses of $A^\times$ is involved,

$$\begin{aligned} h_+(\lambda)^{-1} &= I_m - C(\lambda - A^\times)^{-1}\Pi B \\ &= I_m + C\big(I - \lambda(A^\times)^{-1}\big)^{-1}(A^\times)^{-1}\Pi B \\ &= I_m + \sum_{j=0}^\infty \lambda^j C(A^\times)^{-(j+1)}\Pi B. \end{aligned}$$

Now compare coefficients with $h_+(\lambda)^{-1} = \sum_{j=0}^{\infty} \lambda^j \gamma_j^+$. Similarly, the formulas for γ_j^- are obtained by comparing

$$\begin{aligned} h_-(\lambda)^{-1} &= I_m - C(I-\Pi)(\lambda - A^\times)^{-1}B \\ &= I_m - C(I-\Pi)\sum_{j=1}^{\infty} \frac{1}{\lambda^j}(A^\times)^{j-1}B \\ &= I_m - \sum_{j=-\infty}^{-1} \lambda^j C(I-\Pi)(A^\times)^{-(j+1)}B \end{aligned}$$

with $h_-(\lambda)^{-1} = \sum_{j=-\infty}^{0} \lambda^j \gamma_j^-$.

To obtain the formulas for γ_{ks}, we use formula (1.32). For $s < k$ we have to find

$$\gamma_{ks} = \gamma_{k-s}^+ \gamma_0^- + \sum_{r=0}^{s-1} \gamma_{k-r}^+ \gamma_{r-s}^-,$$

while for $s > k$ we need to calculate

$$\gamma_{ks} = \gamma_0^+ \gamma_{k-s}^- + \sum_{r=0}^{k-1} \gamma_{k-r}^+ \gamma_{r-s}^-.$$

Again by slight abuse of notation

$$\begin{aligned} \gamma_{k-r}^+ \gamma_{r-s}^- &= -C(A^\times)^{-(k-r+1)}\Pi B C(I-\Pi)(A^\times)^{-(r-s+1)}B \\ &= -C(A^\times)^{-(k-r+1)}(A^\times\Pi - \Pi A^\times)(A^\times)^{-(r-s+1)}B \\ &= -C(A^\times)^{-(k-r)}\Pi(A^\times)^{-(r-s+1)}B + \\ &\qquad + C(A^\times)^{-(k-r+1)}\Pi(A^\times)^{-(r-s)}B. \end{aligned}$$

Observe that if we replace r by $r+1$ in the last one of the latter two terms we get the first one. So the summation in the formula for γ_{ks} is telescoping and collapses into just a few terms. We proceed as follows.

For $s < k$ we get

$$\gamma_{ks} = \gamma_{k-s}^+ \gamma_0^- - C(A^\times)^{-(k-s+1)}\Pi B + C(A^\times)^{-(k+1)}\Pi(A^\times)^s B.$$

Since $\gamma_0^- = I$ and $\gamma_{k-s}^+ = C(A^\times)^{-(k-s+1)}\Pi B$, this results into

$$\gamma_{ks} = C(A^\times)^{-(k+1)}\Pi(A^\times)^s B.$$

For $s > k$ the computation is a little more involved as $\gamma_0^+ = I_n + C(A^\times)^{-1}\Pi B$. Using that $\Pi BC\big(I-\Pi\big) = A^\times\Pi - \Pi A^\times$, it runs as follows:

$$\begin{aligned}
\gamma_{ks} &= -\big(I + C(A^\times)^{-1}\Pi B\big)C\big(I-\Pi\big)(A^\times)^{-(k-s+1)}B + \\
&\quad + C(A^\times)^{-(k+1)}\Pi(A^\times)^s B - C(A^\times)^{-1}\Pi(A^\times)^{-(k-s)}B \\
&= -C\big(I-\Pi\big)(A^\times)^{-(k-s+1)}B + \\
&\quad + C(A^\times)^{-1}(\Pi A^\times - A^\times\Pi)(A^\times)^{-(k-s+1)}B + \\
&\quad + C(A^\times)^{-(k+1)}\Pi(A^\times)^s B - C(A^\times)^{-1}\Pi(A^\times)^{-(k-s)}B \\
&= C(A^\times)^{-(k+1)}\Pi(A^\times)^s B - C(A^\times)^{-(k-s+1)}B \\
&= -C(A^\times)^{-(k+1)}(I-\Pi)(A^\times)^s B.
\end{aligned}$$

It remains to consider the case $k = s$. Then we have

$$\gamma_{ss} = \gamma_0^+\gamma_0^- + \sum_{r=0}^{k-1}\gamma_{s-r}^+\gamma_{r-s}^-.$$

Following the line of argument as in the case $s < k$ this yields

$$\begin{aligned}
\gamma_{ss} &= I_m + C(A^\times)^{-1}\Pi B - C(A^\times)^{-1}\Pi B + C(A^\times)^{-(k+1)}\Pi(A^\times)^k B \\
&= I_m + C(A^\times)^{-(k+1)}\Pi(A^\times)^k B,
\end{aligned}$$

which completes the proof. □

The main step in the factorization method for solving the equation (6.7) is to construct a right canonical factorization of the symbol a with respect to the unit circle. In Theorem 6.4 we obtained explicit formulas for the case when a is rational and has the value I_n at ∞. The latter condition is not essential. Indeed, by a suitable Möbius transformation one can transform the symbol $\alpha(\lambda)$ into a function which is invertible at infinity (see Section 3.6). Next one makes the Wiener-Hopf factorization of the transformed symbol relative to the image of the unit circle under the Möbius transformation. Here one can use the same formulas as in Theorem 6.4. Finally, using the inverse Möbius transformation, one can obtain explicit formulas for the factorization with respect to the unit circle, and hence also for the solution of equation (6.7).

Notes

The material in this chapter is taken from [14] and [15]. The notion of a canonical factorization can be viewed as a special case of minimal factorization which we

shall treat later in Chapter 9. In Section 9.2 we shall resume the discussion of canonical factorizations. Theorem 6.2 is a slightly changed version of Theorem I.3.4 in [15]. With natural appropriate modifications Theorem 6.2 is also valid in the infinite-dimensional case. Moreover in the infinite-dimensional case we can sometimes allow for spectrum on the real line or (when the state space operators are unbounded) at infinity. Finally, we note that the results of Section 6.1 extend to operator-valued functions that are analytic on an open neighborhood of the given contour. In fact, for such functions non-canonical Wiener-Hopf factorization relative to the contour can also be described explicitly in terms of realizations. For this and related results we refer to [16].

Part II

Minimal Realization and Minimal Factorization

This part is concerned with minimality of systems and minimality of factorization of rational matrix functions. An analysis of rational matrix functions in terms of spectral data (eigenvalues, eigenvectors and Jordan chains) is also included. This part consists of three chapters (7–9).

Chapter 7 is devoted to minimality of systems. For finite-dimensional systems minimality is equivalent to controllability and observability. For various other classes of systems the notion of minimality is analyzed. In particular, this done for the classes of systems introduced in Chapter 3. Special attention is paid to Hilbert space systems, that is, systems for which the input space, the output space and the state space are (possibly infinite-dimensional) Hilbert spaces. In Chapter 8 finite-dimensional systems are studied in terms of the zero and pole data of their transfer functions. This includes the construction of minimal realizations in terms of the pole data, and a spectral analysis of rational matrix functions in terms of eigenvalues, eigenvectors and Jordan chains. Here the notions of McMillan degree and local degree are introduced. The final chapter (Chapter 9) contains the theory of minimal factorization, with special attention for systems that are not biproper. Also in this chapter, using the notion of local minimality, the concept of a pseudo-canonical factorization relative to a curve is introduced and analyzed for rational matrix functions with singularities on the given curve.

Chapter 7

Minimal Systems

In this chapter the notion of a minimal system is considered. If two systems are similar, then they have the same transfer function. The converse statement is not true. In fact, systems with rather different state spaces may have the same transfer function. For minimal systems this phenomenon does not occur. In Section 7.1 minimal systems are defined as systems that are controllable and observable. The latter two notions are explained in more detail for finite-dimensional systems in Section 7.2. In the finite-dimensional case the connection between a minimal system Θ and its transfer function W_Θ is very close. For example in that case Θ is uniquely determined up to similarity by W_Θ. This result, which is known as the *state space similarity theorem*, will be proved in Section 7.3. Several examples, presented in Section 7.4, show that a generalization of the finite-dimensional theory to an infinite-dimensional setting is not possible in a straightforward way. An appropriate generalization requires a further refinement of the state space theory. In Section 7.5 the notion of minimality is considered for Brodskii systems, Kreĭn systems, unitary systems, monic systems, and polynomial systems.

7.1 Minimality of systems

Two similar systems have the same transfer function. On the other hand, the transfer function will in general not determine the system up to similarity. For example, consider the unital systems $\Theta_1 = (A_1, B_1, 0; X_1, Y)$ and $\Theta_2 = (A_2, 0, C_2; X_2, Y)$. The transfer functions of Θ_1 and Θ_2 are both identically equal to the identity operator on Y, but if either B_1 or C_2 is nonzero, then Θ_1 and Θ_2 will not be similar.

More generally, let $\Theta_0 = (A_0, B_0, C_0, D; X_0, U, Y)$ be a system, and let X_1 and X_2 be arbitrary complex Banach spaces. Put $X = X_1 \dot{+} X_0 \dot{+} X_2$, and let

$A : X \to X$, $B : U \to X$ and $C : X \to Y$ be operators of the form

$$A = \begin{bmatrix} * & * & * \\ 0 & A_0 & * \\ 0 & 0 & * \end{bmatrix}, \qquad B = \begin{bmatrix} * \\ B_0 \\ 0 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & C_0 & * \end{bmatrix}, \tag{7.1}$$

where the stars $*$ denote unspecified operators acting between appropriate spaces. Now consider the system $\Theta = (A, B, C, D; X, U, Y)$. One easily verifies that the transfer function of Θ coincides on a neighborhood of ∞ with the transfer function of Θ_0. However, after a suitable choice of the spaces X_1 and X_2 or of the unspecified operators, the systems Θ and Θ_0 will not be similar. Under certain minimality conditions, to be discussed below, positive results on similarity do exist (cf., Section 7.3)

When the system $\Theta = (A, B, C, D; X, U, Y)$ is related to the system $\Theta_0 = (A_0, B_0, C_0, D; X_0, U, Y)$ as in (7.1), then Θ is called a *dilation* of Θ_0, and, conversely, Θ_0 is called a *restriction* of Θ. If the space X_0 is strictly contained in X or, equivalently, X_1 or X_2 (or both these spaces) contain nonzero vectors, then Θ is called a *proper dilation* of Θ_0, and in this case we also say that Θ_0 is a *proper restriction* of Θ.

Let $\Theta = (A, B, C, D; X, U, Y)$ be a system. In the sequel we let $\operatorname{Ker}(C|A)$ and $\operatorname{Im}(A|B)$ be the linear submanifolds of X defined by

$$\begin{aligned} \operatorname{Ker}(C|A) &= \operatorname{Ker} C \cap \operatorname{Ker} CA \cap \operatorname{Ker} CA^2 \cap \cdots, \\ \operatorname{Im}(A|B) &= \operatorname{Im} B + \operatorname{Im} AB + \operatorname{Im} A^2B + \cdots, \end{aligned}$$

where the right-hand side of the latter expression denotes the linear hull of the linear manifolds $\operatorname{Im} A^jB$, $j = 0, 1, 2, \ldots$. If Θ is a proper dilation of a system Θ_0, then $\operatorname{Ker}(C|A) \neq \{0\}$ or $\operatorname{Im}(A|B)$ is not dense in X (and possibly even both these properties hold true). We call Θ *approximately observable* if $\operatorname{Ker}(C|A) = \{0\}$ and we call Θ *approximately controllable* if $\operatorname{Im}(A|B)$ is dense in X. In this case we also say that the pairs (C, A) and (A, B) are approximately observable and approximately controllable, respectively. In the sequel we shall omit the adverb approximately and simply speak about *observable* and *controllable*. We say that Θ is *minimal* if Θ is both observable and controllable. The terms observable and controllable come from system theory, and for finite-dimensional systems they will be explained in more detail in the next section.

Proposition 7.1. *Let $\Theta = (A, B, C, D; X, U, Y)$ be a biproper system, and assume that the spectra of A and $A^\times = A - BD^{-1}C$ are disjoint. Then the system Θ is minimal.*

Proof. Let $K = \operatorname{Ker}(C|A)$. Obviously, K is a closed subspace of X which is invariant under A. Thus K is a Banach space in its own right, and $A|_K$ is a bounded linear operator on K.

Assume that $K = \operatorname{Ker}(C|A) \neq \{0\}$. Then the spectrum of $A|_K$ is nonempty. Choose λ_0 in the boundary of the spectrum of $A|_K$. Then (see Theorem V.4.1 in [110]) the point λ_0 is in the approximate point spectrum of $A|_K$, that is, there exists a sequence of vectors, $x_1, x_2, \ldots$, in K such that $\|x_n\| = 1$ for each n and $(\lambda_0 - A|_K)x_n \to 0$ if $n \to \infty$. Obviously, $A|_K x_n = Ax_n$ for $n = 1, 2, \ldots$. Since $K \subset \operatorname{Ker} C$, the operators A and $A^\times$ coincide on K. Thus for our sequence $x_1, x_2, \ldots$ we have

$$\|x_n\| = 1, \quad n = 1, 2, \ldots, \qquad \lim_{n\to\infty} (\lambda_0 - A)x_n = 0,$$

$$\|x_n\| = 1, \quad n = 1, 2, \ldots, \qquad \lim_{n\to\infty} (\lambda_0 - A^\times)x_n = 0.$$

From the first part of the above formula we see that $\lambda \in \sigma(A)$, and from the second part that $\lambda \in \sigma(A^\times)$. Thus λ_0 is a common point of the spectra of A and $A^\times$, which contradicts our hypotheses. Hence $\operatorname{Ker}(C|A) = \{0\}$.

Next, we consider the system $\Theta' = (A', C', B', D'; X', Y', U')$, where the prime means that one has to take the Banach space conjugate (see [48], Sections 11.4 and 13.5). Since a Banach space operator is invertible if and only if its Banach dual is invertible, the operators A' and $(A^\times)'$ have no common spectra. But $(A^\times)' = A' - C'(D')^{-1}B'$, and therefore the result of the previous paragraph shows that $\operatorname{Ker}(B'|A') = \{0\}$. Now assume that $\operatorname{Im}(A|B)$ is not dense in X. Then, by the Hahn-Banach theorem (see [48], Section 11.5), there exists a nonzero $f \in X'$ such that $f(A^n Bu) = 0$ for each n and each $u \in U$. It follows that $B'(A')^n f = 0$, and thus $f \in \operatorname{Ker}(B'|A')$. Therefore $f = 0$ which is a contradiction. Hence $\operatorname{Im}(A|B)$ is dense in X.

From the results of the two previous paragraphs we conclude that Θ is minimal. □

The converse of Proposition 7.1 is not true. To see this, take $X = \mathbb{C}$, $U = Y = \mathbb{C}^2$, and

$$A = 1, \qquad B = \begin{bmatrix} 0 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then $\Theta = (A, B, C, D; \mathbb{C}, \mathbb{C}^2, \mathbb{C}^2)$ is a biproper minimal system. However, we have $BD^{-1}C = 0$, and hence $A = A^\times$. In particular, the spectra of A and $A^\times$ are not disjoint.

The fact that minimality is not the same as disjointness of the spectra of the main and associate main operator makes the notion of minimality an interesting one.

By a *minimal realization* of an operator function W we mean a minimal system that is a realization for W. Also, if W is given by

$$W(\lambda) = D + C(\lambda - A)^{-1}B, \tag{7.2}$$

we say that (7.2) is a minimal realization for W if the system determined by the operators A, B, C and D is minimal. In the same way one can define the notions of an *observable* and a *controllable realization*.

Below we present some elementary facts concerning minimal systems. With appropriate modifications the results are also valid for systems that are observable or controllable only.

Suppose $\Theta_1 = (A_1, B_1, C_1, D_1; X_1, Y)$ and $\Theta_2 = (A_2, B_2, C_2, D_2; X_2, Y)$ are similar systems. Then Θ_1 is minimal if and only if Θ_2 is minimal. If S and S' are system similarities between Θ_1 and Θ_2, then

$$\operatorname{Im}(S - S') \subset \operatorname{Ker}(C_2|A_2).$$

So $S = S'$, provided that Θ_1 and Θ_2 are minimal. This proves the following result.

Proposition 7.2. *A system similarity between two minimal systems is uniquely determined by the given two systems.*

If $\Theta = (A, B, C, D; X, Y)$ is a system with an invertible external operator D, then Θ is minimal if and only if $\Theta^\times$ is minimal. This is immediate from the identities

$$\operatorname{Ker}(C|A) = \operatorname{Ker}(-D^{-1}C|A^\times), \qquad \operatorname{Im}(A|B) = \operatorname{Im}(A^\times|BD^{-1}).$$

The product of two minimal systems need not be minimal. To see this, multiply the minimal systems $(0, 1, 1; \mathbb{C}, \mathbb{C})$ and $(-1, 1, -1; \mathbb{C}, \mathbb{C})$. On the other hand, we have the following proposition.

Proposition 7.3. *If the product of two systems Θ_1 and Θ_2 is minimal, then so are the factors Θ_1 and Θ_2.*

Proof. For $j = 1, 2$, write $\Theta_j = (A_j, B_j, C_j, D_1; X_j, U, Y)$. Then the product $\Theta_1\Theta_2$ is given by $\Theta_1\Theta_2 = (A, B, C, D; X_1 \dot{+} X_2, Y)$ with

$$A = \begin{bmatrix} A_1 & B_1C_2 \\ 0 & A_2 \end{bmatrix}, \quad B = \begin{bmatrix} B_1D_2 \\ B_2 \end{bmatrix}, \quad C = \begin{bmatrix} C_1 & D_1C_2 \end{bmatrix}, \quad D = D_1D_2.$$

Assume Θ is minimal. We shall prove that Θ_1 and Θ_2 are both observable and both controllable.

Take x in $\operatorname{Ker}(C_1|A_1)$. Then the column vector $[\,x\ 0\,]^\top$ belongs to the space $\operatorname{Ker}(C|A)$. But $\operatorname{Ker}(C|A) = \{0\}$, and so $x = 0$. This proves that Θ_1 is observable. Next, we show that also Θ_2 is observable. Take x in $\operatorname{Ker}(C_2|A_2)$. This implies that $C_2x = 0$ and $A_2x \in \operatorname{Ker}(C_2|A_2)$. Thus one shows by induction that

$$A^j \begin{bmatrix} 0 \\ x \end{bmatrix} = \begin{bmatrix} 0 \\ A_2^j x \end{bmatrix}.$$

From this it follows that $[\,0\ x\,]^\top$ is in $\operatorname{Ker}(C|A) = \{0\}$. Hence $x = 0$, which proves that Θ_2 is observable.

Next take z in X_1. Then $[\,z\ 0\,]^\top$ is in the closure of $\operatorname{Im}(A|B)$. So the vector $[\,z\ 0\,]^\top$ can be approximated arbitrarily close by sums of the form $\sum_{j=0}^{n-1} A^j B y_j$ with $y_0, \ldots, y_{n-1}$ in Y. The first coordinate of such a sum is easily seen to belong to $\operatorname{Im}(A_1|B_1)$. Thus z is in the closure of $\operatorname{Im}(A_1|B_1)$, and we conclude that Θ_1 is controllable. To show that Θ_2 is controllable, take z in X_2. Then $[\,z\ 0\,]^\top$ is in the closure of $\operatorname{Im}(A|B)$, and it follows from this that z is in the closure of $\operatorname{Im}(A_2|B_2)$. Thus Θ_2 is controllable as well. □

If Π is a supporting projection for a unital system Θ (i.e., the external operator is the identity), then

$$\Theta = \operatorname{pr}_{I-\Pi}(\Theta)\operatorname{pr}_{\Pi}(\Theta).$$

Thus, if Θ is minimal, then so are $\operatorname{pr}_{I-\Pi}(\Theta)$ and $\operatorname{pr}_{\Pi}(\Theta)$. An arbitrary projection of a minimal system need not be minimal, not even when the image of the projection is an invariant subspace for the main operator of the system. Indeed, if $\Theta = (A, B, C; \mathbb{C}^3, \mathbb{C}^2)$ and Π are given by

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},\ B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix},\ C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \Pi = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

then Θ is minimal, but

$$\operatorname{pr}_{\Pi}(\Theta) = \left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \mathbb{C}^2, \mathbb{C}^2 \right)$$

is not. Note that $\operatorname{Im}\Pi$ is an invariant subspace for A.

7.2 Controllability and observability for finite-dimensional systems

In the previous section we defined controllability and observability in a rather formal way. In this section we present alternative definitions of these notions for finite-dimensional systems. The new definitions, which reflect better the system theoretical contents, will be shown to be equivalent to the ones appearing in the previous section. Throughout this section we restrict ourselves to finite-dimensional systems.

A system is called controllable (in systems theoretical sense) if (roughly speaking) starting from an arbitrary initial state x_0 any other state x_1 can be reached by applying a suitable input. To make this more precise, let

$$\Theta = (A, B, C, D; X, \mathbb{C}^p, \mathbb{C}^q)$$

be a finite-dimensional system. For a given x_0 in the state space X and a given input u we let $x(t; x_0, u)$ denote the unique solution of

$$\begin{cases} x'(t) = Ax(t) + Bu(t), & t \geq 0, \\ x(0) = x_0. \end{cases} \tag{7.3}$$

In other words,

$$x(t; x_0, u) = e^{tA}x_0 + \int_0^t e^{(t-s)A}Bu(s)\,ds, \qquad t \geq 0. \tag{7.4}$$

The system Θ is said to be *controllable* (in systems theoretical sense) if for any x_0, x_1 in X there exist $t_1 > 0$ and u in $PCE(\mathbb{C}^p)$ such that $x_1 = x(t_1; x_0, u)$.

Note that controllability does not involve the output operator C. The next proposition shows that for finite-dimensional systems the above definition of controllability coincides with the one given in the previous section.

Proposition 7.4. *Let $\Theta = (A, B, C, D; X, \mathbb{C}^p, \mathbb{C}^q)$ be a finite-dimensional system. Then Θ is controllable (in systems theoretical sense) if and only if* $\operatorname{Im}(A|B) = X$.

Proof. Let $\tau > 0$ be fixed, and let $\mathcal{S}(\tau)$ be the set of states in X that can be reached at time $t = \tau$ starting from the initial state $x_0 = 0$. Thus

$$\mathcal{S}(\tau) = \{x(\tau; 0, u) \mid u \in PCE(\mathbb{C}^p)\}.$$

Obviously, $\mathcal{S}(\tau)$ is a linear subspace of the state space X. Endow X with an inner product, and consider the input

$$u_0(t) = B^* e^{(\tau-t)A^*}x, \qquad t \geq 0,$$

where x is an arbitrary vector in X. Here A^* and B^* denote the adjoints of A and B, respectively. Note that $u \in PCE(\mathbb{C}^p)$. One computes that

$$x(\tau; 0, u_0) = \left(\int_0^\tau e^{tA}BB^*e^{tA^*}\,dt \right) x.$$

Thus $\mathcal{S}(\tau) \supset \operatorname{Im}\big(\int_0^\tau e^{tA}BB^*e^{tA^*}\,dt \big)$. We shall prove that

$$\mathcal{S}(\tau) = \operatorname{Im}\left(\int_0^\tau e^{tA}BB^*e^{tA^*}\,dt \right) = \operatorname{Im}(A|B). \tag{7.5}$$

Let $z \in X$, and assume $z \perp \operatorname{Im}\big(\int_0^\tau e^{tA}BB^*e^{tA^*}\,dt \big)$. To prove the first equality in (7.5), it suffices to show that $z \perp \mathcal{S}(\tau)$. Our hypothesis on z implies that $\int_0^\tau \|B^*e^{tA^*}z\|^2\,dt = 0$, and hence $B^*e^{tA^*}z = 0$ for each $t \in [0, \tau]$. In particular, $z \perp \operatorname{Im} e^{tA}B$ for each $0 \leq t \leq \tau$ Now, take an arbitrary $u \in PCE(\mathbb{C}^p)$. Then

$$x(\tau; 0, u) = \int_0^\tau e^{(\tau-s)A}Bu(s)\,ds = \int_0^\tau e^{tA}Bu(\tau - t)\,dt,$$

and thus $z \perp x(\tau; 0, u)$. Since u is arbitrary, we conclude that $z \perp \mathcal{S}(\tau)$. The first equality in (7.5) is proved.

By definition $A^j B y \in \operatorname{Im}(A|B)$ for each $j \geq 0$ and $y \in \mathbb{C}^p$. Since $e^{tA}B = \sum_{j=0}^{\infty} \frac{1}{j!} A^j B$ and $\operatorname{Im}(A|B)$ is closed because of finite dimensionality, we conclude that $\operatorname{Im}\big(\int_0^\tau e^{tA} B B^* e^{tA^*}\, dt\big) \subset \operatorname{Im}(A|B)$. Again take $z \perp \operatorname{Im}\big(\int_0^\tau e^{tA} B B^* e^{tA^*}\, dt\big)$. To prove the second equality in (7.5) it remains to show that $z \perp \operatorname{Im}(A|B)$. We have already seen that our hypothesis on z implies that

$$B^* e^{tA^*} z = \sum_{j=0}^{\infty} \frac{1}{j!} t^j B^* (A^*)^j z = 0, \qquad 0 \leq t \leq \tau.$$

But then $B^*(A^*)^j z = 0$ and hence $z \perp \operatorname{Im} A^j B$ for $j = 0, 1, 2, \ldots$, which proves that $z \perp \operatorname{Im}(A|B)$.

We have now proved (7.5). Note that (7.5) implies that the space $\mathcal{S}(\tau)$ does not depend on the choice of τ.

Assume that $\Theta = (A, B, C, D; X, \mathbb{C}^p, \mathbb{C}^q)$ is controllable in the systems theoretical sense. Take $x \in X$. According to the definition given above, there exists $t_1 > 0$ and $u \in PCE(\mathbb{C}^p)$ such that $x_1 = x(t_1; 0, u)$. In other words, $x \in \mathcal{S}(t_1)$. But then we can use (7.5) to show that $x \in \operatorname{Im}(A|B)$. Since x is arbitrary, we conclude that $\operatorname{Im}(A|B) = X$.

Next, suppose that $\operatorname{Im}(A|B) = X$. Take x_0, x_1 in X. Choose any $\tau > 0$. According to (7.5) there exists $u \in PCE(\mathbb{C}^p)$ such that $x_1 - e^{\tau A} x_0 = x(\tau; 0, u)$. But then $x_1 = x(\tau; x_0, u)$. Since x_0 and x_1 are arbitrary, we have proved that Θ is controllable. □

We now turn to observability. Roughly speaking a system is observable (in systems theoretical sense) if the output determines uniquely the state of the system at time $t = 0$. To make this more precise, let $\Theta = (A, B, C, D; X, \mathbb{C}^p, \mathbb{C}^q)$ be a finite-dimensional system. Consider the system equations

$$\begin{cases} x'(t) = Ax(t) + Bu(t), \\ y(t) = Cx(t) + Du(t), \qquad t \geq 0, \\ x(0) = x_0. \end{cases} \tag{7.6}$$

For a given input u and initial state x_0 we denote the output of (7.6) by $y(t; x_0, u)$. Thus

$$y(t; x_0, u) = Cx(t; x_0, u) + Du(t),$$

where $x(t; x_0, u)$ is given by (7.4). Note that

$$y(t; x_0, u) = y(t; x_0, 0) + \int_0^t Ce^{(t-s)A} B u(s)\, ds.$$

Hence $y(t; x_0, u) = y(t; \widetilde{x}_0, u)$ if and only if $y(t; x_0, 0) = y(t; \widetilde{x}_0, 0)$. Thus to determine the initial state from the output, the role of the input is irrelevant. This

leads to the following definition. The system Θ is *observable* (in systems theoretical sense) if there exists $\tau > 0$ such that $y(t; x_0, 0) = y(t; \widetilde{x}_0, 0)$ on $0 \le t \le \tau$ implies that $x_0 = \widetilde{x}_0$.

Note that observability does not involve the input operator B and the external operator D. The next proposition shows that for finite-dimensional systems the above definition of observability coincides with the one given in the previous section.

Proposition 7.5. *Let $\Theta = (A, B, C, D; X, \mathbb{C}^p, \mathbb{C}^q)$ be a finite-dimensional system. Then Θ is observable (in systems theoretical sense) if and only if* $\operatorname{Ker}(C|A) = \{0\}$.

Proof. Assume Θ is observable. Take $x_0 \in \operatorname{Ker}(C|A)$. Thus $CA^j x_0 = 0$ for each $j \ge 0$. It follows that

$$y(t; x_0, 0) = Ce^{tA}x_0 = \sum_{j=0}^{\infty} \frac{1}{j!} t^j CA^j x_0 = 0, \qquad t \ge 0.$$

In particular, $y(t; x_0, u) = y(t; 0, 0)$ for all $t \ge 0$. Since Θ is observable, this implies $x_0 = 0$. Hence $\operatorname{Ker}(C|A) = \{0\}$.

To prove the reverse implication, assume that $\operatorname{Ker}(C|A) = \{0\}$. Take an arbitrary $\tau > 0$, and let $y(t; x_0, 0) = y(t; \widetilde{x}_0, 0)$ for $0 \le t \le \tau$. Then

$$Ce^{tA}(x_0 - \widetilde{x}_0) = 0, \qquad 0 \le t \le \tau$$

and hence $CA^j(x_0 - \widetilde{x}_0) = 0$ for $j \ge 0$. In other words, $x_0 - \widetilde{x}_0 \in \operatorname{Ker}(C|A)$. Hence $x_o = \widetilde{x}_0$, and Θ is observable. □

For operators acting between finite-dimensional spaces we shall sometimes use the terms "null kernel pair" and "full range pair" in place of observable pair and controllable pair. Thus a pair (C, A) of finite-dimensional operators is called a *null kernel pair* if $\operatorname{Ker}(C|A) = \{0\}$, and a pair (A, B) of finite-dimensional operators is called a *full range pair* if $\operatorname{Im}(A|B) = X$, where X is the space on which A acts.

7.3 Minimality for finite-dimensional systems

Let $\Theta = (A, B, C, D; X, U, Y)$ be a finite-dimensional system. Thus Θ is a system and the spaces X, U and Y are finite-dimensional. Let n be an integer larger than or equal to the degree of the minimal polynomial of A (for instance $n \ge \dim X$). Then, by the Cayley-Hamilton theorem,

$$\begin{aligned} \operatorname{Ker}(C|A) &= \operatorname{Ker} C \cap \operatorname{Ker} CA \cap \operatorname{Ker} CA^2 \cap \cdots \cap \operatorname{Ker} CA^{n-1}, \\ \operatorname{Im}(A|B) &= \operatorname{Im} B + \operatorname{Im} AB + \operatorname{Im} A^2B + \cdots + \operatorname{Im} A^{n-1}B. \end{aligned}$$

From this it is obvious that Θ is minimal if and only if the right hand sides of these expressions are $\{0\}$ and X, respectively, i.e.,

$$\begin{aligned} \operatorname{Ker} C \cap \operatorname{Ker} CA \cap \operatorname{Ker} CA^2 \cap \cdots \cap \operatorname{Ker} CA^{n-1} &= \{0\}, \\ \operatorname{Im} B + \operatorname{Im} AB + \operatorname{Im} A^2B + \cdots + \operatorname{Im} A^{n-1}B &= X. \end{aligned}$$

An equivalent requirement is that the operators defined by $\operatorname{col}(CA^j)_{j=0}^{n=1}$ and $\operatorname{row}(A^j)_{j=0}^{n-1}$ are left and right invertible, respectively.

Theorem 7.6. *Any finite-dimensional system is a dilation of a finite-dimensional minimal system. In particular, a finite-dimensional system is minimal if and only if it does not have a proper restriction.*

Proof. Let $\Theta = (A, B, C, D; X, U, Y)$ be a finite-dimensional system, and let n be the degree of the minimal polynomial of A. Put $\Omega = \operatorname{col}(CA^j)_{j=0}^{n-1}$ and $\Delta = \operatorname{row}(A^jB)_{j=0}^{n-1}$. Then $\Omega : X \to Y^n$, $\Delta : U^n \to X$ and

$$A[\operatorname{Ker}\Omega] \subset \operatorname{Ker}\Omega, \qquad A[\operatorname{Im}\Delta] \subset \operatorname{Im}\Delta.$$

Put $X_1 = \operatorname{Ker}\Omega$, and let X_0 be a direct complement of $X_1 \cap \operatorname{Im}\Delta$ in $\operatorname{Im}\Delta$. Further, choose X_2 such that

$$X = X_1 \dot{+} X_0 \dot{+} X_2.$$

With respect to this decomposition the operators A, B and C can be written in the form

$$A = \begin{bmatrix} A_1 & * & * \\ 0 & A_0 & * \\ 0 & 0 & A_2 \end{bmatrix}, \qquad B = \begin{bmatrix} * \\ B_0 \\ 0 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & C_0 & * \end{bmatrix}. \tag{7.7}$$

Put $\Theta_0 = (A_0, B_0, C_0, D; X_0, Y)$. Then the transfer functions of Θ_0 and Θ coincide (on a neighborhood of ∞). One verifies without difficulty that

$$\begin{aligned} \operatorname{Ker} C_0 \cap \operatorname{Ker} C_0A_0 \cap \operatorname{Ker} C_0A_0^2 \cap \cdots \cap \operatorname{Ker} C_0A_0^{n-1} &= \{0\}, \\ \operatorname{Im} B_0 + \operatorname{Im} A_0B_0 + \operatorname{Im} A_0^2B_0 + \cdots + \operatorname{Im} A_0^{n-1}B_0 &= X_0. \end{aligned}$$

Thus Θ_0 is a minimal system (obviously finite-dimensional), and Θ is a dilation of Θ_0.

Next we consider the final statement. Let Θ be as in the previous paragraph, and assume that Θ does not have a proper restriction. Then the system Θ_0 constructed in previous paragraph must be equal to Θ. But Θ_0 is observable and controllable. It follows that the same holds true for Θ, and thus Θ is minimal.

Conversely, assume $\Theta = (A, B, C, D; X, U, Y)$ is a finite-dimensional minimal system, and Θ_0 be a restriction of Θ. Then (7.1) shows that $X_1 \subset \operatorname{Ker}(C|A)$ and $\operatorname{Im}(A|B) \subset X_1 \dot{+} X_0$. But, because of minimality, $\operatorname{Ker}(C|A) = \{0\}$ and $\operatorname{Im}(A|B) = X$. It follows that $X_1 = \{0\}$ and $X_2 = \{0\}$. Hence $\Theta = \Theta_0$, and thus Θ does not have a proper restriction. □

Theorem 7.6 can be used to give a simple proof of Proposition 7.1 for finite-dimensional systems. Indeed, let $\Theta = (A, B, C, D; X, U, Y)$ be a biproper finite-dimensional system, and assume that the finite-dimensional operators A and $A^\times = A - BD^{-1}C$ do not have a common eigenvalue. We have to show that Θ is minimal. By Theorem 11.5 the system Θ is a dilation of a minimal system. Thus the state space X admits a direct sum decomposition $X = X_1 \dot{+} X_0 \dot{+} X_2$ such that relative to this decomposition the operators A, B and C can be written in the form (7.7) with the system $\Theta_0 = (A_0, B_0, C_0, D; X_0, Y)$ being minimal. Note that (7.7) implies that

$$A^\times = A - BD^{-1}C = \begin{bmatrix} A_1 & * & * \\ 0 & A_0^\times & * \\ 0 & 0 & A_2 \end{bmatrix},$$

where $A_0^\times = A_0 - B_0 D^{-1} C_0$. From the block matrix representations of A and $A^\times$ we see that each eigenvalue of A_1 and each eigenvalue of A_2 is a common eigenvalue of A and $A^\times$. But, according to our assumptions, A and $A^\times$ do not have a common eigenvalue. Thus the spaces X_1 and X_2 consist of the zero vector only, that is, $X = X_0$. Hence $\Theta = \Theta_0$, and thus Θ is minimal.

The construction of the minimal system Θ_0 presented in the proof of Theorem 7.6 can also be carried out by taking quotients instead of complements. This approach also works in the infinite-dimensional case. The next result is known as the *state space similarity theorem*.

Theorem 7.7. *For $k = 1, 2$, let $\Theta_k = (A_k, B_k, C_k, D_k; X_k, U, Y)$ be a finite-dimensional minimal system. Assume that the transfer functions of Θ_1 and Θ_2 coincide (on some open set and hence on a neighborhood of ∞). Then Θ_1 and Θ_2 are similar. Moreover, the (unique) system similarity S between Θ_1 and Θ_2 is given by*

$$\begin{aligned} S &= \Big(\operatorname{col}\big(C_2 A_2^j\big)_{j=0}^{n-1}\Big)^+ \Big(\operatorname{col}\big(C_1 A_1^j\big)_{j=0}^{n-1}\Big) \\ &= \Big(\operatorname{row}\big(A_2^j B_2\big)_{j=0}^{n-1}\Big)\Big(\operatorname{row}\big(A_1^j B_1\big)_{j=0}^{n-1}\Big)^\dagger, \end{aligned}$$

where n is a positive integer larger than or equal to the degree of the minimal polynomial of A_1, the superscript $+$ indicates a left inverse and the superscript $\dagger$ indicates a right inverse.

Proof. For $k = 1, 2$, put

$$\Omega_k = \operatorname{col}\big(C_k A_k^j\big)_{j=0}^{n-1}, \qquad \Delta_k = \operatorname{row}\big(A_k^j B_k\big)_{j=0}^{n-1},$$

where n is a positive integer larger than or equal to the maximum of the degrees of the minimal polynomials of A_1 and A_2. Since Θ_k is minimal, the operators Ω_k and Δ_k are left and right invertible, respectively. Let Ω_k^+ be a left inverse of Ω_k and let $\Delta_k^\dagger$ be a right inverse of Δ_k.

Comparing the Laurent expansions of the transfer functions of Θ_1 and Θ_2 at ∞, we obtain

$$D_1 = D_2, \qquad C_1 A_1^j B_1 = C_2 A_2^j B_2, \qquad j = 0, 1, 2, \ldots .$$

It follows that $\Omega_1 \Delta_1 = \Omega_2 \Delta_2$. But then $\Omega_2^+ \Omega_1 = \Delta_2 \Delta_1^\dagger$. We denote the operator appearing in this equality by S. Observe that $S : X_1 \to X_2$. A direct computation shows that S is invertible with inverse $\Omega_1^+ \Omega_2 = \Delta_1 \Delta_2^\dagger$ and that $\Omega_2 S = \Omega_1$, $S\Delta_1 = \Delta_2$. The last two identities yield

$$A_2 S = S A_1, \qquad S B_1 = B_2, \qquad C_2 S = C_1.$$

Thus Θ_1 and Θ_2 are similar. Moreover we proved that S is of the form indicated in the theorem for n larger than or equal to the maximum of the degrees of the minimal polynomials of A_1 and A_2. But these polynomials are the same since A_1 and A_2 are similar. So the proof is complete. □

By combining Theorems 7.6 and 7.7 we obtain the following result.

Corollary 7.8. *The transfer functions of two finite-dimensional systems coincide if and only if these systems are dilations of similar systems.*

We conclude this section with a discussion of Möbius transformations of finite-dimensional systems as defined in Section 3.6. We begin with a remark.

Let p, q, r and s be complex numbers. For $n = 1, 2, \ldots$ and $t, j = 0, \ldots, n-1$, let the complex number $a_{t,\,j}^{(n)}$ be given by the expression

$$\sum_{\substack{k = 0, \ldots, n-1-t \\ m = 0, \ldots, t \\ k+m = j}} (-1)^{k+t-m} \binom{n-1-t}{k} \binom{t}{m} p^{n-1-t-k} q^{t-m} r^k s^m.$$

In other words $a_{t,\,0}^{(n)}, \ldots, a_{t,\,n-1}^{(n)}$ are the coefficients of the polynomial

$$(p - rx)^{n-1-t} (sx - q)^t.$$

The $n \times n$ matrix $\left[a_{t,\,j}^{(n)}\right]_{t,j=0}^{n-1}$ will be denoted by $[p, q, r, s]_n$. For what follows it is important to note that

$$\det[p, q, r, s]_n = (ps - qr)^{n(n-1)/2}.$$

The proof goes by an induction argument involving the following recurrence relations

$$\begin{aligned}
a_{t,\,0}^{(n+1)} &= -q a_{t-1,\,0}^{(n)}, \\
a_{t,\,k}^{(n+1)} &= s a_{t-1,\,k-1}^{(n)} - q a_{t-1,\,k}^{(n)}, \qquad k = 1, \ldots, n-1, \\
a_{t,\,n}^{(n+1)} &= s a_{t-1,\,n-1}^{(n)}.
\end{aligned}$$

Theorem 7.9. *Let $\Theta = (A, B, C, D; X, Y)$ be a system, and let*

$$\varphi(\lambda) = \frac{p\lambda + q}{r\lambda + s}$$

be a Möbius transformation. Suppose $rA - p$ is invertible. Then Θ_φ is minimal if and only if Θ is minimal.

Proof. Write $\Theta_\varphi = (A_1, B_1, C_1, D_1; X, Y)$. Then, see formula (3.14),

$$A_1 = -(q - sA)(p - rA)^{-1}, \qquad C_1 = (ps - qr)C(p - rA)^{-1}.$$

A simple computation shows that for $n = 1, 2, \ldots$

$$\left(\operatorname{col}\left(C_1 A_1^j\right)_{j=0}^{n-1}\right)(p - rA)^n = (ps - qr)[p, q, r, s]_n \operatorname{col}\left(CA^j\right)_{j=0}^{n-1}.$$

Since φ is a Möbius transformation, we have $ps - qr \neq 0$. So the matrix $[p, q, r, s]_n$ is invertible. By hypothesis, $p - rA$ is invertible. It follows that $\operatorname{col}\left[C_1 A_1^j\right]_{j=0}^{n-1}$ is left invertible if and only if $\operatorname{col}\left(CA^j\right)_{j=0}^{n-1}$ is left invertible. Thus Θ_φ is observable if and only if Θ is observable. In the same way one can show that Θ_φ is controllable if and only if Θ is controllable. □

7.4 Minimality for Hilbert space systems

In this section we consider Hilbert space systems, that is, systems for which the input space, the output space and the state space are Hilbert spaces. We present an example showing that the state space similarity theorem for finite-dimensional systems does not hold in this (possibly infinite-dimensional) Hilbert space setting. To get an appropriate generalization of this result pseudo-similarity, a weaker form of the usual similarity, has to be used. But even with this weaker similarity it can happen that two minimal Hilbert space systems with the same transfer function in a neighborhood of infinity are pseudo-similar but (in contrast to the finite-dimensional case) the pseudo-similarity does not have to be unique.

A system $\Theta = (A, B, C, D; X, U, Y)$ is said to be a *Hilbert space system* if the underlying spaces X, U, and Y are Hilbert spaces. The class of Hilbert space systems is a subclass of the systems considered in Section 7.1, and hence all terminology and notation introduced in that section applies to Hilbert space systems. For instance, by definition, a Hilbert space systems is *minimal* if and only if it is approximately observable and approximately controllable. The class of Hilbert space systems is closed under taking restrictions, that is, the restriction of a Hilbert space system is again a Hilbert space system. The latter does not hold for dilations. A dilation Θ of a Hilbert space system is again a Hilbert space system only when the state space of Θ is a Hilbert space. The following result is a generalization of Theorem 7.6.

Theorem 7.10. *Any Hilbert space system is a dilation of a Hilbert space system that is minimal. In particular, a Hilbert space system is minimal if and only if it does not have a proper restriction.*

Proof. With some modifications the proof follows the same line of reason as that of the proof of Theorem 7.6. Let $\Theta = (A, B, C, D; X, U, Y)$ be a Hilbert space system. Put $X_1 = \operatorname{Ker}(C|A)$, and let X_0 be the orthogonal complement of $X_1 \cap \overline{\operatorname{Im}(A|B)}$ in the closed subspace $\overline{\operatorname{Im}(A|B)}$. Obviously, X_1 and X_0 are orthogonal closed subspaces of X. We define X_2 to be the orthogonal complement of $X_1 \oplus X_0$ in X. Then X_2 is also a closed subspace of X, and we see that $X = X_1 \oplus X_0 \oplus X_2$. Furthermore, relative to this decomposition A, B, and C partition as follows:

$$A = \begin{bmatrix} * & * & * \\ 0 & A_0 & * \\ 0 & 0 & * \end{bmatrix}, \qquad B = \begin{bmatrix} * \\ B_0 \\ 0 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & C_0 & * \end{bmatrix}.$$

Since X_0 is a closed subspace of the Hilbert space X, the space X_0 is a Hilbert space in its own right. Thus the system $\Theta_0 = (A_0, B_0, C_0, D; X_0, U, Y)$ is a Hilbert system and it is a restriction of Θ.

We claim that Θ_0 is also minimal. To see this, let x_0 be a vector in the space $\operatorname{Ker}(C_0, A_0)$. Thus $C_0 A_0^j x_0 = 0$ for $j = 0, 1, 2 \ldots$. Since

$$CA^j = \begin{bmatrix} 0 & C_0 A_0^j & * \end{bmatrix}, \qquad j = 0, 1, 2 \ldots,$$

it follows that the vector $x = 0 \oplus x_0 \oplus 0$ belongs to $\operatorname{Ker}(C, A) = X_1$. This can only happen when $x_0 = 0$. Hence Θ_0 is approximately observable. Next, take y_0 in X_0 such that $y_0 \perp \operatorname{Im}(A_0, B_0)$. Notice that

$$A^j B = \begin{bmatrix} * \\ A_0^j B_0 \\ 0 \end{bmatrix}, \qquad j = 0, 1, 2 \ldots.$$

Since $y_0 \perp X_1$, we conclude that $y_0 \perp \operatorname{Im}(A, B)$. But X_0 is contained in $\overline{\operatorname{Im}(A|B)}$. So $y_0 = 0$, and Θ_0 is approximately controllable. We have proved that Θ_0 is minimal.

The final statement of the theorem is proved in the same way as the final statement of Theorem 7.6. One only has to replace $\operatorname{Im}(A|B)$ by its closure. □

Next, we present an example showing that two minimal Hilbert space systems of which the transfer functions coincide in a neighborhood of infinity do not have to be similar. For this purpose let ℓ_2 be the Hilbert space of all square summable sequences $x = (x_n)_{n=0}^{\infty}$ with entries in $\mathbb{C}$. By T we denote the backward shift on ℓ_2, that is, T is the operator defined by

$$T(x_n)_{n=0}^{\infty} = (y_n)_{n=0}^{\infty}, \qquad \text{where } y_n = x_{n+1} \text{ for } n = 0, 1, 2, \ldots$$

We shall need the following lemma.

Lemma 7.11. *For a real number $r > 0$, let $\varphi(r)$ be the element in ℓ_2 given by*

$$\varphi(r) = \left(\frac{r^{-n}}{(n+1)!} \right)_{n=0}^{\infty}. \tag{7.8}$$

The element $\varphi(r)$ is cyclic *with respect to the backward shift T, that is, the smallest closed T-invariant subspace containing $\varphi(r)$ is the full space ℓ_2.*

Proof. Let $\varphi_n(r)$ be the nth entry in the sequence $\varphi(r)$. Then

$$\frac{\varphi_{n+k}(r)}{\varphi_k(r)} = \frac{r^{-n}}{(n+k+1)\cdots(1+k+1)} \leq \frac{r^{-n}}{n!}$$

It follows that

$$s_k = \frac{1}{|\varphi_k(r)|^2} \sum_{n=1}^{\infty} |\varphi_{n+k}(r)|^2 < \infty.$$

Moreover the sequence $s_1, s_2, s_3, \ldots$ is decreasing, and hence $\lim_{k\to\infty} s_k$ exists. But then we can use the solution to Problem 160 in [77] to show that $\varphi(r)$ is cyclic with respect to the backward shift on ℓ_2. □

Now consider the Hilbert space system $\Theta_r = (A_r, B_r, C, D; \ell_2, \mathbb{C}, \mathbb{C})$, where

$$\begin{aligned}
&A_r : \ell_2 \to \ell_2, && A_r = rT, \\
&B_r : \mathbb{C} \to \ell_2, && B_r a = a\varphi(r), \qquad a \in \mathbb{C}, \\
&C : \ell_2 \to \mathbb{C}, && C\big((x_n)_{n=0}^{\infty}\big) = x_0, \\
&D : \mathbb{C} \to \mathbb{C}, && Da = a, \qquad a \in \mathbb{C}.
\end{aligned}$$

Proposition 7.12. *The Hilbert space systems $\Theta_r = (A_r, B_r, C, D; \ell_2, \mathbb{C}, \mathbb{C})$, $r > 0$, are all minimal and their transfer functions coincide in a neighborhood of infinity. Nevertheless, the systems Θ_r, $r > 0$, are mutually non-similar.*

Proof. Fix $r > 0$. Note that

$$CA_r^j\big((x_n)_{n=0}^{\infty}\big) = r^j x_j, \qquad j = 0, 1, 2, \ldots.$$

Thus, if $x = (x_n)_{n=0}^{\infty}$ belongs to $\operatorname{Ker}(C|A_r)$, then $x_j = 0$ for each $j = 0, 1, 2, \ldots$. Hence Θ_r is approximately observable. Next, observe that

$$\begin{aligned}
&\operatorname{Im}\left[\, B_r \;\; A_rB_r \;\; \cdots \;\; A_r^nB_r \,\right] \\
&\qquad = \operatorname{span}\{\, \varphi(r),\, rT\varphi(r),\, \ldots,\, r^nT^n\varphi(r) \} \\
&\qquad = \operatorname{span}\{\, \varphi(r),\, T\varphi(r),\, \ldots,\, T^n\varphi(r) \}, \qquad n = 0, 1, 2, \ldots.
\end{aligned}$$

This implies that

$$\operatorname{Im}(A_r|B_r) = \operatorname{span}\{T^n\varphi(r) \mid n = 0, 1, 2, \ldots\}.$$

But the latter space is dense in ℓ_2 by Lemma 7.11. Thus Θ_r is approximately controllable. We have proved that Θ_r is minimal.

Next we compute the transfer function of Θ_r. First note that for each $a \in \mathbb{C}$ we have

$$CA_r^j B_r a = aCA_r^j\varphi(r) = ar^j\frac{r^{-j}}{(j+1)!} = \frac{a}{(j+1)!}, \qquad j = 0, 1, 2, \ldots.$$

Using this and taking $|\lambda| > \|A_r\|$ we get

$$\begin{aligned} W_{\Theta_r}(\lambda) &= D + C(\lambda - A_r)^{-1}B_r = D + \sum_{j=0}^{\infty}\Big(\frac{1}{\lambda}\Big)^{j+1} CA_r^j B_r \\ &= 1 + \sum_{j=0}^{\infty}\Big(\frac{1}{\lambda}\Big)^{j+1}\frac{1}{(j+1)!} = \sum_{j=0}^{\infty}\frac{1}{j!}\Big(\frac{1}{\lambda}\Big)^{j} = e^{1/\lambda}. \end{aligned}$$

We conclude that for any pair r_1 and r_2 of positive numbers the transfer functions of Θ_{r_1} and Θ_{r_2} coincide in a neighborhood of infinity.

Finally, if r_1 and r_2 are different positive numbers, then Θ_{r_1} and Θ_{r_2} are not similar. Indeed, if Θ_{r_1} and Θ_{r_2} would be similar, then their state operators A_{r_1} and A_{r_2} would be similar too, but this can only happen when $r_1 = r_2$. □

To deal with the phenomenon appearing in the previous proposition we introduce a weaker type of system similarity. Consider two systems

$$\Theta_j = (A_j, B_j, C_j, D_j; X_j, U, Y), \qquad j = 1, 2.$$

We say that the systems Θ_1 and Θ_2 are *pseudo-similar* if $D_1 = D_2$, and there exists an injective closed linear operator $S(X_1 \to X_2)$ with domain $\mathcal{D}(S)$ in the Hilbert space X_1 and range in the Hilbert space X_2 such that

$$\overline{\mathcal{D}(S)} = X_1, \qquad \overline{\operatorname{Im}(S)} = X_2, \tag{7.9}$$

$$A_1[\mathcal{D}(S)] \subset \mathcal{D}(S), \qquad SA_1|_{\mathcal{D}(S)} = A_2S, \tag{7.10}$$

$$B_1[U] \subset \mathcal{D}(S), \qquad B_2 = SB_1, \tag{7.11}$$

$$C_1\big|_{\mathcal{D}(S)} = C_2S. \tag{7.12}$$

In this case we call S a *pseudo-similarity from* Θ_1 *to* Θ_2. (Some authors use the term weak similarity, see, e.g., [106], however the term quasi-similarity is usually used for the case when $\mathcal{D}(S)$ is the full space and hence S is bounded.)

Conditions (7.10) and (7.11) imply that $A_1^j B_1 U \subset \mathcal{D}(S)$ and $SA_1^j B_1 = A_2^j B_2$ for each $j \geq 0$, and thus

$$\operatorname{Im}(A_1|B_1) \subset \mathcal{D}(S), \qquad S\big[\operatorname{Im}(A_1|B_1)\big] = \operatorname{Im}(A_2|B_2). \tag{7.13}$$

From (7.10)–(7.12) we get that $C_1 A_1^j B_1 = C_2 S A_1^j B_1 = C_2 A_2^j B_2$ for each $j \geq 0$. Hence, if two Hilbert space systems Θ_1 and Θ_2 are pseudo-similar, then their transfer functions coincide in a neighborhood of infinity. The next theorem shows that the converse is also true.

Theorem 7.13. *Let Θ_1 and Θ_2 be minimal Hilbert space systems, and suppose that their transfer functions coincide in a neighborhood of infinity. Then Θ_1 and Θ_2 are pseudo-similar.*

Proof. Define R from $\operatorname{Im}(A_1|B_1)$ into $\operatorname{Im}(A_2|B_2)$ by setting

$$R\bigg(\sum_{j=0}^{n} A_1^j B_1 u_j\bigg) = \sum_{j=0}^{n} A_2^j B_2 u_j.$$

Then R is well defined. To see this it suffices to show that

$$\sum_{j=0}^{n} A_1^j B_1 u_j = 0 \;\;\Rightarrow\;\; \sum_{j=0}^{n} A_2^j B_2 u_j = 0. \tag{7.14}$$

Assume the left-hand side of (7.14) holds. Then for each $k = 0, 1, 2, \ldots$ we have $\sum_{j=0}^{n} C_1 A_1^{k+j} B_1 u_j = 0$. The fact that the transfer functions of Θ_1 and Θ_2 coincide in a neighborhood of infinity is equivalent to the statement that

$$C_1 A_1^n B_1 = C_2 A_2^n B_2, \qquad n = 0, 1, 2, \ldots . \tag{7.15}$$

Thus

$$C_2 A_2^k \bigg(\sum_{j=0}^{n} A_2^j B_2 u_j\bigg) = 0, \qquad n = 0, 1, 2, \ldots .$$

But $\operatorname{Ker}(C_2|A_2) = \bigcap_{k\geq 0} \operatorname{Ker} C_2 A_2^k = \{0\}$, because Θ_2 is minimal. Thus the right-hand side of (7.14) is proved.

Next, we show that R is closable. Let $x_1, x_2, \ldots$ be a sequence in $\operatorname{Im}(A_1|B_1)$ such that $x_n \to 0$ and $Rx_n \to y$ for $n \to \infty$. We have to show that $y = 0$. Again using (7.15), we see that for each n we have

$$C_1 A_1^k x_n = C_2 A_2^k R x_n, \qquad k = 0, 1, 2, \ldots . \tag{7.16}$$

Fix $k \geq 0$. Then $C_1 A_1^k x_n \to 0$ and $C_2 A_2^k R x_n \to C_2 A_2^k y$ for $n \to \infty$. Thus (7.16) yields $C_2 A_2^k y = 0$ for $k = 0, 1, 2, \ldots$. But $\operatorname{Ker}(C_2|A_2)$ consists of the zero element only, because Θ_2 is minimal. Therefore $y = 0$, and thus R is closable.

Let S be the closure of R. Then S is a closed operator. The operator S is also injective. Indeed, assume $x \in \mathcal{D}(S)$ and $Sx = 0$. Then there exists a sequence $x_1, x_2, \ldots$ in $\operatorname{Im}(A_1|B_1)$ such that $x_n \to x$ and $Rx_n \to 0$ for $n \to \infty$. For these vectors x_n formula (7.16) holds, and hence

$$C_1 A_1^k x = \lim_{n\to\infty} C_1 A_1^k x_n = \lim_{n\to\infty} C_2 A_2^k R x_n = 0, \qquad k = 0, 1, 2, \ldots.$$

Since Θ_1 is minimal, this shows that $x = 0$, and thus S is injective.

We proceed by showing that (7.9)–(7.12) are fulfilled. By definition, we have $\operatorname{Im}(A_1|B_1) \subset \mathcal{D}(S)$, and thus the minimality of Θ_1 yields $\overline{\mathcal{D}(S)} = X_1$. Similarly, $\operatorname{Im} S \supset \operatorname{Im} R = \operatorname{Im}(A_2|B_2)$, and thus $\overline{\operatorname{Im} S} = X_2$ because of the minimality of Θ_2. Thus (7.9) holds. Next, take $x \in \mathcal{D}(S)$. So there exist $x_1, x_2, \ldots$ in $\operatorname{Im}(A_1|B_1)$ such that $x_n \to x$ and $Rx_n \to Sx$ for $n \to \infty$. Now

$$A_1 x_n \in \operatorname{Im}(A_1|B_1) \subset \mathcal{D}(S), \quad A_1 x_n \to A_1 x \qquad (n \to \infty);$$

$$SA_1 x_n = RA_1 x_n = A_2 R x_n \to A_2 S x \qquad (n \to \infty).$$

Since S is closed, this shows that $A_1 x \in \mathcal{D}(S)$ and $SA_1 x = A_2 S x$. Thus (7.10) holds. Since $B_1 U \subset \operatorname{Im}(A_1|B_1)$, we have $B_1 U \subset \mathcal{D}(S)$ and $SB_1 = RB_1 = B_2$, because of the definition of R. Finally, to prove (7.12), take $x \in \mathcal{D}(S)$. So there exist $x_1, x_2, \ldots$ in $\operatorname{Im}(A_1|B_1)$ such that $x_n \to x$ and $Rx_n \to Sx$ for $n \to \infty$. For the vectors x_n formula (7.16) is valid. It follows that

$$C_1 x = \lim_{n\to\infty} C_1 x_n = \lim_{n\to\infty} C_2 R x_n = C_2 S x,$$

which proves (7.12), and we are done. □

We conclude this section with two examples. The first shows that, in contrast to the usual similarity, pseudo-similarity does not necessarily preserve minimality of a Hilbert space system (see Proposition 7.14 below). The second example shows (see Proposition 7.15 below) that, in general, the pseudo-similarity in Theorem 7.13 is not unique. Both examples use the same general setup which we will describe first.

In the sequel $S(X_1 \to X_2)$ is a closed and injective linear operator with domain $\mathcal{D}(S)$ in the Hilbert space X_1 and range in the Hilbert space X_2. We shall assume that

$$\mathcal{D}(S) \neq X_1, \qquad \overline{\mathcal{D}(S)} = X_1, \qquad \operatorname{Im} S \neq X_2. \tag{7.17}$$

Fix $v \in X_1$, $v \notin \mathcal{D}(S)$, and $w \in X_2$, $w \notin \operatorname{Im}(S)$. Let $\hat{S}(X_1 \to X_2)$ be the operator with domain

$$\mathcal{D}(\hat{S}) = \{\lambda v + d \mid \lambda \in \mathbb{C}, \qquad d \in \mathcal{D}(S)\},$$

defined by $\hat{S}(\lambda v + d) = \lambda w + Sd$. We claim the operator $\hat{S}$ is also closed. To see this, let $\mathcal{G}(S)$ and $\mathcal{G}(\hat{S})$ denote the graphs of S and $\hat{S}$, that is,

$$\mathcal{G}(S) = \left\{ \begin{bmatrix} x \\ Sx \end{bmatrix} \mid x \in \mathcal{D}(S) \right\}, \qquad \mathcal{G}(\hat{S}) = \left\{ \begin{bmatrix} x \\ \hat{S}x \end{bmatrix} \mid x \in \mathcal{D}(\hat{S}) \right\}.$$

Since S is closed, its graph $\mathcal{G}(S)$ is a closed subspace of the Hilbert space direct sum $X_1 \oplus X_2$. The definition of $\hat{S}$ implies that

$$G(\hat{S}) \;=\; G(S) \,\dot{+}\, \operatorname{span} \begin{bmatrix} v \\ w \end{bmatrix} \;\subset\; X_1 \oplus X_2.$$

Thus $G(\hat{S})$ is a one-dimensional extension of the closed subspace $G(S)$. It follows that $G(\hat{S})$ is closed too (cf., Theorem XI.2.5 in [48]), and hence $\hat{S}$ is a closed operator. Obviously, we have

$$G(S) \subsetneqq G(\hat{S}), \qquad \mathcal{D}(\hat{S}) \;\text{ is dense in } X_1. \tag{7.18}$$

The operator $\hat{S}$ is also injective, because S is injective and $w \notin \operatorname{Im} S$. Since $\mathcal{D}(S) \subset \mathcal{D}(\hat{S})$ and $\mathcal{D}(S)$ is dense in X_1, we also know that $\mathcal{D}(\hat{S})$ is dense in X_1. However, $\mathcal{D}(\hat{S}) \neq X_1$. Indeed, if $\mathcal{D}(\hat{S}) = X_1$, then $\hat{S}$ is bounded by the closed graph theorem. This implies that the closed operator $S = \hat{S}|_{\mathcal{D}(S)}$ is also bounded. It follows that $\mathcal{D}(S) = \overline{\mathcal{D}(S)} = X_1$, which contradicts the first part of (7.17).

Next we use the operators S and $\hat{S}$ to construct two Hilbert space systems. For both systems the input space U is defined to be the space $\mathcal{D}(S)$ endowed with the graph norm

$$\|x\|_U = (\|x\|^2 + \|Sx\|^2)^{1/2}, \qquad \text{where } x \in \mathcal{D}(S).$$

Analogously, by definition, the output space Y is the space $\mathcal{D}(\hat{S}^*)$ endowed with graph norm

$$\|y\|_Y = (\|\hat{S}^* y\|^2 + \|y\|^2)^{1/2}, \qquad \text{where } y \in \mathcal{D}(\hat{S}^*).$$

Here, as before the * means that one has to take the Hilbert space adjoint. Now define the following operators:

$$B_1 : U \to X_1, \qquad B_1 x = x; \qquad B_2 : U \to X_2, \qquad B_2 x = Sx, \tag{7.19}$$

$$\Gamma_1 : Y \to X_1, \qquad \Gamma_1 y = \hat{S}^* y; \qquad \Gamma_2 : Y \to X_2, \qquad \Gamma_2 y = y, \tag{7.20}$$

and put

$$C_1 = \Gamma_1^* : X_1 \to Y, \qquad C_2 = \Gamma_2^* : X_2 \to Y. \tag{7.21}$$

Obviously, the operators defined by (7.19) and (7.20) are bounded linear operators with operator norm of at most one. It follows that the same holds true for the operators in (7.21). We shall consider the following two Hilbert space systems:

$$\Theta_1 = (0, B_1, C_1, 0; X_1, U, Y), \qquad \Theta_2 = (0, B_2, C_2, 0; X_2, U, Y). \tag{7.22}$$

Let us show that the transfer functions of these two systems coincide in a neighborhood of infinity. Note that the state operators and the external coefficients of Θ_1 and Θ_2 are all zero operators. Thus in order to show that the transfer

functions of Θ_1 and Θ_2 coincide in a neighborhood of infinity, it suffices to show that $B_1C_1 = B_2C_2$. The latter identity follows from

$$\operatorname{Im}\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = G(S) \subset G(\hat{S}) = \operatorname{Ker}\begin{bmatrix} C_1 & C_2 \end{bmatrix}. \tag{7.23}$$

The first equality and first inclusion in (7.23) are trivial. The second equality follows from

$$\begin{aligned} \operatorname{Ker}\begin{bmatrix} C_1 & C_2 \end{bmatrix} &= \left(\operatorname{Im}\begin{bmatrix} C_1^* \\ -C_2^* \end{bmatrix}\right)^{\perp} = \left(\operatorname{Im}\begin{bmatrix} \Gamma_1 \\ -\Gamma_2 \end{bmatrix}\right)^{\perp} \\ &= \left\{-\begin{bmatrix} -\hat{S}^*y \\ y \end{bmatrix} \,\Big|\, y \in \mathcal{D}(\hat{S}^*)\right\}^{\perp} = G(\hat{S}). \end{aligned}$$

The last identity is a well-known property of a densely defined closed linear operator acting in Hilbert spaces (see Proposition XIV.2.1 in [46]). From (7.23) we see that

$$\begin{bmatrix} C_1 & C_2 \end{bmatrix}\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u = 0, \qquad u \in U.$$

Hence $B_1C_1 = B_2C_2$, and the transfer functions of Θ_1 and Θ_2 are both equal to $\lambda^{-1}K$, where $K = B_1C_1 = B_2C_2$.

Proposition 7.14. *In general, minimality of a Hilbert space system is not preserved under pseudo-similarity.*

Proof. We continue to use the notation introduced in the three paragraphs preceding this proposition. Let Θ_1 and Θ_2 be the Hilbert space systems defined by (7.22). Assume additionally that

$$X_2 = \operatorname{span}\{w\} \oplus \overline{\operatorname{Im} S}. \tag{7.24}$$

It is straightforward to construct such an operator S. We claim that in this case Θ_1 is minimal, Θ_2 is not minimal, and $\hat{S}$ is a pseudo-similarity from Θ_1 to Θ_2.

We first show that $\hat{S}$ is a pseudo-similarity. We have already seen that $\hat{S}(X_1 \to X_2)$ is a densely defined injective closed linear operator. The additional assumption (7.24) implies that $\operatorname{Im}\hat{S}$ is dense in X_2. Indeed, since $w \perp \operatorname{Im} S$ and $\operatorname{Im} S \subset \operatorname{Im}\hat{S}$, the space $\overline{\operatorname{Im} S}$ is properly contained in the space $\overline{\operatorname{Im}\hat{S}}$. But then (7.24) yields $\overline{\operatorname{Im}\hat{S}} = X_2$. Thus (7.9) holds with $\hat{S}$ in place of S. Since A_1 and A_2 are both zero operators, condition (7.10) also holds with $\hat{S}$ in place of S. To show that $\hat{S}$ satisfies (7.11) and (7.12), note that according to (7.23) we have

$$\operatorname{Im}\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \subset G(\hat{S}) \subset \operatorname{Ker}\begin{bmatrix} C_1 & C_2 \end{bmatrix}.$$

The first inclusion implies that $B_1U \subset \mathcal{D}(\hat{S})$ and $B_2u = \hat{S}B_1u$ for each $u \in U$. The second inclusion yields $C_1x - C_2\hat{S}x = 0$ for each $x \in \mathcal{D}(\hat{S})$. Thus (7.11) and (7.12) are satisfied for $\hat{S}$ in place of S. Thus $\hat{S}$ is a pseudo-similarity.

Notice that $\operatorname{Im} B_2 = \operatorname{Im} S$, and hence $\operatorname{Im} B_2$ is not dense in X_2 because of (7.24). It follows that Θ_2 is not minimal. On the other hand $\operatorname{Im} B_1 = \mathcal{D}(S)$, and hence by (7.17) the space $\operatorname{Im} B_1$ is dense in X_1. Also

$$\operatorname{Ker} C_1 \;=\; (\operatorname{Im} \Gamma_1)^\perp \;=\; (\operatorname{Im} \hat{S}^*)^\perp \;=\; \operatorname{Ker} \hat{S} = \{0\},$$

because $\hat{S}$ is injective. Thus Θ_1 is minimal too.

Hence the Hilbert space systems Θ_1 to Θ_2 are pseudo-similar, Θ_1 is minimal and Θ_2 is not minimal. We conclude that minimality is not preserved under pseudo-similarity. □

Proposition 7.15. *It can happen that two pseudo-similar minimal systems have two different pseudo-similarities.*

Proof. Again we use notation introduced in the three paragraphs preceding Proposition 7.14. Let Θ_1 and Θ_2 be the Hilbert space systems given by (7.22). In this case we assume additionally that

$$\overline{\operatorname{Im} S} \;=\; X_2. \tag{7.25}$$

We claim that Θ_1 and Θ_2 are both minimal, and that both S and $\hat{S}$ provide a pseudo-similarity from Θ_1 to Θ_2.

As in the one but last paragraph of the proof of the preceding proposition, one shows that Θ_1 is minimal. Since $\operatorname{Im} B_2 = \operatorname{Im} S$, the space $\operatorname{Im} B_2$ is dense in X_2 because of (7.25). Furthermore,

$$\operatorname{Ker} C_2 \;=\; (\operatorname{Im} \Gamma_2)^\perp \;=\; \mathcal{D}(\hat{S}^*)^\perp \;=\; \{0\}.$$

Thus the system Θ_2 is also minimal.

Note that both S and $\hat{S}$ are injective, closed, densely defined, and have dense range. Thus (7.9) is satisfied for both S and $\hat{S}$. Since the state operators of Θ_1 and Θ_2 are both zero operators, condition (7.10) is also satisfied for both S and $\hat{S}$. Using (7.23) it is straightforward to check (again see the proof of Proposition 7.14) that (7.11) and (7.12) hold for both S and $\hat{S}$. Thus S and $\hat{S}$ are pseudo-similarities from Θ_1 to Θ_2. The first part of (7.18) implies that $S \neq \hat{S}$.

We conclude that Θ_1 and Θ_2 are pseudo-similar minimal Hilbert space systems which have two different pseudo-similarities. □

With minor modifications one can transform the example in the above proof into an example of two pseudo-similar minimal systems Θ_1 and Θ_2 for which there exist infinitely many different pseudo-similarities from Θ_1 to Θ_2. In fact, this can already been achieved by choosing $\hat{S}$ in such a way that the quotient space $G(\hat{S})/G(S)$ has dimension two.

7.5 Minimality in special cases

In this section we discuss the notion of minimality for the classes of systems considered in Chapter 3.

7.5.1 Brodskii systems

Let $\Theta = (A, KJ, 2iK^*; H, G)$ be a Brodskii J-system. Following [30] we call Θ *simple* if $\operatorname{Im}(A|K)$ is dense in H. Thus simplicity is here synonymous to controllability. However, in view of the fact that $A - A^* = 2iKJK^*$, we have $\operatorname{Im}(A|K) = \operatorname{Im}(A^*|K)$, and hence $\operatorname{Ker}(K^*|A)$ is the orthogonal complement $\operatorname{Im}(A|K)^\perp$ of $\operatorname{Im}(A|K)$. Therefore, in this particular case, the notions of simplicity (controllability) and minimality coincide.

In [30] it is shown that, given a Brodskii J-system Θ, there exists a simple Brodskii J-system Θ_0 of which the characteristic operator function (transfer function) coincides with that of Θ on a neighborhood of ∞. In fact, if $\Theta = (A, KJ, 2iK^*; H, G)$ and Π is the orthogonal projection of H onto the closure of $\operatorname{Im}(A|K)$, then Π commutes with A and A^* and $\Theta_0 = \operatorname{pr}_\Pi(\Theta)$ has the desired properties. Observe that

$$\Theta = \operatorname{pr}_{I-\Pi}(\Theta)\Theta_0 = \Theta_0 \operatorname{pr}_{I-\Pi}(\Theta).$$

The systems $\operatorname{pr}_\Pi(\Theta)$ and $\operatorname{pr}_{I-\Pi}(\Theta)$ are called the *principal part* and *excess part* of Θ, respectively.

In [30] it is also shown that two simple Brodskii J-systems whose characteristic operator functions coincide on a neighborhood of ∞ are similar, the (unique) similarity transformation being a unitary operator. This fact plays an important role in [30]. For instance, it is used to prove the unicellularity of the Volterra integral operator on $L_2(0, 1)$.

7.5.2 Kreĭn systems

It can be shown that two minimal Kreĭn J-systems whose transfer function coincide on a neighborhood of ∞ are similar, the (unique) similarity transformation being a unitary operator. In fact, this conclusion can be reached under the somewhat weaker assumption that the systems are prime. Following [33], we call a Kreĭn J-system $\Theta = (A, R, -J(K^*)^{-1}R^*A, K; H, G)$ *prime* if

$$\operatorname{Im}(A|R) + \operatorname{Im}(A^*|R)$$

is dense in H. In order to clarify this notion we make some general remarks.

To facilitate the discussion, we introduce a notation. Let $N_j,\ j \in \mathcal{J}$ be a family of linear manifolds in a Banach space indexed with the help of the index set $\mathcal{J}$. The closure of the linear hull of these manifolds will be denoted by $\bigvee_{j\in\mathcal{J}} N_j$. In case the underlying space is finite-dimensional, the linear hull in question is itself already closed.

Suppose $\Theta = (A, B, C, D; X, Y)$ is a system with an invertible main operator A. We say that Θ is *biminimal* if

$$\bigcap_{j=-\infty}^{\infty} \operatorname{Ker} CA^j = 0, \qquad \bigvee_{j=-\infty}^{\infty} \operatorname{Im} A^j B = H.$$

Obviously, if Θ is minimal, then Θ is biminimal too. The converse is also true if, for example, A is an algebraic operator. The latter condition is automatically fulfilled when X is finite-dimensional.

Now, returning to the subject of this subsection, assume that

$$\Theta = (A, R, -J(K^*)^{-1}R^*A, K; H, G)$$

is a Kreĭn J-system. Using the relationship between A, A^* and R appearing in Section 3.2, one can show that

$$\Big(\bigvee_{j=-\infty}^{\infty} \operatorname{Im} A^j R \Big)^{\perp} = \bigcap_{j=-\infty}^{\infty} \operatorname{Ker} R^* A^j \,,$$

while $\operatorname{Im}(A|R) + \operatorname{Im}(A^*|R)$ is the linear hull of $\operatorname{Im} A^j R$, $j = 0, \pm 1, \pm 2, \ldots$. Hence Θ is prime if and only if Θ is biminimal. In particular, if Θ is minimal, then certainly Θ is prime.

Finally we mention that if Θ is a Kreĭn J-system, then there exists a prime Kreĭn J-system Θ_0 whose transfer function coincides with that of Θ on a neighborhood of ∞. The construction of Θ_0 is suggested in [33], Sections 3 and 4.

7.5.3 Unitary systems

Given the unitary system $\Theta = (A, B, C, D; X, U, Y)$, define $\mathcal{R}(\Theta)$ to be the closed linear hull of the vectors $A^n Bu$ and $(A^*)^k C^* y$, where u and y are arbitrary vectors in U and Y, respectively, and $n, k = 0, 1, 2, \ldots$. The space $\mathcal{R}(\Theta)$ is called the *principal subspace* of Θ, and its orthogonal complement in X is called the *excessive* subspace and is denoted by $\mathcal{N}(\Theta)$. Both subspaces are invariant under A.

To explain the terminology, consider the orthogonal direct sum $X = \mathcal{N}(\Theta) \oplus \mathcal{R}(\Theta)$, and write A, B, and C as operator matrices relative to this decomposition. It is straightforward to check that the operator matrices for A, B, and C are of the following form:

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{00} \end{bmatrix} : \mathcal{N}(\Theta) \oplus \mathcal{R}(\Theta) \to \mathcal{N}(\Theta) \oplus \mathcal{R}(\Theta), \tag{7.26}$$

$$B = \begin{bmatrix} 0 \\ B_0 \end{bmatrix} : U \to \mathcal{N}(\Theta) \oplus \mathcal{R}(\Theta), \tag{7.27}$$

$$C = \begin{bmatrix} 0 & C_0 \end{bmatrix} : \mathcal{N}(\Theta) \oplus \mathcal{R}(\Theta) \to Y. \tag{7.28}$$

It follows that the system matrix of Θ partitions as

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A_{11} & 0 & 0 \\ 0 & A_{00} & B_0 \\ 0 & C_0 & D \end{bmatrix} : \mathcal{N}(\Theta) \oplus \mathcal{R}(\Theta) \oplus U \to \mathcal{N}(\Theta) \oplus \mathcal{R}(\Theta) \oplus Y.$$

Since the system matrix is unitary, we conclude that A_{11} on $\mathcal{N}(\Theta)$ and

$$\begin{bmatrix} A_{00} & B_0 \\ C_0 & D \end{bmatrix} : \mathcal{R}(\Theta) \oplus U \to \mathcal{R}(\Theta) \oplus Y$$

are both unitary operators. In particular, the system

$$\Theta_0 = (A_{00}, B_0, C_0, D; \mathcal{R}(\Theta), U, Y)$$

is unitary. Furthermore, from the partitionings (7.26), (7.27) and (7.28) it follows that Θ is a dilation of Θ_0, and hence Θ and Θ_0 have the same transfer function. The system Θ_0 is called the *principal part* of Θ.

A unitary system is called *pure* if its excessive subspace consists of the zero vector only. One can show that the system Θ_0 constructed in the previous paragraph is pure. Hence one can restrict any unitary system to a pure one without changing its transfer function.

The following result is the analogue of Theorem 7.7 for unitary systems; its proof can be found in [47], Section XXVIII.3.

Theorem 7.16. *Two pure unitary systems have the same transfer function if and only if these systems are unitarily equivalent, and in this case the unitary operator establishing the unitary equivalence is unique.*

If a unitary system is observable and controllable, then its excessive part consists of the zero vector only, and hence such a system is pure. Thus a minimal unitary system is pure. For a finite-dimensional unitary system the converse is also true. In other words, a finite-dimensional unitary system is minimal if and only if it is pure. For arbitrary infinite-dimensional unitary systems this result is not true. In fact, it may happen that a pure unitary system with an infinite-dimensional state space is neither observable nor controllable. An example is provided by Corollary 5.3 in Section XXVIII.5 of [47].

7.5.4 Monic systems

From the definition of a monic system it is clear that such a system is always minimal. So it is not surprising that the notion of minimality does not appear in [11], [12]. Note however that a monic system Θ is determined up to similarity by its transfer function (cf., [12], Theorem 1.2). Also, because of linearization, the spectral properties of the main operator of a monic system Θ are determined completely by W_Θ^{-1}.

7.5.5 Polynomial systems

Let P be a comonic polynomial of degree ℓ whose coefficients are $m \times m$ matrices (i.e., operators acting on $\mathbb{C}^m$). Put $L(\lambda) = \lambda^\ell P(\lambda^{-1})$. Then L is a monic polynomial of degree ℓ. Let $\Delta = (T, R, Q, 0; \mathbb{C}^{m\ell}, \mathbb{C}^m)$ be a finite-dimensional monic system whose transfer function coincides with L^{-1}. Then the unital system $\Theta = (T, T^\ell R, Q; \mathbb{C}^{m\ell}, \mathbb{C}^m)$ and its associate $\Theta^\times = (T - T^\ell RQ, T^\ell R, -Q; \mathbb{C}^{m\ell}, \mathbb{C}^m)$ are realizations for $P(\lambda^{-1})^{-1}$ and $P(\lambda^{-1})$, respectively (cf., Subsection 3.5). As $\operatorname{col}(QT^j)_{j=0}^{\ell-1}$ and $\operatorname{row}(T^jR)_{j=0}^{\ell-1}$ are both invertible, we have

$$\operatorname{Ker}(Q \,|\, T) = \{0\}, \qquad \operatorname{Im}(T^\ell R \,|\, T) = \operatorname{Im} T^\ell.$$

So Θ is observable, but generally not controllable. The same is true for $\Theta^\times$.

In order to obtain minimal realizations for $P(\lambda^{-1})^{-1}$ and $P(\lambda^{-1})$, we apply the method indicated in the proof of Theorem 7.6. Put $X_0 = \operatorname{Im} T^\ell$. Then X_0 is invariant under T. Let T_0 be the restriction of T to X_0 considered as an operator on X_0. For convenience we write B instead of $T^\ell R$. Note that B maps $\mathbb{C}^n$ into X_0. Let B_0 be the operator B viewed as an operator from $\mathbb{C}^m$ into X_0. Finally, let Q_0 be the restriction of Q to X_0. Then $\Theta_0 = (T_0, B_0, Q_0; X_0, \mathbb{C}^m)$ and $\Theta_0^\times = (T_0 - B_0Q_0, B_0, -Q_0; X_0, \mathbb{C}^m)$ are minimal realizations for $P(\lambda^{-1})^{-1}$ and $P(\lambda^{-1})$, respectively.

Minimal realizations for $P\big(\lambda^{-1}\big)^{-1}$ and $P(\lambda^{-1})$ can also be obtained by applying the method of Section 8.3 in the next chapter. The alternative construction presented above however is somewhat more direct. Observe that it can also be used to produce a minimal realization for $P(\lambda^{-1})$ when P is an arbitrary, possibly non-comonic, $n \times n$ matrix polynomial. Indeed, one just constructs a minimal realization for $I - P(0) + P(\lambda^{-1})$ and adds $P(0) - I$ to the external operator.

Notes

This chapter is based on the text of Chapter 3 in [14] with Sections 7.2 and 7.4, and Subsection 7.5.3 as new additions. The notions of controllability, observability and minimality are standard in system and control theory; see, e.g., the textbooks [84] and [36]. Section 7.4 is taken from [4]. A full description of all vectors in ℓ_2 that are cyclic with respect to the backward shift can be found in [41]. Theorem 7.13 has appeared as Theorem 3b.1 in [79], and as Theorem 3.2 in [7] (see Theorem 9.2.3 in [106] for a continuous time version). The fact that the pseudo-similarity constructed in the proof of Theorem 7.13 is closed can be found in [2], Proposition 6. Propositions 7.14 and 7.15 can also be viewed as results about minimal representations of an operator as a product of two bounded operators; see [4]. For more information about pseudo-similarity, see Sections 3.2 and 3.3 in [5]). With appropriate modifications Theorem 7.6 and Corollary 7.8 hold for various classes of time varying systems; see, e.g., [53] and [3].

Chapter 8

Minimal Realizations and Pole-Zero Structure

In this chapter finite-dimensional systems are studied in terms of the zero or pole data of their transfer functions. In the first two sections we describe the local zero and pole data, and related Jordan chains, of a meromorphic $m \times m$ matrix function of which the determinant does not vanish identically. In the third section these results are used to construct minimal realizations of rational matrix functions in terms of the zero or pole data of the function. The fourth section deals with the notions of local degree of a transfer function and local minimality of a finite-dimensional system. The global versions of these notions are studied in the final section.

8.1 Zero data and Jordan chains

Throughout this section M is an $m \times m$ matrix function which is meromorphic on the connected open set Ω in $\mathbb{C}$. We assume that M is *regular* on Ω, that is, $\det M(\lambda) \not\equiv 0$. As usual the values of M are identified with their canonical action on $\mathbb{C}^m$.

Let λ_0 be a point in Ω, and let

$$M(\lambda) = \sum_{j=-q}^{\infty} (\lambda - \lambda_0)^j A_j$$

be the Laurent expansion of M. Here it is assumed that $q \geq 0$. Notice that $q = 0$ corresponds to the case when M is analytic at λ_0. Although q is not unique, the definitions given below do not depend on the choice of q.

We call λ_0 a *zero* or *eigenvalue* of M if there exist vectors $x_0, x_1, \ldots, x_q$ in $\mathbb{C}^m$, $x_0 \neq 0$, such that

$$A_{-q}x_j + \cdots + A_{-q+j}x_0 = 0, \qquad j = 0, \ldots, q. \tag{8.1}$$

In that case the vector x_0 is called an *eigenvector* or *root vector* of M at the eigenvalue λ_0.

Proposition 8.1. *The vector $x \neq 0$ is an eigenvector of M at λ_0 if and only if there exists a $\mathbb{C}^m$-valued function φ, analytic at λ_0, such that $\varphi(\lambda_0) = x$, the function $M(\lambda)\varphi(\lambda)$ is analytic at zero, and*

$$\lim_{\lambda \to \lambda_0} M(\lambda)\varphi(\lambda) = 0. \tag{8.2}$$

Proof. Let φ be an arbitrary $\mathbb{C}^m$-valued function which is analytic at λ_0, and consider its Taylor expansion at λ_0:

$$\varphi(\lambda) = \varphi_0 + (\lambda - \lambda_0)\varphi_1 + (\lambda - \lambda_0)^2\varphi_2 + \cdots .$$

Then (8.2) holds if and only if

$$A_{-q}\varphi_j + \cdots + A_{-q+j}\varphi_0 = 0, \qquad j = 0, \ldots, q.$$

By comparing this with (8.1) the proof of the lemma is immediate. □

The linear space of all eigenvectors of M at λ_0 together with the zero vector will be denoted by $\operatorname{Ker}(M; \lambda_0)$. The dimension of the space $\operatorname{Ker}(M; \lambda_0)$ is called the *geometric multiplicity of λ_0 as a zero* of M. If M is analytic at λ_0, we can take $q = 0$, and then λ_0 is an eigenvalue of M if and only if $M(\lambda_0)x_0 = 0$ for some $x_0 \neq 0$ in $\mathbb{C}^m$. Furthermore in that case $\operatorname{Ker}(M; \lambda_0) = \operatorname{Ker} M(\lambda_0)$. In general, we have $\operatorname{Ker}(M; \lambda_0) \subset \operatorname{Ker} L(\lambda_0)$, where $L(\lambda) = (\lambda - \lambda_0)^q M(\lambda)$. For q sufficiently large, this becomes the trivial inclusion $\operatorname{Ker}(M; \lambda_0) \subset \mathbb{C}^m = \operatorname{Ker} L(\lambda_0)$.

An ordered set $(x_0, x_1, \ldots, x_{k-1})$ of vectors in $\mathbb{C}^m$ is called a *Jordan chain* for M at λ_0 if $x_0 \neq 0$ and there exist vectors $x_k, x_{k+1}, \ldots, x_{q+k-1}$ in $\mathbb{C}^m$ such that

$$A_{-q}x_j + \cdots + A_{-q+j}x_0 = 0, \qquad j = 0, \ldots, q + k - 1. \tag{8.3}$$

The number k is the *length* of the chain. Note that x_0 is an eigenvector of M at λ_0 if and only if x_0 is the first vector in a Jordan chain for M at λ_0.

The following observation extends Proposition 8.1.

Proposition 8.2. *The vector $x \neq 0$ is the first vector in a Jordan chain for M at λ_0 of length $k > 0$ if and only if there exists a $\mathbb{C}^m$-valued function φ, analytic at λ_0, such that $\varphi(\lambda_0) = x$ and*

$$\lim_{\lambda \to \lambda_0} \frac{1}{(\lambda - \lambda_0)^{k-1}} M(\lambda)\varphi(\lambda) = 0. \tag{8.4}$$

The proof of Proposition 8.2 is analogous to that of Proposition 8.1. The two propositions show that the definitions given above do not depend on the particular choice of q.

A function φ with the properties described in Proposition 8.2 is called a *root function* of M at λ_0 of *order at least* k. Thus a $\mathbb{C}^m$-valued function φ, analytic at λ_0, is called a root function of M at λ_0 of order at least k if and only if $\varphi(\lambda_0) \neq 0$ and $M(\lambda)\varphi(\lambda)$ has a zero at λ_0 of order al least k. If the order of λ_0 as zero of $M(\lambda)\varphi(\lambda)$ is equal to k, then φ is root function of *order* k.

Given an eigenvector x_0 of M at λ_0, there are in general many Jordan chains for M at λ_0 which have x_0 as their first vector. However, as the next lemma shows, the lengths of these Jordan chains have a finite supremum which we shall call the *rank* of the eigenvector x_0.

Lemma 8.3. *The length of a Jordan chain of M at λ_0 is less than or equal to $\nu - q$, where ν is the order of λ_0 as a zero of* $\det L(\lambda)$ *with* $L(\lambda) = (\lambda - \lambda_0)^q M(\lambda)$.

Proof. Note that $L(\lambda) = (\lambda - \lambda_0)^q M(\lambda)$ is analytic at zero. Also $\det L(\lambda) \not\equiv 0$ and so the order ν of λ_0 as a zero of the analytic scalar function $\det L(\lambda)$ is finite. Put

$$\varphi(\lambda) = x_0 + (\lambda - \lambda_0)x_1 + \cdots + (\lambda - \lambda_0)^{q+k-1} x_{q+k-1},$$

where $x_0, x_1, \ldots, x_{q+k-1}$ satisfy (8.3). It follows that $L(\lambda)\varphi(\lambda)$ is analytic at λ_0, and that the analytic vector function $L(\lambda)\varphi(\lambda)$ has a zero at λ_0 of order at least $q + k$. Since $x_0 \neq 0$, we can choose $y_2, \ldots, y_m$ such that $x_0, y_2, \ldots, y_m$ is a basis of $\mathbb{C}^m$. Let $X(\lambda)$ be the $m \times m$ matrix of which the columns are given by $\varphi(\lambda), y_2, \ldots, y_m$. From $\varphi(\lambda_0) = x_0$ and the choice of the vectors $y_2, \ldots, y_m$, we conclude that $\det X(\lambda) \neq 0$ for λ sufficiently close to λ_0. Next observe that

$$\begin{aligned} \det L(\lambda) \det X(\lambda) &= \det\big(L(\lambda)X(\lambda)\big) \\ &= \det\big[\, L(\lambda)\varphi(\lambda) \;\; L(\lambda)y_2 \;\; \cdots \;\; L(\lambda)y_n \,\big]. \end{aligned}$$

Since $\det X(\lambda) \neq 0$ for λ sufficiently close to λ_0, the order of λ_0 as a zero of the term in the left-hand side is equal to ν. On the other hand, $(\lambda - \lambda_0)^{q+k}$ is a factor of the first column of the matrix in the right-hand side, and therefore also of the determinant. It follows that $q + k \leq \nu$. □

To bring appropriate structure in the collection of Jordan chains corresponding to the eigenvalue λ_0, we proceed as follows. Choose an eigenvector $x_{1,0}$ in $\operatorname{Ker}(M; \lambda_0)$ such that the rank r_1 of $x_{1,0}$ is maximal, and let $(x_{1,0}, \ldots, x_{1,r_1-1})$ be a corresponding Jordan chain of M. Next we choose among all vectors x in $\operatorname{Ker}(M; \lambda_0)$, with x not a multiple of $x_{1,0}$, a vector $x_{2,0}$ of maximal rank, r_2 say, and we select a corresponding Jordan chain $(x_{2,0}, \ldots, x_{2,r_2-1})$. We go on inductively. Assume

$$(x_{1,0}, \ldots, x_{1,r_1-1}), \ldots, (x_{j-1,0}, \ldots, x_{j-1,r_{j-1}-1})$$

have been chosen. Then, among all vectors in $\mathrm{Ker}\,(M;\lambda_0)$ not belonging to the linear space span $\{x_{1,0},\dots,x_{j-1,0}\}$, we pick $x_{j,0}$ having maximal rank, and we let $(x_{j,0},\dots,x_{j,r_j-1})$ be a corresponding Jordan chain. In this way, in a finite number of steps, we obtain a system

$$(x_{1,0},\dots,x_{1,r_1-1}),\ (x_{2,0},\dots,x_{2,r_2-1}),\ \dots,\ (x_{p,0},\dots,x_{p,r_p-1}) \tag{8.5}$$

of Jordan chains for M at λ_0 with the following properties:

(i) the vectors $x_{1,0},\dots,x_{p,0}$ form a basis for $\mathrm{Ker}\,(M;\lambda_0)$ and they have ranks $r_1,\dots,r_p$, respectively,

(ii) for $j=1,\dots,p$, the vector $x_{j,0}$ has maximal rank among all eigenvectors in $\mathrm{Ker}\,(M;\lambda_0)$ that do not belong to span $\{x_{1,0},\dots,x_{j-1,0}\}$; in particular the rank of the eigenvector $x_{1,0}$ in $\mathrm{Ker}\,(M;\lambda_0)$ has the maximal possible value.

A system with these characteristics will be called a *canonical system of Jordan chains* for M at λ_0. The above reasoning shows that such canonical systems of Jordan chains always exist. They are not unique, however, and so it makes sense to ask what can be said about the numbers p and $r_1,\dots,r_p$.

For p the situation is easy: $p=\dim \mathrm{Ker}\,(M;\lambda_0)$, a number which is completely determined by M and independent of certain choices that can be made. With respect to the ranks of the chains, we note the following. Clearly $r_1\ge r_2\ge\cdots\ge r_p$. Further, if

$$x\in \mathrm{span}\,\{x_{1,0},\dots,x_{j,0}\}\setminus \mathrm{span}\,\{x_{1,0},\dots,x_{j-1,0}\},$$

then the rank r of x is equal to r_j. The argument is as follows. The fact that x is a linear combination of $x_{1,0},\dots,x_{j,0}$ implies that there is a Jordan chain for M at λ_0, starting with x, which has length r_j. Just take an appropriate linear combination of Jordan chains starting with the vectors $x_{1,0},\dots,x_{j,0}$. Hence r_j does not exceed r. But clearly we have $r\le r_j$ too, because x is not in span $\{x_{1,0},\dots,x_{j-1,0}\}$. So $r=r_j$.

We can now conclude that the set $\{r_1,\dots,r_p\}$ coincides with the collection of all possible ranks of eigenvectors of M at λ_0, and is thus completely determined by M. In fact, as our next result shows, uniqueness holds even for the (not necessarily distinct) numbers $r_1,\dots,r_p$ themselves.

Proposition 8.4. *Consider a canonical system of Jordan chains for M at λ_0 given by* (8.5)*, and let*

$$(y_{1,0},\dots,y_{1,\rho_1-1}),\ (y_{2,0},\dots,y_{2,\rho_2-1}),\ \dots,\ (y_{\nu,0},\dots,y_{\nu,\rho_\nu-1}) \tag{8.6}$$

be another system of Jordan chains for M at λ_0 with lengths $\rho_1,\dots,\rho_\nu$, respectively, where $\rho_1\ge\rho_2\ge\cdots\ge\rho_\nu$. Assume $y_{1,0},\dots,y_{\nu,0}$ are linearly independent. Then $\nu\le p$ and $\rho_j\le r_j$, $j=1,\dots,\nu$. Moreover, (8.6) *is a canonical system of Jordan chains for M at λ_0 if and only if $\nu=p$ and $\rho_j=r_j$, $j=1,\dots,\nu$.*

Proof. Since $y_{1,0},\dots,y_{\nu,0}$ are linearly independent vectors, they are all nonzero, and hence $\{y_{1,0},\dots,y_{\nu,0}\}$ is a linear independent set in $\operatorname{Ker}(M;\lambda_0)$. It follows that $\nu \le p$.

The eigenvector $x_{1,0}$ of M at λ_0 has maximal possible rank. In particular, $\rho_1 \le r_1$. Next fix $1 < j \le \nu$. The vectors $y_{1,0},\dots,y_{j,0}$ are linearly independent in $\operatorname{Ker}(M;\lambda_0)$ and the dimension of $\operatorname{span}\{x_{1,0},\dots,x_{j-1,0}\}$ is $j-1$. So at least one of the vectors $y_{1,0},\dots,y_{j,0}$ does not belong to $\operatorname{span}\{x_{1,0},\dots,x_{j-1,0}\}$, say $y_{k,0}$. Since $x_{j,0}$ is of maximal rank among all eigenvectors in $\operatorname{Ker}(M;\lambda_0)$ that do not belong to $\operatorname{span}\{x_{1,0},\dots,x_{j-1,0}\}$, we conclude that $r_j \ge \rho_k$. But then $r_j \ge \rho_j$ too as $\rho_k \ge \rho_j$.

Assume that (8.6) is a canonical system of Jordan chains of M at λ_0. Then we may interchange the roles of the systems (8.5) and (8.6), and we can apply the results obtained so far to (8.6) in place of (8.5). This yields, $p \le \nu$ and $r_j \le \rho_j$ for $j = 1,\dots,p$. Hence in this case we have $\nu = p$ and $\rho_j = r_j,\ j = 1,\dots,\nu$.

Finally, suppose $\nu = p$ and $\rho_j = r_j$ for $j = 1,\dots,\nu$. Assume (8.6) is not a canonical system of Jordan chains of M at λ_0. This means that for some k the vector $y_{k,0}$ is not an eigenvector of maximal rank among all vectors in $\operatorname{Ker}(M;\lambda_0)$ that do not belong to $\operatorname{span}\{y_{1,0},\dots,y_{k-1,0}\}$. So we can choose a vector $\widehat{y}_{k,0}$ in $\operatorname{Ker}(M;\lambda_0)$ outside $\operatorname{span}\{y_{1,0},\dots,y_{k-1,0}\}$ such that the rank of $\widehat{y}_{k,0}$ is larger than $\rho_k = r_k$. This allows us to construct a canonical system of Jordan chains of M at λ_0 with ranks $\nu_1 \ge \cdots \ge \nu_p$ such that $\nu_k > r_k$, contrary to the conclusion of the previous paragraph. □

As we have seen now, the numbers $r_1,\dots,r_p$ in a canonical system (8.5) are uniquely determined by M. They are called the *partial zero-multiplicities* of M at λ_0. Their sum $r_1 + \cdots + r_p$ is called the *zero-multiplicity* of M at λ_0. The next result provides a further motivation for this terminology.

Theorem 8.5. *There exist $m \times m$ matrix functions $\Phi(\lambda)$ and $E(\lambda)$, analytic at λ_0, such that $\Phi(\lambda_0)$ and $E(\lambda_0)$ are invertible while, for λ in a neighborhood of λ_0,*

$$M(\lambda)\Phi(\lambda) = E(\lambda)D(\lambda), \tag{8.7}$$

where $D(\lambda)$ is an $m \times m$ diagonal matrix given by

$$D(\lambda) = \operatorname{diag}\big((\lambda-\lambda_0)^{\kappa_1},(\lambda-\lambda_0)^{\kappa_2},\dots,(\lambda-\lambda_0)^{\kappa_m}\big) \tag{8.8}$$

with exponents $\kappa_1 \ge \kappa_2 \ge \cdots \ge \kappa_m$. These exponents are uniquely determined by M and do not depend on the particular choice of Φ and E in (8.7)*. Furthermore, λ_0 is a zero of M if and only if $\kappa_1 > 0$, and in that case the (strictly) positive exponents in $D(\lambda)$ are the partial zero-multiplicities of M at λ_0.*

We shall refer to the diagonal matrix function $D(\lambda)$ in (8.8) as the *local Smith-McMillan form* of M at λ_0.

Proof. We split the proof into three parts. In the first part we derive the identity (8.7). In the second part we prove the uniqueness of the exponents in (8.8). The third part concerns the final statement involving the strictly positive exponents.

Part 1. Let L be an $m \times m$ matrix function which is analytic at λ_0. Later we shall make a particular choice for L, namely $L(\lambda) = (\lambda - \lambda_0)^q M(\lambda)$, which is obviously analytic at λ_0, but for the time being L is arbitrary. Let

$$(x_{1,0}, \dots, x_{1,\ell_1-1}),\ (x_{2,0}, \dots, x_{2,\ell_2-1}),\ \dots,\ (x_{t,0}, \dots, x_{t,\ell_t-1}) \tag{8.9}$$

be a canonical system of Jordan chains for L at λ_0. We know that the vectors $x_{1,0}, \dots, x_{t,0}$ form a basis of $\operatorname{Ker}(L;\lambda_0) = \operatorname{Ker} L(\lambda_0)$, and hence we can choose vectors $x_{t+1,0}, \dots, x_{m,0}$ such that $x_{1,0}, \dots, x_{m,0}$ form a basis of $\mathbb{C}^m$. Write

$$\varphi_j(\lambda) = \begin{cases} x_{j,0} + (\lambda - \lambda_0)x_{j,1} + \cdots + (\lambda - \lambda_0)^{\ell_j - 1} x_{j,\ell_j - 1}, & j = 1, \dots, t, \\ x_{j,0}, & j = t+1, \dots, m, \end{cases}$$

and let $\Phi(\lambda)$ be the $m \times m$ matrix for which the jth column vector is equal to $\varphi_j(\lambda)$. Then Φ is analytic at λ_0, and $\Phi(\lambda_0)$ is invertible because the vectors $\varphi_j(\lambda_0) = x_{j,0}$, $j = 1, \dots, m$, form a basis of $\mathbb{C}^m$. From the definition of a Jordan chain it follows that

$$L(\lambda)\varphi_j(\lambda) = (\lambda - \lambda_0)^{\ell_j} e_j(\lambda), \tag{8.10}$$

where e_j is a $\mathbb{C}^m$-valued function which is analytic at λ_0. Here, in first instance, $j = 1, \dots, t$. For $j = t+1, \dots, m$, we take the index ℓ_j equal to zero, and hence we can use the equality (8.10) to define an $\mathbb{C}^m$-valued function e_j which is analytic at λ_0 and satisfies (8.10). Put $E(\lambda) = \begin{bmatrix} e_1(\lambda) & e_2(\lambda) & \cdots & e_m(\lambda) \end{bmatrix}$. Then

$$L(\lambda)\Phi(\lambda) = E(\lambda)\Delta(\lambda), \tag{8.11}$$

where $\Delta(\lambda)$ is the $m \times m$ diagonal matrix given by

$$\Delta(\lambda) = \operatorname{diag}\big((\lambda - \lambda_0)^{\ell_1}, \dots, (\lambda - \lambda_0)^{\ell_m}\big).$$

Obviously, E is analytic at λ_0. Let us prove that $E(\lambda_0)$ is invertible.

To the contrary, assume $E(\lambda_0)z = 0$ for a vector $z \neq 0$. Without loss of generality we may assume that, for appropriately chosen j,

$$z = (0, \dots, 0, 1, z_{j+1}, \dots, z_m)^\top.$$

Consider the function $\widetilde{\varphi}_j$ given by

$$\widetilde{\varphi}_j(\lambda) = \varphi_j(\lambda) + \sum_{i=j+1}^{m} (\lambda - \lambda_0)^{\ell_j - \ell_i} z_i \varphi_k(\lambda).$$

Note that the vector $\widetilde{\varphi}_j(\lambda_0)$ does not appear as a linear combination of the vectors $\varphi_1(\lambda_0), \dots, \varphi_{j-1}(\lambda_0)$. Furthermore,

$$\begin{aligned} L(\lambda)\widetilde{\varphi}_j(\lambda) &= L(\lambda)\varphi_j(\lambda) + \sum_{i=j+1}^{m} (\lambda-\lambda_0)^{\ell_j-\ell_i} z_i L(\lambda)\varphi_i(\lambda) \\ &= (\lambda-\lambda_0)^{\ell_j}\Big(e_j(\lambda) + \sum_{i=j+1}^{m} z_i e_i(\lambda)\Big) = (\lambda-\lambda_0)^{\ell_j} E(\lambda)z. \end{aligned}$$

Since $E(\lambda_0)z = 0$, it follows that $\widetilde{x}_{j,0} = \widetilde{\varphi}_j(\lambda_0)$ belongs to $\operatorname{Ker} L(\lambda_0)$. This implies $1 \le j \le t$. Indeed, if $j > t$, then the fact that $\widetilde{\varphi}_j(\lambda_0)$ does not belong to $\operatorname{span}\{\varphi_1(\lambda_0), \dots, \varphi_{j-1}(\lambda_0)\}$ implies that

$$\widetilde{x}_{j,0} \notin \operatorname{span}\{\varphi_1(\lambda_0), \dots, \varphi_t(\lambda_0)\} = \operatorname{span}\{x_{1,0}, \dots, x_{t,0}\} = \operatorname{Ker} L(\lambda_0).$$

Contradiction, and thus $1 \le j \le t$. Notice that $\widetilde{x}_{j,0}$ is an eigenvalue of L at λ_0 of rank at least $\ell_j + 1$. But this contradicts the choice of the vector $x_{j,0}$, which is of maximal rank ℓ_j among all vectors in $\operatorname{Ker} L(\lambda_0)$ that do not belong to $\operatorname{span}\{\varphi_1(\lambda_0), \dots, \varphi_{j-1}(\lambda_0)\}$. Thus $E(\lambda_0)$ is invertible.

From the identity (8.11) and the fact that $\Phi(\lambda_0)$ and $E(\lambda_0)$ are invertible we see that

$$\sum_{j=1}^{m} \ell_j = \text{ order of } \lambda_0 \text{ as a zero of } \det L(\lambda). \tag{8.12}$$

Now, we specialize to $L(\lambda) = (\lambda-\lambda_0)^q M(\lambda)$. Since $L(\lambda)$ is analytic at λ_0, we can apply the previous results for this choice of $L(\lambda)$. Put $D(\lambda) = (\lambda-\lambda_0)^{-q}\Delta(\lambda)$. Then (8.7) holds, the matrices $\Phi(\lambda)$ and $E(\lambda)$ have the desired properties, and $D(\lambda)$ is the diagonal $m \times m$ matrix given by (8.8) with $\kappa_j = \ell_j - q,\ j = 1, \dots, m$.

Part 2. Here we prove the uniqueness of the exponents $\kappa_1 \ge \kappa_2 \ge \cdots \ge \kappa_m$ in (8.8). Assume (8.7) is satisfied with Φ and E being analytic and invertible at λ_0. Put $\Gamma(\lambda) = (\lambda-\lambda_0)^\gamma M(\lambda)$, where γ is an integer such that $\rho_j = \kappa_j + \gamma > 0,\ j = 1, \dots, m$. Note that

$$\Gamma(\lambda)\Phi(\lambda) = E(\lambda)\Lambda(\lambda),$$

where $\Lambda(\lambda)$ is the diagonal $m \times m$ matrix

$$\Lambda(\lambda) = \operatorname{diag}\big((\lambda-\lambda_0)^{\rho_1}, (\lambda-\lambda_0)^{\rho_2}, \dots, (\lambda-\lambda_0)^{\rho_m}\big).$$

Since $\rho_1, \dots, \rho_m$ are all positive, Γ is analytic at λ_0 and $\operatorname{Ker}\Gamma(\lambda_0) = \mathbb{C}^m$. Let $\varphi_j(\lambda)$ and $e_j(\lambda)$ denote the jth columns of $\Phi(\lambda)$ and $E(\lambda)$, respectively. The functions φ_j and e_j are analytic at λ_0, and

$$\Gamma(\lambda)\varphi_j(\lambda) = (\lambda-\lambda_0)^{\rho_j} e_j(\lambda). \tag{8.13}$$

Consider the Taylor expansion

$$\varphi_j(\lambda) = \varphi_{j,0} + (\lambda-\lambda_0)\varphi_{j,1} + (\lambda-\lambda_0)^2\varphi_{j,2} + \cdots.$$

As $\Phi(\lambda_0)$ is a invertible, the vectors $\varphi_{1,0},\dots,\varphi_{m,0}$ form a basis of $\mathbb{C}^m = \operatorname{Ker}\Gamma(\lambda_0)$. From (8.13) it follows that

$$(\varphi_{1,0},\dots,\varphi_{1,\rho_1-1}),\ (\varphi_{2,0},\dots,\varphi_{2,\rho_2-1}),\ \dots,\ (\varphi_{n,0},\dots,\varphi_{n,\rho_m-1})$$

is a set of Jordan chains of Γ at λ_0. Let $\gamma_1,\dots,\gamma_r$ be the partial zero-multiplicities of Γ at λ_0. Since $\operatorname{Ker}\Gamma(\lambda_0) = \mathbb{C}^m$, we have $r = m$. Now, apply Proposition 8.4 to Γ at λ_0. It follows that $\rho_j \le \gamma_j$ for $j = 1,\dots,m$. By applying the results of Part 1 to Γ in place of L, we can use (8.12) to show that $\sum_{j=1}^m \gamma_j$ is equal to order of λ_0 as a zero of $\det\Gamma(\lambda)$. But $\Gamma(\lambda)\Phi(\lambda) = E(\lambda)\Lambda(\lambda)$ shows that the same holds true for $\sum_{j=1}^m \rho_j$. Thus $\sum_{j=1}^m \rho_j = \sum_{j=1}^m \gamma_j$. This can only happen when $\rho_j = \gamma_j$ for $j = 1,\dots,m$. We conclude that the numbers $\rho_1,\dots,\rho_m$ are uniquely determined by Γ, and hence the same is true for $\kappa_1,\dots,\kappa_m$.

Part 3. In this part we prove the final statement of the proposition. Since the exponents in (8.8) are uniquely determined by M, we may assume without loss of generality that $\kappa_j = \ell_j - q$, where ℓ_j is defined in Part 1 of the proof. Here $j = 1,\dots,m$ and, as in the last paragraph of Part 1, $L(\lambda) = (\lambda-\lambda_0)^q M(\lambda)$. Note that $x_0,\dots,x_{r-1}$ is a Jordan chain for M at λ_0 if and only if there exist vectors $x_r,\dots,x_{q+r-1}$ such that $x_0,\dots,x_{q+r-1}$ is a Jordan chain for L at λ_0 and, in that case, the ranks of x_0 as an eigenvector of L and as an eigenvector of M differ by q.

Assume λ_0 is a zero of M. Then there exist vectors $x_0,\dots,x_q$, $x_0 \ne 0$, such that (8.1) holds. But $x_0 \ne 0$ and (8.1) are equivalent to the statement that $(x_0,\dots,x_q)$ is a Jordan chain of L at λ_0 of length $q+1$. Since $(x_{1,0},\dots,x_{1,\ell_1-1})$ is a Jordan chain of L at λ_0 of maximal length, we have $\ell_1 \ge q+1$, and thus $\kappa_1 = \ell_1 - q \ge 1$. Therefore κ_1 is (strictly) positive provided that λ_0 is a zero of M.

Conversely, assume $\kappa_1 = \ell_1 - q \ge 1$. Then $\ell_1 \ge q+1$. This implies that the truncation $(x_{1,0},\dots,x_{1,q})$ of the chain $(x_{1,0},\dots,x_{1,\ell_1-1})$ is well defined and is a Jordan chain of L at λ_0 too. As we have seen in the previous paragraph, this implies that $x_{1,0}$ is a zero vector of M at λ_0, and hence λ_0 is a zero of M.

Again assume that λ_0 is a zero of M. Define

$$p = \max\{j = 1,\dots,m,\ \kappa_j = \ell_j - q > 0\}.$$

Then p does not exceed the number t in (8.9) for, by definition, $\ell_{t+1} = \cdots = \ell_m = 0$. Truncating the chains from (8.9) we obtain the following system of Jordan chains for M at λ_0:

$$(x_{1,0},\dots,x_{1,\kappa_1-1}),\ (x_{2,0},\dots,x_{2,\kappa_2-1}),\ \dots,\ (x_{p,0},\dots,x_{p,\kappa_p-1}). \tag{8.14}$$

The vectors $x_{1,0},\dots,x_{p,0} \in \operatorname{Ker}(M;\lambda_0)$ are linearly independent and have ranks $\kappa_1,\dots,\kappa_p$, respectively. Also, for $j = 1,\dots,p$, the vector $x_{j,0}$ has maximal rank among all eigenvectors in $\operatorname{Ker}(M;\lambda_0)$ that do not belong to the linear space $\operatorname{span}\{x_{1,0},\dots,x_{j-1,0}\}$. Suppose $x_{1,0},\dots,x_{p,0}$ is not a basis for $\operatorname{Ker}(M;\lambda_0)$ and take x in the space $\operatorname{Ker}(M;\lambda_0) \subset \operatorname{Ker} L(\lambda_0)$ but outside $\operatorname{span}\{x_{1,0},\dots,x_{p,0}\}$. Clearly the rank r of x as an eigenvector of L does not exceed ℓ_{p+1}, and so

$r - q \leq \ell_{p+1} - q \leq 0$. On the other hand $r - q$ is the rank of x as an eigenvector of M, hence a positive integer, and we have reached a contradiction. We conclude that $x_{1,\,0}, \ldots, x_{p,\,0}$ is a basis for $\operatorname{Ker}(M; \lambda_0)$, and so (8.14) is a canonical system of Jordan chains for M at λ_0. In particular $\kappa_1, \ldots, \kappa_p$ are the partial zero-multiplicities of M at λ_0. But these numbers $\kappa_1, \ldots, \kappa_p$ are precisely the strictly positive exponents in (8.8). □

Theorem 8.5 is a very useful one. To illustrate this, first note that the local Smith-McMillan form of M at λ_0 is equal to the local Smith-McMillan form at λ_0 of the transposed matrix function $M^\top$, $M^\top(\lambda) = M(\lambda)^\top$. To see this, assume (8.7) holds with $D(\lambda)$ given by (8.8) and with Φ and E being analytic and invertible at λ_0. Then in a sufficiently small open neighborhood of λ_0 we have

$$M^\top(\lambda)\Psi(\lambda) = F(\lambda)D(\lambda)^\top,$$

where $\Psi(\lambda) = \big(E(\lambda)^{-1}\big)^\top$ and $F(\lambda) = \big(\Phi(\lambda)^{-1}\big)^\top$. Since $D(\lambda)$ is a diagonal matrix, $D = D^\top$, and hence D is also the the local Smith-McMillan form of $M^\top$ at λ_0. Thus, by Theorem 8.5, if λ_0 is a zero of M it is also a zero of $M^\top$, and conversely. Moreover, in that case, the partial zero-multiplicities of λ_0 as a zero of M are the same as the partial zero-multiplicities of λ_0 as a zero of the transposed matrix function $M^\top$.

The information contained in the canonical system (8.5) can be put into a pair of matrices. In order to do this, let Q_i be the $m \times r_i$ matrix of which the jth column is equal to the column vector $x_{i,\,j-1}$. Thus $Q_i = \big[\, x_{i,\,0} \;\; x_{i,\,1} \;\cdots\; x_{i,\,r_k-1} \,\big]$. Put $Q = [Q_1\; Q_2\; \cdots\; Q_p]$. Further, let J be the block diagonal matrix

$$J = \operatorname{diag}\big(J_1,\, J_2,\, \ldots,\, J_p\big)$$

where J_k stands for the upper triangular $r_i \times r_i$ Jordan block with λ_0 on the main diagonal. Note that J is a Jordan matrix with one single eigenvalue, namely λ_0. The orders of its blocks are equal to the partial zero-multiplicities of M at λ_0. Hence the order of J is precisely equal to the zero-multiplicity of λ_0 as an eigenvalue of M. Furthermore,

$$\dim \operatorname{Ker}(\lambda_0 - J) = p = \dim \operatorname{Ker}(M; \lambda_0).$$

So the geometric multiplicity of λ_0 as a zero of M is equal to the geometric multiplicity of λ_0 as an eigenvalue of J.

The pair (Q, J) is called a *Jordan pair* of M at λ_0. The name Jordan pair will also be used for any pair of matrices which is obtained from (Q, J) by some permutation of the blocks J_k in J and the same permutation of the corresponding blocks in Q. Since the initial vectors $x_{1,\,0}, \ldots, x_{p,\,0}$ are linearly independent and each $J_k - \lambda_0$ is an upper triangular nilpotent $r_k \times r_k$ Jordan block, it is straightforward to show that

$$\bigcap_{j=0}^{\infty} \operatorname{Ker} Q(\lambda_0 - J)^j = \{0\}.$$

It follows that a Jordan pair (Q, J) is an observable pair, that is,

$$\bigcap_{i=0}^{\infty} \operatorname{Ker} QJ^i = \{0\}. \tag{8.15}$$

Next we define the notion of a dual pair. Assume λ_0 is a zero of M, and hence also of the transposed matrix functions $M^\top$. Let

$$(y_{1,0}, \dots, y_{1,r_1-1}),\ (y_{2,0}, \dots, y_{2,r_2-1}),\ \dots,\ (y_{p,0}, \dots, y_{p,r_p-1}) \tag{8.16}$$

be a canonical system of Jordan chains for $M^\top$ at λ_0. The fact that the numbers $p, r_1, \dots, r_p$ are the same numbers as those appearing is (8.5) is justified by the circumstance that the partial zero-multiplicities of M and $M^\top$ at λ_0 are the same (see the paragraph directly after the proof of Theorem 8.5) . For $i = 1, \dots, p$, let R_i be the $r_i \times m$ matrix of which the kth row is formed by the entries of the vector y_{i,r_i-k}. In other words,

$$R_j = \begin{bmatrix} y_{j,r_j-1} & y_{j,r_j-2} & \cdots & y_{j,0} \end{bmatrix}^\top.$$

As before, let J_i be the upper triangular $r_i \times r_i$ Jordan block with λ_0 on the main diagonal. Define

$$R = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_p \end{bmatrix}, \qquad J = \operatorname{diag}\big(J_1, J_2, \dots, J_p\big). \tag{8.17}$$

The pair (J, R) is called a *dual Jordan pair* of M at λ_0. This term will also be used for any pair of matrices which is obtained from (J, R) by some permutation of the blocks in (8.17). From the properties of a canonical system of Jordan chains, it follows that

$$\bigvee_{i=0}^{\infty} \operatorname{Im} J^i R = \mathbb{C}^r, \tag{8.18}$$

where $r = r_1 + \cdots + r_p$ is equal to the order of the matrix J. In other words the pair (J, R) is controllable.

Observe that a Jordan pair (Q, J) and a dual Jordan pair (J, R) have the same Jordan matrix J. We conclude that $(J, Q, R, 0; \mathbb{C}^r, \mathbb{C}^m)$ is a system, and (8.15) and (8.18) imply that this system is minimal.

Theorem 8.6. *Assume that λ_0 is a zero of M. Then given a Jordan pair (Q, J) of M at λ_0, there exists a dual Jordan pair (J, R) of M at λ_0 such that $Q(\lambda - J)^{-1}R$ is a minimal realization of the principal part of the Laurent expansion of $M(\lambda)^{-1}$ at λ_0.*

Proof. Without loss of generality we may assume the Jordan pair (Q, J) is determined by the canonical system of Jordan chains (8.5). Thus

$$Q = \begin{bmatrix} Q_1 & Q_2 & \cdots & Q_p \end{bmatrix}, \qquad J = \operatorname{diag}\big(J_1, J_2, \ldots, J_p\big),$$

where $Q_i = \begin{bmatrix} x_{i,0} & x_{k,1} & \cdots & x_{i,r_i-1} \end{bmatrix}$ and J_i is the upper triangular nilpotent $r_i \times r_i$ Jordan block.

Put $L(\lambda) = (\lambda - \lambda_0)^q M(\lambda)$. From the type of reasoning presented in Part 3 of the proof of Theorem 8.5, we see that L has a canonical system of Jordan chains at λ_0,

$$(x_{1,0}, \ldots, x_{1,\ell_1-1}),\ (x_{2,0}, \ldots, x_{2\ell_2-1}),\ \ldots,\ (x_{t,0}, \ldots, x_{t,\ell_t-1}), \tag{8.19}$$

such that, for $i = 1, \ldots, p$, the chain $(x_{i,0}, \ldots, x_{i,\ell_i-1})$ is a continuation with q vectors of the jth chain in (8.5). In particular,

$$\ell_i - q = r_i, \quad i = 1, \ldots, p, \qquad \ell_i - q \le 0, \quad i = p+1, \ldots, t. \tag{8.20}$$

Now, using the chains in (8.19), we construct, as in Part 1 of the proof of Theorem 8.5, matrix functions $\Phi(\lambda)$ and $E(\lambda)$, analytic and invertible at λ_0, such that

$$M(\lambda)\Phi(\lambda) = E(\lambda)D(\lambda),$$

where $D(\lambda)$ is the local Smith-McMillan form of M at λ_0. Put $\Theta(\lambda) = E(\lambda)^{-1}$, and write

$$\Theta(\lambda) = \begin{bmatrix} \theta_1(\lambda) \\ \theta_2(\lambda) \\ \vdots \\ \theta_m(\lambda) \end{bmatrix}, \qquad \Phi(\lambda) = \begin{bmatrix} \varphi_1(\lambda) & \varphi_2(\lambda) & \cdots & \varphi_m(\lambda) \end{bmatrix}.$$

It follows that

$$M(\lambda)^{-1} = \Phi(\lambda)D(\lambda)^{-1}\Theta(\lambda) = \sum_{i=1}^{m} (\lambda - \lambda_0)^{q-\ell_i} \varphi_i(\lambda)\theta_i(\lambda).$$

Using (8.20), we conclude that the Laurent principal part $P(\lambda)$ of $M(\lambda)^{-1}$ at λ_0 is given by $P(\lambda) = \sum_{i=1}^{p} P_i(\lambda)$, where $P_k(\lambda)$ is the Laurent principal part of $(\lambda - \lambda_0)^{-r_i}\varphi_i(\lambda)\theta_k(\lambda)$. Thus in matrix form $P(\lambda)$ is given by the following expression:

$$P(\lambda) = \sum_{k=1}^{p} \begin{bmatrix} \varphi_{k,0} & \varphi_{k,1} & \cdots & \varphi_{k,r_k-1} \end{bmatrix} \big(\lambda - J_k\big)^{-1} \begin{bmatrix} \theta_{k,r_k-1} \\ \theta_{k,r_k-2} \\ \vdots \\ \theta_{k,0} \end{bmatrix}. \tag{8.21}$$

Here, for $i = 1, \dots, p$, the matrix J_i is the upper triangular $r_i \times r_i$ Jordan block with λ_0 on the main diagonal, and the vectors $\varphi_{i,\,j}$ and $\theta_{i,\,j}$ are the jth coefficients in the Taylor expansions of $\varphi_i(\lambda)$ and $\theta_i(\lambda)$ at λ_0, respectively.

Next, observe that $M^\top(\lambda)\Psi(\lambda) = F(\lambda)D(\lambda)$, where

$$\Psi(\lambda) = \big(E(\lambda)^{-1}\big)^\top = \Theta(\lambda)^\top, \qquad F(\lambda) = \big(\Phi(\lambda)^{-1}\big)^\top.$$

Let $\psi_j(\lambda)$ be the jth column of $\Psi(\lambda)$, and consider the Taylor expansion

$$\psi_j(\lambda) = y_{j,\,0} + (\lambda - \lambda_0)y_{j,\,1} + (\lambda - \lambda_0)^2 y_{j,\,2} + \cdots .$$

Then the chains

$$(y_{1,\,0}, \dots, y_{1,\,r_1-1}),\ (y_{2,\,0}, \dots, y_{2,\,r_2-1}),\ \dots,\ (y_{p,\,0}, \dots, y_{p,\,r_p-1})$$

form a set of Jordan chains for $M^\top$ at λ_0. Since $\Psi(\lambda_0)$ is invertible, the vectors $y_{1,\,0}, \dots, y_{p,\,0}$ are linearly independent. Recall that $r_1, \dots, r_p$ are the partial zero-multiplicities of M and of $M^\top$ at λ_0. It follows that the above set of chains is actually a canonical system of Jordan chains for $M^\top$ at λ_0. Now put,

$$R_j = \begin{bmatrix} y_{j,\,r_j-1}^\top \\ y_{j,\,r_j-2}^\top \\ \vdots \\ y_{j,\,0}^\top \end{bmatrix}, \quad j = 1, \dots, p, \qquad R = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_p \end{bmatrix}$$

Then, by definition, the pair (J, R) is a dual Jordan pair of M at λ_0.

It remains to show that $P(\lambda) = Q(\lambda - J)^{-1}R$. To do this we first observe that

$$Q(\lambda - J)^{-1}R = \sum_{i=1}^{p} Q_i(\lambda - J_i)^{-1}R_i.$$

From the construction of Φ in Part 1 of the proof of Theorem 8.5 we know that for $i = 1, \dots, p$ and $j = 0, \dots, r_i$ the vector $x_{i,\,j}$ is equal to the jth coefficient in the Taylor expansion of φ_i at λ_0. It follows that

$$Q_i = \begin{bmatrix} \varphi_{i,\,0} & \varphi_{i,\,1} & \cdots & \varphi_{i,\,r_i-1} \end{bmatrix}, \qquad i = 1, \dots, p\,.$$

Next, recall that $\Theta(\lambda) = \Psi(\lambda)^\top$. Thus $\theta_j(\lambda) = \psi_j(\lambda)^\top$ for $j = 1, \dots, m$. It follows that, for $j = 1, \dots, p$ and $k = 0, \dots, r_j$, the vector $y_{j,\,k}^\top$ is equal to the kth coefficient in the Taylor expansion of θ_j at λ_0. But then we see that

$$R_j = \begin{bmatrix} \theta_{j,\,r_j-1} \\ \theta_{j,\,r_j-2} \\ \vdots \\ \theta_{j,\,0} \end{bmatrix}, \qquad j = 1, \dots, p\,.$$

Using the above expressions for Q_i and R_i in (8.21) yields

$$P(\lambda) = \sum_{i=1}^{p} Q_i\big(\lambda - J_i\big)^{-1} R_i = Q(\lambda - J)^{-1} R,$$

which completes the proof. □

Let (Q, J) be a Jordan pair and (J, R) a dual Jordan pair of M at λ_0 such that $Q(\lambda - J)^{-1}R$ is equal to the Laurent principal part of $M(\lambda)^{-1}$ at λ_0. Then we refer to the triple (Q, J, R) as a *canonical Jordan triple* of $M(\lambda)^{-1}$ at λ_0. The previous theorem shows that a canonical Jordan triple always exists.

So far λ_0 has been a point in the finite complex plane. We conclude this section by considering the case when $\lambda_0 = \infty$. Thus let M be an $m \times m$ matrix function which is meromorphic on a connected open subset of the Riemann sphere $\mathbb{C} \cup \infty$. In that case M has an expansion at ∞ of the form

$$M(\lambda) = \sum_{j=-\infty}^{q} \lambda^j M_j$$

where q is a non-negative integer. As before, it is assumed that $M(\lambda) \not\equiv 0$. We call $\lambda_0 = \infty$ a *zero* or *eigenvalue* of M if there exist vectors $x_0, \ldots, x_q$ in $\mathbb{C}^m$, $x_0 \neq 0$, such that

$$M_q x_j + \cdots + M_{q-j} x_0 = 0, \qquad j = 0, \ldots, q.$$

In that case the vector x_0 is called an *eigenvector* (or *root vector*) of M at the eigenvalue ∞. Clearly, $\lambda_0 = \infty$ is a zero of M if and only if the origin is an eigenvalue of the function $M^\sharp$ defined by

$$M^\sharp(\lambda) = M(\lambda^{-1}). \tag{8.22}$$

This fact allows us (without any further explanation) to introduce for the point $\lambda_0 = \infty$ all notions defined above for a finite eigenvalue. For example, we define the *partial zero-multiplicities* of M at $\lambda_0 = \infty$ to be equal to the partial zero-multiplicities of $M^\sharp$ at the point 0. Similarly, a triple $(Q_\infty, J_\infty, R_\infty)$ is called a *canonical Jordan triple* of M at $\lambda_0 = \infty$ if $(Q_\infty, J_\infty, R_\infty)$ is a canonical Jordan triple of $M^\sharp$ at the point 0. Observe that in that case J_∞ is a nilpotent matrix. Furthermore we have the following corollary.

Corollary 8.7. *Let $\lambda_0 = \infty$ be a zero of M at ∞, and let $(Q_\infty, J_\infty, R_\infty)$ be a corresponding canonical Jordan triple of M at ∞. Then the system*

$$(J_\infty, Q_\infty, R_\infty, 0; \mathbb{C}^{\rho_\infty}, \mathbb{C}^m)$$

is minimal and $\lambda Q_\infty (I - \lambda J_\infty)^{-1} R_\infty$ is equal to the principal part of $M(\lambda)^{-1}$ at ∞, that is, $M(\lambda)^{-1} - \lambda Q_\infty (I - \lambda J_\infty)^{-1} R_\infty$ is analytic at ∞.

8.2 Pole data

As in the previous section, M is an $m\times m$ matrix function which is meromorphic on a connected open set Ω, $\det M(\lambda) \not\equiv 0$, and λ_0 is a point Ω. Thus in a neighborhood of λ_0 the function M has the following expansion

$$M(\lambda) = \sum_{j=-q}^{\infty} (\lambda - \lambda_0)^j A_j.$$

As before, the definitions given below do not depend on the choice of q which is again assumed to be a non-negative integer.

A nonzero vector $x \in \mathbb{C}^m$ is called a *pole-vector* of M at λ_0 if there exist vectors $\varphi_1, \dots, \varphi_q$ in $\mathbb{C}^m$ such that

$$\begin{bmatrix} A_{-1} & A_{-2} & \cdots & A_{-q} \\ A_{-2} & & \iddots & 0 \\ \vdots & \iddots & \iddots & \vdots \\ A_{-q} & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_q \end{bmatrix} = \begin{bmatrix} x \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{8.23}$$

The linear space consisting of all pole-vectors of M at λ_0 together with the zero vector will be denoted by $\mathrm{Pol}\,(M;\lambda_0)$. Note that there exists a pole-vector of M at λ_0 if and only if M has a pole at λ_0, that is, at least one of the coefficients $A_{-q}, \dots, A_{-1}$ is nonzero.

The pointwise inverse M^{-1} of M is given by $M^{-1}(\lambda) = M(\lambda)^{-1}$. From Cramer's rule for inverting a matrix it is clear that M^{-1} is meromorphic on Ω.

Lemma 8.8. *The vector x is a pole-vector of M at λ_0 if and only if x is an eigenvector of M^{-1} at λ_0. In other words,*

$$\mathrm{Pol}\,(M;\lambda_0) = \mathrm{Ker}\,(M^{-1};\lambda_0).$$

Proof. Assume $x \in \mathbb{C}^m$ is a pole-vector of M at λ_0. Put

$$\varphi(\lambda) = (\lambda - \lambda_0)\varphi_1 + \cdots + (\lambda - \lambda_0)^q \varphi_q,$$

where $\varphi_1, \dots, \varphi_q$ are as in (8.23), and set $\psi(\lambda) = M(\lambda)\varphi(\lambda)$. Consider the Taylor expansion of ψ at λ_0:

$$\psi(\lambda) = \psi_0 + (\lambda - \lambda_0)\psi_1 + (\lambda - \lambda_0)\psi_2 + \cdots.$$

From (8.23) we see that $\psi_0 = x$. Furthermore, we have

$$M^{-1}(\lambda)\psi(\lambda) = M(\lambda)^{-1}M(\lambda)\varphi(\lambda) = \varphi(\lambda).$$

Write

$$M^{-1}(\lambda) = \sum_{j=-\widetilde{q}}^{\infty} \widetilde{A}_j(\lambda - \lambda_0)^j,$$

where $\widetilde{q} \geq 0$. The identity $M^{-1}(\lambda)\psi(\lambda) = \varphi(\lambda)$ implies that

$$\widetilde{A}_{-\widetilde{q}}\psi_j + \cdots + \widetilde{A}_{-\widetilde{q}+j}\psi_0 = 0, \qquad j = 0, \ldots, \widetilde{q}.$$

Thus $x = \psi_0$ is an eigenvector of M^{-1} at λ_0.

Conversely, assume x is an eigenvector of M^{-1} at λ_0. Then we can find vectors $x_0, \ldots, x_{\widetilde{q}}$ in $\mathbb{C}^m$, $x_0 = x$, such that

$$\widetilde{A}_{-\widetilde{q}}x_j + \cdots + \widetilde{A}_{-\widetilde{q}+j}x_0 = 0, \qquad j = 0, \ldots, \widetilde{q}.$$

Put $\psi(\lambda) = x_0 + (\lambda - \lambda_0)x_1 + \cdots + (\lambda - \lambda_0)^{\widetilde{q}}x_{\widetilde{q}}$. Then $M^{-1}(\lambda)\psi(\lambda)$ is analytic at λ_0 and its Taylor expansion at λ_0 is of the form

$$M^{-1}(\lambda)\psi(\lambda) = (\lambda - \lambda_0)y_1 + (\lambda - \lambda_0)^2 y_2 + \cdots .$$

It follows that

$$M(\lambda)\Big(\sum_{j=1}^{\infty}(\lambda - \lambda_0)^j y_j\Big) = x_0 + (\lambda - \lambda_0)x_1 + \cdots .$$

Hence (8.23) holds with $x = x_0$ and $\varphi_j = y_j$, $j = 1, \ldots, q$. Thus x is a pole-vector of M at λ_0. □

By applying the above lemma to M^{-1} in place of M we also see that λ_0 is a zero of M if and only if λ_0 is a pole of M^{-1}.

The dimension of the space $\mathrm{Pol}\,(M; \lambda_0)$ is called the *geometric multiplicity of λ_0 as a pole of M*. By definition the *rank* of a pole-vector x of M at λ_0 is the rank of x as an eigenvector of M^{-1} at λ_0. Similarly, the *partial pole-multiplicities* of M at λ_0 are by definition equal to the partial zero-multiplicities of M^{-1} at λ_0, and their sum is called the *pole-multiplicity* of M at λ_0. Using Lemma 8.8 and the above definitions, the following addition to Theorem 8.5 is immediate.

Proposition 8.9. *Let $D(\lambda) = \mathrm{diag}\,\big((\lambda-\lambda_0)^{\kappa_1}, (\lambda-\lambda_0)^{\kappa_2}, \ldots, (\lambda-\lambda_0)^{\kappa_n}\big)$, $\kappa_1 \geq \kappa_2 \geq \cdots \geq \kappa_n$, be the local Smith-McMillan form of M at λ_0. Then λ_0 is a pole of M if and only if $\kappa_n < 0$, and in that case the absolute values of the strictly negative exponents in $D(\lambda)$ are the partial pole-multiplicities of M at λ_0. In particular, the order of λ_0 as a pole of M is equal to the largest partial pole-multiplicity of M at λ_0.*

Corollary 8.10. *The order of λ_0 as a pole of M is equal to the pole-multiplicity of M at λ_0 if and only if the geometric multiplicity of λ_0 as a pole of M is equal to one.*

Proof. Let $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_r > 0$ be the partial pole-multiplicities of M at λ_0. Then

$$\begin{aligned} \pi_1 &= \text{the order of } \lambda_0 \text{ as a pole of } M, \\ \textstyle\sum_{j=1}^{r} \pi_j &= \text{the pole-multiplicity of } M \text{ at } \lambda_0, \\ r &= \text{the geometric multiplicity of } \lambda_0 \text{ as a pole of } M. \end{aligned}$$

Since each π_j is strictly positive, we see that the order of λ_0 as a pole of M is equal to the pole-multiplicity of M at λ_0 if and only if $r = 1$, that is, if and only if the geometric multiplicity of λ_0 as a pole of M is one. □

The next result, which is a supplement to Theorem 8.6, shows that the pole-multiplicity is equal to the rank of the block matrix in the left-hand side of (8.23).

Proposition 8.11. *Let λ_0 be a pole of M, and let (Q, J, R) be a canonical Jordan triple of M^{-1} at λ_0. Then $Q(\lambda - J)^{-1}R$ is equal to the principal part of the Laurent expansion of M at λ_0, and*

$$\operatorname{rank} \begin{bmatrix} A_{-1} & A_{-2} & \cdots & A_{-q} \\ A_{-2} & & \iddots & 0 \\ \vdots & \iddots & \iddots & \vdots \\ A_{-q} & 0 & \cdots & 0 \end{bmatrix} = \text{pole-multiplicity of } M \text{ at } \lambda_0. \tag{8.24}$$

Proof. The first statement is immediate from the definition of canonical Jordan triple. It implies that the principal part of the the Laurent expansion of M at λ_0 is given by

$$\sum_{i=1}^{\infty} \frac{1}{(\lambda - \lambda_0)^i} Q(J - \lambda_0)^{i-1} R.$$

Hence $Q(J - \lambda_0)^{i-1}R = A_{-i}$ for $i = 1, \ldots, q$, while $Q(J - \lambda_0)^{i-1}R = 0$ for $i = q + 1, q + 2, \ldots$. Write

$$\Omega_1(s) = \operatorname{col}\left(Q(J - \lambda_0)^{i-1}\right)_{i=1}^{s}, \qquad \Omega_2(s) = \operatorname{row}\left(J - \lambda_0)^{i-1}R\right)_{i=1}^{s},$$

where $s = q, q + 1, q + 2, \ldots$. Then the block matrix in the left-hand side of (8.24) can be written as $\Omega_1(q)\Omega_2(q)$. Thus the left-hand side of (8.24) is equal to the rank of the product $\Omega_1(q)\Omega_2(q)$. For $s > q$, the matrix $\Omega_1(s)\Omega_2(s)$ is obtained from $\Omega_1(q)\Omega_2(q)$ by adding $s - q$ zero columns and zero rows. Since these operations do not affect the rank, the left-hand side of (8.24) is equal to rank $\big(\Omega_1(s)\Omega_2(s)\big)$, $s > q$. From (8.15) and (8.18) we see that, for s sufficiently large, $\Omega_1(s)$ is injective and $\Omega_2(s)$ is surjective. Therefore the left-hand side of (8.24) is equal to the order of J, which in turn is equal to the pole-multiplicity of M at λ_0. □

We conclude this section by defining the various pole notions for the case when $\lambda_0 = \infty$. To do this we use again (8.22). So, for example, a nonzero vector x

in $\mathbb{C}^m$ is called a *pole-vector* of M at ∞ if x is a pole vector for $M^\sharp$ at the point 0. Similarly, the *partial pole-multiplicities* of M at ∞ are by definition equal to the partial pole-multiplicities of $M^\sharp$ at 0. In the same way one can define the other notions too.

8.3 Minimal realizations in terms of zero or pole data

In this section W is a rational $m \times m$ matrix function which is assumed to be regular, i.e., $\det W(\lambda) \not\equiv 0$. We apply the results of the preceding two sections to obtain minimal realizations of W^{-1} in terms of the zero data of W, and of W in terms of the pole data. The following results are the main theorems of this section.

Theorem 8.12. *Let W be a regular rational $m \times m$ matrix function. Let $\lambda_1, \dots, \lambda_k$ be the finite zeros of W, let $\rho_1, \dots, \rho_k$ be the corresponding zero-multiplicities and, for $i = 1, \dots, k$, let (Q_i, J_i, R_i) be a canonical Jordan triple of W at λ_i. Put $r = \rho_1 + \cdots + \rho_k$, and set*

$$Q = \operatorname{row}(Q_i)_{i=1}^k, \qquad J = \operatorname{diag}\big(J_1, \dots, J_k\big), \qquad R = \operatorname{col}(R_i)_{i=1}^k.$$

Furthermore, let $(Q_\infty, J_\infty, R_\infty)$ be a canonical Jordan triple of W at ∞. Then

$$W(\lambda)^{-1} = Q(\lambda - J)^{-1}R + D + \lambda Q_\infty (I - \lambda J_\infty)^{-1} R_\infty$$

for a suitable choice of D. Moreover, if W^{-1} is proper, then $D = W^{-1}(\infty)$, and the system $\Theta = (J, R, Q, D; \mathbb{C}^r, \mathbb{C}^m)$ is a minimal realization of W^{-1}.

Theorem 8.13. *Let W be a regular rational $m \times m$ matrix function. Let $\lambda_1, \dots, \lambda_\ell$ be the finite poles of W, let $\sigma_1, \dots, \sigma_\ell$ be the corresponding pole-multiplicities and, for $i = 1, \dots, \ell$, let (Q_i, J_i, R_i) be a canonical Jordan triple of W^{-1} at λ_i. Put $s = \sigma_1 + \cdots + \sigma_\ell$, and set*

$$Q = \operatorname{row}(Q_i)_{i=1}^\ell, \qquad J = \operatorname{diag}(J_1, \dots, J_\ell), \qquad R = \operatorname{col}(R_i)_{i=1}^\ell.$$

Furthermore, let $(Q_\infty, J_\infty, R_\infty)$ be a canonical Jordan triple of W^{-1} at ∞. Then

$$W(\lambda) = Q(\lambda - J)^{-1}R + D + \lambda Q_\infty (I - \lambda J_\infty)^{-1} R_\infty$$

for a suitable choice of D. Moreover, if W is proper, then $D = W(\infty)$ and the system $\Theta = (J, R, Q, D; \mathbb{C}^s, \mathbb{C}^m)$ is a minimal realization of W.

It suffices to prove Theorem 8.12. Indeed, one obtains Theorem 8.13 by applying Theorem 8.12 to W^{-1} in place of W.

Proof of Theorem 8.12. Observe that

$$Q(\lambda - J)^{-1}R = \sum_{i=1}^k Q_i(\lambda - J_i)^{-1}R_i. \tag{8.25}$$

Theorem 8.6 gives that for each $i \in \{1, \dots, k\}$ the function $Q_i(\lambda - J_i)^{-1}R_i$ is a minimal realization of the principal part of the Laurent expansion of $W(\lambda)^{-1}$ at λ_i. In particular, $W(\lambda)^{-1} - Q(\lambda - J)^{-1}R$ has no poles in the complex plane $\mathbb{C}$. According to Corollary 8.7, the function $\lambda Q_\infty(I - \lambda J_\infty)^{-1}R_\infty$ is equal to the principal part of $W(\lambda)^{-1}$ at ∞. Since J_∞ is a nilpotent matrix, the function $\lambda Q_\infty(I - \lambda J_\infty)^{-1}R_\infty$ is analytic on $\mathbb{C}$. We conclude that the function

$$W(\lambda)^{-1} - Q(\lambda - J)^{-1}R - \lambda Q_\infty(I - \lambda J_\infty)^{-1}R_\infty$$

is analytic on $\mathbb{C}$ and at ∞. But then Liouville's theorem implies that this function must be identically equal to a constant matrix, D say. Thus $W(\lambda)^{-1}$ has the desired form. If $W(\lambda)^{-1}$ is proper, the term $\lambda Q_\infty(I - \lambda J_\infty)^{-1}R_\infty$ is missing, and we obtain

$$W(\lambda)^{-1} = D + Q(\lambda - J)^{-1}R.$$

It remains to show that $\Theta = (J, Q, R, D; \mathbb{C}^r, \mathbb{C}^m)$ is minimal. Put

$$\mathcal{N} = \bigcap_{i=0}^{\infty} \operatorname{Ker} QJ^i.$$

We need to show that $\mathcal{N} = \{0\}$. Assume not. Since $\mathcal{N}$ is invariant under J, there exist a nonzero $x \in \mathcal{N}$ and a complex number μ such that $Jx = \mu x$. Clearly μ is an eigenvalue of J. Recall that $J = \operatorname{diag}(J_1, \dots, J_k)$ where, for $i = 1, \dots, k$, the matrix J_i is an upper triangular Jordan matrix which has λ_i as its only eigenvalue. We conclude that $\mu = \lambda_t$ for some $t \in \{1, \dots, k\}$. Write x as a sum $x = x_1 + \cdots + x_m$ corresponding to the partitioning $J = \operatorname{diag}(J_1, \dots, J_k)$ of J. Then, as $\lambda_1, \dots, \lambda_k$ are distinct, $Jx = \lambda_t x$ implies that $x_k = 0$ for $k \neq t$. Thus $x = x_t$, and $QJ^i x = Q_t J_t^i x_t$. But $QJ^i x = 0$, and hence $Q_t J_t^i x_t = 0$, $i = 0, 1, \dots$. Now recall that the system (J_t, Q_t, R_t) is minimal. It follows that $x = x_t = 0$.

We have established that $\bigcap_{i=0}^{\infty} \operatorname{Ker} QJ^i = \{0\}$. In a similar way one can show that $\bigcap_{i=0}^{\infty} \operatorname{Ker} R^\top (J^\top)^i = \{0\}$ too. Thus the system $(J, Q, R, D; \mathbb{C}^r, \mathbb{C}^m)$ is minimal. □

Combining the above results with those about minimal realizations in Section 7.3 we obtain the following corollary.

Corollary 8.14. *Let W be the transfer function of the minimal finite-dimensional system $\Theta = (A, B, C, D; X, Y)$. Suppose D is invertible, and let $\lambda_0 \in \mathbb{C}$. Then λ_0 is an eigenvalue of A if and only if λ_0 is a pole of W and the partial multiplicities of λ_0 as an eigenvalue of A are the same as the partial pole-multiplicities of W at λ_0. Also, λ_0 is an eigenvalue of $A^\times = A - BD^{-1}C$ if and only if λ_0 is a zero of W and the partial multiplicities of λ_0 as an eigenvalue of $A^\times$ are the same as the partial zero-multiplicities of W at λ_0.*

Proof. We apply Theorem 8.13. Since two minimal realizations of W are similar, the system Θ is similar to the system $(J, Q, R, D; \mathbb{C}^\delta, \mathbb{C}^m)$ constructed in Theorem

8.13. In particular, A and J are similar, and hence A and J have the same eigenvalues with the same partial multiplicities. But then the first part of the corollary (about A) is an immediate consequence of the construction of the triple (Q, J, R).

To prove the second part of the theorem, notice that the system $\Theta^\times$ is minimal too. The transfer function of $\Theta^\times$ coincides with W^{-1}. Now apply the first part of the theorem to W^{-1} and $\Theta^\times$. □

The preceding corollary can be used to prove the following addition to Theorem 2.7.

Theorem 8.15. *Let W be a proper rational $m \times m$ matrix function such that $W(\infty) = I$. Assume that W has simple poles only. Then W admits a factorization of the following form*

$$W(\lambda) = \Big(I + \frac{1}{\lambda - \lambda_1} R_1\Big) \cdots \Big(I + \frac{1}{\lambda - \lambda_n} R_n\Big),$$

where $R_1, \ldots, R_n$ are rank one $m \times m$ matrices and n is the state space dimension of a minimal realization of W.

Here we use the convention that a pole of W is said to be *simple* if it is of order one.

Proof. Since W is proper and $W(\infty) = I$, we can choose a unital minimal realization $\Theta = (A, B, C; \mathbb{C}^n, \mathbb{C}^m)$ for W. Note that the state space dimension is n; so finite in particular. As W has first-order poles only, we see from Corollary 8.14 that for each eigenvalue of A the algebraic multiplicity is equal to the geometric multiplicity. It follows that the Jordan matrix for A is diagonal, and hence A is diagonalizable. So we can apply Theorem 2.7 to get the desired result. □

The factorization of W constructed in the proof of Theorem 8.15 is a minimal factorization in the sense of the first section of the next chapter (Section 9.1). This implies that the points $\lambda_1, \ldots, \lambda_n$ are precisely the poles of W counted according to the pole-multiplicity (see Section 9.1). With minor modifications Theorem 8.15 can be extended to the case where $\det W(\lambda)$ does not vanish identically and W has a simple pole at ∞.

For further information of the type of factorizations appearing in Theorem 8.15 we refer to Chapter 10; in particular, see Section 10.3.

8.4 Local degree and local minimality

Let W be a rational $m \times m$ matrix function, and let $\lambda_0 \in \mathbb{C}$. In a deleted neighborhood of λ_0 we have the following expansion

$$W(\lambda) = \sum_{j=-q}^{\infty} (\lambda - \lambda_0)^j W_j. \tag{8.26}$$

Here q is some positive integer. By the *local degree* of W at λ_0 we mean the number $\delta(W;\lambda_0) = \operatorname{rank}\Omega$, where Ω is the block Hankel matrix

$$\Omega = \begin{bmatrix} W_{-1} & W_{-2} & \cdots & W_{-q} \\ W_{-2} & & \iddots & 0 \\ \vdots & \iddots & \iddots & \vdots \\ W_{-q} & 0 & \cdots & 0 \end{bmatrix}. \tag{8.27}$$

Of course this definition does not depend on the choice of q. We also introduce $\delta(W;\infty)$ by putting $\delta(W;\infty) = \delta(W^\sharp;0)$, where $W^\sharp(\lambda) = W(\lambda^{-1})$. Observe that W is analytic at a point μ in the Riemann sphere $\mathbb{C}\cup\{\infty\}$ if and only if $\delta(W;\mu) = 0$. If $\det W(\lambda) \not\equiv 0$, then $\delta(W;\mu)$ is just the pole-multiplicity of W at μ as defined in Section 8.2.

The local degree enjoys a sublogarithmic property. To see this, let W_1 and W_2 be rational $m \times m$ matrix functions, suppose $W = W_1W_2$, and take $\lambda_0 \in \mathbb{C}^m$. Write

$$W_k(\lambda) = \sum_{j=-p}^{\infty} (\lambda - \lambda_0)^j W_j^{(k)}, \qquad k = 1, 2,$$

for some positive integer p. Then W admits an expansion of the form (8.26) with $q = 2p$. Although the definition of the local degree has been given in terms of block Hankel matrices, it is now convenient to change to block Toeplitz matrices. So we introduce

$$\widetilde{\Omega} = \begin{bmatrix} W_{-q} & \cdots & W_{-2} & W_{-1} \\ 0 & W_{-q} & \cdots & W_{-2} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & W_{-q} \end{bmatrix},$$

$$\widetilde{\Omega}_k = \begin{bmatrix} W_{-p}^{(k)} & \cdots & W_{-2}^{(k)} & W_{-1}^{(k)} \\ 0 & W_{-p}^{(k)} & \cdots & W_{-2}^{(k)} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & W_{-p}^{(k)} \end{bmatrix}, \qquad k = 1, 2.$$

Then $\delta(W;\lambda_0) = \operatorname{rank}\widetilde{\Omega}$ and $\delta(W_k;\lambda_0) = \operatorname{rank}\widetilde{\Omega}_k$, $k = 1, 2$. Observe that $\widetilde{\Omega}$ can be written as

$$\widetilde{\Omega} = \begin{bmatrix} \widetilde{\Omega}_1 & * \\ 0 & \widetilde{\Omega}_1 \end{bmatrix} \begin{bmatrix} \widetilde{\Omega}_2 & * \\ 0 & \widetilde{\Omega}_2 \end{bmatrix} = \begin{bmatrix} \widetilde{\Omega}_1 \\ 0 \end{bmatrix} \begin{bmatrix} \widetilde{\Omega}_2 & * \end{bmatrix} + \begin{bmatrix} * \\ \widetilde{\Omega}_1 \end{bmatrix} \begin{bmatrix} 0 & \widetilde{\Omega}_2 \end{bmatrix},$$

where the $*$'s denote matrices that we do not need to specify explicitly. It follows that $\operatorname{rank}\widetilde{\Omega} \leq \operatorname{rank}\widetilde{\Omega}_1 + \operatorname{rank}\widetilde{\Omega}_2$. In other words

$$\delta(W_1W_2;\lambda_0) \;\leq\; \delta(W_1;\lambda_0) + \delta(W_2;\lambda_0). \tag{8.28}$$

Obviously, we also have $\delta(W_1W_2;\infty) \leq \delta(W_1;\infty) + \delta(W_2;\infty)$.

The definitions and statements of the preceding two paragraphs also apply to rational functions of which the values are operators on an arbitrary finite-dimensional space Y. Indeed, if $\dim Y = m$, such functions can be identified with rational $m \times m$ matrix functions.

The next result shows that the difference between the left- and right-hand side in (8.28) is the same for the product $W = W_1W_2$ and for the product $W^{-1} = W_2^{-1}W_1^{-1}$. In fact, the result holds for products of two or more factors.

Theorem 8.16. *Let $W_1, \dots, W_k$ and W be proper rational $m \times m$ matrix functions, all having the value I_m at infinity, and suppose that $W(\lambda) = W_1(\lambda)\cdots W_k(\lambda)$. Then, for each $\alpha \in \mathbb{C}$,*

$$\sum_{j=1}^{k} \delta(W_j;\alpha) - \delta(W;\alpha) \;=\; \sum_{j=1}^{k} \delta(W_j^{-1};\alpha) - \delta(W^{-1};\alpha). \tag{8.29}$$

Roughly speaking, this theorem says the following. The poles of W (pole-multiplicities counted) are among the poles of $W_1, \dots, W_k$, the zeros of W (zero-multiplicities counted) are among the zeros of $W_1, \dots, W_k$, and the additional poles in the factorization $W = W_1 \cdots W_k$ coincide with the additional zeros (again the appropriate multiplicities counted).

Proof. For $j = 1, \dots, k$, let $W_j(\lambda) = I_m + C_j(\lambda - A_j)^{-1}B_j$ be minimal realization of W_j. Define the matrices A, B and C by

$$A = \begin{bmatrix} A_1 & B_1C_2 & \cdots & B_1C_k \\ 0 & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & B_{k-1}C_k \\ 0 & \cdots & 0 & A_k \end{bmatrix}, \tag{8.30}$$

$$B = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_k \end{bmatrix}, \qquad C = \begin{bmatrix} C_1 & C_2 & \cdots & C_k \end{bmatrix}.$$

Then, as can be seen by a repeated application of Theorem 2.2,

$$W(\lambda) = I_m + C(\lambda - A)^{-1}B$$

is a realization of W. The matrix A is block upper triangular while $A^\times = A - BC$, being of the form

$$A^\times = \begin{bmatrix} A_1^\times & 0 & \cdots & 0 \\ -B_2C_1 & A_2^\times & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ -B_kC_1 & \cdots & -B_kC_{k-1} & A_n^\times \end{bmatrix}, \tag{8.31}$$

is block lower triangular.

As an intermediate step, it is useful to fix some notation. If α is a complex number and T is a square matrix, then $m_T(\alpha)$ will denote the algebraic multiplicity of α as an eigenvalue of T when $\alpha \in \sigma(T)$, and $m_T(\alpha) = 0$ otherwise. We are now prepared to make the connection with the poles and zeros of W.

Recall from the material on dilation of Section 7.3 that there exists an invertible $n \times n$ matrix S such that $S^{-1}AS$, $S^{-1}B$ and CS have the form

$$S^{-1}AS = \begin{bmatrix} A_- & * & * \\ 0 & A_0 & * \\ 0 & 0 & A_+ \end{bmatrix}, \qquad S^{-1}B = \begin{bmatrix} * \\ B_0 \\ 0 \end{bmatrix}, \qquad CS = \begin{bmatrix} 0 & C_0 & * \end{bmatrix},$$

where $W(\lambda) = I_m + C_0(\lambda - A_0)^{-1}B_0$ is a minimal realization of W. The block matrix representation of $S^{-1}AS$ implies that the characteristic polynomial of A is the product of the characteristic polynomials of A_-, A_0 and A_+, i.e.,

$$\det(\lambda - A) = \det(\lambda - A_-)\det(\lambda - A_0)\det(\lambda - A_+).$$

Thus, for α a complex number,

$$m_A(\alpha) = m_{A_-}(\alpha) + m_{A_0}(\alpha) + m_{A_+}(\alpha). \tag{8.32}$$

From the block upper triangular form of A in (8.30) it is clear that

$$m_A(\alpha) = \sum_{j=1}^{k} m_{A_j}(\alpha).$$

Also, as the realization of W involving A_0, B_0 and C_0 is minimal, $m_{A_0}(\alpha) = \delta(W;\alpha)$, and likewise $m_{A_j}(\alpha) = \delta(W_j;\alpha)$, $j = 1, \ldots, k$ (cf., Section 8.3). But then (8.32) can be rewritten as

$$\sum_{j=1}^{k} \delta(W_j;\alpha) - \delta(W;\alpha) = m_{A_-}(\alpha) + m_{A_+}(\alpha). \tag{8.33}$$

Turning to $A^\times$, we note that

$$m_{A^\times}(\alpha) = \sum_{j=1}^{k} m_{A_j^\times}(\alpha).$$

This is obvious from (8.31). Further, with $A_0^\times = A_0 - B_0C_0$, we have $m_{A_0^\times}(\alpha) = \delta(W^{-1};\alpha)$. Now $S^{-1}A^\times S = S^{-1}AS - S^{-1}BCS$ has the form

$$S^{-1}A^\times S = \begin{bmatrix} A_- & * & * \\ 0 & A_0^\times & * \\ 0 & 0 & A_+ \end{bmatrix}.$$

Hence $m_{A^\times}(\alpha) = m_{A_-}(\alpha) + m_{A_0^\times}(\alpha) + m_{A_+}(\alpha)$, and it follows that

$$\sum_{j=1}^{k} \delta(W_j^{-1};\alpha) - \delta(W^{-1};\alpha) = m_{A_-}(\alpha) + m_{A_+}(\alpha). \tag{8.34}$$

Combining (8.33) and (8.34), we see that the left-hand side and the right-hand side of (8.29) are both equal to $m_{A_-}(\alpha) + m_{A_+}(\alpha)$. □

We shall be interested in factorizations W_1W_2 such that

$$\delta(W_1W_2;\lambda_0) = \delta(W_1;\lambda_0) + \delta(W_2;\lambda_0) \tag{8.35}$$

regardless of the choice of $\lambda_0 \in \mathbb{C}\cup\{\infty\}$. Such factorizations are called minimal. We shall come back to this concept in the next chapter. To understand the meaning of condition (8.35), we introduce the notion of local minimality of a system.

Let $\Theta = (A, B, C, D; X, Y)$ be a finite-dimensional system and $\lambda_0 \in \mathbb{C}$. We say that Θ is *minimal at the point* λ_0 if

$$\bigcap_{j=0}^{\infty} \operatorname{Ker} CA^jP = \operatorname{Ker} P, \qquad \bigvee_{j=0}^{\infty} \operatorname{Im} PA^jB = \operatorname{Im} P, \tag{8.36}$$

where P is the Riesz projection of A at λ_0. Note that Θ is minimal at each point in the resolvent set $\rho(A)$ of A. If the external operator of Θ is the identity operator on Y, then Θ is minimal at λ_0 if and only if the projection $\operatorname{pr}_P(\Theta)$ of Θ is a minimal system.

For later purposes we present the following local version of Proposition 7.1.

Proposition 8.17. *Let $\Theta = (A, B, C, D; X, Y)$ be a biproper finite-dimensional system, and assume that $\lambda_0 \in \mathbb{C}$ is not a common eigenvalue of A and $A^\times = A - BD^{-1}C$. Then Θ is minimal at the point λ_0.*

Proof. We already noted that Θ is minimal at each point in the resolvent set $\rho(A)$ of A. Thus we may assume that λ_0 is an eigenvalue of A but not of $A^\times$.

Put $X_1 = \operatorname{Ker}(C|A)$, and let X_0 be some linear complement of X_1 in X. Relative to the decomposition $X = X_1 \dot{+} X_0$ the operators A, B and C can be written as block matrices

$$A = \begin{bmatrix} A_1 & * \\ 0 & A_0 \end{bmatrix}, \qquad B = \begin{bmatrix} * \\ B_0 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & C_0 \end{bmatrix}.$$

Here we used that $X_1 \subset \operatorname{Ker} C$ and that X_1 is invariant under A. The $*$'s denote operators that will not be specified any further. From the block matrix representations of A, B and C we see that

$$A^\times = A - BD^{-1}C = \begin{bmatrix} A_1 & * \\ 0 & A_0^\times \end{bmatrix},$$

where, following our standard convention, $A_0^\times = A_0 - B_0D^{-1}C_0$. Since λ_0 an eigenvalue of A but not of $A^\times$, we know that λ_0 is not an eigenvalue of A_1. It follows that the Riesz projection P of A at λ_0 partitions as

$$P = \begin{bmatrix} 0 & R \\ 0 & P_0 \end{bmatrix},$$

where P_0 is the Riesz projection of A_0 at λ_0. The fact that P is a projection implies that the operator R in the block matrix representation of P satisfies $RP_0 = R$. Thus $\operatorname{Ker} P = X_1 \dot{+} \operatorname{Ker} P_0$. Next, note that

$$CA^jP = \begin{bmatrix} 0 & C_0A_0^jP_0 \end{bmatrix}, \qquad j = 0, 1, 2, \ldots.$$

Since $X_1 = \operatorname{Ker}(C|A)$, we have $\operatorname{Ker}(C_0|A_0) = \{0\}$. Using this in the previous formula, we see that

$$\bigcap_{j=0}^{\infty} \operatorname{Ker} CA^jP = X_1 \dot{+} \bigcap_{j=0}^{\infty} \operatorname{Ker} C_0A_0^jP_0 = X_1 \dot{+} \operatorname{Ker} P_0 = \operatorname{Ker} P.$$

Thus the first identity in (8.36) is proved.

To prove the second identity in (8.36) put $X_2 = \operatorname{Im}(A|B)$, and let X_0 be some linear complement of X_2 in X. Note that X_2 is invariant under A and contains $\operatorname{Im} B$. Hence relative to the decomposition $X = X_0 \dot{+} X_2$ the operators A, B and C can be written as block matrices

$$A = \begin{bmatrix} A_0 & 0 \\ * & A_2 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ B_2 \end{bmatrix}, \qquad C = \begin{bmatrix} * & C_2 \end{bmatrix}.$$

As before $*$'s denote operators that will not be specified any further. From these block matrix representations it follows that

$$A^{\times} = A - BD^{-1}C = \begin{bmatrix} A_0 & 0 \\ * & A_2^{\times} \end{bmatrix},$$

with $A_2^{\times} = A_2 - B_2D^{-1}C_2$. Since λ_0 an eigenvalue of A but not of $A^{\times}$, we know that λ_0 is not an eigenvalue of A_0. It follows that the Riesz projection P of A at λ_0 partitions as

$$P = \begin{bmatrix} 0 & 0 \\ Q & P_2 \end{bmatrix},$$

where P_2 is the Riesz projection of A_2 at λ_0 and $Q = P_2Q$. It follows that

$$\operatorname{Im} PA^jB = \operatorname{Im} \begin{bmatrix} 0 \\ P_2A_2^jB_2 \end{bmatrix}, \qquad j = 0, 1, 2, \ldots .$$

As $X_2 = \operatorname{Im}(A_2|B_2)$, we see that

$$\bigvee_{j=0}^{\infty} \operatorname{Im} PA^jB = \bigvee_{j=0}^{\infty} \operatorname{Im} \begin{bmatrix} 0 \\ P_2A_2^jB_2 \end{bmatrix} = \{0\} \dot{+} \operatorname{Im} P_2 = \operatorname{Im} P.$$

This proves the second identity in (8.36). □

The connection between local minimality and local degree is expressed by the following theorem.

Theorem 8.18. *Let W be the transfer function of the finite-dimensional system $\Theta = (A, B, C, D; X, Y)$, let $\lambda_0 \in \mathbb{C}$, and let P be the Riesz projection of A at λ_0. Then $\delta(W; \lambda_0) \leq \operatorname{rank} P$, equality occurring if and only if Θ is minimal at λ_0.*

Proof. Write W in the form (8.26) and note that

$$W_{-j} = CP(A - \lambda_0)^jB, \qquad j = 1, 2, \ldots .$$

Here $W_{-j} = 0$ for $j = q+1, q+2, \ldots$. It follows that

$$\operatorname{col}\left(C(\lambda_0 - A)^{j-1}\right)_{j=1}^{q} P \operatorname{row}\left((\lambda_0 - A)^{q-j}B\right)_{j=1}^{q} = \Omega, \tag{8.37}$$

where Ω is as in (8.27). From this and the definition of $\delta(W_1; \lambda_0)$ we conclude $\delta(W; \lambda_0) \leq \operatorname{rank} P$. We may assume that q is larger than or equal to the degree of the minimal polynomial of A. In that case one has that Θ is minimal at λ_0 if and only if the rank of the operator

$$\operatorname{col}(CA^{j-1})_{j=1}^{q} P \operatorname{row}(A^{q-j}B)_{j=1}^{q} \tag{8.38}$$

is equal to the rank of P. Moreover, the operator appearing in the left-hand side of (8.37) has the same rank as the operator (8.38). Hence Θ is minimal at λ_0 if and only $\delta(W; \lambda_0) = \operatorname{rank} P$. □

Suppose Θ_1 and Θ_2 are systems with the same input/output space. Let $\lambda_0 \in \mathbb{C}$. If the product $\Theta_1\Theta_2$ is minimal at λ_0, then so are the factors Θ_1 and Θ_2. The converse of this is not true. In the next theorem we present a necessary and sufficient condition for $\Theta_1\Theta_2$ to be minimal at λ_0. Part of the condition involves the logarithmic property of the local degree; cf., formula (8.28).

Theorem 8.19. *For $j=1,2$, let W_j be the transfer function of the finite-dimensional system $\Theta_j = (A_j, B_j, C_j, D_j; X_j, Y)$, and let $\lambda_0 \in \mathbb{C}$. Then $\Theta_1\Theta_2$ is minimal at λ_0 if and only if Θ_1 and Θ_2 are minimal at λ_0 and, in addition,*

$$\delta(W_1W_2;\lambda_0) \;=\; \delta(W_1;\lambda_0) + \delta(W_2;\lambda_0). \tag{8.39}$$

Proof. Recall that $\Theta_1\Theta_2 = (A, B, C, D; X_1 \dotplus X_2, Y)$, where A, B, C, and D are given by

$$A = \begin{bmatrix} A_1 & B_1C_2 \\ 0 & A_2 \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}, \quad D = D_1D_2.$$

Let P, P_1 and P_2 be the Riesz projections of A, A_1 and A_2 at λ_0, respectively. Then, for a sufficiently small circle Γ around λ_0, we have

$$P = \begin{bmatrix} P_1 & \dfrac{1}{2\pi i}\displaystyle\int_\Gamma (\lambda - A_1)^{-1}B_1C_2(\lambda - A_2)^{-1}\,d\lambda \\ 0 & P_2 \end{bmatrix}.$$

Let Q be the operator given by the integral in the right upper corner of the block matrix for P. Since $P^2 = P$, we have $Q = P_1Q + QP_2$. It follows that

$$P = \begin{bmatrix} I & Q \\ 0 & I \end{bmatrix} \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} \begin{bmatrix} I & Q \\ 0 & I \end{bmatrix}. \tag{8.40}$$

But then

$$\operatorname{rank} P \;=\; \operatorname{rank} P_1 + \operatorname{rank} P_2, \tag{8.41}$$

because the first and the last factor in the right-hand side of (8.40) are invertible.

Suppose now that Θ_1 and Θ_2 are minimal at λ_0 and that (8.39) is satisfied. Then $\delta(W_1;\lambda_0) = \operatorname{rank} P_1$ and $\delta(W_2;\lambda_0) = \operatorname{rank} P_2$ by Theorem 8.18. Together with the above identity and (8.41) this gives $\operatorname{rank} P = \delta(W_1W_2;\lambda_0)$. By applying Theorem 8.18 again, we arrive at the conclusion that $\Theta_1\Theta_2$ is minimal at λ_0.

Conversely, assume that $\Theta_1\Theta_2$ is minimal at λ_0. Combining Theorem 8.18 and formulas (8.28) and (8.41), we get

$$\begin{aligned} \operatorname{rank} P &= \delta(W_1W_2;\lambda_0) \\ &\le \delta(W_1;\lambda_0) + \delta(W_2;\lambda_0) \\ &\le \operatorname{rank} P_1 + \operatorname{rank} P_2 \;=\; \operatorname{rank} P. \end{aligned}$$

Since the first and last quantity in this expression are the same, all the inequalities are in fact equalities. In particular, we have (8.39). Moreover, it is clear from Theorem 8.18 that $\delta(W_1;\lambda_0) = \operatorname{rank} P_1$ and $\delta(W_2;\lambda_0) = \operatorname{rank} P_2$. So, by the same theorem, Θ_1 and Θ_2 are minimal at λ_0. □

If Θ is a minimal system, then clearly Θ is minimal at each point of $\mathbb{C}$. The converse of this is also true.

Theorem 8.20. *Let $\Theta = (A, B, C, D; X, Y)$ be a finite-dimensional system, and suppose that Θ is minimal at each eigenvalue of A. Then Θ is a minimal system.*

Proof. Let λ be an eigenvalue of A, and let P be the corresponding Riesz projection. In view of (8.36), we have

$$P^{-1}[\operatorname{Ker}(C|A)] = \operatorname{Ker} P, \qquad P[\operatorname{Im}(A|B)] = \operatorname{Im} P.$$

Observe that $\operatorname{Ker}(C|A)$ and $\operatorname{Im}(A|B)$ are invariant subspaces for A. Since X is finite-dimensional, it follows that they are invariant for P too. Hence

$$\operatorname{Ker}(C|A) \subset P^{-1}[\operatorname{Ker}(C|A)] = \operatorname{Ker} P$$

$$\operatorname{Im} P = P[\operatorname{Im}(A|B)] \subset \operatorname{Im}(A|B).$$

If A has just a single eigenvalue $\operatorname{Ker} P$ is the zero space, and $\operatorname{Im} P = X$, so in this case Θ is minimal. If A has more than one eigenvalue the intersections of the kernels of the corresponding Riesz projections is zero, proving that Θ is observable. Also, the direct sum of the images of the corresponding Riesz projections is the whole state space X, proving that Θ is controllable. Hence Θ is minimal. □

Let $\Theta = (A, B, C, D; X, Y)$ be a finite-dimensional system and $\lambda_0 \in \mathbb{C}$. Denote the Riesz projection of A at λ_0 by P. Then for k sufficiently large $\operatorname{Im} P = \operatorname{Ker}(\lambda_0 - A)^k$ and $\operatorname{Ker} P = \operatorname{Im}(\lambda_0 - A)^k$. Using this one easily verifies that Θ is minimal at the point λ_0 if and only if the operators

$$\begin{bmatrix} C \\ \lambda_0 - A \end{bmatrix} : X \to Y \dot{+} X, \qquad \begin{bmatrix} \lambda_0 - A & B \end{bmatrix} : X \dot{+} Y \to X$$

are injective and surjective, respectively. Applying now the preceding theorem we obtain the so-called *Hautus test* for minimality: The system Θ is minimal if and only if

$$\operatorname{rank} \begin{bmatrix} C \\ \lambda - A \end{bmatrix} = \operatorname{rank} \begin{bmatrix} \lambda - A & B \end{bmatrix} = \dim X$$

for each eigenvalue λ of A.

Let W be as in (8.26), and let φ be a $\mathbb{C}^m$-valued function which is analytic at λ_0 and such that $\varphi(\lambda_0) = 0$. We call φ a *co-pole function* of W at λ_0 if $W(\lambda)\varphi(\lambda)$

is analytic at λ_0 and $y = \lim_{\lambda\to\lambda_0} W(\lambda)\varphi(\lambda)$ is nonzero. Note that in this case the vector y is a pole vector of W at λ_0 (see Section 8.2). Furthermore, if $\det W(\lambda) \not\equiv 0$ in a neighborhood of λ_0, then the function ψ, given by $\psi(\lambda) = W(\lambda)\varphi(\lambda)$, is a root function of W^{-1} at λ_0. A root function of W^{-1} at λ_0 is also referred to as a *pole function* of W at λ_0; see [8], page 67. The next proposition shows how co-pole functions of W at λ_0 are related to Jordan chains of the main operator in a realization for W which is minimal at λ_0.

Proposition 8.21. *Let W be the transfer function of the finite-dimensional system $\Theta = (A, B, C, D; X, \mathbb{C}^m)$, and let λ_0 be an eigenvalue of A. Assume Θ is minimal at λ_0. Let $k \geq 1$, and let*

$$\varphi(\lambda) \;=\; (\lambda - \lambda_0)^k\varphi_k + (\lambda - \lambda_0)^{k+1}\varphi_{k+1} + \cdots$$

be a co-pole function of W at λ_0. Put

$$x_j \;=\; \sum_{i=k}^{\infty} P(A - \lambda_0)^{i-j-1}B\varphi_i, \qquad j = 0, \ldots, k-1, \tag{8.42}$$

where P is the Riesz projection of A corresponding to λ_0. Then $x_0, \ldots, x_{k-1}$ is a Jordan chain of A at λ_0, that is, $x_0 \neq 0$ and

$$(A - \lambda_0)x_0 = 0, \quad (A - \lambda_0)^r x_{k-1} = x_{k-1-r}, \qquad r = 0, \ldots, k-1. \tag{8.43}$$

Moreover, each Jordan chain of A at λ_0 is obtained in this way. Finally, if the chain (8.42) is maximal, that is, $x_{k-1} \notin \operatorname{Im}(A - \lambda_0)$, then $\varphi_k \neq 0$.

Proof. Since φ is a co-pole function of W at λ_0, we know that

$$\sum_{i=k}^{\infty} W_{-j-i}\varphi_i = 0, \qquad j = 1, 2, \ldots . \tag{8.44}$$

Here W_{-j} is the coefficient of $(\lambda - \lambda_0)^{-j}$ in the Laurent expansion of W at λ_0. Observe that only a finite number of the terms in (8.44) are nonzero. Since W is the transfer function of the system Θ, we have

$$W_{-j} = CP(A - \lambda_0)^{j-1}B, \qquad j = 1, 2, \ldots . \tag{8.45}$$

Using that P and $A - \lambda_0$ commute we see that

$$\begin{aligned} C(A - \lambda_0)^r x_0 \;&=\; \sum_{i=k}^{\infty} C(A - \lambda_0)^r P(A - \lambda_0)^{i-1}B\varphi_i \\ &=\; \sum_{i=k}^{\infty} CP(A - \lambda_0)^{i+r-1}B\varphi_i \\ &=\; \sum_{i=k}^{\infty} W_{-(i+r)}\varphi_i, \qquad r = 1, 2, \ldots . \end{aligned}$$

We conclude that $(A-\lambda_0)x_0 \in \operatorname{Ker}(C|A)$. By definition, $x_0 \in \operatorname{Im} P$, and thus $(A-\lambda_0)x_0 \in \operatorname{Im} P$. It follows that $(A-\lambda_0)x_0 \in \operatorname{Ker}\big(C|_{\operatorname{Im} P}, A|_{\operatorname{Im} P}\big)$, and the fact that Θ is minimal at λ_0 yields $(A-\lambda_0)x_0 = 0$. This proves the first part of (8.43).

To prove the second part of (8.43), we note that for $1 \le j \le k$ we have

$$(A-\lambda_0)x_j \;=\; \sum_{i=k}^{\infty} P(A-\lambda_0)^{i-(j-1)-1}\varphi_i = x_{j-1}.$$

Thus $(A-\lambda_0)^r x_{k-1} = x_{k-1-r}$ for $r = 0, \dots, k-1$. From (8.42) we also see that

$$x_{k-1} - B\varphi_k \;=\; \sum_{i=k+1}^{\infty} P(A-\lambda_0)^{i-k}B\varphi_i \;\in\; \operatorname{Im}(A-\lambda_0).$$

Thus $x_{k-1} \notin \operatorname{Im}(A-\lambda_0)$ implies $B\varphi_k \neq 0$, and hence $\varphi_k \neq 0$.

To deal with the converse statement, let $x_0, \dots, x_{k-1}$ be a Jordan chain of A at λ_0. In particular, the vectors $x_0, \dots, x_{k-1}$ belong to $\operatorname{Im} P$. Since Θ is minimal at λ_0, we have $\operatorname{Im}(A|_{\operatorname{Im} P}, PB) = \operatorname{Im} P$, and hence there exists $\varphi_k, \varphi_{k+1}, \dots$ in $\mathbb{C}^m$, with only a finite number φ_j's being nonzero, such that

$$x_{k-1} = \sum_{i=k}^{\infty} P(A-\lambda_0)^{i-k}\varphi_i. \tag{8.46}$$

Put $\varphi(\lambda) = \sum_{j=k}^{\infty}(\lambda-\lambda_0)^j\varphi_j$. Using (8.45) it follows that for $j = 0, 1, 2, \dots$ we have

$$\begin{aligned}
\sum_{i=k}^{\infty} W_{-j-i}\varphi_i \;&=\; \sum_{i=k}^{\infty} CP(A-\lambda_0)^{j+i-1}B\varphi_i \\
&=\; CP(A-\lambda_0)^{j+k-1}\sum_{i=k}^{\infty} P(A-\lambda_0)^{i-k}B\varphi_i \\
&=\; CP(A-\lambda_0)^{j+k-1}x_{k-1}.
\end{aligned}$$

Since $x_0, \dots, x_{k+1}$ is a Jordan chain of A at λ_0, we conclude that

$$\sum_{i=k}^{\infty} W_{-j-i}\varphi_i \;=\; \begin{cases} Cx_0, & j = 0, \\ 0, & j = 1, 2, \dots . \end{cases} \tag{8.47}$$

Thus $W(\lambda)\varphi(\lambda)$ is analytic at λ_0, and $\lim_{\lambda\to\lambda_0} W(\lambda)\varphi(\lambda) = Cx_0$. To prove that φ is a co-pole function of W at λ_0, it remains to show that $Cx_0 \neq 0$. Assume not, i.e., $Cx_0 = 0$. Since $(A-\lambda_0)x_0 = 0$ and $x_0 \in \operatorname{Im} P$, it follows that x_0 is in $\operatorname{Ker}\big(C|_{\operatorname{Im} P}, A|_{\operatorname{Im} P}\big)$. Using again the minimality of Θ at λ_0, this yields $x_0 = 0$. But $x_0, \dots, x_{k-1}$ is a Jordan chain of A at λ_0, and thus $x_0 \neq 0$ by definition. Therefore, $Cx_0 \neq 0$.

Finally, since $x_0, \dots, x_{k-1}$ is a Jordan chain of A at λ_0 and x_{k-1} is given by (8.46) it is clear that (8.42) holds. □

Let $W(\lambda) = D + C(\lambda - A)^{-1}B$. Given a Jordan chain $x_0, \dots, x_{k-1}$ of A at λ_0, any co-pole function $\varphi(\lambda) = \sum_{j=k}^{\infty}(\lambda - \lambda_0)^j \varphi_j$ satisfying (8.42) will be called a *co-pole function corresponding to the Jordan chain* $x_0, \dots, x_{k-1}$. Using (8.42) and (8.45) it is clear that in this case Cx_j $(j = 0, \dots, k-1)$ is precisely the coefficient of $(\lambda - \lambda_0)^r$ in the Taylor expansion of $W(\lambda)\varphi(\lambda)$ at λ_0. This yields the following corollary.

Corollary 8.22. *Let W be the transfer function of the finite-dimensional system $\Theta = (A, B, C, D; X, \mathbb{C}^m)$, and assume $\det W(\lambda) \not\equiv 0$. Let λ_0 be an eigenvalue of A, and assume Θ is minimal at λ_0. If $x_0, \dots, x_{k-1}$ is a Jordan chain of A at λ_0, then $Cx_0, \dots, Cx_{k-1}$ is a Jordan chain of W^{-1} at λ_0, and each Jordan chain of W^{-1} at λ_0 is obtained in this way. Furthermore, C maps $\operatorname{Ker}(\lambda_0 - A)$ in a one-to-one way onto $\operatorname{Pol}(W; \lambda_0)$.*

Proof. Let $x_0, \dots, x_{k-1}$ be a Jordan chain of A at λ_0. Let φ be a corresponding co-pole function. Put $\psi(\lambda) = W(\lambda)\varphi(\lambda)$, and for $j = 0, \dots, k-1$ let y_j be the coefficient of $(\lambda - \lambda_0)^j$ in the Taylor expansion of ψ at λ_0. From the remark made in the paragraph preceding this proposition we know that $y_j = Cx_j$ for $j = 0, \dots, k-1$. As φ is a co-pole function, ψ is analytic at λ_0, the vector $\psi(\lambda_0)$ is nonzero, and

$$\lim_{\lambda \to \lambda_0} \frac{1}{(\lambda - \lambda_0)^{k-1}} W(\lambda)^{-1}\psi(\lambda) = 0.$$

Thus we know from Propositions 8.1 and 8.2 that ψ is a root function of W^{-1} at λ_0 of order at least k. This implies that $y_0 \dots, y_{k-1}$ is a Jordan chain of W^{-1} at λ_0. Thus $Cx_0, \dots, Cx_{k-1}$ is a Jordan chain of W^{-1} at λ_0.

Conversely, assume $y_0 \dots, y_{k-1}$ is a Jordan chain of W^{-1} at λ_0. Then there exists a root function ψ of W^{-1} at λ_0 of order at least k such that

$$\psi(\lambda) = y_0 + (\lambda - \lambda_0)y_1 + \cdots + (\lambda - \lambda_0)^{k-1}y_{k-1} + \cdots .$$

Define $\varphi(\lambda) = W(\lambda)^{-1}\psi(\lambda)$. Then φ is analytic at λ_0, has a zero of order at least k at λ_0, and $W(\lambda)\varphi(\lambda) = \psi(\lambda)$. Thus φ is a co-pole function. Let $x_0, \dots, x_{k-1}$ be the vectors defined by φ via formula (8.42). Then by Proposition 8.21, the vectors $x_0, \dots, x_{k-1}$ form a Jordan chain of A at λ_0, and φ is a corresponding co-pole function. Moreover, we have $Cx_j = y_j$ for $j = 0, \dots, k-1$.

Next we prove the final statement. First, let us recall from Lemma 8.8 that $\operatorname{Pol}(W; \lambda_0) = \operatorname{Ker}(W^{-1}; \lambda_0)$. Thus it suffices to prove that C maps the space $\operatorname{Ker}(\lambda_0 - A)$ in a one-to-one way onto $\operatorname{Ker}(W^{-1}; \lambda_0)$. By specifying the results obtained so far for $k = 1$ we see that $C[\operatorname{Ker}(\lambda_0 - A)] = \operatorname{Ker}(W^{-1}; \lambda_0)$. Hence it remains to show that C is one-to-one on $\operatorname{Ker}(\lambda_0 - A)$. To do this, take $x_0 \in \operatorname{Ker}(\lambda_0 - A)$, and assume that $Cx_0 = 0$. Then $CA^j x_0 = \lambda_0^j Cx_0 = 0$ for each $j = 0, 1, 2 \dots$. But the system Θ is minimal at λ_0, and $x_0 \in \operatorname{Im} P$, where P is the Riesz projection of A at λ_0. Note that $Px_0 = x_0$. But then the first identity in (8.36) shows that $Px_0 = x_0 \in \operatorname{Ker} P$. This can only happen when $x_0 = 0$. Thus C is one-to-one on $\operatorname{Ker}(\lambda_0 - A)$. □

Let W be a proper rational $m \times m$ matrix function with realization

$$W(\lambda) = D + C(\lambda - A)^{-1}B. \tag{8.48}$$

Fix $\lambda_0 \in \mathbb{C}$. As before, $m_A(\lambda_0)$ denotes the algebraic multiplicity of λ_0 as an eigenvalue of A and $m_A(\lambda_0) = 0$ when λ_0 is not an eigenvalue of A. Note that always $m_A(\lambda_0) \geq \delta(W;\lambda_0)$, and we have $m_A(\lambda_0) = \delta(W;\lambda_0)$ if and only if the realization (8.48) is minimal at λ_0 (see Theorem 8.18). Now let λ_0 be a pole of W in (8.48). Then necessarily λ_0 is an eigenvalue of A, and the order of λ_0 as a pole of W does not exceed the order of λ_0 as pole of $(\lambda - A)^{-1}$. When the latter two numbers are equal, we say that the realization (8.48) is *pole order preserving at* λ_0. The phrase "the realization (8.48) is pole order preserving at λ_0" will also be used when λ_0 is neither a pole of W nor a pole of $(\lambda - A)^{-1}$.

Proposition 8.23. *Let W be a proper rational $m \times m$ matrix function with realization* (8.48)*, and let λ_0 be a pole of W. Then the realization* (8.48) *is minimal at λ_0 if and only λ_0 is an eigenvalue of A and the partial multiplicities of λ_0 as a pole of W coincide with the partial multiplicities of λ_0 as an eigenvalue of A. Moreover, in that case, the realization* (8.48) *is pole order preserving at λ_0.*

Proof. If the realization (8.48) is minimal at λ_0, then we know from Corollary 8.22 (cf., Theorem 8.6) that the partial multiplicities of λ_0 as a pole of W coincide with the partial multiplicities of λ_0 as an eigenvalue of A.

For the converse, recall that $\delta(W;\lambda_0)$ is equal to the sum of the partial multiplicities of λ_0 as a pole of W (see the first paragraph of Section 8.4; also Proposition 8.11). On the other hand, the sum of the partial multiplicities of λ_0 as an eigenvalue of A is equal to the algebraic multiplicity $m_A(\lambda_0)$ of λ_0 as an eigenvalue of A. Thus, if the partial multiplicities of λ_0 as a pole of W and as an eigenvalue of A coincide, then $\delta(W;\lambda_0) = m_A(\lambda_0)$, and hence the realization is minimal at λ_0. Since the order of λ_0 as pole of W is equal to the largest partial multiplicity of λ_0 as a pole of W and the order of λ_0 as pole of $(\lambda - A)^{-1}$ is equal to the largest partial multiplicity of λ_0 as an eigenvalue of A, the final statement is trivially true. □

Corollary 8.24. *Let W be a proper rational $m \times m$ matrix function with realization* (8.48)*, and let λ_0 be an eigenvalue of A of geometric multiplicity one. Then the realization* (8.48) *is minimal at λ_0 if and only if the realization* (8.48) *is pole order preserving at λ_0.*

Proof. Since λ_0 is an eigenvalue of A of geometric multiplicity one, we have $\dim \operatorname{Ker}(\lambda_0 - A) = 1$ and $m_A(\lambda_0)$ is equal to the order of λ_0 as pole of $(\lambda - A)^{-1}$.

Assume the realization (8.48) is minimal at λ_0. Then $\dim \operatorname{Pol}(W;\lambda_0) = \dim \operatorname{Ker}(\lambda_0 - A)$, and hence the geometric multiplicity of λ_0 as a pole of W is equal to one. It follows (Corollary 8.10 above) that the order of λ_0 as a pole of W is equal to the pole-multiplicity, which in turn is equal to $\delta(W;\lambda_0)$. By minimality, $\delta(W;\lambda_0) = m_A(\lambda_0)$. We conclude that the order λ_0 as a pole of W is equal to the order of λ_0 as a pole of $(\lambda - A)^{-1}$.

To prove the converse implication, assume the realization (8.48) is pole order preserving at λ_0. Since λ_0 is an eigenvalue of A, the function $(\lambda - A)^{-1}$ has a pole at λ_0. Our pole order preserving assumption implies that λ_0 as a pole of W and the orders of λ_0 as a pole of W and as a pole of $(\lambda - A)^{-1}$ coincide. By the result of the first paragraph, the order of λ_0 as a pole of $(\lambda - A)^{-1}$ is equal to $m_A(\lambda_0)$. Recall from Theorem 8.18 that $m_A(\lambda_0) \geq \delta(W; \lambda_0)$. On the other hand it is clear from the definition given in the first paragraph of the present section that the order of λ_0 as a pole of W does not exceed $\delta(W; \lambda_0)$. Hence $\delta(W; \lambda_0) \geq m_A(\lambda_0)$, and it follows that $\delta(W; \lambda_0)$ and $m_A(\lambda_0)$ coincide. But then the realization is minimal at λ_0 by Theorem 8.18. □

By applying the previous corollary to each eigenvalue of the state matrix A we get the following result.

Corollary 8.25. *Let W be a proper rational $m \times m$ matrix function with realization* (8.48)*, and assume A is nonderogatory. Then the realization* (8.48) *is minimal if and only if it is pole order preserving at each eigenvalue of A.*

8.5 McMillan degree and minimality of systems

Let W be a rational $m \times m$ matrix function. Recall that the local degree $\delta(W; \lambda)$ of W at λ vanishes if and only if W is analytic at λ. Therefore it makes sense to put

$$\delta(W) = \sum_{\lambda \in \mathbb{C} \cup \{\infty\}} \delta(W; \lambda).$$

This number is known in the literature as the *McMillan degree* of W. It plays an important role in network theory. Of course the definition applies to any rational operator function of which the values act on a finite-dimensional space. A change of parameter involving a Möbius transformation does not affect the McMillan degree. Therefore we concentrate on the case when W is analytic at ∞. The next theorem is an immediate consequence of Theorems 8.18 and 8.20.

Theorem 8.26. *Let W be the transfer function of the finite-dimensional system Θ, and let X be the state space of Θ. Then $\delta(W) \leq \dim X$, equality occurring if and only if Θ is minimal.*

Let W be analytic at ∞. From Theorem 7.6 it is clear that the minimal realizations for W are just the realizations with smallest possible state space dimension. Theorem 8.26 adds to this that the smallest possible state space dimension is equal to the McMillan degree of W.

Suppose $W(\lambda)$ is invertible for some $\lambda \in \mathbb{C} \cup \{\infty\}$. Then W^{-1} is also a rational $m \times m$ matrix function. We claim that

$$\delta(W) = \delta\big(W^{-1}\big). \tag{8.49}$$

To see this, we may assume that W is analytic at ∞ and $W(\infty)$ invertible. Otherwise we apply a suitable Möbius transformation. Let $\Theta = (A, B, C, D; X, \mathbb{C}^m)$ be a minimal system of which the transfer function coincides with W. Since $D = W(\infty)$, we have that D is invertible. But then $\Theta^\times$ is well defined and its transfer function is W^{-1}. The minimality of Θ implies that of $\Theta^\times$. Formula (8.49) is now an immediate consequence of Theorem 8.26.

As a second application of Theorem 8.26, we deduce another description of the McMillan degree for the case when W is analytic at ∞. Let

$$W(\lambda) \;=\; D + \frac{1}{\lambda}D_1 + \frac{1}{\lambda^2}D_2 + \cdots$$

be the Laurent expansion of W at ∞, and let H_m be the block Hankel matrix given by

$$H_m \;=\; \big[D_{i+j-1}\big]_{i,j=1}^{m}.$$

Then, for m sufficiently large, we have $\delta(W) = \operatorname{rank} H_m$. To prove this, choose a minimal system $\Theta = (A, B, C, D; X, \mathbb{C}^m)$ of which the transfer function coincides with W. Then

$$D_j \;=\; CA^{j-1}B, \qquad j = 1, 2, \ldots .$$

Hence $H_m = \operatorname{col}(CA^{j-1})_{j=1}^{m} \operatorname{row}(A^{j-1}B)_{j=1}^{m}$. As Θ is minimal, we see that for m sufficiently large $\operatorname{rank} H_m = \dim X$. But $\dim X = \delta(W)$ by Theorem 8.26, and the proof is complete.

From (8.28) we know that δ has the sublogarithmic property, that is

$$\delta(W_1W_2) \;\leq\; \delta(W_1) + \delta(W_2). \tag{8.50}$$

The next theorem is the global analogue of Theorem 8.19; it is one of the main tools for studying minimal factorizations (see the next chapter).

Theorem 8.27. *For $j = 1,2$, let W_j be the transfer function of the finite-dimensional system $\Theta_j = (A_j, B_j, C_j, D_j; X_j, Y)$. Then $\Theta_1\Theta_2$ is minimal if and only if Θ_1 and Θ_2 are minimal and $\delta(W_1W_2) = \delta(W_1) + \delta(W_2)$.*

Proof. Combine the results of Section 8.4. □

Notes

The results in the first two sections are taken from [74]. For analytic matrix functions the results of the first section can also be found in the Appendix of [56]. Section 8.3 is based on Section 2.1 in [14]. For earlier material related to Theorem 8.15 see [39], the references therein, and [104]. Sections 8.4 and 8.5 are based on Sections 4.1 and 4.2 in [14]. Theorem 8.16 and the final part of Section 8.4 (starting from Proposition 8.21) are new.

Chapter 9

Minimal Factorization of Rational Matrix Functions

In this chapter the notion of minimal factorization of rational matrix functions, which has its origin in mathematical system theory, is introduced and analyzed. In Section 9.1 minimal factorizations are identified as those factorizations that do not admit pole zero cancellation. Canonical factorization is an example of minimal factorization but the converse is not true. In Section 9.2 we use minimal factorization to extend the notion of canonical factorization to rational matrix functions that are allowed to have poles and zeros on the curve. In Section 9.3 (the final section of this chapter) the concept of a supporting projection is extended to finite-dimensional systems that are not necessarily biproper. This allows us to prove one of the main theorems of the first section also for proper rational matrix functions of which the value at infinity is singular.

9.1 Minimal factorization

Let W, W_1 and W_2 be rational $n \times n$ matrix functions, and assume that

$$W(\lambda) = W_1(\lambda)W_2(\lambda). \tag{9.1}$$

Then we know from Section 8.5 that $\delta(W) \leq \delta(W_1) + \delta(W_2)$. The factorization (9.1) is called *minimal* if $\delta(W) = \delta(W_1) + \delta(W_2)$. An equivalent requirement is that this equality holds pointwise

$$\delta(W;\lambda) \,=\, \delta(W_1;\lambda) + \delta(W_2;\lambda), \qquad \lambda \in \mathbb{C} \cup \{\infty\}. \tag{9.2}$$

Minimal factorizations are important in network theory (see [113] and the references therein).

Applying, if necessary, a suitable Möbius transformation, we may assume that W is analytic at ∞. But then, if (9.1) is a minimal factorization, the factors W_1 and W_2 are analytic at ∞ too. Indeed, from $0 = \delta(W;\infty) = \delta(W_1;\infty) + \delta(W_2;\infty)$ it follows that $\delta(W_1;\infty) = \delta(W_2;\infty) = 0$. In view of this we shall concentrate on the case when the rational matrix functions are analytic at ∞. In other words we assume that they appear as transfer functions of finite-dimensional systems.

For such functions there is an alternative way of defining the notion of a minimal factorization. The definition is suggested by Theorem 8.27 and reads as follows: The factorization (9.1) is minimal if (and only if) from Θ_1 and Θ_2 being minimal realizations for W_1 and W_2, respectively, it follows that $\Theta_1\Theta_2$ is a minimal realization for W.

It is of interest to note that this alternative definition makes sense in a more general context. One just has to specify a suitable class of systems together with the corresponding transfer functions. For the Livsic-Brodskii characteristic operator function this leads to the concept of a factorization into regular factors. For details, see [30] Section I.5. One could also consider Kreĭn systems and the corresponding transfer functions. However, for such systems biminimality rather than minimality seems to be the natural property. As a final special case we mention the class of transfer functions of monic systems with given fixed (possibly infinite-dimensional) input/output space Y. This class coincides with that of the inverses of monic operator polynomials having as coefficients bounded linear operators on Y. Recall that monic systems are always minimal. So in this context every factorization is minimal. For the finite-dimensional case this can also be seen directly from the behavior of the McMillan degree. The argument may then be based on the fact that the McMillan degrees of a function and its inverse coincide and the observation that if L is a monic $n\times n$ matrix polynomial, then $\delta(L) = \delta(L;\infty) = n\ell$ where ℓ is the degree of L.

Now let us return to the study of minimal factorizations of rational matrix functions. So suppose W, W_1 and W_2 are rational $m \times m$ matrix functions. In the remainder of this section we shall always suppose that $\det W(\lambda) \not\equiv 0$. This implies the existence of a scalar $a \in \mathbb{C}$ such that $W(a)$ is invertible. Put $\widetilde{W}(\lambda) = W(a)^{-1}W(\lambda^{-1} + a)$. Then $\widetilde{W}(\infty) = I_m$. There is a one-to-one correspondence between the (minimal) factorizations of W and those of $\widetilde{W}$. Therefore there is no loss of generality in assuming that $W(\infty) = I_m$.

Suppose $W(\infty) = I_m$. We are interested in the minimal factorizations of W. We claim that it suffices to consider only those factorizations (9.1) for which $W_1(\infty) = W_2(\infty) = I_m$. To make this claim more precise, assume that (9.1) is a minimal factorization of W. Then W_1 and W_2 are analytic at ∞, and we have $W_1(\infty)W_2(\infty) = I_m$. So $W_1(\infty)$ and $W_2(\infty)$ are each others inverse. By multiplying W_1 from the right with $W_2(\infty)$ and W_2 from the left by $W_1(\infty)$, we obtain a minimal factorization of W of which the factors have the value I_m at ∞.

These considerations justify the fact that in this section, *from now on, without further notice, we only deal with rational matrix functions that are analytic at*

∞ *with value the identity matrix*. In other words the rational matrix functions considered below appear as transfer functions of unital finite-dimensional systems.

Intuitively, formula (9.2) means that in the product W_1W_2 pole-zero cancellations do not occur. The following theorem makes this statement more precise. Recall that $A^\top$ stands for the transpose of the matrix A. The meaning of the symbols $\mathrm{Ker}\,(W;\lambda)$ and $\mathrm{Pol}\,(W;\lambda)$ has been explained in Sections 8.1 and 8.2, respectively.

Theorem 9.1. *The factorization $W = W_1W_2$ is a minimal factorization if and only if for each λ in $\mathbb{C}$ we have*

(i) $\mathrm{Ker}\,(W_1;\lambda) \cap \mathrm{Pol}\,(W_2;\lambda) = \{0\}$,

(ii) $\mathrm{Pol}\,(W_1^\top;\lambda) \cap \mathrm{Ker}\,(W_2^\top;\lambda) = \{0\}$.

To prove this theorem we need the following lemma.

Lemma 9.2. *Let W be the transfer function of the unital minimal system $\Theta = (A, B, C; \mathbb{C}^n, \mathbb{C}^m)$. Then C maps* $\mathrm{Ker}\,(A-\lambda)$ *in a one-one manner onto* $\mathrm{Pol}\,(W;\lambda)$.

Proof. Using a similarity transformation we may assume without loss of generality that $A = J$, $B = R$ and $C = Q$, where J, R and Q are the operators constructed in Theorem 8.12. But for the minimal system $(J, R, Q; \mathbb{C}^n, \mathbb{C}^m)$ the lemma is trivial. □

Let W and Θ be as in the preceding lemma, and apply this lemma to the associate system $\Theta^\times$. Then one sees that C maps $\mathrm{Ker}\,(A^\times-\lambda)$ in a one-one manner onto $\mathrm{Ker}\,(W;\lambda_0)$.

Proof of Theorem 9.1. Let

$$\Theta_1 = (A_1, B_1, C_1; \mathbb{C}^{n_1}, \mathbb{C}^m), \qquad \Theta_2 = (A_2, B_2, C_2; \mathbb{C}^{n_2}, \mathbb{C}^m)$$

be minimal realizations for W_1 and W_2, respectively, and write

$$\Theta = (A, B, C; \mathbb{C}^{n_1} \dotplus \mathbb{C}^{n_2}, \mathbb{C}^m)$$

for the product $\Theta_1\Theta_2$. So

$$A = \begin{bmatrix} A_1 & B_1C_2 \\ 0 & A_2 \end{bmatrix}, \qquad \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \qquad \begin{bmatrix} C_1 & C_2 \end{bmatrix}.$$

Fix $k \geq 1$. By induction one proves that

$$A^k = \begin{bmatrix} A_1^k & T_m \\ 0 & A_2^k \end{bmatrix}, \qquad T_k = \sum_{j=0}^{k-1} A_1^{k-1-j} B_1 C_2 A_2^j.$$

It follows that $CA^k = \begin{bmatrix} C_1A_1^k & Z_k \end{bmatrix}$, where

$$Z_k \;=\; C_2A_2^k + \sum_{j=0}^{k-1} C_1A_1^{k-1-j}B_1C_2A_2^j. \tag{9.3}$$

Again employing induction one shows that

$$C_1A_1^j \;=\; C_1(A_1^\times)^j + \sum_{i=0}^{j-1} C_1(A_1^\times)^{j-1-i}B_1C_1A_1^i, \qquad j = 1, 2, \ldots . \tag{9.4}$$

Using this in (9.3) one obtains

$$Z_k \;=\; C_2A_2^k + \sum_{j=0}^{k-1} C_1(A_1^\times)^{k-1-j}B_1Z_j. \tag{9.5}$$

As $CA^k = \begin{bmatrix} C_1A_1^k & Z_k \end{bmatrix}$, we see from (9.4) and (9.5) that

$$\begin{bmatrix} I & 0 & 0 & \cdots & 0 \\ -C_1B_1 & I & 0 & & \vdots \\ -C_1A_1^\times B_1 & -C_1B_1 & I & & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -C_1(A_1^\times)^{k-1}B_1 & -C_1(A_1^\times)^{k-2}B_1 & \cdots & -C_1B_1 & I \end{bmatrix} \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^k \end{bmatrix}$$

$$= \begin{bmatrix} C_1 & C_2 \\ C_1A_1^\times & C_2A_2 \\ C_1(A_1^\times)^2 & C_2(A_2)^2 \\ \vdots & \vdots \\ C_1(A_1^\times)^k & C_2(A_2)^k \end{bmatrix}.$$

In particular,

$$\bigcap_{j=0}^{k} \operatorname{Ker} CA^j \;=\; \bigcap_{j=0}^{k} \operatorname{Ker} \begin{bmatrix} C_1(A_1^\times)^j & C_2(A_2)^j \end{bmatrix}. \tag{9.6}$$

Using (9.6) we shall prove that the system Θ is observable if and only if $\operatorname{Ker}(W_1;\lambda) \cap \operatorname{Pol}(W_2;\lambda) = \{0\}$ for each $\lambda \in \mathbb{C}$. First, assume

$$0 \neq y_0 \in \operatorname{Ker}(W_1;\lambda_0) \cap \operatorname{Pol}(W_2;\lambda_0). \tag{9.7}$$

Note that $\mathrm{Ker}\,(W_1;\lambda_0) = \mathrm{Pol}\,(W_1^{-1};\lambda_0)$. So applying Lemma 9.2 to $\Theta_1^\times$ and Θ_2, we see that there exist $x_1 \in \mathrm{Ker}\,(A_1^\times - \lambda_0)$ and $x_2 \in \mathrm{Ker}\,(A_2 - \lambda_0)$ such that $C_1x_1 = C_2x_2 = y_0$. As $y_0 \neq 0$, we have $x_1, x_2 \neq 0$. Furthermore,

$$C_1(A_1^\times)^j x_1 \;=\; \lambda_0^j y_0 \;=\; C_2(A_2)^j x_2, \qquad j = 0, 1, 2, \ldots .$$

But then we can use (9.6) to show that

$$x_0 = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix}$$

is a nonzero element in $\mathrm{Ker}\,(C|A)$, and it follows that Θ is not observable.

Next, assume that Θ is not observable. Applying (9.6) we conclude that the space $K = \bigcap_{j=0}^\infty \mathrm{Ker}\begin{bmatrix} C_1(A_1^\times)^j & C_2(A_2)^j \end{bmatrix}$ is non-trivial. By [54] Lemma 2.2 (see also [55], Theorem 9.1) we have

$$K = \left\{ \begin{bmatrix} x_1 \\ -Sx_1 \end{bmatrix} \mid x_1 \in M \right\},$$

where M is a non-trivial $A_1^\times$-invariant subspace of $\mathbb{C}^{n_1}$ and $S : M \to \mathbb{C}^{n_2}$ is a linear map such that

$$C_1|_M = C_2S, \qquad S(A_1^\times|_M) = A_2S. \tag{9.8}$$

Since M is non-trivial, the operator $A_1^\times$ has an eigenvector, x_1 say, in M. Let λ_0 be the corresponding eigenvalue. Put $x_2 = Sx_1$ and $y_0 = C_1x_1$. Employing (9.8), we see that $A_2x_2 = \lambda_0x_2$ and $y_0 = C_2x_2$. But then, we can apply Lemma 9.7 to both $\Theta_1^\times$ and Θ_2 to show that

$$0 \neq y_0 \in \mathrm{Pol}\,(W_1^{-1};\lambda_0) \cap \mathrm{Pol}\,(W_2;\lambda_0).$$

As $\mathrm{Pol}\,(W_1^{-1};\lambda_0) = \mathrm{Ker}\,(W_1;\lambda_0)$, we obtain (9.7). So we have proved that Θ is observable if and only if condition (i) is satisfied for each $\lambda \in \mathbb{C}$.

To finish the proof, observe that Θ is controllable if and only if $\Theta^\top = (A^\top, C^\top, B^\top; \mathbb{C}^n, \mathbb{C}^m)$ is observable. Since $W^\top = W_2^\top W_1^\top$, we see from the first part of the proof that $\Theta^\top$ is observable if and only if

$$\mathrm{Ker}\,(W_2^\top;\lambda) \cap \mathrm{Pol}\,(W_1^\top;\lambda) \;=\; \{0\}, \qquad \lambda \in \mathbb{C}.$$

In other words Θ is controllable if and only if condition (ii) is satisfied for each $\lambda \in \mathbb{C}$. This completes the proof □

From Theorem 9.1 it follows that canonical factorization is a special case of minimal factorization. Indeed, if $W(\lambda) = W_-(\lambda)W_+(\lambda)$ is a canonical factorization of an $m \times m$ rational matrix function, then W_+ has its pole and zeros in the open lower half-plane and W_- has its pole and zeros in the open upper half-plane. Hence there is no pole-zero cancellation between the factors W_+ and W_-, and thus the factorization is minimal.

We now come to the main theorem of this section. It gives a complete description of minimal factorizations in terms of supporting projections (see Section 2.4) of minimal systems.

Theorem 9.3. *Let the unital system Θ be a minimal realization of the rational $m \times m$ matrix function W.*

(i) *If Π is a supporting projection for Θ, W_1 is the transfer function of $\mathrm{pr}_{I-\Pi}(\Theta)$ and W_2 is the transfer function of $\mathrm{pr}_{\Pi}(\Theta)$, then $W = W_1W_2$ is a minimal factorization of W.*

(ii) *If $W = W_1W_2$ is a minimal factorization of W, then there exists a unique supporting projection Π for the system Θ such that W_1 and W_2 are the transfer functions of $\mathrm{pr}_{I-\Pi}(\Theta)$ and $\mathrm{pr}_{\Pi}(\Theta)$, respectively.*

Proof. Let Π be a supporting projection for Θ. Then

$$\Theta \;=\; \mathrm{pr}_{I-\Pi}(\Theta)\mathrm{pr}_{\Pi}(\Theta).$$

Since Θ is minimal, it follows that $\mathrm{pr}_{I-\Pi}(\Theta)$ and $\mathrm{pr}_{\Pi}(\Theta)$ are minimal. But then one can apply Theorems 2.2 and 8.27 to show that $W = W_1W_2$ is a minimal factorization. This proves (i).

Next assume that $W = W_1W_2$ is a minimal factorization. For $i = 1, 2$, let Θ_i be a minimal realization of W_i with state space $\mathbb{C}^{n_i}$. Here $n_i = \delta(W_i)$ is the McMillan degree of W_i (see Theorem 8.26). By Theorem 8.27 the product $\Theta_1\Theta_2$ is minimal. Note that $\Theta_1\Theta_2$ is a realization for W. Hence $\Theta_1\Theta_2$ and Θ are similar, say with system similarity $S : \mathbb{C}^{n_1} \dot{+} \mathbb{C}^{n_2} \to \mathbb{C}^n$, where $n = n_1 + n_2 = \delta(W)$. Let Π be the projection of $\mathbb{C}^n$ along $S[\mathbb{C}^{n_1}]$ onto $S[\mathbb{C}^{n_2}]$. Then Π is a supporting projection for Θ. Moreover $\mathrm{pr}_{I-\Pi}(\Theta)$ is similar to Θ_1 and $\mathrm{pr}_{\Pi}(\Theta)$ is similar to Θ_2. It remains to prove the unicity of Π.

Suppose P is another supporting projection of Θ such that $\mathrm{pr}_{I-P}(\Theta)$ and $\mathrm{pr}_{P}(\Theta)$ are realizations of W_1 and W_2, respectively. Then $\mathrm{pr}_{I-P}(\Theta)$ and $\mathrm{pr}_{P}(\Theta)$ are minimal again. Hence $\mathrm{pr}_{I-\Pi}(\Theta)$ and $\mathrm{pr}_{I-P}(\Theta)$ are similar, say with system similarity $U : \mathrm{Ker}\,\Pi \to \mathrm{Ker}\,P$, and $\mathrm{pr}_{\Pi}(\Theta)$ and $\mathrm{pr}_{P}(\Theta)$ are similar, say with system similarity $V : \mathrm{Im}\,\Pi \to \mathrm{Im}\,P$. Define T on $\mathbb{C}^\delta$ by

$$T = \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} : \mathrm{Ker}\,\Pi \dot{+} \mathrm{Im}\,\Pi \;\to\; \mathrm{Ker}\,P \dot{+} \mathrm{Im}\,P.$$

Then T is a system similarity between Θ and itself. Since Θ is minimal it follows that T is the identity operatory on $\mathbb{C}^n$. But then $\Pi = P$, as desired. □

Theorem 9.3 may be viewed as an analogue of Theorem 5.4 and Theorem 5.6 in [30], where the one-one correspondence between regular divisors of the Livsic-Brodskii characteristic operator function and the left divisors of a simple Brodskii system is described. The one-one correspondence between the supporting subspaces of a monic system Θ and the right divisors of $L = W_\Theta^{-1}$ (cf., Subsection 3.4

and the references given there) is the variant of Theorem 9.3 for monic operator polynomials.

We conclude this section with an example that will be useful in later chapters. Let n be an integer, $n \geq 2$, and consider the rational 2×2 matrix function

$$W(\lambda) = \begin{bmatrix} 1 & \dfrac{1}{\lambda^n} \\ 0 & 1 \end{bmatrix}. \tag{9.9}$$

We shall show that this rational matrix function does not admit any non-trivial minimal factorization.

Note that W is proper and has only one pole, namely at zero. Using the definition of the local degree (see Section 8.4) we see that the local degree of W at zero is equal to n. It follows that McMillan degree of W is equal to n too. Next, consider the following matrices (which have sizes $n \times n$, $n \times 2$ and $2 \times n$, respectively):

$$A = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

The empty spots in the matrix for A stand for zero entries. It is straightforward to check the system $\Theta = (A, B, C, I_2; \mathbb{C}^n, \mathbb{C}^2)$ is a realization of W. Since $\delta(W)$ is equal to the order of A, it follows that this realization is minimal, by Theorem 8.26. Now observe $BC = 0$, and so the matrices A and $A^\times$ coincide. From the special form of A it follows that the non-trivial invariant subspaces are the space $M_j = \operatorname{span}\{e_1, \ldots, e_j\}$, $j = 1, \ldots, n$, where $e_1, \ldots, e_n$ is the standard basis of $\mathbb{C}^n$. Since A and $A^\times$ have the same invariant subspaces, the only supporting projections for Θ are the zero operator and the identity operator on $\mathbb{C}^n$. Hence, by Theorem 9.3, the matrix function W has no non-trivial minimal factorization.

9.2 Pseudo-canonical factorization

Let Γ be a Cauchy contour in $\mathbb{C}$. Denote by F_+ the interior domain of Γ, and by F_- the exterior domain of Γ. As usual, if Γ is the closure of the real line F_+ is the open upper half-plane, and F_- is the open lower half-plane, while if Γ is the closure of the imaginary line, F_+ is the open left half-plane and F_- is the open right half-plane.

Let W be an $m \times m$ rational matrix function, analytic on an open neighborhood of Γ and at infinity, and having invertible values on Γ and at infinity. By a

right pseudo-canonical factorization of W with respect to Γ we mean a factorization

$$W(\lambda) = W_-(\lambda)W_+(\lambda), \qquad \lambda \in \Gamma, \tag{9.10}$$

where W_- and W_+ are rational $m \times m$ matrix functions such that W_- has no poles and zeros in F_-, W_+ has no poles and zeros in F_+, and the factorization (9.10) is locally minimal at each point of Γ. If in (9.10) the factors W_- and W_+ are interchanged, we speak of a *left pseudo-canonical factorization.*

Observe the differences with a right canonical factorization of W. In that case W has no poles and zeros on Γ, while in the case of pseudo-canonical factorization it is allowed that W has poles and zeros on Γ. Further, in a canonical factorization the factors are required to have no poles and zeros on Γ as well. If W has no poles and zeros on Γ, then a pseudo-canonical factorization is a canonical factorization because of the minimality condition: it follows from this condition that the factors will not have poles and zeros on Γ either.

Also observe that since W_- has no poles and zeros in F_-, and W_+ has no poles and zeros in F_+, the factorization (9.10) is minimal at each point in F_-, as well as at each point of F_+. Hence the condition that the factorization (9.10) is locally minimal at each point of Γ can be replaced by the condition that the factorization is minimal.

The following theorem describes all right pseudo-canonical factorizations of W in terms of a minimal realization of W.

Theorem 9.4. *Let $W(\lambda) = D + C(\lambda I_n - A)^{-1}B$ be the transfer function of the minimal system $\Theta = (A, B, C, D; \mathbb{C}^n, \mathbb{C}^m)$, and let Γ be a Cauchy contour. Let $D = D_1 D_2$, with D_1 and D_2 square matrices. Then there is a one-to-one correspondence between the right pseudo-canonical factorizations $W = W_-W_+$ of W with respect to Γ with $W_-(\infty) = D_1$ and $W_+(\infty) = D_2$, and the set of pairs of subspaces $(M, M^\times)$ with the following properties*

(i) *M is an A-invariant subspace such that the restriction $A|_M$ of A to M has no eigenvalues in F_-, and M contains the span of all eigenvectors and generalized eigenvectors of A corresponding to eigenvalues in F_+,*

(ii) *$M^\times$ is an $A^\times$-invariant subspace such that the restriction $A^\times|_{M^\times}$ of $A^\times$ to $M^\times$ has no eigenvalues in F_+, and $M^\times$ contains the span of all eigenvectors and generalized eigenvectors of $A^\times$ corresponding to eigenvalues in F_-,*

(iii) *$M \dot{+} M^\times = \mathbb{C}^n$.*

The correspondence is as follows: given a pair of subspaces $(M, M^\times)$ with the properties (i), (ii) *and* (iii), *a right pseudo-canonical factorization of W with respect to Γ is given by $W(\lambda) = W_-(\lambda)W_+(\lambda)$, where*

$$W_-(\lambda) = D_1 + C(\lambda - A)^{-1}(I - \Pi)BD_2^{-1}, \tag{9.11}$$

$$W_+(\lambda) = D_2 + D_1^{-1}\Pi(\lambda - A)^{-1}B, \tag{9.12}$$

where Π is the projection along M onto $M^\times$.

Conversely, given a right pseudo-canonical factorization of W with respect to Γ and with $W_-(\infty) = D_1$, $W_+(\infty) = D_2$, there exists a unique pair of subspaces M and $M^\times$ with the properties (i), (ii) *and* (iii) *above, such that the factors W_- and W_+ are given by* (9.11) *and* (9.12)*, respectively.*

Observe that in general a pair of subspaces M, $M^\times$ for which (i), (ii) and (iii) hold need not be unique. Consequently, also pseudo-canonical factorizations need not be essentially unique. An example of this phenomenon will be given after the proof.

In comparison with the main theorem on canonical factorization (Theorem 6.1) we restrict attention here to minimal systems. This condition may be relaxed to systems that are locally minimal at each point of Γ; the same result holds for such systems. For details we refer to Theorem 3.1 in [103].

Proof. First assume that M and $M^\times$ are subspaces with the properties (i), (ii) and (iii), and denote by Π the projection onto $M^\times$ along M. Define W_- and W_+ by (9.11) and (9.12). Then the factorization $W = W_-W_+$ is a minimal factorization because of (iii) and Theorem 9.3, and the factorization is a right pseudo-canonical factorization because of (i) and (ii). Indeed, the poles of W_- are the eigenvalues of $(I-\Pi)A(I-\Pi)$, while the zeros of W_- are the eigenvalues of $(I-\Pi)A^\times(I-\Pi)$. So, the poles of W_- are in $F_+ \cup \Gamma$, and the zeros of W_- are in the same set. Likewise, the poles of W_+ are the eigenvalues of $\Pi A\Pi$, while the zeros of W_+ are the eigenvalues of $\Pi A^\times\Pi$. So, the poles and zeros of W_+ are in the set $F_- \cup \Gamma$. Hence the factorization is a right pseudo-canonical factorization.

Conversely, assume that the factorization is a right pseudo-canonical factorization. As the factorization is minimal, by Theorem 9.3 there exist two subspaces M and $M^\times$ such that (iii) holds, and such that M is A-invariant and $M^\times$ is $A^\times$-invariant. Because the factorization is a right pseudo-canonical factorization the poles of W_- lie in $F_+ \cup \Gamma$, while the poles of W_+ lie in $F_- \cup \Gamma$. As the poles of W_- are precisely the eigenvalues of $A|_M$ (the appropriate multiplicities counted), this proves (i). A similar argument applied to the zeros of W_- and W_+ proves (ii). □

Let $W = W_-W_+$ and $W = \tilde{W}_-\tilde{W}_+$ be two right pseudo-canonical factorizations with respect to Γ. These two factorizations are called equivalent if $W_-(\infty) = \tilde{W}_-(\infty)$, and there exists an invertible matrix E such that $W_-(\lambda) = \tilde{W}_-(\lambda)E$, $W_+(\lambda) = E^{-1}\tilde{W}_+(\lambda)$. Canonical factorizations, when they exist, are equivalent in this sense. The following example shows that this is not the case for pseudo-canonical factorizations.

Example. Let

$$W(\lambda) = \begin{bmatrix} \dfrac{\lambda}{\lambda+2i} & \dfrac{3i\lambda}{(\lambda-i)(\lambda+2i)} \\ 0 & \dfrac{\lambda}{\lambda-i} \end{bmatrix}.$$

We are interested in right pseudo-canonical factorization with respect to the real line of W. A minimal realization of W is given by

$$A = \begin{bmatrix} -2i & 0 \\ 0 & i \end{bmatrix}, \quad B = \begin{bmatrix} -i & i \\ 0 & i \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

One checks that $A^\times = 0$. Hence, the poles of W are $-2i$ and i, and the only zero is 0, with double multiplicity.

For M we have only one choice, being the space spanned by the column vector $[\,0\ 1\,]^\top$. For $M^\times$ we can take the space spanned by any column vector of the form $[\,1\ \alpha\,]^\top$. One computes that the corresponding factorization is given by

$$W(\lambda) \ = \ W_-^{(\alpha)}(\lambda) W_+^{(\alpha)}(\lambda),$$

where

$$W_-^{(\alpha)}(\lambda) \ = \ \begin{bmatrix} \dfrac{\lambda - i(1+\alpha)}{\lambda - i} & \dfrac{i(1+\alpha)}{\lambda - i} \\ \\ \dfrac{-i\alpha}{\lambda - i} & \dfrac{\lambda + i\alpha}{\lambda - i} \end{bmatrix},$$

$$W_+^{(\alpha)}(\lambda) \ = \ \begin{bmatrix} \dfrac{\lambda + i\alpha}{\lambda + 2i} & \dfrac{i(2-\alpha)}{\lambda + 2i} \\ \\ \dfrac{i\alpha}{\lambda + 2i} & \dfrac{\lambda + i(2-\alpha)}{\lambda + 2i} \end{bmatrix}.$$

Thus, for any complex number α we have a right pseudo-canonical factorization. Now suppose that the two factorizations corresponding to α and β are equivalent. Then $W_+^{(\alpha)}(\lambda) W_+^{(\beta)}(\lambda)^{-1} = W_-^{(\alpha)}(\lambda)^{-1} W_-^{(\beta)}(\lambda)$ would have to be a constant matrix. However, this product is given by

$$W_+^{(\alpha)}(\lambda) W_+^{(\beta)}(\lambda)^{-1} = \begin{bmatrix} \dfrac{\lambda + i(\alpha-\beta)}{\lambda} & \dfrac{-i(\alpha-\beta)}{\lambda} \\ \\ \dfrac{i(\alpha-\beta)}{\lambda} & \dfrac{\lambda - i(\alpha-\beta)}{\lambda} \end{bmatrix},$$

which clearly is not a constant when $\alpha \neq \beta$.

9.3 Minimal factorization in a singular case

In Section 2.4 we have introduced the notion of a supporting projection for systems that are biproper. In this section we extend this notion to finite-dimensional systems with equal input and output space that are not necessarily biproper. More

precisely, we shall consider systems of the form

$$\Theta = (A, B, C, D; X, \mathbb{C}^m), \tag{9.13}$$

where $\dim X$ is finite, and D is allowed to be singular.

Let Θ be as in (9.13). Given a projection Π of X we define $M(\Pi; \Theta)$ to be the operator from $\operatorname{Im}\Pi \dot{+} \mathbb{C}^m$ into $\operatorname{Ker}\Pi \dot{+} \mathbb{C}^m$ given by

$$M(\Pi; \Theta) = \begin{bmatrix} (I-\Pi)A\Pi & (I-\Pi)B \\ C\Pi & D \end{bmatrix}. \tag{9.14}$$

As we have seen in Section 2.4 (see the proof of Proposition 2.4), if Θ is biproper, i.e., the external coefficient D is non-singular, then

$$\operatorname{rank} M(\Pi; \Theta) = \operatorname{rank} D + \operatorname{rank}\big((I-\Pi)A^{\times}\Pi\big),$$

where $A^{\times} = A - BD^{-1}C$. Since $\operatorname{rank} D = m$, it follows that for D non-singular

$$\operatorname{rank} M(\Pi; \Theta) \leq m \;\Leftrightarrow\; A^{\times}[\operatorname{Im}\Pi] \subset \operatorname{Im}\Pi.$$

This proves the following proposition (which is a reformulation of Proposition 2.4).

Proposition 9.5. *Let Θ in (9.13) be biproper. Then a projection Π of X is a supporting projection for Θ if and only if*

$$A[\operatorname{Ker}\Pi] \subset \operatorname{Ker}\Pi, \qquad \operatorname{rank} M(\Pi; \Theta) \leq m. \tag{9.15}$$

Note that in (9.15) the inverse of D does not appear. Hence we can use (9.15) to extend the notion of supporting projections to systems Θ that are not biproper. In the sequel, if Θ is as in (9.13), then a projection Π of X is called a *supporting projection* for Θ whenever (9.15) is satisfied.

The first condition in (9.15) allows us to write the operators in Θ in block form, namely

$$\Theta = \left(\begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix}, \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \begin{bmatrix} C_1 & C_2 \end{bmatrix}, D\,; X, \mathbb{C}^m \right), \tag{9.16}$$

where the block matrix representations are taken with respect to the decomposition $X = \operatorname{Ker}\Pi \dot{+} \operatorname{Im}\Pi$. We shall refer to (9.16) as the *block matrix representation of Θ induced by the supporting projection* Π. Notice that in the notation of (9.16) we have

$$M(\Pi; \Theta) = \begin{bmatrix} A_{12} & B_1 \\ C_2 & D \end{bmatrix}. \tag{9.17}$$

From the second condition in (9.15) it follows that $M(\Pi; \Theta)$ admits a factorization over $\mathbb{C}^m$, that is, there exist operators

$$\begin{bmatrix} F \\ D_1 \end{bmatrix} : \mathbb{C}^m \to \begin{bmatrix} \operatorname{Ker}\Pi \\ \mathbb{C}^m \end{bmatrix}, \qquad \begin{bmatrix} G & D_2 \end{bmatrix} : \begin{bmatrix} \operatorname{Im}\Pi \\ \mathbb{C}^m \end{bmatrix} \to \mathbb{C}^m \tag{9.18}$$

such that

$$M(\Pi;\Theta) = \begin{bmatrix} F \\ D_1 \end{bmatrix} \begin{bmatrix} G & D_2 \end{bmatrix}. \tag{9.19}$$

We shall refer to (9.19) as a $\mathbb{C}^m$-*factorization associated with the supporting projection* Π.

Now, let Π be a supporting projection for Θ, let (9.16) be the block matrix representation of Θ induced by Π, and let (9.19) be a $\mathbb{C}^m$-factorization associated with Π. Consider the systems

$$\Theta_1 = (A_1, F, C_1, D_1; \operatorname{Ker}\Pi, \mathbb{C}^m), \tag{9.20}$$

$$\Theta_2 = (A_2, B_2, G, D_2; \operatorname{Im}\Pi, \mathbb{C}^m), \tag{9.21}$$

We call Θ_1 the *left factor* of Θ associated with Π and the given $\mathbb{C}^m$-factorization of $M(\Pi;\Theta)$. Similarly, Θ_2 is called the *right factor* of Θ associated with M and the given $\mathbb{C}^m$-factorization. The next theorem justifies the use of this terminology.

Theorem 9.6. *Let Π be a supporting projection for Θ, let* (9.16) *be the block matrix representation of Θ induced by Π, and let* (9.19) *be a $\mathbb{C}^m$-factorization associated with Π and Θ. Then Θ admits the factorization $\Theta = \Theta_1\Theta_2$, where Θ_1 and Θ_2 are given by* (9.20) *and* (9.21)*, respectively. Conversely, if $\Theta = \Theta_1\Theta_2$ is a factorization of Θ, with*

$$\Theta_1 = (A_1, B_1, C_1, D_1; X_1, \mathbb{C}^m),$$

$$\Theta_2 = (A_2, B_2, C_2, D_2; X_2, \mathbb{C}^m),$$

then the state space of X of Θ is given by $X = X_1 \dot{+} X_2$, the projection Π of X along X_1 onto X_2 is a supporting projection of Θ,

$$M(\Pi;\Theta) = \begin{bmatrix} B_1 \\ D_1 \end{bmatrix} \begin{bmatrix} C_2 & D_2 \end{bmatrix} \tag{9.22}$$

is a $\mathbb{C}^m$-factorization associated with Π and Θ, and the left and right factors of Θ associated with Π and the $\mathbb{C}^m$-factorization (9.22) *are equal to Θ_1 and Θ_2, respectively.*

Proof. Let Θ_1 and Θ_2 be given by (9.20) and (9.21), respectively. By definition (see Section 2.3), the product $\Theta_1\Theta_2$ is given by

$$\Theta_1\Theta_2 = \left(\begin{bmatrix} A_1 & FG \\ 0 & A_2 \end{bmatrix}, \begin{bmatrix} FD_2 \\ B_2 \end{bmatrix}, \begin{bmatrix} C_1 & D_1G \end{bmatrix}, D_1D_2 ; \operatorname{Ker}\Pi \dot{+} \operatorname{Im}\Pi, \mathbb{C}^m \right).$$

The $\mathbb{C}^m$-factorization (9.19) of $M(\Pi;\Theta)$ yields

$$FG = A_{12}, \qquad FD_2 = B_1, \qquad D_1G = C_2, \qquad D_1D_2 = D.$$

Since $X = \operatorname{Ker}\Pi \dot{+} \operatorname{Im}\Pi$, it follows that $\Theta_1\Theta_2$ is precisely equal to the system given by (9.16), that is, $\Theta_1\Theta_2 = \Theta$.

To prove the converse statement, let $\Theta = \Theta_1\Theta_2$, where Θ_1 and Θ_2 are as in the second part of the theorem. Again using the product definition we have

$$\Theta = \left(\begin{bmatrix} A_1 & B_1C_2 \\ 0 & A_2 \end{bmatrix}, \begin{bmatrix} B_1D_2 \\ B_2 \end{bmatrix}, \begin{bmatrix} C_1 & D_1C_2 \end{bmatrix}, D_1D_2 \,;\, \operatorname{Ker}\Pi \dot{+} \operatorname{Im}\Pi, \mathbb{C}^m \right).$$

Then (9.22) holds, and it is straightforward to check that Θ_1 is the left factor and Θ_2 is the right factor of Θ associated with Π and the $\mathbb{C}^m$-factorization (9.22). □

Let Θ be a finite-dimensional system as in (9.13), and let Π be a supporting projection of Θ. Consider the block matrix representation (9.16) of Θ induced by Π. Then the transfer function W_Θ of Θ admits the following representation

$$W_\Theta(\lambda) = \begin{bmatrix} C_1(\lambda - A_1)^{-1} & I_m \end{bmatrix} M(\Pi;\Theta) \begin{bmatrix} (\lambda - A_2)^{-1}B_2 \\ I_m \end{bmatrix}. \tag{9.23}$$

To see this, let (9.19) be a $\mathbb{C}^m$-factorization of $M(\Pi;\Theta)$, and let Θ_1 and Θ_2 be the finite-dimensional systems defined by (9.20) and (9.21), respectively. From Theorem 9.6 we know that $\Theta = \Theta_1\Theta_2$, and hence the transfer function W_Θ of Θ admits the factorization $W_\Theta(\lambda) = W_{\Theta_1}(\lambda)W_{\Theta_2}(\lambda)$. Notice that

$$\begin{aligned} W_{\Theta_1}(\lambda) &= D_1 + C_1(\lambda - A_1)^{-1}F, \\ W_{\Theta_2}(\lambda) &= D_2 + G(\lambda - A_1)^{-1}B_2. \end{aligned}$$

Using (9.19), we see that $W_{\Theta_1}(\lambda)W_{\Theta_2}(\lambda)$ is precisely equal to the right-hand side of (9.23), and hence (9.23) is proved.

In general, the matrix $M(\Pi,\Theta)$ has many different $\mathbb{C}^m$-factorizations. Hence a single supporting projection of Θ yields many different factorizations of the transfer function W_Θ. Given a supporting projection Π of Θ we say that a pair of proper rational $m \times m$ matrix functions $\{W_1, W_2\}$ is a *pair of factors of* W_Θ *induced* by Π if there exists a $\mathbb{C}^m$-factorization of $M(\Pi;\Theta)$ such that $W_1 = W_{\Theta_1}$, where Θ_1 is the left factor of Θ associated with Π and the given factorization, and $W_2 = W_{\Theta_2}$, where Θ_2 is the right factor of Θ associated with Π and the given factorization. In that case, by Theorem 9.6, we have $W_\Theta(\lambda) = W_1(\lambda)W_2(\lambda)$. If W_Θ is regular (that is, when there exists a complex number λ_0, not a pole of W_Θ, such that $\det W_\Theta(\lambda_0) \neq 0$), then the set of factors $\{W_1, W_2\}$ induced by a single supporting projection is relatively simple to describe. We have the following result.

Proposition 9.7. *Let* Θ *be a finite-dimensional system as in* (9.23)*, let* Π *be a supporting projection of* Θ*, let* (9.19) *be a given* $\mathbb{C}^m$*-factorization of* $M(\Pi;\Theta)$*, and*

let Θ_1, Θ_2 *be given by* (9.20) *and* (9.21)*, respectively. Assume* W_Θ *is regular. Then the set of all pairs of factors* $\{W_1, W_2\}$ *of* W_Θ *induced by* Π *is given by*

$$\Big\{\big\{W_{\Theta_1}(\cdot)E,\ E^{-1}W_{\Theta_2}(\cdot)\big\}\ \Big|\ E \text{ is a non-singular } m\times m \text{ matrix}\Big\}. \tag{9.24}$$

Proof. We first show that the regularity of W_Θ implies that

$$\operatorname{rank} M(\Pi;\Theta) = m. \tag{9.25}$$

To do this we use the representation (9.23). Since W_Θ is regular, we can choose $\lambda \in \mathbb{C}$ in such a way that the three matrices $W_\Theta(\lambda)$, $\lambda - A_1$, and $\lambda - A_2$ are all non-singular. For such a choice of λ the equality (9.23) is valid, and its left-hand side has rank equal to m. The latter can only happen when $\operatorname{rank} M(\Pi;\Theta) \geq m$. But Π is a supporting projection. Thus, by definition, $\operatorname{rank} M(\Pi;\Theta) \leq m$. It follows that (9.25) holds.

Next we use the given $\mathbb{C}^m$-factorization (9.19). Since (9.25) holds, any other $\mathbb{C}^m$-factorization of $M(\Pi;\Theta)$ is of the form

$$M(\Pi;\Theta) = \begin{bmatrix} FE \\ D_1E \end{bmatrix} \begin{bmatrix} E^{-1}G & E^{-1}D_2 \end{bmatrix}, \tag{9.26}$$

where E is an arbitrary non-singular $m\times m$ matrix. The representation (9.24) now follows from (9.26) and Theorem 9.6. □

Next we consider minimal factorization. The following result is the analogue of Theorem 9.3 for finite-dimensional systems that are not necessarily biproper.

Theorem 9.8. *Let* Θ *in* (9.13) *be a minimal realization of the rational* $m\times m$ *matrix function* W*.*

(i) *Assume that* Π *is a supporting projection for* Θ*, and let* (9.19) *be a* $\mathbb{C}^m$*-factorization associated with* Π *and* Θ*. Then* $W = W_{\Theta_1}W_{\Theta_2}$*, where* Θ_1 *is the left factor and* Θ_2 *is the right factor associated with* Π *and the given* $\mathbb{C}^m$*-factorization* (9.19)*, and the factorization* $W = W_{\Theta_1}W_{\Theta_2}$ *is minimal.*

(ii) *Assume that* $W = W_1W_2$ *is a minimal factorization of* W*. Then there exists a unique supporting projection* Π *for* Θ *and a unique* $\mathbb{C}^m$*-factorization associated with* Π *and* Θ *such that* $W_1 = W_{\Theta_1}$ *and* $W_2 = W_{\Theta_2}$*, where* Θ_1 *is the left factor and* Θ_2 *is the right factor of* Θ *associated with* Π *and the given* $\mathbb{C}^m$*-factorization.*

Proof. We split the proof in three parts. In the first part we prove statement (i), the two other parts concern item (ii).

Part 1. Let the conditions of (i) be satisfied. Then it follows from the first part of Theorem 9.6 that $\Theta = \Theta_1\Theta_2$, and hence $W_\Theta = W_{\Theta_1}W_{\Theta_2}$. Since $W = W_\Theta$ and Θ is minimal, we can apply Theorem 8.27 to show that $W = W_{\Theta_1}W_{\Theta_2}$ is a minimal factorization.

Part 2. Let $W = W_1 W_2$ be a minimal factorization. Since $W = W_\Theta$, we know that W is analytic at infinity. The minimality of the factorization then implies that the same holds true for factors W_1 and W_2. Let Θ_1' and Θ_2' be minimal realizations of W_1 and W_2, respectively. Since the factorization $W = W_1 W_2$ is minimal, it follows (see Theorem 8.27) that $\Theta' = \Theta_1' \Theta_2'$ is a minimal realization of $W = W_\Theta$. Using the minimality of Θ, it follows that Θ and Θ' are similar, and the similarity from Θ to Θ' is unique. Write

$$\Theta_j' \;=\; (A_j', B_j', C_j', D_j'; X_j', \mathbb{C}^m), \qquad j = 1, 2.$$

Then

$$\Theta' = \left(\begin{bmatrix} A_1' & B_1' C_2' \\ 0 & A_2' \end{bmatrix}, \begin{bmatrix} B_1' D_2 \\ B_2' \end{bmatrix}, \begin{bmatrix} C_1' & D_1 C_2' \end{bmatrix}, D_1 D_2 \,;\, X_1' \dot{+} X_2', \mathbb{C}^m \right).$$

Put $X' = X_1' \dot{+} X_2'$, and let Π' be the projection of X' along X_1' unto X_2'. Notice that

$$M(\Pi'; \Theta') = \begin{bmatrix} B_1' \\ D_1 \end{bmatrix} \begin{bmatrix} C_2' & D_2 \end{bmatrix}.$$

Now, let S from X to X' be the (unique) similarity from Θ to Θ', and put $\Pi = S^{-1} \Pi' S$. Then Π is a supporting projection for Θ, and

$$M(\Pi, \Theta) = \begin{bmatrix} S_1^{-1} & 0 \\ 0 & I_m \end{bmatrix} M(\Pi'; \Theta') \begin{bmatrix} S_2 & 0 \\ 0 & I_m \end{bmatrix}$$

where S_1 is the restriction of S to $\operatorname{Ker} \Pi$ viewed as a map from $\operatorname{Ker} \Pi$ into $\operatorname{Ker} \Pi'$, and S_2 is the restriction of S to $\operatorname{Im} \Pi$ viewed as a map from $\operatorname{Im} \Pi$ into $\operatorname{Im} \Pi'$. Both S_1 and S_2 are invertible. Put

$$F = S_1^{-1} B_1', \qquad G = C_2' S_2.$$

Then

$$M(\Pi; \Theta) = \begin{bmatrix} F \\ D_1 \end{bmatrix} \begin{bmatrix} G & D_2 \end{bmatrix}. \tag{9.27}$$

Let Θ_1 be the left factor and Θ_2 the right factor associated with Π and the $\mathbb{C}^m$-factorization (9.27). Then it is straightforward to check that S_1 is a similarity from Θ_1 to Θ_1', and that S_2 is a similarity from Θ_2 to Θ_2'. We conclude that $W_{\Theta_1} = W_{\Theta_1'}$ and $W_2 = W_{\Theta_2}$, as desired.

Part 3. It remains to prove the uniqueness of Π and the $\mathbb{C}^m$-factorization associated with Π and Θ. To prove this uniqueness, let $\widetilde{\Pi}$ be a supporting projection for Θ, and let

$$M(\widetilde{\Pi}; \Theta) = \begin{bmatrix} \widetilde{F} \\ \widetilde{D}_1 \end{bmatrix} \begin{bmatrix} \widetilde{G} & \widetilde{D}_2 \end{bmatrix} \tag{9.28}$$

be a corresponding $\mathbb{C}^m$-factorization such that $W_1 = W_{\widetilde{\Theta}_1}$ and $W_2 = W_{\widetilde{\Theta}_2}$, where $\widetilde{\Theta}_1$ is the left factor and $\widetilde{\Theta}_2$ the right factor associated with $\widetilde{\Pi}$ and the $\mathbb{C}^m$-factorization (9.28). First of all, note that

$$D_1 = W_1(\infty) = W_{\widetilde{\Theta}_1}(\infty) = \widetilde{D}_1, \qquad D_2 = W_2(\infty) = W_{\widetilde{\Theta}_2}(\infty) = \widetilde{D}_2.$$

For $j = 1, 2$ the systems Θ_j and $\widetilde{\Theta}_j$ are minimal realizations of W_j. Hence there exists a similarity S_j from Θ_j to $\widetilde{\Theta}_j$. From the first part of the proof we know that $\Theta = \widetilde{\Theta}_1\widetilde{\Theta}_2$ and $\Theta = \Theta_1\Theta_2$. Thus

$$S = \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}$$

is a self-similarity of Θ. For minimal systems the similarity is unique. Thus S is an identity operator. Hence both S_1 and S_2 are identity operators. It follows that $\Theta_1 = \widetilde{\Theta}_1$ and $\Theta_2 = \widetilde{\Theta}_2$. We conclude that $\Pi = \widetilde{\Pi}$ and that the $\mathbb{C}^m$ factorizations (9.27) and (9.28). □

Theorem 9.8 contains Theorem 9.3 as a special case. To see this, recall that in Theorem 9.3 the systems are required to be unital, and the rational matrix functions are assumed to be proper and to have the value I_m at infinity. Now let Θ be a unital finite-dimensional system, and let Π be a supporting projection for Θ. Then the transfer function $W = W_\Theta$ is regular, and hence Proposition 9.7 applies. It follows that the set of all pairs of factors $\{W_1, W_2\}$ of W_Θ induced by Π is given by (9.24). But in Theorem 9.3 we are only interested in factors that have the value I_m at infinity. This restricts the choice of the invertible matrix E in the set (9.24) to one matrix only, namely to the $m \times m$ identity matrix. It follows that (9.24) contains only one pair of factors $\{W_1, W_2\}$ with W_1 and W_2 having the value I_m at infinity. From these remarks we see that Theorem 9.3 is covered by Theorem 9.8.

We conclude this section with an example which is not covered by Proposition 9.7. Let W be the 2×2 rational matrix function given by

$$W(\lambda) = \begin{bmatrix} 0 & 0 \\ \dfrac{1-\lambda^2}{\lambda^2} & \dfrac{1-\lambda^2}{\lambda^2} \end{bmatrix}.$$

A minimal realization of W is provide by the following system

$$\Theta = \left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ -1 & -1 \end{bmatrix} ; \mathbb{C}^2, \mathbb{C}^2 \right).$$

Let Π be the projection of $\mathbb{C}^2$ along the first coordinate space onto the second. Identifying both $\operatorname{Ker}\Pi$ and $\operatorname{Im}\Pi$ with $\mathbb{C}$, the matrix $M(\Pi;\Theta)$ is given by

$$M(\Pi;\Theta) = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & -1 & -1 \end{bmatrix}$$

Note that $\operatorname{rank} M(\Pi;\Theta) = 1$, and hence (9.25) is not fulfilled in this case. Since W is not regular, Proposition 9.7 also does not apply. To illustrate this further we consider the following $\mathbb{C}^2$-factorizations of $M(\Pi;\Theta)$:

$$M(\Pi;\Theta) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

These two $\mathbb{C}^2$ factorizations of $M(\Pi;\Theta)$ yield, respectively, the following factorizations of W:

$$W(\lambda) = \begin{bmatrix} 0 & 0 \\ \dfrac{\lambda+1}{\lambda} & \dfrac{\lambda+1}{\lambda} \end{bmatrix} \begin{bmatrix} \dfrac{1}{\lambda} & \dfrac{1}{\lambda} \\ -1 & -1 \end{bmatrix}, \tag{9.29}$$

$$W(\lambda) = \begin{bmatrix} 0 & 1 \\ \dfrac{\lambda+1}{\lambda} & 0 \end{bmatrix} \begin{bmatrix} \dfrac{1-\lambda}{\lambda} & \dfrac{1-\lambda}{\lambda} \\ 0 & 0 \end{bmatrix}. \tag{9.30}$$

Since Θ is a minimal realization of W, the two factorizations above are both minimal. Note that the first factor in (9.30) is regular while the first factor in (9.30) is not. Thus the set of all pairs of factors of $W = W_\Theta$ induced by Π cannot not be described by a formula like (9.24).

Notes

The material in Section 9.1 is taken from the first part of Chapter 4 in [14], in particular from Section 4.3. Section 9.2 is based on the paper [102]; see also the dissertation [103] which contains some additional applications of pseudo-canonical factorization. For connections between contractive matrix functions and pseudo-canonical factorization, see [72]. Section 9.3 originates from [35].

Part III

Degree One Factors, Companion Based Rational Matrix Functions, and Job Scheduling

This part is devoted to the study of factorization into degree one factors, that is, into factors that have a minimal realization with a state space of dimension one. A second main theme is the connection between the problem of degree one factorization and a problem of job scheduling from operations research.

There are three chapters (10, 11 and 12) in this part. In Chapter 10 the problem to factorize a rational matrix function in degree one factors is analyzed in a state space setting. The notions of complete and quasicomplete degree one factorizations are introduced. In general, the latter factorizations are non-minimal. The results are specified further for so-called companion based matrix functions in Chapter 11. Finally, in Chapter 12 it is shown that the issue of quasicomplete degree one factorization of companion based matrix functions is intimately connected to a particular job scheduling problem, namely the two machine flow shop problem. Maple procedures to calculate degree one factorizations for companion based matrix functions complement the text.

Chapter 10

Factorization into Degree One Factors

In this chapter we study the factorization of a proper rational $m \times m$ matrix function having the value I_m at infinity into elementary factors satisfying the same constraints. These elementary factors are of McMillan degree one by definition. It turns out, by using realization, that the problem of factorizing a function in such degree one factors is intimately connected with the issue of simultaneous reduction to complementary triangular forms of pairs of matrices. We prove that factorization into elementary factors is always possible.

In general, however, pole-zero cancellations occur so that the factorizations in question are non-minimal. This is further underlined by the fact that one has to allow for the introduction of new poles and zeros not present in the given function. Such new poles and zeros do not occur in the situation where the factorization has the additional property of being minimal. In that case the factorization is called complete and the number of elementary factors in it is equal to the McMillan degree of the function that is factorized. In general, complete factorization is not possible. A quasicomplete factorization is one where the number of elementary factors is as small as possible. The number of factors involved is called the quasidegree and we give an upper bound for it. Examples are presented to illustrate the material.

This chapter consists of four sections. The main topic of the first section is simultaneous reduction to complementary triangular forms of pairs of matrices. A number of conditions for such reductions to exist are given. In the second section factorization into elementary factors is studied in terms of realizations, and the connection with simultaneous reduction to complementary triangular forms is described. The third section is devoted to complete factorizations and the final section deals with quasicomplete factorizations.

10.1 Simultaneous reduction to complementary triangular forms

We say that two complex $n \times n$ matrices A and Z *admit simultaneous reduction to complementary triangular forms* if there exists an invertible $n \times n$ matrix S such that $S^{-1}AS$ is an upper triangular matrix and $S^{-1}ZS$ is a lower triangular matrix. As will become clear in the next section, this definition is inspired by Theorem 2.6. It will turn out to be useful in the study of factorizations of rational matrix functions involving factors of McMillan degree one only.

We begin with a proposition presenting a number of equivalent conditions for the above notion. The first is geometric in nature and requires the following terminology. A chain $M_0 \subset M_1 \subset \cdots \subset M_n$ of subspaces of $\mathbb{C}^n$ is called *complete* if $\dim M_j = j$ for $j = 0, 1, \ldots, n$ (in particular, $M_0 = \{0\}$ and $M_n = \mathbb{C}^n$).

Proposition 10.1. *Let A and Z be complex $n \times n$ matrices. Then A and Z admit simultaneous reduction to complementary triangular forms if and only if one (or all) of the following conditions is (are) satisfied.*

(a) *There exist complete chains $M_0 \subset M_1 \subset \cdots \subset M_n$ and $N_0 \subset N_1 \subset \cdots \subset N_n$ of subspaces of $\mathbb{C}^n$ such that, for $j = 1, \ldots, n$,*

$$AM_j \subset M_j, \qquad ZN_j \subset N_j, \qquad M_j \dotplus N_{n-j} = \mathbb{C}^n.$$

(b) *There exist bases $f_1, \ldots, f_n$ and $g_1, \ldots, g_n$ of $\mathbb{C}^n$ such that, for $j = 1, \ldots, n$, $f_1, \ldots, f_j, g_{j+1}, \ldots, g_n$ is a basis for $\mathbb{C}^n$ while, in addition,*

$$Af_j \in \operatorname{span}\{f_1, \ldots, f_j\}, \qquad Zg_j \in \operatorname{span}\{g_j, \ldots, g_n\}.$$

(c) *There exist invertible $n \times n$ matrices F and G such that $F^{-1}AF$ is upper triangular, $G^{-1}ZG$ is lower triangular, and $G^{-1}F$ admits a lower-upper factorization.*

Proof. We split the proof into four parts corresponding to the following list of implications: (b) $\Rightarrow$ (a) $\Rightarrow$ (SR) $\Rightarrow$ (c) $\Rightarrow$ (b). Here (SR) is a shorthand notation for the property that A and Z admit simultaneous reduction to complementary triangular forms.

Part 1. We prove (b) $\Rightarrow$ (a). Let $f_1, \ldots, f_n$ and $g_1, \ldots, g_n$ be bases of $\mathbb{C}^n$ as in (b). For $j = 1, \ldots n$, put

$$M_j = \operatorname{span}\{f_1, \ldots, f_j\}, \qquad N_j = \operatorname{span}\{g_{n-j+1}, \ldots, g_n\}.$$

Then $M_0 \subset M_1 \subset \cdots \subset M_n$ and $N_0 \subset N_1 \subset \cdots \subset N_n$ are complete chains of subspaces of $\mathbb{C}^n$. Note that for each $j = 1, \ldots n$, the statement that the vectors $f_1, \ldots, f_j, g_{j+1}, \ldots, g_n$ form a basis for $\mathbb{C}^n$ is equivalent to the equality $M_j \dotplus N_{n-j} = \mathbb{C}^n$. Finally, the inclusion properties in (b) show that the spaces M_j and N_j are invariant under A and Z, respectively. Thus (a) holds.

Part 2. We prove (a) $\Rightarrow$ (SR). Let $M_0 \subset M_1 \subset \cdots \subset M_n$ and $N_0 \subset N_1 \subset \cdots \subset N_n$ be complete chains of subspaces of $\mathbb{C}^n$ with the additional properties mentioned in (a). It is an elementary matter to check that

$$\dim\big(M_j \cap N_{n-j+1}\big) = 1, \qquad j = 1, \ldots, n.$$

Pick $s_j \neq 0$ from $M_j \cap N_{n-j+1}$, and let $S = \begin{bmatrix} s_1 & s_2 & \cdots & s_n \end{bmatrix}$, that is, S is the matrix of which the jth column is equal to s_j where $j = 1, \ldots, n$. Then $s_1, \ldots, s_n$ form a basis for $\mathbb{C}^n$ and S is invertible. The invariance conditions on the subspaces M_j and N_j imply that $S^{-1}AS$ is upper triangular and $S^{-1}ZS$ is lower triangular. Thus (SR) holds.

Part 3. We prove (SR) $\Rightarrow$ (c). Let S be an invertible matrix such that $S^{-1}AS$ is upper triangular and $S^{-1}ZS$ is lower triangular. Put $F = G = S$. Then all conditions of (c) are fulfilled.

Part 4. We prove (c) $\Rightarrow$ (b). Let F and G be invertible $n \times n$ matrices such that $F^{-1}AF$ is upper triangular and $G^{-1}ZG$ is lower triangular. Suppose $G^{-1}F$ can be written as LU where L is an invertible lower triangular matrix and U is an invertible upper triangular matrix. Put $S = FU^{-1}$. Then $S^{-1}AS = U\big(F^{-1}AF\big)U^{-1}$ is upper triangular. Since $S = GL$, the matrix product $S^{-1}ZS$ is equal to the matrix $L^{-1}\big(G^{-1}ZG\big)L$, and hence $S^{-1}ZS$ is lower triangular. Now, for $j = 1, \ldots, n$, take $f_j = g_j = s_j$, where s_j is the jth column of S. Then $f_1, \ldots, f_n$ and $g_1, \ldots, g_n$ have all the properties required in (b). □

Next we present two theorems with sufficient conditions for a pair of matrices to admit simultaneous reduction to complementary triangular forms.

Theorem 10.2. *Let A and Z be complex $n \times n$ matrices, one of which is diagonalizable. Then A and Z admit simultaneous reduction to complementary triangular forms. In fact, if A is diagonalizable, then, given an ordering $\zeta_1, \ldots, \zeta_n$ of the eigenvalues of Z (algebraic multiplicities taken into account) there exists an invertible $n \times n$ matrix S such that $S^{-1}AS$ is upper triangular and $S^{-1}ZS$ is lower triangular with diagonal elements $\zeta_1, \ldots, \zeta_n$ (again algebraic multiplicities taken into account).*

Here and elsewhere diagonal elements of matrices are read from top left to bottom right.

It suffices to prove the second part of the theorem. Indeed, if two $n \times n$ matrices A and Z admit simultaneous reduction to complementary triangular forms, then so do Z and A. To see this, use the $n \times n$ reversed identity matrix having ones on the antidiagonal and zeros everywhere else (cf., the proof of Corollary 10.7 below).

Proof of Theorem 10.2. For $n = 1$, there is nothing to prove. So assume n is at least two. Whenever convenient, we shall view A and Z as linear operators on $\mathbb{C}^n$.

Let $a_1, \dots, a_n$ be a basis in $\mathbb{C}^n$ such that the matrix representation of A with respect to this basis has diagonal form. So

$$Aa_j = \alpha_j a_j, \qquad j = 1, \dots, n,$$

where $\alpha_1, \dots, \alpha_n$ are the eigenvalues of A counted according to algebraic multiplicity. Also, let $z_1, \dots, z_n$ be a basis in $\mathbb{C}^n$ such that the matrix representation of Z with respect to this basis has lower triangular form. So, for $j = 1, \dots, n$, the vector Zz_j is in the linear hull of $z_j, \dots, z_n$. In particular z_n is an eigenvector of Z. In fact, the basis $z_1, \dots, z_n$ can be chosen in such a way that $Zz_n = \zeta_n z_n$. We may assume too that the basis $a_1, \dots, a_n$ is ordered in such a way that the vectors $a_1, \dots, a_{n-1}, z_n$ form a basis of $\mathbb{C}^n$ as well. Let X_0 be the linear hull of $a_1, \dots, a_{n-1}$ and let X_1 be the linear hull of the single vector z_n. Then $\mathbb{C}^n = X_0 \dot{+} X_1$ and with respect to this decomposition, the matrix representations of A and Z have the form

$$\begin{bmatrix} A_0 & A_+ \\ 0 & A_1 \end{bmatrix}, \qquad \begin{bmatrix} Z_0 & 0 \\ Z_- & Z_1 \end{bmatrix}.$$

The space X_0 has dimension $n-1$ and with respect to its basis $a_1, \dots, a_{n-1}$, the linear operator A_0 has diagonal form. Further, the eigenvalues of Z_0 are $\zeta_1, \dots, \zeta_{n-1}$. We may assume (using induction) that there is a basis $u_1, \dots, u_{n-1}$ in X_0 with respect to which A_0 has upper triangular and Z_0 has lower triangular form with $\zeta_1, \dots, \zeta_{n-1}$ on the diagonal (read from top left to bottom right). But then $u_1, \dots, u_{n-1}, z_n$ is a basis of $\mathbb{C}^n$ for which the matrix representations of A and Z are upper and lower triangular, respectively, and the proof is complete. □

Theorem 10.3. *Let A and Z be complex $n \times n$ matrices. Suppose A and Z have no common eigenvalue and, in addition,* $\operatorname{rank}(A - Z) = 1$. *Then A and Z admit simultaneous reduction to complementary triangular forms. In fact, given an ordering $\alpha_1, \dots, \alpha_n$ of the eigenvalues of A and an ordering $\zeta_1, \dots, \zeta_n$ of the eigenvalues of Z (in both cases algebraic multiplicities taken into account), there exists an invertible $n \times n$ matrix S such that $S^{-1}AS$ is upper triangular with diagonal elements $\alpha_1, \dots, \alpha_n$ and $S^{-1}ZS$ is lower triangular with diagonal elements $\zeta_1, \dots, \zeta_n$.*

The above theorem will be proved in the next section, by using the connection between reduction to complementary triangular forms and factorization into elementary factors.

Another sufficient condition for a pair of matrices to admit simultaneous reduction to complementary triangular forms is presented in Theorem 10.14 of Section 10.3.

The equivalence of statements (b) and (c) in Proposition 10.1 can also be obtained as a corollary of the following more general result. For the sake of completeness we present the result with a full proof, although it will not play a role in the sequel.

Proposition 10.4. *Let F and G be invertible $n \times n$ complex matrices, and for $j = 1, \ldots, n$, let f_j and g_j be the jth column of F and G, respectively. Then $G^{-1}F$ admits a lower-upper factorization if and only if for j running from 1 to $n-1$, the vectors $f_1, \ldots, f_j, g_{j+1}, \ldots, g_n$ form a basis for $\mathbb{C}^n$.*

Proof. Suppose $G^{-1}F$ admits a lower-upper factorization, say $G^{-1}F = LU$ with L and U invertible matrices, L lower triangular, U upper triangular. Putting $H = FU^{-1} = GL$ we obtain an invertible $n \times n$ matrix. Clearly $F = HU$ and $G = HL^{-1}$. Let j be an integer among $1, \ldots, n-1$, let U_+ be the $j \times j$ matrix obtained from U by omitting the last $n-j$ rows and columns, and let the $(n-j) \times (n-j)$ matrix L_- be obtained from L^{-1} by omitting the first j rows and columns. Since U is invertible and upper triangular, the matrix U_+ is invertible. Similarly, as L^{-1} is invertible and lower triangular, the matrix L_- is invertible. Now

$$\begin{bmatrix} f_1 & \cdots & f_j & g_{j+1} & \cdots & g_n \end{bmatrix} = H \begin{bmatrix} U_+ & 0 \\ 0 & L_- \end{bmatrix},$$

and it follows that the matrix $\begin{bmatrix} f_1 & \cdots & f_j & g_{j+1} & \cdots & g_n \end{bmatrix}$ is invertible. This proves the only if part of the proposition.

Moving on to the if part, we assume that for j running from 1 to n, the n vectors $f_1, \ldots, f_j, g_{j+1}, \ldots, g_n$ are linearly independent, i.e., the matrix

$$R_j = \begin{bmatrix} f_1 & \cdots & f_j & g_{j+1} & \cdots & g_n \end{bmatrix}$$

is invertible. Putting $R_0 = G$ and $R_n = F$, we have

$$G^{-1}F = \big(R_0^{-1}R_1\big) \cdots \big(R_{n-2}^{-1}R_{n-1}\big)\big(R_{n-1}^{-1}R_n\big).$$

Now, by a straightforward argument,

$$R_{j-1}^{-1}R_j = I_n + r_j e_j^\top, \qquad j = 1, \ldots, n,$$

where $r_j = R_{j-1}^{-1}f_j$ and e_j is the jth unit vector in $\mathbb{C}^n$ (having 1 on the jth position and zeros everywhere else). Note that $1 + e_j^\top r_j \neq 0, \;\; j = 1, \ldots, n$ as the matrix $R_{j-1}^{-1}R_j = I_n + r_j e_j^\top$ is invertible and its determinant is equal to $1 + e_j^\top r_j$. It now suffices to prove that the kth leading principal minor of the matrix

$$\big(I_n + r_1 e_1^\top\big) \cdots \big(I_n + r_{n-1}e_{n-1}^\top\big)\big(I_n + r_n e_n^\top\big) \tag{10.1}$$

is given by $(1 + e_1^\top r_1) \cdots (1 + e_k^\top r_k), \;\; k = 1, \ldots, n$. Indeed, this follows from the well-known fact that a matrix allows lower-upper factorization if and only if all its leading principal minors are invertible.

The argument for this employs induction (on n). For $n = 1$, the statement is obviously correct. Suppose $n \geq 2$. The nth leading principal minor and the determinant of (10.1) coincide and the value in question is

$$(1 + e_1^\top r_1) \cdots (1 + e_{n-1}^\top r_{n-1})(1 + e_n^\top r_n),$$

as desired. Now let us look at the other leading principal minors of (10.1). These coincide with the leading principal minors of the matrix R obtained from (10.1) by omitting its last column and row. For $j = 1, \dots, n-1$, the last column of $I_n + r_j e_j^\top$ is e_n. Also the matrix obtained from $I_n + r_n e_n^\top$ by omitting the last column and row is I_{n-1}. It follows that

$$R = \big(I_{n-1} + \widehat{r}_1 \widehat{e}_1^\top\big) \cdots \big(I_{n-1} + \widehat{r}_{n-1} \widehat{e}_{n-1}^\top\big),$$

where $\widehat{r}_j \in \mathbb{C}^{n-1}$ is obtained from r_j by omitting the last component and the vectors $\widehat{e}_1, \dots, \widehat{e}_{n-1}$ are the unit vectors in $\mathbb{C}^{n-1}$. For $k = 1, \dots, n-1$, (by induction hypothesis) the kth leading principal minor of R is $(1 + \widehat{e}_1^\top \widehat{r}_1) \cdots (1 + \widehat{e}_k^\top \widehat{r}_k)$. But this is clearly equal to $(1 + e_1^\top r_1) \cdots (1 + e_k^\top r_k)$, again as desired. □

10.2 Factorization into elementary factors and realization

A rational $m \times m$ matrix function E is called *elementary* whenever the McMillan degree of E is equal to one. Throughout this chapter an elementary rational $m \times m$ matrix function is also assumed to be proper and to have the value I_m at infinity. This allows us to write such a function E in the form

$$E(\lambda) = I_m + \frac{1}{\lambda - \alpha} R, \tag{10.2}$$

where R is a rank one $m \times m$ matrix and α is the unique pole of E which is necessarily simple (i.e., of order one). To describe the inverse $E^{-1}(\lambda) = E(\lambda)^{-1}$ of $E(\lambda)$, we put $\alpha^\times = \alpha - \operatorname{trace} R$. Then $E(\lambda)$ is invertible if and only if $\lambda \neq \alpha^\times$, and in that case

$$E^{-1}(\lambda) = I_m - \frac{1}{\lambda - \alpha^\times} R. \tag{10.3}$$

This identity can be verified by direct computation, but one can also make a connection with the material developed in Section 2.2. Here are the details. Write R in the form $R = cb^\top$, where b and c are nonzero vectors in $\mathbb{C}^m$ and the superscript $\top$ denotes the operation of taking the transpose. Then $E(\lambda) = I_m + c(\lambda - \alpha)^{-1} b^\top$, which is a minimal realization of E with state space $\mathbb{C}$ and main operator (the multiplication by) α. Now $\operatorname{trace} R = b^\top c$, so $\alpha^\times = \alpha - b^\top c$, and we can apply Theorem 2.1. Clearly the function E^{-1} is again elementary and has $\alpha^\times = \alpha - \operatorname{trace} R$ as its (unique) pole.

As we shall see in Section 10.4 below (see Theorem 10.15) any proper rational $m \times m$ matrix function W with $W(\infty) = I_m$ can be written as a product of elementary factors. For the scalar case ($m = 1$) this fact is easy to prove.

Indeed, let w be a scalar rational function with $w(\infty) = 1$. Then w is the quotient of two scalar polynomials $a^\times$ and a with the same leading coefficient 1 and of the same degree, n say. We shall assume, as we may do without loss of

generality, that these polynomials are relatively prime, i.e., they have no common zero. Writing

$$a^\times(\lambda) = (\lambda - a_1^\times)\cdots(\lambda - a_n^\times), \qquad a(\lambda) = (\lambda - a_1)\cdots(\lambda - a_n),$$

we have

$$w(\lambda) = \frac{a^\times(\lambda)}{a(\lambda)} = \left(1 + \frac{\alpha_1 - \alpha_1^\times}{\lambda - \alpha_1}\right)\cdots\left(1 + \frac{\alpha_n - \alpha_n^\times}{\lambda - \alpha_n}\right), \tag{10.4}$$

and this is a factorization of w into n scalar rational functions of McMillan degree one, that is, a factorization into n scalar elementary factors. Note that in this case n is precisely equal to the McMillan degree of w, and hence the factorization is a minimal one. Factorizations into elementary factors that are minimal are of special interest and will be studied in Section 10.3.

In the present section we analyze factorization into elementary factors in terms of realizations. The following theorem, which is inspired by Theorem 2.6, is our first main result. It clarifies the connection between simultaneous reduction to complementary triangular forms (for pairs of matrices) and factorization into elementary factors.

Theorem 10.5. *Let W be a proper rational $m \times m$ matrix function with $W(\infty) = I_m$, and let n be a non-negative integer. The following statements are equivalent:*

(i) *W admits a factorization into at most n elementary factors,*

(ii) *W admits a factorization into precisely n elementary factors,*

(iii) *W has a realization $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ such that A and $A^\times (= A - BC)$ are upper and lower triangular, respectively,*

(iv) *W has a realization $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ such that A and $A^\times (= A - BC)$ admit simultaneous reduction to complementary triangular forms.*

We begin with a lemma that will be needed in the proof.

Lemma 10.6. *Each elementary rational $m \times m$ matrix function can be written as the product of two functions of the same type.*

Proof. Consider E given by (10.2) with $\operatorname{rank} R = 1$. Write R in the form $R = cb^\top$, where b and c are nonzero vectors in $\mathbb{C}^m$. Let β be a complex number different from α and $\alpha^\times = \alpha - b^\top c$, and choose $f \in \mathbb{C}^m$ such that $f^\top c = \beta - \alpha^\times$. As $\beta - \alpha^\times \neq 0$, the vector f is a nonzero vector in $\mathbb{C}^m$. Also $(f^\top - b^\top)c = \beta - \alpha$ and $f^\top - b^\top \neq 0$. Now introduce the rank one matrices $R_1 = c(b^\top - f^\top)$ and $R_2 = cf^\top$. Then

$$R_1 + R_2 = c(b^\top - f^\top) + cf^\top = cb^\top = R,$$

$$R_1R_2 = c(b^\top - f^\top)cf^\top = c[(b^\top - f^\top)c]f^\top = (\alpha - \beta)cf^\top = (\alpha - \beta)R_2.$$

This yields

$$I_m + \frac{1}{\lambda-\alpha}R = \left(I_m + \frac{1}{\lambda-\alpha}R_1\right)\left(I_m + \frac{1}{\lambda-\beta}R_2\right), \tag{10.5}$$

as can be verified via a straightforward computation. □

Note that each of the statements (i)-(iv) in Theorem 10.5 implies that n is larger than or equal to $\delta(W)$, the McMillan degree of W. For (i) and (ii) this follows from the sublogarithmic property (8.50) of the McMillan degree, for (iii) and (iv) from Theorem 8.26. Further relevant details will be provided in the proof below.

Proof of Theorem 10.5. The case $n = 0$ corresponds to the trivial situation where W is constant with value I_m. Therefore we assume n to be positive. The proof will be divided into three parts.

Part 1. First let us note some simple relations between the statements (i)–(iv). The implications (ii) $\Rightarrow$ (i) and (iii) $\Rightarrow$ (iv) are trivial. The implication (iv) $\Rightarrow$ (iii) comes about by applying the appropriate similarity transformation to the given realization. Hence (iii) and (iv) amount to the same. If W can be written as the product of at most n elementary factors, then Lemma 10.6 guarantees that (at the possible expense of introducing additional poles) W also admits a factorization into precisely n elementary factors. Thus (i) $\Rightarrow$ (ii), and we conclude that (i) and (ii) are equivalent. It remains to prove (ii) $\Rightarrow$ (iii) and (iii) $\Rightarrow$ (i).

Part 2. We prove (ii) $\Rightarrow$ (iii). Suppose W can be written as a product of elementary factors,

$$W(\lambda) = \left(I_m + \frac{1}{\lambda-\alpha_1}R_1\right)\cdots\left(I_m + \frac{1}{\lambda-\alpha_n}R_n\right). \tag{10.6}$$

For $j = 1, \ldots, n$, write $R_j = c_j b_j^\top$ with b_j and c_j nonzero vectors in $\mathbb{C}^m$, so that (10.6) becomes

$$W(\lambda) = \left(I_m + c_1(\lambda-\alpha_1)^{-1}b_1^\top\right)\cdots\left(I_m + c_n(\lambda-\alpha_n)^{-1}b_n^\top\right).$$

Define matrices A, B and C by

$$A = \begin{bmatrix} \alpha_1 & b_1^\top c_2 & \cdots & b_1^\top c_n \\ 0 & \alpha_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & b_{n-1}^\top c_n \\ 0 & \cdots & 0 & \alpha_n \end{bmatrix}, \tag{10.7}$$

$$B = \begin{bmatrix} b_1^\top \\ b_2^\top \\ \vdots \\ b_n^\top \end{bmatrix}, \qquad C = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix}.$$

Then $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$, as can be seen by a repeated application of Theorem 2.2. Next, put $\alpha_j^\times = \alpha_j - \operatorname{trace} R_j$ $(j = 1, \dots, n)$. Then

$$A^\times = A - BC = \begin{bmatrix} \alpha_1^\times & 0 & \cdots & 0 \\ -b_2^\top c_1 & \alpha_2^\times & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ -b_n^\top c_1 & \cdots & -b_n^\top c_{n-1} & \alpha_n^\times \end{bmatrix}. \tag{10.8}$$

Note that the matrix A is upper triangular while $A^\times$ is lower triangular. Thus (ii) implies (iii).

Part 3. We prove (iii) $\Rightarrow$ (i) We begin by relating the notion of simultaneous reduction to complementary triangular forms with Theorem 2.6. The matrices A and Z admit simultaneous reduction to complementary triangular forms if and only if there exist a chain

$$\{0\} \subset M_1 \subset \cdots \subset M_{n-1} \subset \mathbb{C}^n$$

of invariant subspaces for A and a chain

$$\{0\} \subset N_1 \subset \cdots \subset N_{n-1} \subset \mathbb{C}^n$$

of invariant subspaces for Z such that, for $j = 1, \dots, n-1$, the spaces M_j and N_j have (the same) dimension j while, moreover, $\mathbb{C}^n = M_j \dot{+} N_{n-j}$. But this, in turn, is equivalent to the existence of mutually disjoint rank one projections $\Pi_1, \dots, \Pi_n$ of $\mathbb{C}^n$ such that $\Pi_1 + \cdots + \Pi_n = \mathbb{C}^n$ and, for $j = 1, \dots, n-1$,

$$\begin{aligned} A[\operatorname{Ker}(\Pi_{j+1} + \cdots + \Pi_n)] &\subset \operatorname{Ker}(\Pi_{j+1} + \cdots + \Pi_n), \\ Z[\operatorname{Im}(\Pi_{j+1} + \cdots + \Pi_n)] &\subset \operatorname{Im}(\Pi_{j+1} + \cdots + \Pi_n). \end{aligned}$$

From this the connection with Theorem 2.6 is obvious.

Now assume that (iii) is satisfied, and let

$$W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B \tag{10.9}$$

be a realization with A and $A^\times$ upper and lower triangular, respectively. From what we saw in the previous paragraph and Theorem 2.6, it is already clear that (i) holds, i.e., W admits a factorization into at most n elementary factors. For later reference, however, it is useful to give some details.

Write

$$A = \begin{bmatrix} \alpha_1 & a_{12} & \cdots & a_{1n} \\ 0 & \alpha_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1,n} \\ 0 & \cdots & 0 & \alpha_n \end{bmatrix}, \tag{10.10}$$

$$A^\times = \begin{bmatrix} \alpha_1^\times & 0 & \cdots & 0 \\ a_{21}^\times & \alpha_2^\times & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{m1}^\times & \cdots & a_{m,m-1}^\times & \alpha_n^\times \end{bmatrix}, \tag{10.11}$$

$$B = \begin{bmatrix} b_1^\top \\ b_2^\top \\ \vdots \\ b_n^\top \end{bmatrix}, \qquad C = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix}, \tag{10.12}$$

where $b_1, \ldots, b_n$ and $c_1, \ldots, c_n$ are in $\mathbb{C}^m$. Since $A^\times = A - BC$, it follows that

$$\begin{array}{rcll} \alpha_j^\times & = & \alpha_j - b_j^\top c_j, & j = 1, \ldots, n, \\ a_{ij} & = & b_i^\top c_j, & i, j = 1, \ldots, n, \ \ i < j, \\ a_{ij}^\times & = & -b_i^\top c_j, & i, j = 1, \ldots, n, \ \ i > j. \end{array}$$

From this we conclude that

$$W(\lambda) = \left(I_m + \frac{1}{\lambda - \alpha_1} c_1 b_1^\top\right) \cdots \left(I_m + \frac{1}{\lambda - \alpha_n} c_n b_n^\top\right), \tag{10.13}$$

$$W^{-1}(\lambda) = \left(I_m - \frac{1}{\lambda - \alpha_n^\times} c_n b_n^\top\right) \cdots \left(I_m - \frac{1}{\lambda - \alpha_1^\times} c_1 b_1^\top\right). \tag{10.14}$$

Some of the factors in these expressions may coincide with I_m. Thus (10.13) and (10.14) are factorizations of W and W^{-1} into at most n elementary factors. In particular (i) is satisfied. □

Notice that the eigenvalues $\alpha_1, \ldots, \alpha_n$ of A in (10.10) and the eigenvalues $\alpha_1^\times, \ldots, \alpha_n^\times$ of $A^\times$ in (10.11) are related by

$$\alpha_j - \alpha_j^\times = b_j^\top c_j = \operatorname{trace}(c_j b_j^\top), \qquad j = 1, \ldots, n.$$

It follows that the jth factor in (10.14) read from left to right is the inverse of the $(n+1-j)$th factor in (10.13).

Combining Theorems 10.5 and 10.2 we arrive at the following result.

Corollary 10.7. *Let W be a proper rational $m \times m$ matrix function, let*

$$W(\lambda) = I_m + C(\lambda I_n - A)^{-1} B$$

be a realization of W, and assume $A^\times = A - BC$ is diagonalizable. Then, given an ordering $\alpha_1, \ldots, \alpha_n$ of the eigenvalues of A (algebraic multiplicities taken into

account), there exist factorizations of W *and* W^{-1} *of the form*

$$W(\lambda) = \left(I_m + \frac{1}{\lambda - \alpha_1} R_1\right) \cdots \left(I_m + \frac{1}{\lambda - \alpha_n} R_n\right), \tag{10.15}$$

$$W^{-1}(\lambda) = \left(I_m - \frac{1}{\lambda - \alpha_n^\times} R_n\right) \cdots \left(I_m - \frac{1}{\lambda - \alpha_1^\times} R_1\right), \tag{10.16}$$

where $R_1, \ldots, R_n$ *are* $m \times m$ *matrices of rank at most one and* $\alpha_1^\times, \ldots, \alpha_n^\times$ *are the eigenvalues of* $A^\times$ *(again algebraic multiplicities taken into account).*

Proof. Apply Theorem 10.2 with A replaced by $A^\times$, Z by A and $\zeta_1, \ldots, \zeta_n$ by $\alpha_n, \ldots, \alpha_1$. This gives an invertible $n \times n$ matrix T such that $T^{-1}A^\times T$ is upper triangular and $T^{-1}AT$ is lower triangular with diagonal elements $\alpha_n, \ldots, \alpha_1$ (read, as always, from top left to bottom right). Let E be the $n \times n$ reversed identity matrix, and put $S = TE$. Then $S^{-1}A^\times S$ is lower triangular and $S^{-1}AS$ is upper triangular with diagonal elements $\alpha_1, \ldots, \alpha_n$. In other words, $S^{-1}AS$ and $S^{-1}A^\times S$ are of the form appearing in the right-hand sides of (10.10) and (10.11), respectively. From (the final part of) the proof of Theorem 10.5 we know that this implies the existence of factorizations of the desired type. □

Using the remark made in the paragraph after the proof of Theorem 10.5, we see that the jth factor in (10.16) read from left to right is the inverse of the $(n + 1 - j)$th factor in (10.15). Deleting in (10.15) and (10.16) possible factors that are constant with value I_m, one obtains factorizations of W and W^{-1} into elementary factors.

There is an alternative version of Corollary 10.7 where it is assumed that A (instead of $A^\times$) is diagonalizable. In this case we just consider

$$W^{-1}(\lambda) = I_m - C(\lambda I_n - A^\times)^{-1}B$$

in place of W. In this alternative version, which clearly is a refinement of Theorem 2.7, the order of the eigenvalues of $A^\times$ can be chosen at will.

To illustrate another (reverse) way of using Theorem 10.5 we prove now Theorem 10.3 which is stated at the end of the previous section.

Proof of Theorem 10.3. Let A and Z be complex $n \times n$ matrices with $\operatorname{rank}(A - Z) = 1$, and assume that A and Z have no common eigenvalue. Write $A - Z = bc^\top$ with $b, c \in \mathbb{C}^n$, and put

$$w(\lambda) = 1 + c^\top(\lambda I_n - A)^{-1}b. \tag{10.17}$$

The associate main operator for this realization of the rational scalar function w is $A^\times = A - bc^\top = Z$ and, by assumption, this matrix has no eigenvalue in common with A. But then we know from Theorem 7.6 that the realization is minimal. In particular, $n = \delta(w)$.

Now, let $\alpha_1, \ldots, \alpha_n$ and $\zeta_1, \ldots, \zeta_n$ be given orderings of the eigenvalues of A and Z, respectively. By a well-known identity for the determinant, we have

$$w(\lambda) = \det\big(1 + c^\top(\lambda I_n - A)^{-1}b\big) = \det\big(I_n + (\lambda I_n - A)^{-1}bc^\top\big).$$

It follows that

$$\begin{aligned} w(\lambda) &= \det\big(I_n + (\lambda I_n - A)^{-1}(A - Z)\big) \\ &= \det\Big(I_n + (\lambda I_n - A)^{-1}\big((\lambda I_n - Z) - (\lambda I_n - A)\big)\Big) \\ &= \det\Big((\lambda I_n - A)^{-1}(\lambda I_n - Z)\Big) \\ &= \frac{\det(\lambda I_n - Z)}{\det(\lambda I_n - A)} = \frac{(\lambda - \zeta_1)\cdots(\lambda - \zeta_n)}{(\lambda - \alpha_1)\cdots(\lambda - \alpha_n)} \\ &= \left(1 + \frac{\alpha_1 - \zeta_1}{\lambda - \alpha_1}\right)\cdots\left(1 + \frac{\alpha_n - \zeta_n}{\lambda - \alpha_n}\right). \end{aligned}$$

Using this factorization of w into elementary factors, we see from Part 2 of the proof of Theorem 10.5, with $R_j = \alpha_j - \zeta_j$ for $j = 1, \ldots, n$, that w has a realization

$$w(\lambda) = 1 + \widetilde{c}^\top(\lambda I_n - \widetilde{A})^{-1}\widetilde{b}, \tag{10.18}$$

such that $\widetilde{A}$ is upper triangular with diagonal elements $\alpha_1, \ldots, \alpha_n$, and the matrix $\widetilde{A}^\times (= \widetilde{A} - \widetilde{b}\widetilde{c}^\top)$ is lower triangular with diagonal elements $\zeta_1, \ldots, \zeta_n$. Since $n = \delta(W)$, the realizations in (10.17) and (10.18) are both minimal realizations of the same function, and hence they are similar. Thus there exists an invertible $n \times n$ matrix S such that

$$\widetilde{A} = S^{-1}AS, \qquad \widetilde{b} = S^{-1}b, \qquad \widetilde{c}^\top = c^\top S.$$

It follows that $S^{-1}ZS = S^{-1}(A - bc^\top)S = \widetilde{A}^\times$. Since $\widetilde{A}$ is upper triangular and $\widetilde{A}^\times$ is lower triangular, we see that A and Z admit simultaneous reduction to complementary forms. Moreover, the matrices $S^{-1}AS$ and $S^{-1}ZS$ have the desired triangular structure with desired diagonal entries. □

We conclude this section with a few remarks related to Lemma 10.6. The factorization (10.5) is non-minimal and has been obtained at the expense of introducing a "non-essential" pole β (cf., the systematic analysis of "pole-zero cancellation" presented in Chapter 8). Taking inverses in (10.5) leads to

$$\begin{aligned} \left(I_m + \frac{1}{\lambda - \alpha}R\right)^{-1} &= I_m - \frac{1}{\lambda - \alpha^\times}R \\ &= \left(I_m - \frac{1}{\lambda - \alpha^\times}R_2\right)\left(I_m + \frac{1}{\lambda - \beta}R_1\right). \end{aligned}$$

Note that this factorization features the same "non-essential" pole β too.

The phenomenon referred to above can be put in a more general context. Indeed, by specifying Theorem 8.16 in Section 8.4 to factorization in elementary factors, we obtain the following result.

Proposition 10.8. *Let W be a proper rational $m \times m$ matrix function with $W(\infty) = I_m$. Suppose W is given as a product of elementary factors,*

$$W(\lambda) = \left(I_m + \frac{1}{\lambda - \alpha_1} R_1\right) \cdots \left(I_m + \frac{1}{\lambda - \alpha_n} R_n\right). \tag{10.19}$$

Put $\alpha_j^{\times} = \alpha_j - \operatorname{trace} R_j$ $(j = 1, \ldots, n)$. Then, for each $\alpha \in \mathbb{C}$, we have

$$\sharp\{j \mid \alpha_j = \alpha\} - \delta(W; \alpha) = \sharp\{j \mid \alpha_j^{\times} = \alpha\} - \delta(W^{-1}; \alpha), \tag{10.20}$$

and these coinciding numbers are non-negative integers.

Here $\sharp$ stands for number of elements, $\delta(W; \alpha)$ is the pole-multiplicity of W at α, and $\delta(W^{-1}; \alpha)$ is the pole-multiplicity of W^{-1} at α, which is equal to the zero-multiplicity of W at α. If α is not a pole of W, respectively not a zero of W, then $\delta(W; \alpha)$, respectively $\delta(W^{-1}; \alpha)$, is zero by definition.

Roughly speaking, the second conclusion in Proposition 10.8 says the following. The poles of W (pole-multiplicities counted) are among $\alpha_1, \ldots, \alpha_n$, the zeros of W (zero-multiplicities counted) are among $\alpha_1^{\times}, \ldots, \alpha_n^{\times}$, and the additional poles in the factorization coincide with the additional zeros (again multiplicities counted).

The question arises whether, in general, for an arbitrary biproper rational $m \times m$ matrix function one can obtain factorizations into elementary factors without adding new poles and new zeros. We shall prove later that the answer is negative; see the final example in Section 12.4.

10.3 Complete factorization (general)

Let W be a rational $m \times m$ matrix function having the value I_m at infinity. A *complete factorization* of W is a minimal factorization of W involving elementary factors only. Thus a factorization of W is complete if it has the form

$$W(\lambda) = \left(I_m + \frac{1}{\lambda - \alpha_1} R_1\right) \cdots \left(I_m + \frac{1}{\lambda - \alpha_n} R_n\right), \tag{10.21}$$

where $n = \delta(W)$ is the McMillan degree of W, $\alpha_1, \ldots, \alpha_n$ are complex numbers, and $R_1, \ldots, R_n$ are rank one $m \times m$ matrices. Recall from the discussion in the paragraph preceding the proof of Theorem 10.5 that in a factorization of the type (10.21) the number of elementary factors is always at least $\delta(W)$

In view of (10.3), the factorization (10.21) of W brings with it the factorization

$$W^{-1}(\lambda) = \left(I_m - \frac{1}{\lambda - \alpha_n^{\times}} R_n\right) \cdots \left(I_m - \frac{1}{\lambda - \alpha_1^{\times}} R_1\right). \tag{10.22}$$

Here $\alpha_j^\times = \alpha_j - \operatorname{trace} R_j$, $j = 1, \ldots, n$. As W and W^{-1} have the same McMillan degree, the factorization (10.22) is complete if and only if this is the case for (10.21). We conclude that W^{-1} admits a complete factorization if and only if W does.

From the result presented in the third paragraph of the previous section (see (10.4)) we know that each proper rational scalar function (with the value at infinity being one) admits a complete factorization. For matrix functions this result does not hold true. Indeed, as we know from the example mentioned at the end of Section 9.1, the McMillan degree two function

$$\begin{bmatrix} 1 & \dfrac{1}{\lambda^2} \\ 0 & 1 \end{bmatrix} \tag{10.23}$$

does not admit any non-trivial minimal factorization. In particular, (10.23) cannot be written as the product of two elementary functions. On the other hand, as we shall see in the next section, dropping the requirement of minimality, one can write (10.23) as the product of three elementary functions.

In this section we use the results of the previous sections to present four theorems on complete factorization. The first makes the connection with simultaneous reduction to complementary triangular forms.

Theorem 10.9. *Let W be a rational $m \times m$ matrix function, and let*

$$W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$$

be a minimal realization of W. Then W admits a complete factorization if and only if A and $A^\times$ admit simultaneous reduction to complementary triangular forms.

Proof. Since the given realization of W is minimal, we have $n = \delta(W)$. Suppose W admits a complete factorization, i.e., a factorization into n elementary factors. Then, by the implication (ii) $\Rightarrow$ (iii) in Theorem 10.5, there is a realization

$$W(\lambda) = I_m + \widehat{C}(\lambda I_n - \widehat{A})^{-1}\widehat{B} \tag{10.24}$$

with upper triangular $\widehat{A}$ and lower triangular $\widehat{A}^\times = \widehat{A} - \widehat{B}\widehat{C}$. As $n = \delta(W)$, the realization (10.24) is minimal. Applying Theorem 7.7 (the state space isomorphism theorem), one sees that A and $A^\times$ admit simultaneous reduction to complementary triangular forms. This proves the only if part of the theorem.

The if part can be immediately related to the implication (iv) $\Rightarrow$ (ii) in Theorem 10.5, but it is instructive to follow a slightly different path. Assume A and $A^\times$ admit simultaneous reduction to complementary triangular forms. Then the proof of the implication (iv) $\Rightarrow$ (ii) in Theorem 10.5 yields a factorization (10.13) of W into at most n elementary factors. However, as already mentioned before (and because of the sublogarithmic property of the McMillan degree), the number of factors cannot be smaller than $n = \delta(W)$. So the factorization in question is complete. □

The next two theorems can be viewed as additions to Theorem 8.15. As in Theorem 8.15, let W be a rational $m \times m$ matrix function with $W(\infty) = I_m$. Since the state space dimension of a minimal realization of W is equal to the McMillan degree of W, Theorem 8.15 tells us that W admits a complete factorization whenever the poles of W are all simple. Using the fact that W admits a complete factorization if and only if W^{-1} admits a complete factorization, the role of the poles in the previous result can be taken over by the zeros. Indeed, defining a zero z of W to be *simple* whenever z is a simple pole of W^{-1}, we see that W admits a complete factorization whenever either all poles of W are simple or all zeros of W are simple. The next two theorems add to this statement that the poles of the elementary factors in a complete factorization of W or W^{-1} can be chosen in prescribed order.

Theorem 10.10. *Let W be a proper $m \times m$ rational matrix function with $W(\infty) = I_m$. Assume W has simple poles only. Then, given an ordering $\alpha_1^\times, \dots, \alpha_n^\times$ of the zeros of W (zero-multiplicities taken into account), there exists a complete factorization of W^{-1} of the form*

$$W^{-1}(\lambda) = \left(I_m + \frac{1}{\lambda - \alpha_1^\times} R_1\right) \cdots \left(I_m + \frac{1}{\lambda - \alpha_n^\times} R_n\right),$$

where $R_1, \dots, R_n$ are rank one $m \times m$ matrices.

Theorem 10.11. *Let W be a proper rational $m \times m$ matrix function with $W(\infty) = I_m$. Assume W has simple zeros only. Then, given an ordering $\alpha_1, \dots, \alpha_n$ of the poles of W (pole-multiplicities taken into account), there exists a complete factorization of W of the form*

$$W(\lambda) = \left(I_m + \frac{1}{\lambda - \alpha_1} R_1\right) \cdots \left(I_m + \frac{1}{\lambda - \alpha_n} R_n\right),$$

where $R_1, \dots, R_n$ are rank one $m \times m$ matrices.

Proofs. Let $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ be a minimal realization of W, so n is the McMillan degree of W. Suppose W has simple zeros only. Then we see from Corollary 8.14 that $A^\times = A - BC$ is diagonalizable. By the same corollary, $\alpha_1, \dots, \alpha_n$ are the eigenvalues of A counted according to algebraic multiplicity. Apply now Corollary 10.7 to obtain Theorem 10.11. Applying the latter to W^{-1} one arrives at Theorem 10.10. □

Theorem 10.12. *Let W be a rational $m \times m$ matrix function, and let*

$$W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B \tag{10.25}$$

be a realization of W such that $\operatorname{rank} BC = 1$ or, what amounts to the same, $\operatorname{rank}(A - A^\times) = 1$. Suppose, in addition, that A and $A^\times$ have no common eigenvalue. Then the given realization is minimal (i.e., $\delta(W) = n$) and W admits a complete factorization. In fact, given an ordering $\alpha_1, \dots, \alpha_n$ of the eigenvalues of A and an ordering $\alpha_1^\times, \dots, \alpha_n^\times$ of the eigenvalues of $A^\times$ (in both cases algebraic

multiplicities taken into account), there exist complete factorizations of W and W^{-1} of the form

$$\begin{aligned} W(\lambda) &= \left(I_m + \frac{1}{\lambda - \alpha_1} R_1\right) \cdots \left(I_m + \frac{1}{\lambda - \alpha_n} R_n\right), \\ W^{-1}(\lambda) &= \left(I_m - \frac{1}{\lambda - \alpha_n^\times} R_n\right) \cdots \left(I_m - \frac{1}{\lambda - \alpha_1^\times} R_1\right), \end{aligned}$$

where $R_1, \ldots, R_n$ are rank one $m \times m$ matrices.

Since in the above theorem A and $A^\times$ have no common eigenvalue, it follows from Theorem 7.6 that the realization (10.25) of W is minimal. Hence, by Corollary 8.14, the poles of W are the eigenvalues of A and the poles of W^{-1} are the eigenvalues of $A^\times$, the appropriate multiplicities taken into account. Thus the condition in the above theorem that A and $A^\times$ have no common eigenvalue implies that W and W^{-1} have no common pole. Conversely, if W and W^{-1} have no common pole and the realization (10.25) is minimal, then A and $A^\times$ have no common eigenvalue. Thus Theorem 10.12 remains true if the phrase *A and $A^\times$ have no common eigenvalue* is replaced by *W and W^{-1} have no common pole and the realization* (10.25) *is minimal*. Moreover, in that case one can take for $\alpha_1, \ldots, \alpha_n$ any ordering of the poles of W and for $\alpha_1^\times, \ldots, \alpha_n^\times$ any ordering of the poles of W^{-1}.

Proof. As mentioned in the preceding paragraph, since A and $A^\times$ have no common eigenvalue, we know from Theorem 7.6 that the given realization of W is minimal. By Theorem 10.3, the matrices A and $A^\times$ admit simultaneous reduction to complementary triangular forms. Applying the if part of Theorem 10.9 we now see that W admits a compete factorization. This proves the first part of the theorem. The more detailed second part can be obtained by combining the second part of Theorem 10.3 with Part 3 of the proof of Theorem 10.5. □

Rational matrix functions of the type appearing in the above theorem form a subclass of the so-called companion based matrix functions. We shall come back to this fact in the next chapter, Section 11.3.

As was remarked (and made more precise) in Section 9.1, minimal factorization amounts to the absence of pole-zero cancellations. The next proposition underlines this point for factorizations into elementary factors.

Proposition 10.13. *Let W be a rational $m \times m$ matrix function, and let* (10.21) *be a factorization of W into elementary factors. The following statements are equivalent:*

(i) *the factorization* (10.21) *is complete,*

(ii) *$\alpha_1, \ldots, \alpha_n$ are the poles of W counted according to pole-multiplicity,*

(iii) *$\alpha_1^\times, \ldots, \alpha_n^\times$ are the zeros of W counted according to zero-multiplicity.*

Here $\alpha_j^\times = \alpha_j - \operatorname{trace} R_j$, $j = 1, \ldots, n$.

Proof. From Part 2 of the proof of Theorem 10.5 we know that the factorization (10.21) induces a realization

$$W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$$

of W such that A is upper triangular with $\alpha_1, \ldots, \alpha_n$ on the diagonal. If (10.21) is complete, $n = \delta(W)$ and the realization is minimal. But then Corollary 8.14 gives that the poles of W and the eigenvalues of A coincide, taking the appropriate multiplicities for poles and eigenvalues into account. From the special form of A indicated above, it is clear however that $\alpha_1, \ldots, \alpha_n$ are the eigenvalues of A counted according to algebraic multiplicity.

Thus (i) implies (ii). To establish the converse, let $\alpha_1, \ldots, \alpha_n$ be the poles of W counted according to pole-multiplicity. Then Theorem 8.13 shows that there is a minimal realization of W with state space dimension n. It follows that $n = \delta(W)$, and (10.21) is complete.

The factorization (10.21) of W induces the factorization (10.22) of W^{-1}. Applying what we established above to W^{-1}, we get that (10.22) is complete if and only if $\alpha_1^\times, \ldots, \alpha_n^\times$ are the poles of W^{-1} counted according to pole-multiplicity. But this is the same as saying that (10.22) is complete if and only if $\alpha_1^\times, \ldots, \alpha_n^\times$ are the zeros of W counted according to zero-multiplicity. The equivalence of (i) and (iii) now follows from our earlier observation that the factorization (10.22) is complete provided this is the case for (10.21). □

10.4 Quasicomplete factorization (general)

As before, let W be a proper $m \times m$ rational matrix function having the value I_m at infinity. In this section we show that there is always a factorization of W into elementary factors, and we give an estimate for the minimal number of factors in such a factorization (see Theorem 10.15 and Corollary 10.16 below). The case of an empty product of such factors corresponds to the trivial situation $\delta(W) = 0$. Therefore, in what follows, it will be assumed that the McMillan degree of W is positive.

We begin with some preparations. Let T be an $n \times n$ matrix. By the *spectral polynomial* of T we mean the (scalar) polynomial

$$p_T(\lambda) = (\lambda - \tau_1) \cdots (\lambda - \tau_s)$$

where $\tau_1, \ldots, \tau_s$ are the distinct eigenvalues of T. It is the monic (scalar) polynomial of minimal degree vanishing on the spectrum of T. Along with the spectral polynomial comes the matrix

$$p_T(T) = (T - \tau_1 I_n) \cdots (T - \tau_s I_n)$$

which will play an important role in what follows. Note that $p_T(T)$ is nilpotent, and if T is nilpotent, then $p_T(T) = T$. Also $p_T(T) = 0$ if and only if T is diagonalizable. Finally, the subspace $\operatorname{Ker} p_T(T)$ of $\mathbb{C}^n$ is spanned by the eigenvectors

of T. In particular, it has a basis consisting of eigenvectors of T. To see this, note that, whenever S is an invertible $n \times n$ matrix,

$$p_{S^{-1}TS} = p_T, \qquad p_{S^{-1}TS}(S^{-1}TS) = S^{-1}p_T(T)S,$$

and pass to the Jordan form. With an eye on later use, we also observe that $p_T(T)^* = p_{T^*}(T^*)$.

Theorem 10.14. *Let A and Z be $n \times n$ matrices, and suppose that (at least) one of the following identities is satisfied*

$$\operatorname{Ker} p_A(A) + \operatorname{Ker} p_Z(Z) = \mathbb{C}^n, \qquad \operatorname{Im} p_A(A) \cap \operatorname{Im} p_Z(Z) = \{0\}. \tag{10.26}$$

Then A and Z admit simultaneous reduction to complementary triangular forms.

If A is diagonalizable, then $p_A(A) = 0$ and both identities in (10.26) are trivially satisfied. Thus the first part of Theorem 10.2 is a special case of Theorem 10.14.

Proof. Assume that the first identity in (10.26) holds true. We shall prove that this implies that A and Z admit simultaneous reduction to complementary triangular forms. Let $u_1, \dots, u_k$ be a basis of $\operatorname{Ker} p_A(A)$ consisting of eigenvectors of A, and let $\alpha_1, \dots, \alpha_k$ in $\mathbb{C}$ be such that

$$Au_j = \alpha_j u_j, \qquad j = 1, \dots, k.$$

Similarly, let $v_1, \dots, v_m$ be a basis of $\operatorname{Ker} p_Z(Z)$ consisting of eigenvectors of Z, and let $\zeta_1, \dots, \zeta_m$ in $\mathbb{C}$ be such that

$$Zv_j = \zeta_j v_j, \qquad j = 1, \dots, m.$$

Then the first identity in (10.26) implies that $\mathbb{C}^n$ is spanned by the vectors $u_1, \dots, u_k, v_1, \dots, v_m$. From this collection of vectors, we now extract a basis $w_1, \dots, w_n$ for $\mathbb{C}^n$ in such a way that, for an appropriate choice of s, the elements $w_1, \dots, w_s$ are eigenvectors of A, while $w_{s+1}, \dots, w_n$ are eigenvectors of Z. Now let S_0 be the $n \times n$ matrix having w_j as its jth column. Then $S_0^{-1}AS_0$ and $S_0^{-1}ZS_0$ have the form

$$S_0^{-1}AS_0 = \begin{bmatrix} A_1 & A_0 \\ 0 & A_2 \end{bmatrix}, \qquad S_0^{-1}ZS_0 = \begin{bmatrix} Z_1 & 0 \\ Z_0 & Z_2 \end{bmatrix}$$

with A_1 an $s \times s$ diagonal matrix and Z_2 an $(n-s) \times (n-s)$ diagonal matrix. By Theorem 10.2, applied to A_1, Z_1 and A_2, Z_2, there exist an invertible $s \times s$ matrix S_1 and an invertible $(n-s) \times (n-s)$ matrix S_2 such that $S_1^{-1}A_1S_1$ and $S_2^{-1}A_2S_2$ are upper triangular while $S_1^{-1}Z_1S_1$ and $S_2^{-1}Z_2S_2$ are lower triangular. Now put

$$S = S_0 \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}.$$

Then S is an invertible $n \times n$ matrix, $S^{-1}AS$ is upper triangular and $S^{-1}ZS$ is lower triangular.

Next, assume that the second identity in (10.26) holds. Passing to orthogonal complements and adjoints, we see that

$$\operatorname{Ker} p_{A^*}(A^*) + \operatorname{Ker} p_{Z^*}(Z^*) = \mathbb{C}^n.$$

Thus, by what we just proved, A^* and Z^* admit simultaneous reduction to complementary triangular forms. Let T be an invertible $n \times n$ matrix such that $T^{-1}A^*T$ and $T^{-1}Z^*T$ are upper and lower triangular,respectively. Put $S = (T^*)^{-1}E$, where E is the $n \times n$ reversed identity matrix. Then $S^{-1}AS$ is upper triangular and $S^{-1}ZS$ is lower triangular. □

Neither of the identities in (10.26) implies the other. To that see this, consider the matrices

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad Z = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{10.27}$$

and their adjoints.

To state the main theorem of this section about factorization into elementary factors we introduce the following notation. Let W be an $m \times m$ rational matrix function with $W(\infty) = I_m$. Write W in the form

$$W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B,$$

and assume that this realization is minimal, i.e., $n = \delta(W)$. By $\kappa(W)$ we now mean the integer

$$\kappa(W) = \min\{\kappa_-(W), \kappa_+(W)\},$$

where

$$\kappa_-(W) = \delta(W) + \operatorname{codim}\big(\operatorname{Ker} p_A(A) + \operatorname{Ker} p_{A^\times}(A^\times)\big), \tag{10.28}$$

$$\kappa_+(W) = \delta(W) + \dim\big(\operatorname{Im} p_A(A) + \operatorname{Im} p_{A^\times}(A^\times)\big), \tag{10.29}$$

or, what amounts to the same,

$$\kappa_-(W) = 2\delta(W) - \dim\big(\operatorname{Ker} p_A(A) + \operatorname{Ker} p_{A^\times}(A^\times)\big)$$

$$\kappa_+(W) = 2\delta(W) - \operatorname{codim}\big(\operatorname{Im} p_A(A) + \operatorname{Im} p_{A^\times}(A^\times)\big).$$

The state space isomorphism theorem guarantees that $\kappa(W)$ does not depend on the choice of the minimal realization for W. An example where $\kappa_-(W)$ and $\kappa_+(W)$ do not coincide will be given at the end of the section.

Theorem 10.15. *Let W be an $m \times m$ rational matrix function having the value I_m at infinity. Then W admits a factorization into $\kappa(W)$ elementary factors.*

Proof. We split the proof into two parts. In the first part we consider the case $\kappa(W) = \kappa_-(W) \le \kappa_+(W)$; in the second the situation $\kappa(W) = \kappa_+(W) \le \kappa_-(W)$.

Part 1. In this part we suppose that $\kappa_-(W) \le \kappa_+(W)$ and so $\kappa(W) = \kappa_-(W)$. Write W in the form

$$W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B,$$

and assume that this realization is minimal, i.e., $n = \delta(W)$. As (by definition; cf., Section 7.1) minimality implies controllability, the spectral assignment theorem (Theorem 6.5.1 in [70]) applies to the pair of matrices A, B. Therefore, given an $n-$tuple of complex numbers $\zeta_1, \dots, \zeta_n$, there exists an $m \times n$ matrix F, such that $A + BF$ has eigenvalues $\zeta_1, \dots, \zeta_n$. Henceforth we will assume these eigenvalues to be distinct, so that $A + BF$ is diagonalizable. We will also assume that none of $\zeta_1, \dots, \zeta_n$ is an eigenvalue of A or $A^\times$.

Put $k = \kappa_-(W) - \delta(W) = \operatorname{codim}\big(\operatorname{Ker} p_A(A) + \operatorname{Ker} p_{A^\times}(A^\times)\big)$. As $A + BF$ is diagonalizable, there exist k eigenvectors $x_1, \dots, x_k$ of $A + BF$ such that

$$\big(\operatorname{Ker} p_A(A) + \operatorname{Ker} p_{A^\times}(A^\times)\big) \dot{+} \operatorname{span}\{x_1, \dots, x_k\} = \mathbb{C}^n.$$

Renumbering the eigenvalues of $A + BF$ (if necessary), we may write

$$(A + BF)x_j = \zeta_j x_j, \qquad j = 1, \dots, k.$$

Now introduce the $n \times k$ matrix $X = -\begin{bmatrix} x_1 & \cdots & x_k \end{bmatrix}$. Then $(A+BF)X = XG$, where G is the $k \times k$ diagonal matrix with diagonal elements $\zeta_1, \dots, \zeta_k$. With $K = FX$, we arrive at $XG - AX = BK$.

Consider the matrices

$$\widehat{A} = \begin{bmatrix} A & BK \\ 0 & G \end{bmatrix}, \qquad \widehat{B} = \begin{bmatrix} B \\ 0 \end{bmatrix}, \qquad \widehat{C} = \begin{bmatrix} C & K \end{bmatrix}.$$

Note that $W(\lambda) = I_m + \widehat{C}(\lambda I_{n+k} - \widehat{A})^{-1}\widehat{B}$. We have

$$\begin{aligned} \widehat{A} &= \begin{bmatrix} A & XG - AX \\ 0 & G \end{bmatrix} \\ &= \begin{bmatrix} I_n & X \\ 0 & I_k \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & G \end{bmatrix} \begin{bmatrix} I_n & -X \\ 0 & I_k \end{bmatrix} \\ &= \begin{bmatrix} I_n & X \\ 0 & I_k \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & G \end{bmatrix} \begin{bmatrix} I_n & X \\ 0 & I_k \end{bmatrix}^{-1}. \end{aligned}$$

It follows that

$$p_{\widehat{A}}(\widehat{A}) = \begin{bmatrix} I_n & X \\ 0 & I_k \end{bmatrix} \begin{bmatrix} p_{\widehat{A}}(A) & 0 \\ 0 & p_{\widehat{A}}(G) \end{bmatrix} \begin{bmatrix} I_n & X \\ 0 & I_k \end{bmatrix}^{-1},$$

and, consequently,

$$\operatorname{Ker}\ p_{\widehat{A}}(\widehat{A}) = \begin{bmatrix} I_n & X \\ 0 & I_k \end{bmatrix} \operatorname{Ker} \begin{bmatrix} p_{\widehat{A}}(A) & 0 \\ 0 & p_{\widehat{A}}(G) \end{bmatrix}.$$

From the expression for $\widehat{A}$, it also follows that $\sigma(\widehat{A}) = \sigma(G) \cup \sigma(A)$. Furthermore, note that our choice of $\zeta_1, \ldots, \zeta_k$ is such that $\sigma(G)$ and $\sigma(A)$ are disjoint. Thus $p_{\widehat{A}}(\lambda) = p_G(\lambda) p_A(\lambda)$. But then

$$p_{\widehat{A}}(A) = p_G(A) p_A(A), \qquad p_{\widehat{A}}(G) = p_G(G) p_A(G) = 0.$$

In the latter identity we used that $p_G(G) = 0$ which follows from the fact that G is a diagonal matrix. Note also that $p_G(A)$ is invertible. A straightforward argument now gives

$$\operatorname{Ker} \begin{bmatrix} p_{\widehat{A}}(A) & 0 \\ 0 & p_{\widehat{A}}(G) \end{bmatrix} = \operatorname{Ker} \left(\begin{bmatrix} p_A(A) & 0 \\ 0 & 0 \end{bmatrix} \right),$$

and we arrive at

$$\operatorname{Ker}\ p_{\widehat{A}}(\widehat{A}) = \begin{bmatrix} I_m & X \\ 0 & I_k \end{bmatrix} \operatorname{Ker} \left(\begin{bmatrix} p_A(A) & 0 \\ 0 & 0 \end{bmatrix} \right).$$

Next consider the matrix

$$\widehat{A}^{\times} = \widehat{A} - \widehat{B}\widehat{C} = \begin{bmatrix} A^{\times} & 0 \\ 0 & G \end{bmatrix}.$$

A similar reasoning as the one given above yields

$$\operatorname{Ker} p_{\widehat{A}^{\times}}(\widehat{A}^{\times}) = \operatorname{Ker} \left(\begin{bmatrix} p_{A^{\times}}(A^{\times}) & 0 \\ 0 & 0 \end{bmatrix} \right).$$

Combining this with the description of $\operatorname{Ker}\ p_{\widehat{A}}(\widehat{A})$ obtained in the previous paragraph, we see that

$$\operatorname{Ker}\ p_{\widehat{A}}(\widehat{A}) + \operatorname{Ker} p_{\widehat{A}^{\times}}(\widehat{A}^{\times}) = \mathbb{C}^{n+k}$$

if and only if

$$\operatorname{Ker} p_A(A) + \operatorname{Ker} p_{A^{\times}}(A^{\times}) + \operatorname{Im} X = \mathbb{C}^n.$$

By construction, the latter is the case. Theorem 10.14 now yields that the matrices $\widehat{A}$ and $\widehat{A}^{\times}$ admit simultaneous reduction to complementary triangular forms. Next, use Theorem 10.5 to see that W admits a factorization into $n + k$ elementary factors. Since $\kappa_-(W) = n + k$ (by definition), we can conclude that W admits a factorization into $\kappa(W) = \kappa_-(W)$ factors.

Part 2. In this part we suppose that $\kappa_+(W) \le \kappa_-(W)$ and so $\kappa(W) = \kappa_+(W)$. The argument goes by taking adjoints. Thus we consider the rational $m \times m$ matrix function W^* given by $W^*(\lambda) = W(\bar{\lambda})^*$. Obviously

$$W^*(\lambda) = I_m + B^*(\lambda I_n - A^*)^{-1}C^*,$$

and this is a minimal realization. Also $A^* - C^*B^* = (A^\times)^*$. But then

$$\kappa_-(W^*) = n + \operatorname{codim}\Big(\operatorname{Ker} p_{A^*}(A^*) + \operatorname{Ker} p_{(A^\times)^*}\big((A^\times)^*\big)\Big).$$

Since $p_{A^*}(A^*) = p_A(A)^*$ and, similarly, $p_{(A^\times)^*}\big((A^\times)^*\big) = p_{A^\times}(A^\times)^*$, we have

$$\begin{aligned}
\kappa_-(W^*) &= n + \operatorname{codim}\big(\operatorname{Ker} p_A(A)^* + \operatorname{Ker} p_{A^\times}(A^\times)^*\big) \\
&= n + \operatorname{codim}\big(\operatorname{Im} p_A(A)^\perp + \operatorname{Im} p_{A^\times}(A^\times)^\perp\big) \\
&= n + \operatorname{codim}\big(\operatorname{Im} p_A(A) \cap \operatorname{Im} p_{A^\times}(A^\times)\big)^\perp \\
&= n + \dim\big(\operatorname{Im} p_A(A) \cap \operatorname{Im} p_{A^\times}(A^\times)\big) \\
&= \kappa_+(W).
\end{aligned}$$

In a similar fashion, $\kappa_+(W^*) = \kappa_-(W)$. Hence

$$\kappa_-(W^*) = \kappa_+(W) \le \kappa_-(W) = \kappa_+(W^*),$$

and it follows that W^* admits a factorization into $\kappa_-(W^*) = \kappa_+(W) = \kappa(W)$ elementary factors. But then so does W. □

Again, let W be a proper $m \times m$ rational matrix function with $W(\infty) = I_m$. A factorization of W into elementary factors will be called *quasicomplete* if the number of factors involved is minimal. This minimal number of factors will be denoted by $\delta_{\mathrm{q}}(W)$. We call it the *quasidegree* of W. Evidently the quasidegree is sublogarithmic. Note that W admits a complete factorization if and only if $\delta_{\mathrm{q}}(W) = \delta(W)$. We have the following estimates for $\delta_{\mathrm{q}}(W)$.

Corollary 10.16. *Let W be a proper $m \times m$ rational matrix function with $W(\infty) = I_m$ and $\delta(W) > 0$. Then*

$$\delta(W) \le \delta_{\mathrm{q}}(W) \le \kappa(W) \le 2\delta(W) - 1. \tag{10.30}$$

Proof. From the sublogarithmic property of the McMillan degree it is obvious that $\delta(W) \le \delta_{\mathrm{q}}(W)$. From Theorem 10.15 it is clear that $\delta_{\mathrm{q}}(W) \le \kappa(W)$. In the case of positive McMillan degree considered here, it follows that $\delta_{\mathrm{q}}(W) \le 2\delta(W) - 1$. Indeed, $p_A(A)$, being nilpotent, has a non-trivial kernel, so $\kappa(W) \le 2\delta(W) - 1$. □

We conclude this section with three illuminating examples. The first illustrates Theorem 10.15 and its proof.

Example. Consider the rational 2×2 matrix function

$$W(\lambda) = \begin{bmatrix} 1 & \dfrac{1}{\lambda^2} \\ 0 & 1 \end{bmatrix}. \tag{10.31}$$

Introducing the matrices

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

we have the representation $W(\lambda) = I_2 + C(\lambda I_2 - A)^{-1}B$, and this realization is minimal (both observable and controllable). Hence $\delta(W) = 2$. From Section 9.1 we know that the function W does not admit a complete factorization. Thus $\delta_{\mathrm{q}}(W) \geq 3$. Following up on what was already announced in the paragraph in the previous section containing (10.23), we shall now prove that equality holds, i.e., $\delta_{\mathrm{q}}(W) = 3$. This will be done by concretely factorizing W into three elementary factors along the lines suggested by the proof of Theorem 10.15. Note that here $\kappa(W) = \kappa_-(W) = \kappa_+(W) = 3$. Indeed, $A = A^\times$ and both $\operatorname{Ker} A$ and $\operatorname{Im} A$ are one-dimensional.

Put

$$F = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Then

$$A + BF = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

and $A + BF$ is diagonalizable with eigenvalues 1 and -1 both different from the unique eigenvalue 0 of $A = A^\times$. The number $k = \kappa_-(W) - \delta(W)$ in the proof of Theorem 10.15 is here equal to 1. So the matrix X appearing there is a vector in $\mathbb{C}^2$ and the matrix G can be identified with a scalar. In fact, with

$$X = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad G = 1, \qquad K = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

we have $XG - AX = BK$ as desired. Now construct

$$\widehat{A} = \begin{bmatrix} A & BK \\ 0 & G \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \widehat{B} = \begin{bmatrix} B \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},$$

$$\widehat{C} = \begin{bmatrix} C & K \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then $W(\lambda) = I_2 + \widehat{C}(\lambda I_3 - \widehat{A})^{-1}\widehat{B}$ and

$$\widehat{A}^{\times} = \widehat{A} - \widehat{B}\widehat{C} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The matrices $\widehat{A}$ and $\widehat{A}^{\times}$ admit simultaneous reduction to complementary triangular forms. In fact, with

$$S = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \qquad S^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

we have

$$S^{-1}\widehat{A}S = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad S^{-1}\widehat{A}^{\times}S = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

As explained in the proof of Theorem 10.5, the realization

$$W(\lambda) = I_2 + \widehat{C}S(\lambda I_3 - S^{-1}\widehat{A}S)^{-1}S^{-1}\widehat{B}$$

can now be used to obtain a factorization $W(\lambda) = W_1(\lambda)W_2(\lambda)W_3(\lambda)$ into three elementary factors W_1, W_2 and W_3. Using that

$$S^{-1}\widehat{B} = \begin{bmatrix} 0 & 1 \\ 0 & -1 \\ 0 & -1 \end{bmatrix}, \qquad \widehat{C}S = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

these factors can be computed as follows:

$$W_1(\lambda) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{1}{\lambda - 1} \begin{bmatrix} 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \dfrac{1}{\lambda - 1} \\ 0 & \dfrac{\lambda}{\lambda - 1} \end{bmatrix},$$

$$W_2(\lambda) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \frac{1}{\lambda} \begin{bmatrix} 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -\dfrac{1}{\lambda} \\ 0 & 1 \end{bmatrix},$$

$$W_3(\lambda) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \frac{1}{\lambda} \begin{bmatrix} 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{\lambda - 1}{\lambda} \end{bmatrix}.$$

For W given by (10.31), we have $\delta_{\mathrm{q}}(W) = 2\delta(W) - 1$. As we shall see in Section 12.4, this equality remains true when the term λ^2 in (10.31) is replaced by λ^n, where n is any positive integer. This shows that the estimate $\delta_{\mathrm{q}}(W) \leq 2\delta(W) - 1$ is sharp in the sense that for every value of the McMillan degree $\delta(W)$ equality can occur. On the other hand, there are situations with strict inequality $\delta_{\mathrm{q}}(W) < 2\delta(W) - 1$. The next example presents such a case.

Example. In this example we show that it may happen that $\delta_{\mathrm{q}}(W) < \kappa(W) = \kappa_-(W) = \kappa_+(W) < 2\delta(W) - 1$. Consider the 2×2 rational matrix function

$$W(\lambda) = \begin{bmatrix} 1 + \dfrac{1}{\lambda^2} & \dfrac{2}{\lambda} + \dfrac{1}{\lambda^3} \\ \dfrac{1}{\lambda} & 1 + \dfrac{1}{\lambda^2} \end{bmatrix}.$$

Introducing the matrices

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

we have the representation $W(\lambda) = I_2 + C(\lambda I_3 - A)^{-1}B$, and this realization is easily seen to be minimal (both observable and controllable). Hence $\delta(W) = 3$. Note that A is nilpotent, and so is

$$A^\times = A - BC = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}.$$

Thus $p_A(A) = A$ and $p_{A^\times}(A^\times) = A^\times$. It is now easy to check that $\kappa_-(W) = \kappa_+(W) = 4$. However $\delta_{\mathrm{q}}(W) = \delta(W) = 3$ because W admits the complete factorization

$$\begin{bmatrix} 1 + \dfrac{1}{\lambda^2} & \dfrac{2}{\lambda} + \dfrac{1}{\lambda^3} \\ \dfrac{1}{\lambda} & 1 + \dfrac{1}{\lambda^2} \end{bmatrix} = \begin{bmatrix} 1 & \dfrac{1}{\lambda} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \dfrac{1}{\lambda} & 1 \end{bmatrix} \begin{bmatrix} 1 & \dfrac{1}{\lambda} \\ 0 & 1 \end{bmatrix},$$

which is readily obtained from the fact the A is upper and $A^\times$ is lower triangular.

Recall from (10.30) that $\delta_{\mathrm{q}}(W) \leq 2\delta(W) - 1$. This inequality can be sharpened to

$$\delta_{\mathrm{q}}(W) \leq 2\delta(W) - \nu(W), \tag{10.32}$$

where $\nu(W)$ stands for the maximal number of non-trivial factors that can occur in a minimal factorization of W. To verify (10.32), consider a minimal factorization $W = W_1 \cdots W_{\nu(W)}$ of W involving $\nu(W)$ non-trivial factors. For $j = 1, \ldots, \nu(W)$, we then have $\delta_{\mathrm{q}}(W_j) \leq 2\delta(W_j) - 1$, and so

$$\delta_{\mathrm{q}}(W) \leq \sum_{j=1}^{\nu(W)} \delta_{\mathrm{q}}(W_j) \leq \sum_{j=1}^{\nu(W)} \Big(2\delta(W_j) - 1\Big) = 2\delta(W) - \nu(W).$$

For completeness, note that $1 \leq \nu(W) \leq \delta(W)$. We conclude with an example featuring a situation where the inequality (10.32) is strict.

Example. This example shows that it may happen that $\kappa_-(W) < \kappa_+(W)$ and $\delta_{\mathrm{q}}(W) \leq 2\delta(W) - \nu(W)$. Consider the 3×3 rational matrix function

$$W(\lambda) = \begin{bmatrix} 1 & 0 & 0 \\ \dfrac{1}{\lambda^2} - \dfrac{1}{\lambda^3} & 1 - \dfrac{1}{\lambda} & \dfrac{2}{\lambda} - \dfrac{1}{\lambda^2} \\ \dfrac{1}{\lambda} - \dfrac{1}{\lambda^3} & -\dfrac{1}{\lambda} & 1 + \dfrac{1}{\lambda} - \dfrac{1}{\lambda^2} \end{bmatrix}.$$

Introducing the matrices

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 & -1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix},$$

we have the representation $W(\lambda) = I_3 + C(\lambda I_3 - A)^{-1}B$, and this realization is easily seen to be minimal (both observable and controllable). Thus $\delta(W) = 3$. Also

$$A^{\times} = A - BC = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = SAS^{-1},$$

where

$$S = S^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Obviously both A and $A^{\times}$ are nilpotent. Hence $p_A(A) = A$ and $p_{A^{\times}}(A^{\times}) = A^{\times}$. It is now easy to check that $\kappa_-(W) = 4$ and $\kappa_+(W) = 5$. So this is a case where $\kappa_-(W)$ and $\kappa_+(W)$ are different. The matrices A and $A^{\times}$ are unicellular. The

non-trivial invariant subspaces of A are $\operatorname{span}\{e_1\}$ and $\operatorname{span}\{e_1, e_2\}$, where e_j is the vector in $\mathbb{C}^3$ with the jth entry equal to one and the other two entries equal to zero. The non-trivial invariant subspaces of $A^\times$ are $\operatorname{span}\{Se_1\} = \operatorname{span}\{e_2\}$ and $\operatorname{span}\{Se_1, Se_2\} = \operatorname{span}\{e_2, e_1\}$. Hence there are no non-trivial supporting projections for the minimal realization of W under consideration. But then W does not admit any non-trivial minimal factorization (see Theorem 9.3), i.e., $\nu(W) = 1$. It also follows that $\delta_{\mathrm{q}}(W) > 3$. On the other hand, $\delta_{\mathrm{q}}(W) \le \kappa(W) \le \kappa_-(W) = 4$, and we arrive at $\delta_{\mathrm{q}}(W) = 4 < 5 = 2\delta(W) - \nu(W)$.

Notes

This chapter has its roots in a number of theorems on factorization of rational $m \times m$ matrix functions into elementary factors appearing in [14]. However, the problem of simultaneous reduction of two matrices into complementary triangular forms, which is a core element in constructing such factorizations, appears in [14] only implicitly. This second problem was introduced and studied in [19]; see also the survey paper [10]. Section 10.1 is based on [19]. Sections 10.2 and 10.3 originate from Sections 1.1, 1.3 and 3.2 of [14]; see also [39], the references therein, and [104] for earlier material in this direction.

Theorem 10.2, which appears implicitly in the proof of Theorem 1.6 in [14], is taken from Section 7.2 in the survey paper [10] (see also Section 1 in [19], where an alternative proof using lower-upper factorization is given). For an extension of the first part of Theorem 10.2 to commuting families of matrices, see Theorem 1.4 in [19]. Theorem 10.3 can be found in Section 7 of [19], though in a slightly different form. Other results on reduction to complementary triangular forms, mainly concerned with special classes of matrices, can be found in various publication: see [24], [111], [121], [26], and references therein. Section 2 in [19] contains a complete analysis of the case where the given matrices have order 2. In [119] and [120] some aspects of the infinite-dimensional case are treated.

A non-trivial general characterization of simultaneous reduction to complementary triangular forms is as yet not available. This differs from the situation where one looks for simultaneous reduction to the same (say upper) triangular form. For that type of reduction an algebraic characterization exists in the form of McCoy's theorem, see [95]. We note that there does exist a rather straightforward connection between simultaneous reduction to complementary triangular forms and simultaneous reduction to upper (or lower) triangular form. This has been established in [26], but the result given there does not combine with McCoy's theorem so as to produce an effective non-trivial general characterization of simultaneous reduction to complementary triangular forms.

There is no direct reference for Theorem 10.5, but it is closely related to Theorem 1.3 in [14], and the material presented in Section 6 of [19]. Corollary 10.7 is a refined version of Theorem 1.6 in [14]; see also Section 7.2 in [10]. The fact that in Section 10.3 the elementary factors are required to be square is important.

Indeed, minimal factorization involving elementary factors only is always possible if one allows for non-square elementary factors. This result has been established in [109]. We also note that it can happen that a real rational $m \times m$ matrix function which has simple pole only or simple zeros only, does not have a complete factorization with real factors. In other words, Theorems 10.10 and 10.11 do not have real counterparts. An example illustrating this point is given at the end of Section 15.1.

Theorem 10.9 is a reformulation of Theorem 6.1 in [19]; see also Section 7.2 in [10]. Theorem 10.10 is a somewhat stronger version of Theorem 3.4 in [14], and Theorem 10.11 is a modification of the same result (the role of the poles being taken over by the zeros). Theorem 10.12 goes back to the material of Sections 6 and 7 in [19] (cf., the remark made at the end of Section 7 in [19], in particular).

The result that any proper rational $m \times m$ matrix function W with $W(\infty) = I_m$ can be written as a product of elementary factors is due to [39]. The upper bound for the minimal number of factors in such a factorization given in Theorem 10.15 was proved in [119]; see also [121]. An example in [121] shows that this upper bound is sharper than the one which is obtained via the approach presented in [39]. The analysis in Section 10.4 follows the lines set out in [119] and [121]. In particular, Theorem 10.14 is identical to Proposition 2.3.3 in [119], and Theorem 3.3 in [121]. The example in Section 10.4 involving the function

$$\begin{bmatrix} 1 & \dfrac{1}{\lambda^2} \\ 0 & 1 \end{bmatrix}$$

appears in [119] and [121]. A concrete factorization of this function into three elementary factors (as given in the example) was already known to G.Ph.A. Thijsse (personal communication).

Chapter 11

Complete Factorization of Companion Based Matrix Functions

In this chapter results of the previous chapter are specified further for rational matrix functions of a special type, namely for the so-called companion based functions. These are characterized by the fact that they are rational matrix function having a minimal realization in which both the main matrix and the associate main matrix are (first) companions. A description of such functions is presented for the 2×2 case, and necessary and sufficient conditions are given for such functions to admit a complete factorization. The factorization results in this chapter are based on a detailed analysis of simultaneous reduction to complementary forms of pairs of companion matrices.

The present chapter consists of seven sections. The first contains preliminaries about companion matrices, including a description of all complete chains of invariant subspaces for such a matrix. The second section deals with simultaneous reduction to complementary forms of companion matrices. Companion based matrix functions are introduced and studied in the third and fourth section. The fifth section is devoted to complete factorization of companion based matrix functions, and the six section presents Maple procedures to calculate such factorizations explicitly. The final section has the character of an appendix; in this section, as a preparation for the next chapter, detailed information is given about the lattice of invariant subspaces of a companion matrix.

11.1 Companion matrices: preliminaries

A matrix is called an $n \times n$ *first companion* (matrix) if it has the form

$$\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix}, \tag{11.1}$$

where $a_0, \ldots, a_{n-1}$ are complex numbers. More specifically, we sometimes call (11.1) the first companion (matrix) associated with the monic polynomial $a(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0$. This polynomial is precisely the characteristic polynomial of (11.1). First companion matrices are nonderogatory, i.e., their eigenvalues have geometric multiplicity one. In fact, a matrix is nonderogatory if and only if it is similar to a first companion matrix. If α is an eigenvalue of the $n \times n$ first companion matrix A, then $\operatorname{Ker}(\alpha I_n - A)$ is spanned by the column vector $(1, \alpha, \ldots, \alpha^{n-1})^\top$. Here the symbol $^\top$ stands for taking the transpose.

A matrix is called an $n \times n$ *second companion* (matrix) if it has the form

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & & 0 & -a_2 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & 1 & -a_{n-1} \end{bmatrix}. \tag{11.2}$$

Clearly, second companion matrices are just the transposes of first companions. Hence what is said above for first companion matrices holds, with appropriate modifications, for second companions. For instance, if α is an eigenvalue of the $n \times n$ second companion matrix A, then $\operatorname{Ker}(\alpha I_n - A)$ is spanned by the column vector $(x_1, \ldots, x_{n-1}, x_n)^\top$ if and only

$$\begin{bmatrix} 1 & \lambda & \cdots & \lambda^{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \frac{a(\lambda)}{\lambda - \alpha}\, x_n, \quad \text{and } x_n \neq 0, \tag{11.3}$$

where $a(\lambda) = a_0 + \cdots + a_{n-1}\lambda^{n-1} + \lambda^n$.

It is well known that a square matrix and its transpose are always similar. For companion matrices, this statement can be made more explicit. Indeed, if A

and $A^\top$ are given by (11.1) and (11.2), respectively, then $HA = A^\top H$ where

$$H = \begin{bmatrix} a_1 & a_2 & \cdots & a_{n-1} & 1 \\ a_2 & & \iddots & \iddots & 0 \\ \vdots & \iddots & \iddots & & \vdots \\ a_{n-1} & \iddots & & & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{11.4}$$

Since H is invertible, this proves that A and $A^\top$ are similar. The matrix H is a Hankel matrix and it is called the *symmetrizer* of A.

We shall now discuss similarity to the transpose for pairs of first companion matrices. There is an analogous result for second companion matrices; just take transposes.

Proposition 11.1. *Let A and Z be $n \times n$ first companion matrices. Then there exists an invertible $n \times n$ matrix S such that $S^{-1}A^\top S = A$ and $S^{-1}Z^\top S = Z$ if and only if either A and Z are identical or (the other extreme) A and Z do not have a common eigenvalue. In the latter case, the similarity S is unique up to multiplication with a nonzero scalar.*

Proof. First we deal with the only if part. So let S be an invertible $n \times n$ matrix such that $S^{-1}A^\top S = A$ and $S^{-1}Z^\top S = Z$, and assume that α is a common eigenvalue of A and Z. Then the vector $v = (1, \alpha, \ldots, \alpha^{n-1})^\top$ is a common eigenvector of A and Z corresponding to the eigenvalue α. Hence Sv is a common eigenvector of the second companion matrices $A^\top$ and $Z^\top$ corresponding to the eigenvalue α. Let x_n is be the nth entry of the column vector $x = Sv$. Using the remark made at end of the second paragraph of this section (see formula (11.3)) we know that $x_n \neq 0$ and

$$\begin{aligned} \begin{bmatrix} 1 & \lambda & \cdots & \lambda^{n-1} \end{bmatrix} Sv &= \frac{a(\lambda)}{\lambda - \alpha} x_n, \\ \begin{bmatrix} 1 & \lambda & \cdots & \lambda^{n-1} \end{bmatrix} Sv &= \frac{z(\lambda)}{\lambda - \alpha} x_n, \end{aligned}$$

But then the fact that $x_n \neq 0$ yields that a and z coincide. Hence A (the first companion associated with a) and Z (the first companion associated with z) are identical too. This proves the only if part of the proposition.

Next we turn to the if part. The case when $A = Z$ is covered by the material on the symmetrizer presented prior to the proposition. So we assume that A and Z have no common eigenvalue.

As before, let a and z be the characteristic polynomials of A and Z, respectively, and let $\mathrm{Bez}(a, z)$ be the Bezoutian associated with a and z. The latter

means (see, e.g., [92], Section 13.3) that $\mathrm{Bez}(a,z)$ is the $n\times n$ matrix $(b_{ij})_{i,j=1}^n$ determined by

$$\frac{a(\lambda)z(\mu)-a(\mu)z(\lambda)}{\lambda-\mu}=\sum_{i,j=1}^n b_{ij}\lambda^{i-1}\mu^{j-1}.$$

By the Barnett factorization theorem (Proposition 13.3.2 in [92]), the Bezoutian admits the factorization $\mathrm{Bez}(a,z)=Hz(A)$ where H is the symmetrizer of A given by (11.4). Using that $HA=A^\top H$, we obtain

$$\mathrm{Bez}(a,z)A=Hz(A)A=HAz(A)=A^\top Hz(A)=A^\top\mathrm{Bez}(a,z).$$

Interchanging the roles of A and Z, we get $\mathrm{Bez}(z,a)Z=Z^\top\mathrm{Bez}(z,a)$. But, as is obvious from the definition, $\mathrm{Bez}(z,a)=-\mathrm{Bez}(a,z)$, and it follows that $\mathrm{Bez}(a,z)Z=Z^\top\mathrm{Bez}(a,z)$. Recall now that that $\mathrm{Bez}(a,z)$ is invertible if (and only if) a and z do not have a common zero. In other words, $\mathrm{Bez}(a,z)$ is invertible if (and only if) A and Z do not have a common eigenvalue. The latter has been assumed in this part of the proof.

This proves the following: under the assumption that A and Z do not have a common eigenvalue, the matrix $S=\mathrm{Bez}(a,z)$ is invertible and indeed transforms the pair A,Z into the pair $A^\top,Z^\top$. Of course, every nonzero scalar multiple of $\mathrm{Bez}(a,z)$ will do too. It remains to prove that this is all the freedom there is.

Let S be any invertible $n\times n$ matrix such that $S^{-1}A^\top S=A$ and $S^{-1}Z^\top S=Z$. Write $A-Z=bc^\top$ where b and c are nonzero vectors in $\mathbb{C}^n$ such that the $n\times n$ matrix $B=\begin{bmatrix} b & Ab & \cdots & A^{n-1}b\end{bmatrix}$ is invertible. This is possible, since A and Z are different first companion matrices. In fact, one can take for b the nth unit vector in $\mathbb{C}^n$ with last entry one and all others equal to zero. Now $cb^\top=A^\top-Z^\top=S(A-Z)S^{-1}=(Sb)(c^\top S^{-1})$. Since $c^\top S^{-1}$ is a nonzero vector, it makes sense to put

$$\sigma=\frac{b^\top\left(c^\top S^{-1}\right)^\top}{c^\top S^{-1}\left(c^\top S^{-1}\right)^\top},$$

so that $Sb=\sigma c$ with σ necessarily nonzero because $Sb\neq 0$. It follows that

$$\begin{aligned} SB &= \begin{bmatrix} Sb & SAb & \cdots & SA^{n-1}b\end{bmatrix}\\ &= \begin{bmatrix} Sb & A^\top Sb & \cdots & \left(A^\top\right)^{n-1}Sb\end{bmatrix}\\ &= \sigma\begin{bmatrix} c & A^\top c & \cdots & (A^\top)^{n-1}c\end{bmatrix}.\end{aligned}$$

Replacing S by $\mathrm{Bez}(a,z)$, we also get

$$\mathrm{Bez}(a,z)B=\tau\begin{bmatrix} c & A^\top c & \cdots & (A^\top)^{n-1}c\end{bmatrix},$$

where τ is again a nonzero scalar. Hence $\tau SB=\sigma\mathrm{Bez}(a,z)B$. Since B is invertible, this gives $S=\frac{\sigma}{\tau}\mathrm{Bez}(a,z)$, so indeed S is a nonzero scalar multiple of the Bezoutian $\mathrm{Bez}(a,z)$. □

Note that the above proof shows that for the case when A and Z do not have a common eigenvalue, the essentially unique similarity S appearing in Proposition 11.1 can be identified as the Bezoutian $\mathrm{Bez}(a, z)$ associated with the characteristic polynomials a and z of A and Z, respectively.

In general, for $A = Z$ the similarity S appearing in Proposition 11.1 is not unique up to multiplication by a nonzero scalar. To see this, let $A = Z$ be the upper triangular 2×2 Jordan block. Then $S^{-1}A^\top S = A$ and $S^{-1}Z^\top S = Z$ for any S of the form

$$\begin{bmatrix} 0 & a \\ a & b \end{bmatrix}, \qquad a \neq 0.$$

The next proposition will be useful in the next two sections.

Proposition 11.2. *Let A and Z be $n \times n$ matrices with* $\operatorname{rank}(A - Z) = 1$. *Then A and Z have no common eigenvalue if and only if there exist invertible $n \times n$ matrices S_1 and S_2 such that*

(i) *$S_1^{-1}AS_1$ and $S_1^{-1}ZS_1$ are first companion matrices, and*
(ii) *$S_2^{-1}AS_2$ and $S_2^{-1}ZS_2$ are second companion matrices.*

Proof. Suppose there exist S_1 and S_2 with the properties (i) and (ii). Introduce

$$A_1 = S_1^{-1}AS_1, \qquad Z_1 = S_1^{-1}ZS_1, \qquad A_2 = S_2^{-1}AS_2, \qquad Z_2 = S_2^{-1}ZS_2.$$

Then A_1 and Z_1 are first companion matrices. Also, A_2 and Z_2 are second companion matrices. Since the characteristic polynomials of A_1 and A_2 are the same (both equal to that of A), we have that $A_2 = A_1^\top$. Similarly $Z_2 = Z_1^\top$. Put $S = S_2^{-1}S_1$. Then S is invertible and

$$A_1 = S^{-1}A_2S = S^{-1}A_1^\top S, \qquad Z_1 = S^{-1}Z_2S = S^{-1}Z_1^\top S.$$

The rank condition on A and Z implies that A_1 and Z_1 are different. Proposition 11.1 now gives that A_1 and Z_1 do not have a common eigenvalue. But then the same is true for A and Z. This settles the if part of the proposition.

Next we focus on the only if part. So assume that, besides the rank condition $\operatorname{rank}(A - Z) = 1$, the matrices A and Z have no common eigenvalue. Write $A - Z = bc^\top$ with $b, c \in \mathbb{C}^n$. The expression $w(\lambda) = 1 + c^\top(\lambda I_n - A)^{-1}b$ is a minimal realization as, by assumption, A and $A^\times = Z$ have no common eigenvalue (cf., the proof of Theorem 10.3). In particular, the $n \times n$ matrix

$$V = \begin{bmatrix} c^\top \\ c^\top A \\ \vdots \\ c^\top A^{n-1} \end{bmatrix}$$

is invertible, with inverse $V^{-1} = \begin{bmatrix} v_0 & v_1 & \cdots & v_{n-1} \end{bmatrix}$. Notice that here the entries $v_0, v_2, \ldots, v_{n-1}$ belong to $\mathbb{C}^n$. Let A_1 and A_2 be the companion matrices (11.1) and (11.2), respectively, where

$$a_j = -c^\top A^n v_j, \qquad j = 0, \ldots, n-1.$$

A straightforward computation, using that

$$\sum_{j=0}^{n-1} v_j c^\top A^j = I_n,$$

shows that $A_1 V = VA$. With H as in (11.4), we have $HA_1 = HA_1^\top = A_2 H$. Put $S_2 = V^{-1}H^{-1}$. Then S_2 is invertible and

$$S_2^{-1} A S_2 = HVAV^{-1}H^{-1} = HA_1H^{-1} = A_2.$$

Thus $S_2^{-1}AS_2$ is a second companion matrix.

The matrix $S_2^{-1}ZS_2$ is second companion too. To see this we argue as follows. Clearly

$$S_2^{-1} Z S_2 = S_2^{-1} A S_2 - S_2^{-1} b c^\top S_2$$

and we need to show that the first $n-1$ columns of the matrix $S_2^{-1}bc^\top S_2$ have only zero entries. For this it is sufficient to establish that the first $n-1$ entries in the row vector $c^\top S_2$ are equal to zero. This, however, is clear from the identity

$$c^\top V^{-1} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} H.$$

This proves (ii).

Statement (i) can be now be obtained in several ways. One way is to employ that the matrix $\begin{bmatrix} b & Ab & \cdots & A^{n-1}b \end{bmatrix}$ is invertible and mimic the above reasoning. Another way is to consider transposes, using that a matrix and its transposed have the same eigenvalues and rank. Finally, one can resort to Proposition 11.1. Indeed, with S_2 as above, $(S_2^{-1}AS_2)^\top$ and $(S_2^{-1}ZS_2)^\top$ are different first companion matrices without a common eigenvalue, and hence there is an invertible matrix T such that $T^{-1}(S_2^{-1}AS_2)T = (S_2^{-1}AS_2)^\top$ and $T^{-1}(S_2^{-1}ZS_2)T = (S_2^{-1}ZS_2)^\top$ are first companions. Now put $S_1 = S_2 T$, and we are done. □

11.2 Simultaneous reduction to complementary triangular forms

In this section we deal with simultaneous reduction to complementary triangular forms of pairs of companion matrices (of the same type). The first main results are the following two theorems.

Theorem 11.3. *Let A and Z be $n\times n$ first companion matrices. Then A and Z admit simultaneous reduction to complementary triangular forms if and only if there exist orderings $\alpha_1,\dots,\alpha_n$ of the eigenvalues of A and $\zeta_1,\dots,\zeta_n$ of the eigenvalues of Z (in both cases algebraic multiplicities taken into account)such that*

$$\alpha_k \neq \zeta_j, \qquad k,j = 1,\dots,n,\ k<j. \tag{11.5}$$

Theorem 11.4. *Let A and Z be $n\times n$ second companion matrices. Then A and Z admit simultaneous reduction to complementary triangular forms if and only if there exist orderings $\alpha_1,\dots,\alpha_n$ of the eigenvalues of A and $\zeta_1,\dots,\zeta_n$ of the eigenvalues of Z (in both cases algebraic multiplicities taken into account) such that*

$$\alpha_k \neq \zeta_j, \qquad k,j = 1,\dots,n,\ k>j. \tag{11.6}$$

One can derive the first theorem from the second and conversely. In other words, the two theorems are equivalent. To see this we make a few observations. First, by the remark made in the paragraph preceding the proof of Theorem 10.2, two $n\times n$ matrices A and Z admit simultaneous reduction to complementary triangular forms if and only if Z and A admit simultaneous reduction to complementary triangular forms. But the latter is equivalent to $A^\top$ and $Z^\top$ admitting simultaneous reduction to complementary triangular forms. Next, note that A and Z are first (second) companion matrices if and only if $A^\top$ and $Z^\top$ are second (first) companion matrices. Finally, conditions (11.5) and (11.6) on the orderings of the eigenvalues of A and Z are equivalent in the following sense. There exist orderings of the eigenvalues of A and of the eigenvalues of Z such that (11.6) holds if and only there are orderings for which (11.5) is satisfied. This follows by just reversing the order. Since the eigenvalues of A (of Z) are the same as those of $A^\top$ ($Z^\top$) with algebraic multiplicities taking into account, we see that Theorems 11.4 and 11.3 are equivalent. Thus it suffices to prove Theorem 11.4.

In order to prove Theorem 11.4 we first introduce some notation and prove a few auxiliary results. Let n be positive integer. Given complex numbers

$$\mu_1,\dots,\mu_{n-1},\ \nu_1,\dots,\nu_{n-1},$$

we let $U(\mu_1,\dots,\mu_{n-1};\nu_1,\dots,\nu_{n-1})$ be the $n\times n$ matrix $\big[u_{k,j}\big]_{k,j=0}^{n-1}$ determined by

$$\sum_{k=0}^{n-1} u_{k,j}\lambda^k = (\lambda-\nu_1)\cdots(\lambda-\nu_j)(\lambda-\mu_{j+1})\cdots(\lambda-\mu_{n-1}). \tag{11.7}$$

For $j=0$ the right side of the above formula reduces to $(\lambda-\mu_1)\cdots(\lambda-\mu_{n-1})$. An analogous interpretation holds for $j=n-1$. Note that the $n\times n$ matrix $U=U(\mu_1,\dots,\mu_{n-1};\nu_1,\dots,\nu_{n-1})$ is uniquely determined by the equation

$$\begin{aligned}\begin{bmatrix} 1 & \lambda & \cdots & \lambda^{n-1}\end{bmatrix} U \\ = \begin{bmatrix} u_0(\lambda) & u_1(\lambda) & \cdots & u_{n-1}(\lambda)\end{bmatrix},\end{aligned} \qquad \lambda\in\mathbb{C}, \tag{11.8}$$

where for $j = 0, \ldots, n-1$ the entry $u_j(\lambda)$ is the polynomial defined by the right-hand side of (11.7).

Lemma 11.5. *For any choice of $\mu_1, \ldots, \mu_{n-1}$ and $\nu_1, \ldots, \nu_{n-1}$ we have*

$$\det U(\mu_1, \ldots, \mu_{n-1}; \nu_1, \ldots, \nu_{n-1}) = \prod_{j=1}^{n-1} \prod_{k=j}^{n-1} (\nu_j - \mu_k). \tag{11.9}$$

In particular, $U(\mu_1, \ldots, \mu_{n-1}; \nu_1, \ldots, \nu_{n-1})$ is non-singular if and only if

$$\mu_k \neq \nu_j, \qquad k, j = 1, \ldots, n-1, \ k \geq j. \tag{11.10}$$

Proof. Note that both sides of the identity (11.9) depend continuously on the parameters $\mu_1, \ldots, \mu_{n-1}$ and $\nu_1, \ldots, \nu_{n-1}$. Thus, in order to prove the equality (11.9), we may assume without loss of generality that the numbers $\nu_1, \ldots, \nu_{n-1}$ are all different. Let ν_n be any complex number different from the numbers $\nu_1, \ldots, \nu_{n-1}$, and let V be the $n \times n$ matrix defined by

$$V = \begin{bmatrix} 1 & \nu_1 & \cdots & \nu_1^{n-1} \\ 1 & \nu_2 & \cdots & \nu_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & \nu_{n-1} & \cdots & \nu_{n-1}^{n-1} \\ 1 & \nu_n & \cdots & \nu_n^{n-1} \end{bmatrix}.$$

The matrix V is a Vandermonde matrix, and hence

$$\det V = \prod_{j=1}^{n} \prod_{k=1}^{j-1} (\nu_j - \nu_k).$$

Put $U = U(\mu_1, \ldots, \mu_{n-1}; \nu_1, \ldots, \nu_{n-1})$. Using (11.8) we see that

$$VU = \begin{bmatrix} u_0(\nu_1) & u_1(\nu_1) & \cdots & u_{n-1}(\nu_1) \\ u_0(\nu_2) & u_1(\nu_2) & \cdots & u_{n-1}(\nu_2) \\ \vdots & \vdots & & \vdots \\ u_0(\nu_{n-1}) & u_1(\nu_{n-1}) & \cdots & u_{n-1}(\nu_{n-1}) \\ u_0(\nu_n) & u_1(\nu_n) & \cdots & u_n(\nu_{n-1}) \end{bmatrix}.$$

Hence VU is a lower triangular $n \times n$ matrix, and for $j = 1, \ldots, n$ the jth diagonal

entry Δ_j of VU is given by

$$\Delta_j = \begin{cases} \displaystyle\prod_{k=1}^{n-1}(\nu_1 - \mu_k), & j = 1, \\ \displaystyle\Big(\prod_{k=1}^{j-1}(\nu_j - \nu_k)\Big)\prod_{k=j}^{n-1}(\nu_j - \mu_k), & j = 2, \dots, n-1, \\ \displaystyle\prod_{k=1}^{n-1}(\nu_n - \nu_k), & j = n. \end{cases}$$

It follows that

$$\det VU = \Big(\prod_{j=1}^{n}\prod_{k=1}^{j-1}(\nu_j - \nu_k)\Big)\prod_{j=1}^{n-1}\prod_{k=j}^{n-1}(\nu_j - \mu_k).$$

Since $\det V \neq 0$, we have $\det U = \det VU / \det V$. This, together with the formulas for $\det VU$ and $\det V$, yields the desired expression for $\det U$. □

Lemma 11.6. *Let A be an $n \times n$ second companion matrix, let $\alpha_1, \dots, \alpha_n$ be the eigenvalues of A (algebraic multiplicities taken into account), let $\zeta_1, \dots, \zeta_n$ be complex numbers, and let $\widetilde{A}$ be the upper triangular $n \times n$ matrix given by*

$$\widetilde{A} = \begin{bmatrix} \alpha_1 & \alpha_1 - \zeta_1 & \cdots & \cdots & \alpha_1 - \zeta_1 & \alpha_1 - \zeta_1 \\ 0 & \alpha_2 & & \cdots & \alpha_2 - \zeta_2 & \alpha_2 - \zeta_2 \\ \vdots & & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & & \alpha_{n-1} & \alpha_{n-1} - \zeta_{n-1} \\ 0 & 0 & \cdots & \cdots & 0 & \alpha_n \end{bmatrix}. \tag{11.11}$$

Put $S = U(\alpha_2, \dots, \alpha_n; \zeta_1, \dots, \zeta_{n-1})$, that is, $S = \big[s_{k,j}\big]_{k,j=0}^{n-1}$ is the $n \times n$ matrix determined by

$$\sum_{k=0}^{n-1} s_{k,j}\lambda^k = (\lambda - \zeta_1)\cdots(\lambda - \zeta_j)(\lambda - \alpha_{j+2})\cdots(\lambda - \alpha_n). \tag{11.12}$$

Then $S\widetilde{A} = AS$.

Proof. We begin with a few observations. Introduce the $n \times n$ matrix

$$F = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & & & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \\ 0 & 0 & & & 1 & -1 \\ 0 & 0 & & \cdots & 0 & 1 \end{bmatrix}.$$

Clearly F is invertible. Furthermore,

$$\widetilde{A}F = \begin{bmatrix} \alpha_1 & -\zeta_1 & 0 & \cdots & 0 & 0 \\ 0 & \alpha_2 & -\zeta_2 & & & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \\ 0 & 0 & & & \alpha_{n-1} & -\zeta_{n-1} \\ 0 & 0 & & \cdots & 0 & \alpha_n \end{bmatrix}.$$

Thus both F and $\widetilde{A}F$ are simple two-diagonal matrices. Also, note that

$$SF = \begin{bmatrix} s_{00} & s_{01} - s_{00} & \cdots & s_{0,\,n-1} - s_{0,\,n-2} \\ s_{10} & s_{11} - s_{10} & \cdots & s_{1,\,n-1} - s_{1,\,n-2} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ s_{n-2,\,0} & s_{n-2,\,1} - s_{n-2,\,0} & \cdots & s_{n-2,\,n-1} - s_{n-2,\,n-2} \\ s_{n-1,\,0} & s_{n-1,\,1} - s_{n-1,\,0} & \cdots & s_{n-1,\,n-1} - s_{n-1,\,n-2} \end{bmatrix}.$$

Finally, since F is invertible, the identity $S\widetilde{A} = AS$ is equivalent to $S\widetilde{A}F = ASF$.

Next, for each $\lambda \in \mathbb{C}$ let $\Lambda(\lambda)$ be the one row matrix appearing as the first factor in the left-hand side of (11.8), that is, $\Lambda(\lambda) = \begin{bmatrix} 1 & \lambda & \cdots & \lambda^{n-1} \end{bmatrix}$. It is sufficient to prove that, regardless of the choice of λ, the one row matrices $\Lambda(\lambda)S\widetilde{A}F$ and $\Lambda(\lambda)ASF$ are the same. Write A in the form (11.2). Then

$$\lambda^n + \sum_{k=0}^{n-1} \lambda^k a_k = \det(\lambda - A) = (\lambda - \alpha_1)(\lambda - \alpha_2) \cdots (\lambda - \alpha_n). \tag{11.13}$$

Using this identity, we have

$$\Lambda(\lambda)A = \begin{bmatrix} \lambda & \lambda^2 & \cdots & \lambda^{n-1} & \lambda^n - a(\lambda) \end{bmatrix},$$

where $a(\lambda) = (\lambda - \alpha_1)(\lambda - \alpha_2)\cdots(\lambda - \alpha_n)$. Furthermore,

$$\Lambda(\lambda)S = \begin{bmatrix} s_0(\lambda) & s_1(\lambda) & \cdots & s_{n-1}(\lambda) \end{bmatrix},$$

where for $j = 0, \ldots, n-1$ the element $s_j(\lambda)$ is equal to the right-hand side of (11.12). Now using the matrix representations of $\widetilde{A}F$ and SF in the first paragraph of the proof it is straightforward to show that $\Lambda(\lambda)S\widetilde{A}F = \Lambda(\lambda)ASF$ for any choice of λ. □

Lemma 11.7. *Let Z be an $n \times n$ second companion matrix, let $\zeta_1, \ldots, \zeta_n$ be the eigenvalues of Z (algebraic multiplicities taken into account), let $\alpha_1, \ldots, \alpha_n$ be complex numbers, and let $\widetilde{Z}$ be the lower triangular $n \times n$ matrix given by*

$$\widetilde{Z} = \begin{bmatrix} \zeta_1 & 0 & \cdots & \cdots & 0 & 0 \\ \zeta_2 - \alpha_2 & \zeta_2 & & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots & \vdots \\ \vdots & & & \ddots & & \vdots \\ \zeta_{n-1} - \alpha_{n-1} & \zeta_{n-1} - \alpha_{n-1} & \cdots & & \zeta_{n-1} & 0 \\ \zeta_n - \alpha_n & \zeta_n - \alpha_n & \cdots & \cdots & \zeta_n - \alpha_n & \zeta_n \end{bmatrix}. \tag{11.14}$$

Also let $S = U(\alpha_2, \ldots, \alpha_n; \zeta_1, \ldots, \zeta_{n-1})$, that is, $S = \left[s_{k,j}\right]_{k,j=0}^{n-1}$ is the $n \times n$ matrix determined by (11.12)*. Then $S\widetilde{Z} = ZS$.*

Proof. Clearly, it is possible to give a direct argument along the lines of the proof of Lemma 11.6. However, we shall follow another approach and derive the lemma as a corollary of Lemma 11.6.

Introduce the $n \times n$ matrix $T = \left[t_{k,j}\right]_{k,j=0}^{n-1}$ by stipulating that

$$\sum_{k=0}^{n-1} t_{k,j}\lambda^k = (\lambda - \alpha_n)\cdots(\lambda - \alpha_{n+1-j})(\lambda - \zeta_{n-1-j})\cdots(\lambda - \zeta_1). \tag{11.15}$$

Thus T is defined in the same way as S in Lemma 11.6 with the understanding that the eigenvalues $\alpha_1, \ldots, \alpha_n$ of A and the complex numbers $\zeta_1, \ldots, \zeta_n$ there are replaced here by the eigenvalues $\zeta_n, \ldots, \zeta_1$ of Z and the complex numbers

$\alpha_n, \ldots, \alpha_1$. It follows that $T\widehat{Z} = ZT$, where

$$\widehat{Z} = \begin{bmatrix} \zeta_n & \zeta_n - \alpha_n & \cdots & \cdots & \zeta_n - \alpha_n & \zeta_n - \alpha_n \\ 0 & \zeta_{n-1} & & \cdots & \zeta_{n-1} - \alpha_{n-1} & \zeta_{n-1} - \alpha_{n-1} \\ \vdots & & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & & \zeta_2 & \zeta_2 - \alpha_2 \\ 0 & 0 & \cdots & \cdots & 0 & \zeta_1 \end{bmatrix}.$$

Observe now that (11.15) can be rewritten as

$$\sum_{k=0}^{n-1} t_{k,j}\lambda^k = (\lambda - \zeta_1)\cdots(\lambda - \zeta_{n-1-j})(\lambda - \alpha_{n+1-j})\cdots(\lambda - \alpha_n).$$

Comparing this with the defining expression for S, we see that

$$t_{k,j} = s_{k,n-1-j}, \qquad k, j = 0, \ldots, n-1.$$

In other words $T = SE$, where E is the $n \times n$ reversed identity matrix (having ones on the antidiagonal and zeros everywhere else). Combining this with $T\widehat{Z} = ZT$, we get $SE\widehat{Z} = ZSE$, and it follows that $S(E\widehat{Z}E) = ZS$. The argument is now completed by observing that $E\widehat{Z}E = \widetilde{Z}$. □

Proof of Theorem 11.4. Let A and Z be second companion matrices. Assume that there exist an ordering $\alpha_1, \ldots, \alpha_n$ of the eigenvalues of A and an ordering $\zeta_1, \ldots, \zeta_n$ of the eigenvalues of Z such that (11.6) is satisfied. Put $S = U(\alpha_2, \ldots, \alpha_n; \zeta_1, \ldots, \zeta_{n-1})$. Using the final part of Lemma 11.5, we see that (11.6) is equivalent to the invertibility of S. But then we can use Lemmas 11.6 and 11.7 to show that $S^{-1}AS$ and $S^{-1}ZS$ are upper triangular and lower triangular, respectively. Thus A and Z admit simultaneous reduction to complementary triangular forms.

Next, we prove the reverse implication. Assume S is an invertible $n \times n$ matrix such that $S^{-1}AS$ is upper triangular with diagonal elements $\alpha_1, \ldots, \alpha_n$ and $S^{-1}ZS$ is lower triangular with diagonal elements $\zeta_1, \ldots, \zeta_n$. Suppose also, contrary to (11.6), that $\alpha_k = \zeta_j$ for some $k > j$. Put $T = (S^\top)^{-1}$. Then $T^{-1}A^\top T = (S^{-1}AS)^\top$ is lower triangular with diagonal elements $\alpha_1, \ldots, \alpha_n$. Write $T^{-1}A^\top T$ in the form

$$T^{-1}A^\top T = \begin{bmatrix} A_1 & 0 \\ * & A_2 \end{bmatrix},$$

with A_1 a $(k-1) \times (k-1)$ matrix and A_2 an $(n-k+1) \times (n-k+1)$ matrix. Clearly A_2 is lower triangular with α_k on the diagonal (actually as first entry).

Hence α_k is an eigenvalue of A_2. Note that a corresponding eigenvector can be transformed into an eigenvector for the full matrix $T^{-1}A^\top T$, again corresponding to the eigenvalue α_k, by adding $k-1$ zeros at the beginning. The upshot of this is that there exists a nonzero vector $a = (a_1, \ldots, a_n)^\top$ in $\operatorname{Ker}(\alpha_k I_n - T^{-1}A^\top T)$ such that

$$a_i = 0, \qquad i = 1, \ldots, k-1. \tag{11.16}$$

Analogously, by virtue of the upper triangularity of $T^{-1}Z^\top T$, there is a nonzero vector $z = (z_1, \ldots, z_n)^\top$ in $\operatorname{Ker}(\zeta_j I_n - T^{-1}A^\top T)$ for which

$$z_i = 0, \qquad i = j+1, \ldots, n. \tag{11.17}$$

Now $Ta \in \operatorname{Ker}(\alpha_k I_n - A^\top)$ and $Tz \in \operatorname{Ker}(\zeta_j I_n - Z^\top)$. Since $A^\top$ is a first companion matrix, the space $\operatorname{Ker}(\alpha_k I_n - A^\top)$ is spanned by the vector $(1, \alpha_k, \ldots, \alpha_k^{n-1})^\top$. Similarly, the space $\operatorname{Ker}(\zeta_j I_n - Z^\top)$ is spanned by the vector $(1, \zeta_j, \ldots, \zeta_j^{n-1})^\top$. But we have assumed that $\alpha_k = \zeta_j$. It follows that the one-dimensional spaces $\operatorname{Ker}(\alpha_k I_n - A^\top)$ and $\operatorname{Ker}(\zeta_j I_n - Z^\top)$ are spanned by one and the same vector. Hence Ta and Tz are scalar multiples of each other. As T is invertible, we conclude that a and z are scalar multiples of each other. Combining this with (11.16) and (11.17), and using that $k > j$, we see that $a = z = 0$, contradicting the fact that a and z are nonzero vectors. □

The above proof provides some additional information on both parts of Theorem 11.4. Indeed, for $n \times n$ second companion matrices A and Z, we have proved the following two facts.

(a) If $\alpha_1, \ldots, \alpha_n$ is an ordering of the eigenvalues of A and $\zeta_1, \ldots, \zeta_n$ is an ordering of the eigenvalues of Z such that the conditions in (11.6) are satisfied, then $S = U(\alpha_2, \ldots, \alpha_n; \zeta_1, \ldots, \zeta_{n-1})$ is invertible, $S^{-1}AS = \widetilde{A}$ is upper triangular with diagonal elements $\alpha_1, \ldots, \alpha_n$, and $S^{-1}ZS = \widetilde{Z}$ is lower triangular with diagonal elements $\zeta_1, \ldots, \zeta_n$. Here $\widetilde{A}$ and $\widetilde{Z}$ are as in (11.11) and (11.14).

(b) If there exists an invertible $n \times n$ matrix S such that $S^{-1}AS$ is upper triangular with diagonal elements $\alpha_1, \ldots, \alpha_n$ and $S^{-1}ZS$ is lower triangular with diagonal elements $\zeta_1, \ldots, \zeta_n$, then the inequalities (11.6) hold (i.e., the matrix $U(\alpha_2, \ldots, \alpha_n; \zeta_1, \ldots, \zeta_{n-1})$ is invertible).

In this form Theorem 11.4 can be seen as a generalizations of Theorem 10.3. Indeed, the hypotheses of Theorem 10.3 imply (by Proposition 11.2) that the given matrices A and Z admit simultaneous reduction to second companion forms, which allows us to derive Theorem 10.3 as a corollary of Theorem 11.4.

In a similar way Theorem 11.3 can be specified further. In fact, by taking transposes (cf., the observations made in the paragraph directly after Theorem 11.4) we see that for $n \times n$ first companion matrices A and Z, the following two statements hold true.

(c) If $\alpha_1, \ldots, \alpha_n$ is an ordering of the eigenvalues of A and $\zeta_1, \ldots, \zeta_n$ is an ordering of the eigenvalues of Z such that the conditions in (11.5) are satisfied, then $T = U(\zeta_2, \ldots, \zeta_n; \alpha_1, \ldots, \alpha_{n-1})^\top$ is invertible, TAT^{-1} is upper triangular with diagonal elements $\alpha_1, \ldots, \alpha_n$, and TZT^{-1} is lower triangular with diagonal elements $\zeta_1, \ldots, \zeta_n$.

(d) If there exists an invertible $n \times n$ matrix T such that TAT^{-1} is upper triangular with diagonal elements $\alpha_1, \ldots, \alpha_n$ and TZT^{-1} is lower triangular with diagonal elements $\zeta_1, \ldots, \zeta_n$, then the inequalities (11.5) hold (i.e., the matrix $U(\zeta_2, \ldots, \zeta_n; \alpha_1, \ldots, \alpha_{n-1})$ is invertible).

The next theorem is the third main result of this section. Its proof will provide further insight in orderings of the eigenvalues of A and Z satisfying (11.5) or (11.6).

Theorem 11.8. *Let A and Z be $n \times n$ companion matrices of the same type (so either both first or both second companions). Then A and Z admit simultaneous reduction to complementary triangular forms if and only if there exists an ordering $\mu_1, \ldots, \mu_s$ of the (different) elements of $\sigma(A) \cup \sigma(Z)$ such that*

$$\sum_{i=1}^{t} m_Z(\mu_i) \leq 1 + \sum_{i=1}^{t-1} m_A(\mu_i), \qquad t = 1, \ldots, s. \tag{11.18}$$

Here, as before, $m_A(\mu)$ denotes the algebraic multiplicity of μ as an eigenvalue of A (taken to be zero when μ is not in the spectrum of A), and likewise with A replaced by Z. At first sight, the theorem seems to be non-symmetric in A and Z, but in fact it is not. This can be seen by taking the elements of $\sigma(A) \cup \sigma(Z)$ in the reversed order $\mu_s, \ldots, \mu_1$ and using that the algebraic multiplicities for A, as well as those for Z, add up to n. Indeed, for $t = 1, \ldots, s$ we have

$$\begin{aligned} \sum_{i=1}^{t} m_A(\mu_{s+1-i}) &= n - \sum_{i=1}^{s-t} m_A(\mu_i) \\ &\leq n + 1 - \sum_{i=1}^{s-t+1} m_Z(\mu_i) = 1 + \sum_{i=1}^{t-1} m_Z(\mu_{s+1-i}). \end{aligned}$$

Proof. As has been established earlier in this section, A and Z admit simultaneous reduction to complementary triangular forms if and only if there exist orderings $\alpha_1, \ldots, \alpha_n$ of the eigenvalues of A and $\zeta_1, \ldots, \zeta_n$ of the eigenvalues of Z such that (11.5) is satisfied. So we need to prove the equivalence of this condition on the eigenvalues of A and Z with the one mentioned in the present theorem which concerns the algebraic multiplicities of the eigenvalues of A and Z. In view of our needs later (see the proof of Theorem 12.8), we shall establish a somewhat more general result. In fact we shall prove that, for h an arbitrary non-negative integer, the following two statements are equivalent:

(A) There exist orderings $\alpha_1, \ldots, \alpha_n$ of the eigenvalues of A and $\zeta_1, \ldots, \zeta_n$ of the eigenvalues of Z such that

$$\alpha_k \neq \zeta_j, \qquad k, j = 1, \ldots, n, \ k \leq j - h. \tag{11.19}$$

(B) There exists an ordering $\mu_1, \ldots, \mu_s$ of the different elements of $\sigma(A) \cup \sigma(Z)$ such that

$$\sum_{i=1}^{t} m_Z(\mu_i) \ \leq \ h + \sum_{i=1}^{t-1} m_A(\mu_i), \qquad t = 1, \ldots, s. \tag{11.20}$$

Note that (11.19) reduces to (11.5) and, likewise, (11.20) boils down to (11.18) by taking $h = 1$. We split the argument into two parts.

Part 1. In this part we prove that (B) implies (A). Assume that there is an ordering $\mu_1, \ldots, \mu_s$ of the (different) elements of $\sigma(A) \cup \sigma(Z)$ such that (11.20) is satisfied. First we introduce an ordering $\alpha_1, \ldots, \alpha_n$ of the eigenvalues of A. Take k among the integers $1, \ldots, n$. Then there exists a unique integer $t(k)$ among $1, \ldots, s$ such that

$$1 + \sum_{i=1}^{t(k)-1} m_A(\mu_i) \ \leq k \ \leq \sum_{i=1}^{t(k)} m_A(\mu_i),$$

and we put $\alpha_k = \mu_{t(k)}$. Note that, for $t = 1, \ldots, s$,

$$\alpha_k = \mu_t \quad \text{for} \quad k = 1 + \sum_{i=1}^{t-1} m_A(\mu_i), \ldots, \sum_{i=1}^{t} m_A(\mu_i).$$

In this way, indeed, $\alpha_1, \ldots, \alpha_n$ is an ordering of the eigenvalues of A (algebraic multiplicities taken into account). This ordering can also be written as

$$\underbrace{\mu_1, \ldots, \mu_1}_{m_A(\mu_1)} \ \underbrace{\mu_2, \ldots, \mu_2}_{m_A(\mu_2)} \ \cdots \ \underbrace{\mu_{s-1}, \ldots, \mu_{s-1}}_{m_A(\mu_{s-1})} \ \underbrace{\mu_s, \ldots, \mu_s}_{m_A(\mu_s)} \tag{11.21}$$

of course with the (natural) convention that an underbraced subsequence of length zero is just absent. In the same vein,

$$\underbrace{\mu_1, \ldots, \mu_1}_{m_Z(\mu_1)} \ \underbrace{\mu_2, \ldots, \mu_2}_{m_Z(\mu_2)} \ \cdots \ \underbrace{\mu_{s-1}, \ldots, \mu_{s-1}}_{m_Z(\mu_{s-1})} \ \underbrace{\mu_s, \ldots, \mu_s}_{m_Z(\mu_s).} \tag{11.22}$$

is an ordering $\zeta_1, \ldots, \zeta_n$ of the eigenvalues of Z (algebraic multiplicities taken into account) with

$$\zeta_k = \mu_t \quad \text{for} \quad k = 1 + \sum_{i=1}^{t-1} m_Z(\mu_i), \ldots, \sum_{i=1}^{t} m_Z(\mu_i).$$

We claim that (11.19) is satisfied.

Let k and j be integers among $1,\ldots,n$, and suppose that $\alpha_k = \zeta_j$. Let t be the unique integer among $1,\ldots,s$ for which $\mu_t = \alpha_k = \zeta_j$. Then

$$1 + \sum_{i=1}^{t-1} m_A(\mu_i) \leq k \leq \sum_{i=1}^{t} m_A(\mu_i)$$

and

$$1 + \sum_{i=1}^{t-1} m_Z(\mu_i) \leq j \leq \sum_{i=1}^{t} m_Z(\mu_i).$$

Combining the appropriate parts of these inequalities with (11.20) gives

$$j \leq \sum_{i=1}^{t} m_Z(\mu_i) \leq h + \sum_{i=1}^{t-1} m_A(\mu_i) \leq k + h - 1,$$

so $k > j - h$, as desired.

Part 2. Next we prove that (A) implies (B). Suppose that $\alpha_1,\ldots,\alpha_n$ is an ordering of the eigenvalues of A and $\zeta_1,\ldots,\zeta_n$ is an ordering of the eigenvalues of Z such that (11.19) is satisfied. We shall first make clear that there is no loss of generality in assuming that the ordering $\alpha_1,\ldots,\alpha_n$ of the eigenvalues of A has the clustered form

$$\underbrace{\mu_1,\ldots,\mu_1}_{m_A(\mu_1)} \quad \underbrace{\mu_2,\ldots,\mu_2}_{m_A(\mu_2)} \quad \ldots \quad \underbrace{\mu_{p-1},\ldots,\mu_{p-1}}_{m_A(\mu_{p-1})} \quad \underbrace{\mu_p,\ldots,\mu_p}_{m_A(\mu_p)} \tag{11.23}$$

where $\mu_1,\ldots,\mu_p$ are the different eigenvalues of A. As there is nothing to prove when $p = 1$, we will consider the case $p > 1$.

Put $\mu_1 = \alpha_1$. Then μ_1 appears at exactly $m_A(\mu_1)$ positions in $\alpha_1,\ldots,\alpha_n$. Let l be the largest integer among $1,\ldots,n$ such that $\alpha_l = \mu_1$. Then $l \geq m_A(\mu_1)$. If $l = m_A(\mu_1)$, the sequence $\alpha_1,\ldots,\alpha_n$ has the form

$$\underbrace{\mu_1,\ldots,\mu_1}_{m_A(\mu_1)} \quad \underbrace{*\,\ldots\ldots\ldots\ldots\ldots\ldots\,*}_{n - m_A(\mu_1)}, \tag{11.24}$$

with μ_1 not appearing among the entries denoted by a star. Now suppose $l > m_A(\mu_1)$. Then $m_A(\mu_1) > 1$ (hence $l > 2$) and there must be an integer t among $2,\ldots,l-1$ such that $\alpha_t \neq \mu_1$. With the help of such a t – which in practice is best taken as small as possible – we produce a new ordering $\widehat{\alpha}_1,\ldots,\widehat{\alpha}_n$ of the eigenvalues of A. Namely by putting μ_1 on the tth position, α_t on the lth position, and leaving the rest intact. One verifies easily that

$$\widehat{\alpha}_k \neq \zeta_j, \qquad k,j = 1,\ldots,n,\ k \leq j - h.$$

Also the largest integer $\widehat{l}$ among $1,\ldots,n$ such that $\alpha_{\widehat{l}} = \mu_1$ is strictly smaller than l. Proceeding in this way, one arrives in a finite number of steps at an ordering –

written again as $\alpha_1, \ldots, \alpha_n$ by slight abuse of notation – of the eigenvalues of A of the form (11.24) and still satisfying (11.19).

Next, put $\mu_2 = \alpha_{1+m_A(\mu_1)}$. Using the same type of reasoning as in the previous paragraph, one sees that it may be assumed that the ordering $\alpha_1, \ldots, \alpha_n$ of the eigenvalues of A has the form

$$\underbrace{\mu_1, \ldots, \mu_1}_{m_A(\mu_1)} \quad \underbrace{\mu_2, \ldots, \mu_2}_{m_A(\mu_2)} \quad \underbrace{* \cdots\cdots\cdots\cdots\cdots\cdots *}_{n - m_A(\mu_1) - m_A(\mu_2)}$$

with μ_1 and μ_2 not among the entries denoted by a star. In case A has only two distinct eigenvalues (so $p = 2$) we are ready. In the situation where $p > 2$, we continue the process, thereby arriving at (11.23) after a finite number of steps. The reasoning (as well as certain arguments given below) can be formalized by using finite induction.

For what follows it is relevant to note that the construction can be arranged in such a way that (11.23) starts with the possible eigenvalues of A not belonging to the spectrum of Z. Indeed, if necessary shift these eigenvalues to the left. Thus, from now on, we assume that the ordering $\alpha_1, \ldots, \alpha_n$ has the form

$$\begin{array}{ccccccc} \underbrace{\mu_1, \ldots, \mu_1}_{m_A(\mu_1)} & \cdots & \underbrace{\mu_r, \ldots, \mu_r}_{m_A(\mu_r)} & \underbrace{\mu_{r+1}, \ldots, \mu_{r+1}}_{m_A(\mu_{r+1})} & \cdots & \underbrace{\mu_p, \ldots, \mu_p}_{m_A(\mu_p),} \end{array} \tag{11.25}$$

with $\mu_1, \ldots, \mu_r$ the r different eigenvalues of A that are not in $\sigma(Z)$, and with $\mu_{r+1}, \ldots, \mu_p$ the $p - r$ different common eigenvalues of $\sigma(A)$ and $\sigma(Z)$. Here we have $0 \leq r \leq p$, with the cases $r = 0$ and $r = p$ corresponding to the situation where $\sigma(A) \subset \sigma(Z)$ and $\sigma(A) \cap \sigma(Z) = \emptyset$, respectively. The eigenvalues $\mu_1, \ldots, \mu_r$ of A (but not of Z) can be taken in any order.

Next we turn to Z. Carrying out the procedure described above, with the necessary alteration of details, the ordering $\zeta_1, \ldots, \zeta_n$ of the eigenvalues of Z can be brought in clustered form too. This time the construction can be carried out in such a way that the ordering ends with the (possible) eigenvalues of Z that do not belong to $\sigma(A)$ and starts with the common eigenvalues of A and Z. In first instance, however, these common eigenvalues do not necessarily appear in the order $\mu_{r+1}, \ldots, \mu_p$ in which they come in (11.25). Thus the clustered ordering for Z has the form

$$\begin{array}{ccccccc} \underbrace{\mu_{\sigma(r+1)}, \ldots, \mu_{\sigma(r+1)}}_{m_Z\left(\mu_{\sigma(r+1)}\right)} & \cdots & \underbrace{\mu_{\sigma(p)}, \ldots, \mu_{\sigma(p)}}_{m_Z\left(\mu_{\sigma(p)}\right)} & \underbrace{\mu_{p+1}, \ldots, \mu_{p+1}}_{m_Z(\mu_{p+1})} & \cdots & \underbrace{\mu_s, \ldots, \mu_s}_{m_Z(\mu_s)} \end{array} \tag{11.26}$$

where σ is a suitable permutation of $r+1, \ldots, p$, and $\mu_{p+1}, \ldots, \mu_s$ are the different eigenvalues of Z that do not belong to $\sigma(A)$. Here $r \leq p \leq s$, with the cases $p = r$ and $p = s$ corresponding to the (extreme) situation where $\sigma(A) \cap \sigma(Z) = \emptyset$ and $\sigma(Z) \subset \sigma(A)$, respectively. The eigenvalues $\mu_{p+1}, \ldots, \mu_s$ of Z (but not of A) can

be taken in any order. As we have the disjoint union

$$\sigma(A) \cup \sigma(Z) = \big(\sigma(A) \setminus \sigma(Z)\big) \cup \Big(\sigma(A) \cap \sigma(Z)\Big) \cup \big(\sigma(Z) \setminus \sigma(A)\big),$$

the number of different elements in $\sigma(A) \cup \sigma(Z)$ is $r + (p-r) + (s-p)$, hence this number is equal to s. In line with this, $\mu_1, \ldots, \mu_s$ is an ordering of the different elements of $\sigma(A) \cup \sigma(Z)$. This ordering satisfies (11.20). However, to see this, the above ordering of the eigenvalues of Z needs to be cleaned up first.

Suppose the permutation σ is not the identity mapping on $r+1, \ldots, p$ (so in particular $p > r+1$), and let k be the unique integer among $r+1, \ldots, p$ such that $\sigma(j) = j,\ j = k+1, \ldots, p$ and $\sigma(k) \neq k$ (in particular $k = p$ when $\sigma(p) \neq p$). Write $\sigma(k) = q$. Then $r+1 \leq q < k$ and the ordering (11.25) of the eigenvalues of A has the form

$$\begin{array}{ccccccc}
\underbrace{\mu_1, \ldots, \mu_1}_{m_A(\mu_1)} & \cdots\cdots & \underbrace{\mu_r, \ldots, \mu_r}_{m_A(\mu_r)} & & & & \\
\underbrace{\mu_{r+1}, \ldots, \mu_{r+1}}_{m_A(\mu_{r+1})} & \cdots\cdots & \underbrace{\mu_{q-1}, \ldots, \mu_{q-1}}_{m_A(\mu_{q-1})} & & & & \\
\underbrace{\mu_q, \ldots, \mu_q}_{m_A(\mu_q)} & \underbrace{\mu_{q+1}, \ldots, \mu_{q+1}}_{m_A(\mu_{q+1})} & \cdots & \underbrace{\mu_{k-1}, \ldots, \mu_{k-1}}_{m_A(\mu_{-1})} & \underbrace{\mu_k, \ldots, \mu_k}_{m_A(\mu_k)} & & \\
& \underbrace{\mu_{k+1}, \ldots, \mu_{k+1}}_{m_A(\mu_q)} & \cdots\cdots & \underbrace{\mu_p, \ldots, \mu_p}_{m_A(\mu_p)} & & &
\end{array}.$$

Also, $k = \sigma(l)$ for some l among the numbers $(r+1), \ldots, (k-1)$, and the ordering of the eigenvalues of Z obtained in the preceding paragraph looks like

$$\begin{array}{ccccccc}
\underbrace{\mu_{\sigma(r+1)}, \ldots, \mu_{\sigma(r+1)}}_{m_Z\left(\mu_{\sigma(r+1)}\right)} & \cdots\cdots & \underbrace{\mu_{\sigma(l-1)}, \ldots, \mu_{\sigma(l-1)}}_{m_Z\left(\mu_{\sigma(l-1)}\right)} & & & & \\
\underbrace{\mu_k, \ldots, \mu_k}_{m_Z(\mu_k)} & \underbrace{\mu_{\sigma(l+1)}, \ldots, \mu_{\sigma(l+1)}}_{m_Z\left(\mu_{\sigma(l+1)}\right)} & \cdots & \underbrace{\mu_{\sigma(k-1)}, \ldots, \mu_{\sigma(k-1)}}_{m_Z\left(\mu_{\sigma(k-1)}\right)} & \underbrace{\mu_q, \ldots, \mu_q}_{m_Z(\mu_q)} & & \\
& \underbrace{\mu_{\sigma(k+1)}, \ldots, \mu_{\sigma(k+1)}}_{m_Z\left(\mu_{\sigma(k+1)}\right)} & \cdots\cdots & \underbrace{\mu_{\sigma(p)}, \ldots, \mu_{\sigma(p)}}_{m_Z\left(\mu_{\sigma(p)}\right)} & & & \\
& \underbrace{\mu_{p+1}, \ldots \mu_{p+1}}_{m_Z(\mu_{p+1})} & \cdots\cdots, & \underbrace{\mu_s, \ldots, \mu_s}_{m_Z(\mu_s)} & & &
\end{array}.$$

At this point it is crucial to observe that, without violating the property embodied in (11.19), one can replace

$$\begin{array}{ccccccc}
\underbrace{\mu_k, \ldots, \mu_k}_{m_Z(\mu_k)} & \underbrace{\mu_{\sigma(l+1)}, \ldots, \mu_{\sigma(l+1)}}_{m_Z\left(\mu_{\sigma(l+1)}\right)} & \cdots & \underbrace{\mu_{\sigma(k-1)}, \ldots, \mu_{\sigma(k-1)}}_{m_Z\left(\mu_{\sigma(k-1)}\right)} & \underbrace{\mu_q, \ldots, \mu_q}_{m_Z(\mu_q)} & &
\end{array}$$

in the ordering for Z by

$$\begin{array}{cccccc} \underbrace{\mu_{\sigma(l+1)},\dots,\mu_{\sigma(l+1)}} & \cdots & \underbrace{\mu_{\sigma(k-1)},\dots,\mu_{\sigma(k-1)}} & \underbrace{\mu_q,\dots,\mu_q} & \underbrace{\mu_k,\dots,\mu_k} \\ m_Z\left(\mu_{\sigma(l+1)}\right) & \dots & m_Z\left(\mu_{\sigma(k-1)}\right) & m_Z(\mu_q) & m_Z(\mu_k) \end{array}$$

and it can be concluded that we may change the permutation σ, already satisfying $\sigma(j) = j,\ j = k+1,\dots,p$, so that $\sigma(k) = k$ too. To be precise, if $\widehat{\sigma}$ is the permutation of $r+1,\dots,p$ given by

$$\widehat{\sigma}(j) = \begin{cases} \sigma(j), & j = r+1,\dots,l-1, \\ \sigma(j+1), & j = l,\dots,k-1, \\ \sigma(l), & j = k, \\ \sigma(j), & j = k+1,\dots,p \end{cases}$$

(so $\widehat{\sigma}(k-1) = \sigma(k) = q$ and $\widehat{\sigma}(k) = \sigma(l) = k$), then the ordering (11.26) of the eigenvalues of Z may be replaced by

$$\begin{array}{cccccccc} \underbrace{\mu_{\widehat{\sigma}(r+1)},\dots,\mu_{\widehat{\sigma}(r+1)}} & \cdots & \underbrace{\mu_{\widehat{\sigma}(p)},\dots,\mu_{\widehat{\sigma}(p)}} & \underbrace{\mu_{p+1},\dots,\mu_{p+1}} & \cdots & \underbrace{\mu_s,\dots,\mu_s} \\ m_Z\left(\mu_{\widehat{\sigma}(r+1)}\right) & \dots & m_Z\left(\mu_{\widehat{\sigma}(p)}\right) & m_Z(\mu_{p+1}) & \dots & m_Z(\mu_s) \end{array}$$

where $\widehat{\sigma}(j) = j,\ j = k,\dots,p$. It follows, formally by finite induction, that σ can be taken to be the identity mapping on $r+1,\dots,p$. Thus, from now on, we assume that the ordering $\zeta_1,\dots,\zeta_n$ of the eigenvalues of Z has the form

$$\begin{array}{cccccccc} \underbrace{\mu_{r+1},\dots,\mu_{r+1}} & \cdots & \underbrace{\mu_p,\dots,\mu_p} & \underbrace{\mu_{p+1},\dots,\mu_{p+1}} & \cdots & \underbrace{\mu_s,\dots,\mu_s} \\ m_Z(\mu_{r+1}) & \dots & m_Z(\mu_p) & m_Z(\mu_{p+1}) & \dots & m_Z(\mu_s) \end{array} \tag{11.27}$$

(cf., the ordering (11.25) of the eigenvalues of A), and we are ready to establish (11.20).

Take t from the integers $1,\dots,s$. If $t \le r$, the complex numbers $\mu_1,\dots,\mu_t$ do not belong to $\sigma(Z)$, and the inequality in (11.20) holds trivially because its left-hand side vanishes (and h is non-negative by assumption). If $t > p$, the complex numbers $\mu_{p+1},\dots,\mu_{t-1}$ do not belong to $\sigma(A)$. Hence the right-hand side of the inequality in (11.20) comes down to

$$h + \sum_{i=1}^{t-1} m_A(\mu_i) = h + \sum_{i=1}^{p} m_A(\mu_i) = h + n.$$

As the left-hand side of the inequality in (11.20) certainly does not exceed n, the desired inequality is again trivial. Suppose now that $r+1 \le t \le p$. Then $\mu_t \in \sigma(A) \cap \sigma(Z)$ and μ_t appears in (11.25) and (11.27) at the positions

$$k = 1 + \sum_{i=1}^{t-1} m_A(\mu_i), \qquad j = \sum_{i=1}^{t} m_Z(\mu_i), \tag{11.28}$$

respectively. But then, as the condition (11.19) is met, it is impossible that $k \leq j - h$. Thus $j \leq h + k - 1$, which is exactly what (11.20) says. □

Part 2 of the above proof actually contains an algorithm. Given orderings $\alpha_1, \ldots, \alpha_n$ of the eigenvalues of A and $\zeta_1, \ldots, \zeta_n$ of the eigenvalues of Z such that (11.5) is satisfied, the algorithm produces new orderings featuring a special structure, and it does so keeping the property embodied in (11.5) intact. Roughly speaking, the special structure in question comes down to the following: the eigenvalues are clustered in blocks (possibly empty), and these come in the same order for A as for Z. For a graphic depiction, see the expressions (11.21) and (11.22). In the next chapter (Section 12.3) we shall encounter a basically identical situation in the context of job scheduling. It is for that reason that Part 2 of the proof of Theorem 11.8 has been given in considerable detail. This enables us to be brief on the point in question later. We complete the discussion with an example illustrating the algorithm.

Example. Let A and Z be the 10×10 (first) companion matrices associated with the polynomials

$$\begin{aligned} a(\lambda) &= (\lambda-1)(\lambda-3)^4(\lambda-4)(\lambda-5)^3(\lambda-6), \\ z(\lambda) &= (\lambda-2)^2(\lambda-3)^3(\lambda-4)(\lambda-5)^2(\lambda-6)^2, \end{aligned}$$

respectively. The eigenvalues of A are the zeros of a, and the eigenvalues of Z are the zeros of z (the appropriate multiplicities counted in both cases). Here are orderings of the eigenvalues of A and Z respecting (11.19) with $h = 2$:

$$\begin{array}{lcccccccccc} A: & 1 & 3 & 3 & 3 & 5 & 5 & 3 & 4 & 5 & 6, \\ Z: & 3 & 3 & 3 & 5 & 4 & 5 & 2 & 6 & 2 & 6. \end{array}$$

We now follow the algorithm developed in Part 2 of the proof of Theorem 11.8. Accordingly, we begin by dealing with the eigenvalues of A. First, the eigenvalue 3 in the seventh position is interchanged with the eigenvalue 5 in the fifth:

$$\begin{array}{lcccccccccc} A: & 1 & 3 & 3 & 3 & 3 & 5 & 5 & 4 & 5 & 6, \\ Z: & 3 & 3 & 3 & 5 & 4 & 5 & 2 & 6 & 2 & 6. \end{array}$$

Next interchange the eigenvalue 5 in the ninth position with the eigenvalue 4 in the eighth:

$$\begin{array}{lcccccccccc} A: & 1 & 3 & 3 & 3 & 3 & 5 & 5 & 5 & 4 & 6, \\ Z: & 3 & 3 & 3 & 5 & 4 & 5 & 2 & 6 & 2 & 6. \end{array}$$

With this, our operations on the eigenvalues of A are completed, and we turn to those of Z. Here, consecutively, we interchange the eigenvalue 6 in the eighth position with the eigenvalue 2 in the ninth position, and the eigenvalue 5 in the fourth position with the eigenvalue 4 in the fifth:

$$\begin{array}{lcccccccccc} A: & 1 & 3 & 3 & 3 & 3 & 5 & 5 & 5 & 4 & 6, \\ Z: & 3 & 3 & 3 & 4 & 5 & 5 & 2 & 2 & 6 & 6. \end{array}$$

Although both orderings now feature the desired block structure, the process has not been finished yet. One issue is the position of the eigenvalues having zero-multiplicity for either A or Z. These are the eigenvalue 1 of A and 2 of Z. Now 1 is already in the desired leftmost position, but the eigenvalues 2 do not yet appear at the far right. This, however, can easily be arranged by shifting the two eigenvalues 2 to the right, i.e., by replacing the sequence 2 2 6 6 in the ordering for Z by 6 6 2 2:

$$\begin{array}{lcccccccccc} A: & 1 & 3 & 3 & 3 & 3 & 5 & 5 & 5 & 4 & 6, \\ Z: & 3 & 3 & 3 & 4 & 5 & 5 & 6 & 6 & 2 & 2. \end{array}$$

The next (and final) point to be addressed is that in the ordering of the eigenvalues of Z, the eigenvalue 4 precedes the eigenvalues 5, while in the ordering of the eigenvalues of A this is the other way around. The remedy consists of replacing the sequence 4 5 5 in the ordering for Z by 5 5 4:

$$\begin{array}{lcccccccccc} A: & 1 & 3 & 3 & 3 & 3 & 5 & 5 & 5 & 4 & 6, \\ Z: & 3 & 3 & 3 & 5 & 5 & 4 & 6 & 6 & 2 & 2. \end{array}$$

With $\mu_1 = 1$, $\mu_2 = 3$, $\mu_3 = 5$, $\mu_4 = 4$, $\mu_5 = 6$ and $\mu_6 = 2$, we now have an ordering $\mu_1, \ldots, \mu_6$ (i.e, 1, 3, 5, 4, 6, 2) of the different elements of $\sigma(A) \cup \sigma(Z)$ satisfying (11.20) with $h = 2$.

This is not the only ordering of this type. Indeed, another one is obtained by interchanging the eigenvalues 3 and 5:

$$\begin{array}{lcccccccccc} A: & 1 & 5 & 5 & 5 & 3 & 3 & 3 & 3 & 4 & 6, \\ Z: & 5 & 5 & 3 & 3 & 3 & 4 & 6 & 6 & 2 & 2. \end{array} \quad . \tag{11.29}$$

In fact, the ordering in question (i.e., 1, 5, 3, 4, 6, 2) of the different elements of $\sigma(A) \cup \sigma(Z)$ even satisfies (11.20) with $h = 1$, and it follows from Theorem 11.8 that the companion matrices A and Z admit simultaneous reduction to complementary triangular forms (cf., the Examples in Sections 11.5 and 12.3.

11.3 Preliminaries about companion based matrix functions

From the material presented in Section 4.1 it is clear that a proper rational operator function can be realized in such a way that both the main operator and the associate main operator have the form of an operator companion (cf., the expressions (4.1) for A, B and C in Theorem 4.1). This holds in particular for proper rational matrix functions, and there block matrix companions take the role of operator companions. In general, however, one cannot make realizations with ordinary companions of the type discussed in the preceding section. Here we shall study the case when one can, even when the extra condition of minimality is imposed.

A rational $m \times m$ matrix function W will be called (*first*) *companion based* if it admits a minimal realization

$$W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B \tag{11.30}$$

with A and $A^\times = A - BC$ first companion matrices.

Clearly, there is an alternative concept where first companions are replaced by their transposes, the second companions, but there is no need to pursue this issue here. Therefore in the following we will usually drop the qualifier "first" and simply speak about companion based matrix functions meaning all the time first companion based matrix functions.

We begin with two preliminary propositions. The first is a simple uniqueness result.

Proposition 11.9. *Let W be a companion based $m \times m$ matrix function, and for $j = 1, 2$, let*

$$W(\lambda) = I_m + C_j(\lambda I_n - A_j)^{-1}B_j$$

be a minimal realization of W such that A_j and $A_j^\times = A_j - B_jC_j$ are first companions. Then $A_1 = A_2$ and $A_1^\times = A_2^\times$.

Proof. By the state space isomorphism theorem, the matrices A_1 and A_2 are similar. Hence the characteristic polynomials of A_1 and A_2 are the same. As A_1 and A_2 are first companions, $A_1 = A_2$ follows. The identity $A_1^\times = A_2^\times$ is obtained in the same way using the similarity of $A_1^\times$ and $A_2^\times$. □

Next we investigate in how far some simple operations leave the property of being companion based intact. If W is a rational $m \times m$ matrix function, then the functions W^{-1} and $W^\top$ are given by $W^{-1}(\lambda) = W(\lambda)^{-1}$ (when $\det W(\lambda)$ does not vanish identically) and $W^\top(\lambda) = W(\lambda)^\top$. Also, for T an invertible $m \times m$ matrix, W_T denotes the function defined by $W_T(\lambda) = T^{-1}W(\lambda)T$. Note that a companion based matrix function W is biproper, and hence W^{-1} exists.

Proposition 11.10. *Assume W is a companion based $m \times m$ matrix function. Then the following holds.*

(i) *The matrix function W^{-1} is companion based.*

(ii) *The matrix function $W^\top$ is companion based if and only if either W and W^{-1} have no common pole, or (the other extreme) W and W^{-1} have the same poles, pole-multiplicities taken into account.*

(iii) *For T an invertible $m \times m$ matrix, W_T is companion based.*

Recall that the poles of W^{-1} coincide with the zeros of W, with pole-multiplicities and zero-multiplicities corresponding to each other (see Chapter 8).

Proof. Let (11.30) be a minimal realization of W with A and $A^\times = A - BC$ first companion matrices. Then

$$W^{-1}(\lambda) = I_m - C(\lambda I_n - A^\times)^{-1}B$$

is a minimal realization of W^{-1} for which both $A^\times$ and $(A^\times)^\times = A^\times + BC = A$ are first companions. This proves the first item of the proposition.

Turning to the second item, we first assume that the poles of W and W^{-1} meet the requirement mentioned in the theorem. In terms of the minimal realization (11.30), involving first companions A and $A^\times$, this means that either A and $A^\times$ have no common eigenvalue, or (the other extreme) A and $A^\times$ have the same characteristic polynomial and are therefore the same. But then Proposition 11.1 guarantees that there exists an invertible $n \times n$ matrix S such that $S^{-1}A^\top S = A$ and $S^{-1}(A^\times)^\top S = A^\times$. Now

$$W^\top(\lambda) = I_m + B^\top(\lambda I_n - A^\top)^{-1}C^\top \tag{11.31}$$

is a minimal realization of $W^\top$. Replacing $A^\top$ by SAS^{-1}, we obtain another minimal realization for $W^\top$, namely

$$W^\top(\lambda) = I_m + B^\top S(\lambda I_n - A)^{-1}S^{-1}C^\top.$$

As $A - S^{-1}C^\top B^\top S = S^{-1}(A^\top - C^\top B^\top)S = S^{-1}(A^\times)^\top S = A^\times$, we can conclude that $W^\top$ is companion based.

Next, assume that $W^\top$ is companion based. Then $W^\top$ admits a minimal realization

$$W^\top(\lambda) = I_n + \widehat{C}(\lambda I_m - \widehat{A})^{-1}\widehat{B} \tag{11.32}$$

with $\widehat{A}$ and $\widehat{A}^\times = \widehat{A} - \widehat{B}\widehat{C}$ first companions. The state space similarity theorem, applied to (11.31) and (11.32), guarantees the existence of an invertible $n \times n$ matrix S such that

$$S\widehat{A}S^{-1} = A^\top, \qquad \widehat{A}^\times S^{-1} = S\big(A^\top - C^\top B^\top\big)S^{-1} = (A^\times)^\top,$$

In particular, the characteristic polynomials of $\widehat{A}$ and A coincide, and the same is true for those of $\widehat{A}^\times$ and $A^\times$. As we are dealing here with first companions, we may conclude that $\widehat{A} = A$ and $\widehat{A}^\times = A^\times$. But then $SAS^{-1} = A^\top$ and $SA^\times S^{-1} = (A^\times)^\top$. Proposition 11.1 now gives that either A and $A^\times$ do not have a common eigenvalue, or (the other extreme) A and $A^\times$ are identical. In view of the minimality of (11.30), this amounts exactly to what should be established for the poles of W and W^{-1}. This completes the proof of item (ii).

To prove that W_T is companion based, we start with the minimal realization (11.30) for W, assuming (as before) that A and $A^\times = A - BC$ are first companions. The function W_T can be represented as

$$W_T(\lambda) = I_m + T^{-1}C(\lambda I_n - A)^{-1}BT,$$

and this again is a minimal realization. The desired conclusion is now immediate from $A - (BT)(T^{-1}C) = A - BC = A^\times$. □

In Section 10.3, we already mentioned that rational matrix functions of the type featuring in Theorem 10.12 form a subclass of the companion based matrix functions introduced in the present section. The next two propositions make this statement explicit.

Proposition 11.11. *Let W be a rational $m \times m$ matrix function, and let*

$$W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$$

be a realization of W such that $\operatorname{rank} BC = 1$ *or, what amounts to the same,* $\operatorname{rank}(A - A^{\times}) = 1$. *Suppose, in addition, that A and $A^{\times}$ have no common eigenvalue. Then W is companion based.*

Proof. Since A and $A^{\times}$ have no common eigenvalue, it follows from Theorem 7.6 that the given realization of W is minimal. Also, by Proposition 11.2 there exists an invertible $n \times n$ matrix such that $S^{-1}AS$ and $S^{-1}A^{\times}S$ are first companion matrices. Write $\widehat{A} = S^{-1}AS$, $\widehat{B} = S^{-1}B$ and $\widehat{C} = CS$. Then

$$W(\lambda) = I_m + \widehat{C}(\lambda I_n - \widehat{A})^{-1}\widehat{B}$$

is a minimal realization of W for which $\widehat{A} = S^{-1}AS$ and $\widehat{A}^{\times} = \widehat{A} - \widehat{B}\widehat{C} = S^{-1}A^{\times}S$ are first companions. Hence W is companion based. □

Proposition 11.12. *Let W be a rational $m \times m$ matrix function, let*

$$W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$$

be a minimal realization of W, and assume $\operatorname{rank} BC = 1$ *or, what amounts to the same,* $\operatorname{rank}(A - A^{\times}) = 1$. *Suppose, in addition, that W and W^{-1} have no common pole or, equivalently, that the set of poles of W is disjoint from the set of zeros of W. Then W is companion based.*

Proof. As the given realization is minimal, the poles of W and W^{-1} coincide with the eigenvalues of A and $A^{\times}$, respectively. Hence A and $A^{\times}$ have no common eigenvalue. With this we are back in the situation of the previous proposition. □

11.4 Companion based matrix functions: poles and zeros

In this section we deal with the problem of describing companion based matrix functions with prescribed poles and zeros. We start with some terminology.

Let W be a rational $m \times m$ matrix function having the value I_m at infinity.

By the *pole-polynomial* of W we mean the (monic) scalar polynomial $(\lambda - \alpha_1) \cdots (\lambda - \alpha_n)$, where $\alpha_1, \ldots, \alpha_n$ are the poles of W, pole-multiplicities (see Section 8.2) taken into account. The pole-polynomial of W^{-1} will be referred to as

the *zero-polynomial* of W. It has the form $(\lambda-\alpha_1^\times)\cdots(\lambda-\alpha_n^\times)$, where $\alpha_1^\times,\ldots,\alpha_n^\times$ are the zeros of W, zero-multiplicities (see Section 8.1) taken into account. The pole-polynomial and the zero-polynomial of W have the same degree, namely the McMillan degree $\delta(W)$ of W. Actually, if (11.30) is a minimal realization of W, then the pole-polynomial of W coincides with the characteristic polynomial of A, and the zero-polynomial of W is identical to the characteristic polynomial of $A^\times$.

Suppose W is a companion based $m\times m$ matrix function. By Proposition 11.9, a companion based matrix function uniquely determines a pair of first companion matrices. As a matter of fact, if (11.30) is a minimal realization of W with A and $A^\times$ first companions, the pair in question is $A, A^\times$. These companions are completely determined by the pole-polynomial and zero-polynomial of W; the converse is also true.

We now turn to the issue of describing the companion based matrix functions having a given pole and zero-polynomial. So the problem is the following: given two monic scalar polynomials p and $p^\times$ of the same degree, find all companion based $m\times m$ matrix functions having p as pole-polynomial and $p^\times$ as zero-polynomial. In this connection, we will especially pay attention to the size m. One more preliminary remark (showing that with any m each larger integer will do) is in order here.

Suppose W is a companion based $m\times m$ matrix function having pole-polynomial p and zero-polynomial $p^\times$, let I be an identity matrix of arbitrary size, k say, and define $W_{\rm ext}$ by

$$W_{\rm ext}(\lambda)=\begin{bmatrix} I & 0 \\ 0 & W(\lambda)\end{bmatrix}. \tag{11.33}$$

Then $W_{\rm ext}$ is again companion based. Indeed, if (11.30) is a minimal realization of W with A and $A^\times$ first companions, then,

$$W_{\rm ext}(\lambda)=I_{m+k}+\begin{bmatrix}0\\ C\end{bmatrix}(\lambda I_n-A)^{-1}\begin{bmatrix}0 & B\end{bmatrix}.$$

This is again a minimal realization, and the desired conclusion comes from

$$A-\begin{bmatrix}0 & B\end{bmatrix}\begin{bmatrix}0\\ C\end{bmatrix}=A-BC=A^\times.$$

It also follows that $W_{\rm ext}$ has p as its pole-polynomial and $p^\times$ as its zero-polynomial.

We are now ready to deal with the problem formulated above. In light of the remark contained in the preceding paragraph, we add as an additional requirement that the size m of the companion based functions sought for should be as small as possible. As we will see, this brings us to the low-dimensional cases $m=1$ (scalar functions) and $m=2$ (2×2 matrix functions).

Theorem 11.13. *Let p and $p^\times$ be monic scalar polynomials of the same positive degree, n say. Then there exists a scalar companion based function having p as pole-polynomial and $p^\times$ as zero-polynomial if and only if p and $p^\times$ have no common zero. In that situation, the function w given by*

$$w(\lambda) = \frac{p^\times(\lambda)}{p(\lambda)}. \tag{11.34}$$

is the unique scalar companion based function having p as pole-polynomial and $p^\times$ as zero-polynomial.

Proof. Assume p and $p^\times$ have no common zero. As is explained in the first paragraph of Section 10.3, the scalar function w given by (11.34) admits a minimal realization of the type $w(\lambda) = 1 + c^\top(\lambda I_n - A)^{-1}b$ with $b, c \in \mathbb{C}^n$, A an upper triangular $n \times n$ matrix having the zeros of p on the diagonal, and $A^\times = A - bc^\top$ a lower triangular $n \times n$ matrix having the zeros of $p^\times$ on the diagonal. In particular A and $A^\times$ have no common eigenvalue, and it also follows that $A - A^\times = bc^\top$ has rank one. Proposition 11.11 now gives that w is companion based. Also w has p as its pole-polynomial and $p^\times$ as its zero-polynomial. This can be seen directly from (11.34), but it is clear as well from the fact that the characteristic polynomial of A is p and that for $A^\times$ is $p^\times$.

Now suppose that w is a a scalar companion based function having p as pole-polynomial and $p^\times$ as zero-polynomial. We shall prove that p and $p^\times$ have no common zero and, in addition, that w is necessarily given by (11.34).

Clearly, the McMillan degree of W is n. Let

$$w(\lambda) = 1 + c^\top(\lambda I_n - A)^{-1}b, \tag{11.35}$$

with $b, c \in \mathbb{C}^n$, be a minimal realization such that A and $A^\times = A - bc^\top$ are first companions. From the proof of Theorem 10.3 we know that

$$w(\lambda) = \frac{\det(\lambda I_n - A^\times)}{\det(\lambda I_n - A)}.$$

However, the characteristic polynomials of A and $A^\times$ are p and $p^\times$, respectively, and it follows that w is given by (11.34). Assume now that p and $p^\times$ do have a common eigenvalue. Then w can be written as a quotient of two polynomials of degree less than n. Again referring to the first paragraph of Section 10.3, we conclude that w has a realization with state space dimension less than n. But that is impossible in view of the minimality of (11.35). □

Next we consider the more complicated situation when the given polynomials p and $p^\times$ do have a common zero. In that case scalar companion based functions are ruled out (by Theorem 11.13). The next result shows that one can always make do with 2×2 matrix functions.

Theorem 11.14. *Let W be a rational 2×2 matrix function, let p and $p^\times$ be monic polynomials of the same positive degree, n say, and suppose p and $p^\times$ have at least one common zero. Then the following statements are equivalent:*

(i) *W is companion based with pole-polynomial p and zero-polynomial $p^\times$,*

(ii) *W is of the form*

$$W(\lambda) = T^{-1} \begin{bmatrix} 1 & \dfrac{r(\lambda)}{p(\lambda)} \\ 0 & \dfrac{p^\times(\lambda)}{p(\lambda)} \end{bmatrix} T, \tag{11.36}$$

where T is an invertible 2×2 matrix, r is a scalar polynomial of degree not exceeding $n-1$ while, moreover, p, $p^\times$ and r do not have a common zero.

The latter requirement implies, given the assumption that p and $p^\times$ have a zero in common, that r cannot be the zero-polynomial. Scalar polynomials of degree zero (so nonzero constants) are not ruled out, however.

In order to prove Theorem 11.14 it will be convenient to prove first the following auxiliary result.

Lemma 11.15. *Let W be a proper rational 2×2 matrix function of the form*

$$W(\lambda) = \begin{bmatrix} 1 & w_{12}(\lambda) \\ 0 & w_{22}(\lambda) \end{bmatrix} \quad \text{and} \quad \lim_{\lambda \to \infty} W(\lambda) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{11.37}$$

Then W is companion based, its pole-polynomial $p(\lambda)$ is equal to the least common multiple of the denominators of $w_{12}(\lambda)$ and $w_{22}(\lambda)$, and its zero-polynomial $p^\times(\lambda)$ is given by $p^\times(\lambda) = p(\lambda)w_{22}(\lambda)$.

Proof. Let $q(\lambda)$ be the least common multiple of the denominators of $w_{12}(\lambda)$ and $w_{22}(\lambda)$. Thus $q(\lambda)$ is a monic polynomial, $q(\lambda)w_{12}(\lambda)$ and $q(\lambda)w_{22}(\lambda)$ are polynomials, and there is no monic polynomial of smaller degree with the same properties. Put

$$r(\lambda) = q(\lambda)w_{12}(\lambda), \qquad q^\times(\lambda) = q(\lambda)w_{22}(\lambda). \tag{11.38}$$

We split the (remaining part of the) proof into three parts. The first part has a preliminary character. In the second part we prove that q is the pole-polynomial of W, and the final part we show that W is companion based and that $q^\times$ is its zero-polynomial.

Part 1. We claim that the polynomials q, $q^\times$ and r do not have a common zero. Indeed, assume that α is a common zero of these three polynomials. Then there exists polynomials q_1, $q_1^\times$ and r_1 such that

$$q(\lambda) = q_1(\lambda)(\lambda - \alpha), \qquad q^\times(\lambda) = q_1^\times(\lambda)(\lambda - \alpha), \qquad r(\lambda) = r_1(\lambda)(\lambda - \alpha).$$

It follows that

$$w_{12}(\lambda) = \frac{r(\lambda)}{q(\lambda)} = \frac{r_1(\lambda)}{q_1(\lambda)}, \qquad w_{22}(\lambda) = \frac{q^\times(\lambda)}{q(\lambda)} = \frac{q_1^\times(\lambda)}{q_1(\lambda)}.$$

Thus $q_1(\lambda)w_{12}(\lambda)$ and $q_1(\lambda)w_{22}(\lambda)$ are polynomials, and $q_1(\lambda)$ is a monic polynomial of degree strictly less than the degree of $q(\lambda)$, which is impossible. Thus q, $q^\times$ and r do not have a common zero.

From (11.38) and the first identity in (11.37) we see that

$$W(\lambda) = \begin{bmatrix} 1 & \dfrac{r(\lambda)}{q(\lambda)} \\ 0 & \dfrac{q^\times(\lambda)}{q(\lambda)} \end{bmatrix}. \tag{11.39}$$

The second identity in (11.37) implies that $q(\lambda)$ and $q^\times(\lambda)$ have the same degree and that the degree of $r(\lambda)$ is strictly less than the degree of $q(\lambda)$.

Part 2. In this part we show that q is the pole-polynomial of W. The poles of W are certainly zeros of q. Hence (cf., Chapter 8) the McMillan degree $\delta(W)$ of W is given by

$$\delta(W) = \sum_{\alpha \text{ zero of } q} \delta(W, \alpha),$$

where $\delta(W, \alpha)$ is the local degree of W at α. Write $n(\alpha)$ for the multiplicity of α as a zero of q. It suffices to prove that $\delta(W, \alpha) = n(\alpha)$.

Let α be a zero of q. It is clear from (11.39) that the order of α as a (possible) pole of W does not exceed $n(\alpha)$, and so the Laurent expansion of W at α has the form

$$W(\lambda) = \sum_{j=-n(\alpha)}^{\infty} (\lambda - \alpha)^j W_j. \tag{11.40}$$

By definition (cf., Section 8.4), the local degree $\delta(W; \alpha)$ of W at α is the rank of the block upper triangular block matrix

$$\begin{bmatrix} W_{-1} & W_{-2} & \cdots & W_{-n(\alpha)} \\ W_{-2} & & \iddots & 0 \\ \vdots & \iddots & \iddots & \vdots \\ W_{-n(\alpha)} & 0 & \cdots & 0 \end{bmatrix}$$

Now, obviously, $W_1, \ldots, W_{-n(\alpha)}$ are 2×2 matrices with vanishing first column. Hence $\delta(W, \alpha) \leq n(\alpha)$, equality holding if and only if $W_{-n(\alpha)} \neq 0$. The latter, however, is the case because α is not a common zero of q, $q^\times$ and r. Thus q is the pole-polynomial of W.

Part 3. Let n be the degree of $q(\lambda)$. Then the degree of $q^\times(\lambda)$ is also n and the degree of r is strictly less than n (see the last paragraph of the first part of the proof). Since both $q(\lambda)$ and $q^\times(\lambda)$ are monic, the degree of $q(\lambda) - q^\times(\lambda)$ is also strictly less than n.

Since q is the pole-polynomial of W, we know that the McMillan degree of W is equal to n. Now, let A denote the $n \times n$ first companion matrix associated with the polynomial p, and put

$$B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} r_0 & r_1 & \dots & r_{n-1} \\ v_0 & v_1 & \dots & v_{n-1} \end{bmatrix}, \tag{11.41}$$

where $r_0, \dots, r_{n-1}$ are the coefficients of $r(\lambda)$, and $v_0, \dots, v_{n-1}$ are the coefficients of $q(\lambda) - q^\times(\lambda)$. As A is first companion, we have (cf., the proof of Theorem 4.1)

$$(\lambda I_n - A)^{-1} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = \frac{1}{q(\lambda)} \begin{bmatrix} 1 \\ \lambda \\ \vdots \\ \lambda^{n-2} \\ \lambda^{n-1} \end{bmatrix}.$$

Using the above identity, a straightforward calculation shows that

$$C(\lambda - A)^{-1}B = \begin{bmatrix} 0 & \dfrac{r(\lambda)}{q(\lambda)} \\ 0 & \dfrac{q^\times(\lambda) - q(\lambda)}{q(\lambda)} \end{bmatrix}.$$

It follows that $W(\lambda) = I_2 + C(\lambda I_n - A)^{-1}B$ and this realization is minimal because W has McMillan degree n. Note that BC has $[v_0\, v_1\, \dots\, v_{n-1}]$ as its last row and zeros everywhere else. Hence, along with A, the matrix $A^\times = A - BC$ is a first companion. In fact, $A^\times$ is the first companion associated with $q^\times$. Thus W is companion based and $q^\times$ is the zero-polynomial of W. □

Proof of Theorem 11.14. We split the proof into five parts. The first part concerns the implication (ii) ⇒ (i). The other four parts deal with the reverse implication.

Part 1. From Proposition 11.10 we know that it suffices to prove the implication (ii) ⇒ (i) for

$$\widetilde{W}(\lambda) = TW(\lambda)T^{-1} = \begin{bmatrix} 1 & \dfrac{r(\lambda)}{p(\lambda)} \\ 0 & \dfrac{p^\times(\lambda)}{p(\lambda)} \end{bmatrix}.$$

But then we can apply Lemma 11.15 with

$$w_{12}(\lambda) = \frac{r(\lambda)}{p(\lambda)}, \qquad w_{22}(\lambda) = \frac{p^{\times}(\lambda)}{p(\lambda)}.$$

The fact that the three polynomials $p(\lambda)$, $p^{\times}(\lambda)$ and $r(\lambda)$ do not have a common zero implies that p is the least common multiple of the denominators of $w_{12}(\lambda)$ and $w_{22}(\lambda)$. Hence Lemma 11.15 yields (i).

Part 2. In the remainder of the proof it is assumed that (i) is satisfied, and we show that (i) $\Rightarrow$ (ii). We begin with some preliminary observations. The McMillan degree of W is n and there is a minimal realization $W(\lambda) = I_2 + C(\lambda I_n - A)^{-1}B$ of W such that both A and $A^{\times}$ are first companion matrices. The latter implies that $BC = A - A^{\times}$ has rank at most one. Now B is an $n \times 2$ and C is a $2 \times n$ matrix. Hence, if B and C have both rank 2, then BC has rank 2 as well. Thus either B or C has rank at most one. On the other hand, none of these matrices can be the zero matrix because this would conflict with $\delta(W) = n$ and the assumed positivity of n. Thus either B or C has rank one.

Part 3. Suppose rank $B = 1$, and write $B = b\beta^{\top}$ with b and β nonzero vectors in $\mathbb{C}^n$ and $\mathbb{C}^2$, respectively. On account of Cramer's rule, and using that the characteristic polynomial of A is p, we can write $C(\lambda - A)^{-1}b$ in the form

$$C(\lambda - A)^{-1}b = \frac{1}{p(\lambda)} \begin{bmatrix} w(\lambda) \\ \\ \widetilde{w}(\lambda) \end{bmatrix},$$

where w and $\widetilde{w}$ are scalar polynomials of degree at most $n-1$. Let T be an invertible 2×2 matrix having the non zero row vector $\beta^{\top}$ as its last row, i.e., $\beta^{\top} = \begin{bmatrix} 0 & 1 \end{bmatrix} T$. Define the scalar polynomials r and $\widetilde{r}$ by

$$\begin{bmatrix} r(\lambda) \\ \\ \widetilde{r}(\lambda) \end{bmatrix} = T \begin{bmatrix} w(\lambda) \\ \\ \widetilde{w}(\lambda) \end{bmatrix}.$$

Then r and $\widetilde{r}$ have degree at most $n-1$. Also

$$\begin{aligned} W(\lambda) &= I_2 + C(\lambda I_n - A)^{-1}b\beta^{\top} = I_2 + \frac{1}{p(\lambda)} \begin{bmatrix} w(\lambda) \\ \\ \widetilde{w}(\lambda) \end{bmatrix} \beta^{\top} \\ &= I_2 + \frac{1}{p(\lambda)} T^{-1} \begin{bmatrix} r(\lambda) \\ \\ \widetilde{r}(\lambda) \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} T \\ &= T^{-1} \begin{bmatrix} 1 & \dfrac{r(\lambda)}{p(\lambda)} \\ \\ 0 & \dfrac{p(\lambda) + \widetilde{r}(\lambda)}{p(\lambda)} \end{bmatrix} T. \end{aligned}$$

Taking determinants, we get

$$\frac{p(\lambda)+\widetilde{r}(\lambda)}{p(\lambda)} = \det W(\lambda) = \frac{\det(\lambda I_n - A^\times)}{\det(\lambda I_n - A)}, \tag{11.42}$$

where for the latter identity we refer to the proof of Theorem 10.3. As the characteristic polynomials of A and $A^\times$ are p and $p^\times$, respectively, it follows that $p(\lambda)+\widetilde{r}(\lambda) = p^\times(\lambda)$. So W has the form (11.36). Now the matrix function $\widetilde{W}(\lambda) = TW(\lambda)T^{-1}$ has the same McMillan degree as W, that is n. But

$$\widetilde{W}(\lambda) = \begin{bmatrix} 1 & \dfrac{r(\lambda)}{p(\lambda)} \\ 0 & \dfrac{p^\times(\lambda)}{p(\lambda)} \end{bmatrix}.$$

It follows that p, $p^\times$ and r cannot have a common zero.

Part 4. In this part we consider the situation $\operatorname{rank} C = 1$. We shall show that this condition implies $p = p^\times$. In the previous paragraph we did not use the fact that A and $A^\times$ are first companions. The reasoning only depended on the fact that $\operatorname{rank} B = 1$ and that the characteristic polynomials of A and $A^\times$ are p and $p^\times$, respectively. Thus, under the assumption $\operatorname{rank} C = 1$, the arguments can be repeated for $W^\top$, which is given by the realization $W^\top(\lambda) = I_2 + B^\top(\lambda - A^\top)^{-1}C^\top$. Hence $W^\top$ has the form

$$W^\top(\lambda) = T^{-1}\begin{bmatrix} 1 & \dfrac{r(\lambda)}{p(\lambda)} \\ 0 & \dfrac{p^\times(\lambda)}{p(\lambda)} \end{bmatrix} T,$$

where T is an invertible 2×2 matrix, r is a scalar polynomial of degree not exceeding $n-1$ while, moreover, p, $p^\times$ and r do not have a common zero. By the implication (ii) $\Rightarrow$ (i) which has already been established, this implies that $W^\top$ is companion based. Proposition 11.10 now gives that either W and W^{-1} have no common pole, or (the other extreme) W and W^{-1} have the same poles, pole-multiplicities taken into account. In other words, either p and $p^\times$ have no common zero, or (the other extreme) p and $p^\times$ have the same zeros, multiplicities taken into account. The first possibility is ruled out by the hypothesis that p and $p^\times$ do have a common zero. So $\operatorname{rank} C = 1$ implies $p = p^\times$, and we have to find out what happens in this special case. This we do in the next and final part of the proof.

Part 5. Assume $p = p^\times$. Since A and $A^\times$ are first companions with characteristic polynomial p and $p^\times$, respectively, it follows that $A = A^\times$ or, what amounts to the same, $BC = 0$. Let T be an invertible 2×2 matrix such that the second row of the $2\times n$ matrix TC is zero. Then the first row of the rank one matrix TC does not vanish. Now $(BT^{-1})(TC) = BC = 0$, and it follows that the first

column in the $n \times 2$ matrix BT^{-1} is zero. Once again applying Cramer's rule, we conclude that there exists a polynomial r of degree not exceeding $n-1$, such that $TW(\lambda)T^{-1} = I_2 + TC(\lambda - A)^{-1}BT^{-1}$ has the form

$$TW(\lambda)T^{-1} = \begin{bmatrix} 1 & \dfrac{r(\lambda)}{p(\lambda)} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \dfrac{r(\lambda)}{p(\lambda)} \\ 0 & \dfrac{p^{\times}(\lambda)}{p(\lambda)} \end{bmatrix}.$$

This can be rewritten in the desired form

$$W(\lambda) = T^{-1} \begin{bmatrix} 1 & \dfrac{r(\lambda)}{p(\lambda)} \\ 0 & \dfrac{p^{\times}(\lambda)}{p(\lambda)} \end{bmatrix} T = T^{-1}\widetilde{W}(\lambda)T.$$

As before, $\widetilde{W}$ has McMillan degree n, and it follows that $p = p^{\times}$ and r cannot have a common zero. □

Implicitly the material presented above contains complete information about 2×2 companion based matrix functions, also for the case when the pole and zero-polynomial do not have a common zero. The details for the latter case are covered by the following result.

Theorem 11.16. *Let W be a rational 2×2 matrix function, let p and $p^{\times}$ be monic polynomials of the same positive degree, n say, and suppose p and $p^{\times}$ have no common zero. Then the following statements are equivalent:*

(i) *W is companion based with pole-polynomial p and zero-polynomial $p^{\times}$,*

(ii) *W or $W^{\top}$ is of the form*

$$T^{-1} \begin{bmatrix} 1 & \dfrac{r(\lambda)}{p(\lambda)} \\ 0 & \dfrac{p^{\times}(\lambda)}{p(\lambda)} \end{bmatrix} T, \tag{11.43}$$

where T is an invertible 2×2 matrix and r is a scalar polynomial of degree not exceeding $n-1$.

Taking for r the zero-polynomial, (11.43) gets the form

$$T^{-1} \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{p^{\times}(\lambda)}{p(\lambda)} \end{bmatrix} T.$$

This is in line with Proposition 11.10, Theorem 11.13 and the remark made in connection with (11.33).

Proof. The implication (i) $\Rightarrow$ (ii) is contained in Parts 2–4 of the proof of Theorem 11.14.

Next, assume (ii) is satisfied. Note that the conditions of the Theorem imply that the polynomials p, $p^\times$, and r have no common zero. Thus, if W is of the form (11.43), the argument given in Part 1 of the proof of Theorem 11.14 shows that W is companion based. Suppose $W^\top$ is of the form (11.43). Then (by same argument) $W^\top$ is companion based. By hypothesis, p and $p^\times$ do not have a common zero. However, from the expression (11.43) for $W^\top$, it is clear that the poles of $W^\top$ are zeros of p and that those of $(W^\top)^{-1}$ are zeros of $p^\times$. Hence $W^\top$ and $(W^\top)^{-1}$ do not have a common pole. The second part of Proposition 11.10 applied to $W^\top$ now gives that $W = (W^\top)^\top$ is companion based too. □

We close this section with a few comments about rational 2×2 matrix functions of the form

$$\begin{bmatrix} 1 & w_{12}(\lambda) \\ 0 & w_{22}(\lambda) \end{bmatrix}. \tag{11.44}$$

As we have seen, such functions appear in a prominent way in Theorem 11.14 and its proof, in Lemma 11.15, and in Theorem 11.16.

In what follows we say that a rational matrix function W has a λ-*independent fixed point* if there exists a nonzero vector u such that

$$W(\lambda)u = u, \qquad \lambda \text{ not a pole of } W. \tag{11.45}$$

Note that a rational 2×2 matrix function W has the form (11.44) if and only if (11.45) holds with u equal to the first unit vector $e = [1\ 0]^\top$ in $\mathbb{C}^2$. We shall write $\mathcal{FP}$ for the class of all proper rational 2×2 matrix functions W such that $W(\infty) = I_2$ and W has a λ-independent fixed point.

First, we note that $W \in \mathcal{FP}$ if and only if there exists an invertible matrix T such that $T^{-1}W(\lambda)T$ has the form (11.44). Indeed, if such an operator T exists, then clearly $u = Te$ is a λ-independent fixed point of W. Conversely, if u is a λ-independent fixed point of W, then we can choose $v \in \mathbb{C}^2$ so that the vectors u and v form a basis of $\mathbb{C}^2$. Given such a vector v, put $T = [u\ v]$, i.e., T is given by the 2×2 matrix of which the first column is given by u and the second by v. Then T is invertible, and $e = T^{-1}u$ is a λ-independent fixed point of the function $T^{-1}W(\lambda)T$.

Given the result of the previous paragraph, we can use Lemma 11.15 and item (iii) of Proposition 11.10 to show that $W \in \mathcal{FP}$ implies that W is companion based. The converse implication is not true. To see this, take

$$W(\lambda) = \begin{bmatrix} 1 & 0 \\ \dfrac{1}{\lambda^2} & \dfrac{(\lambda-1)^2}{\lambda^2} \end{bmatrix}.$$

Note that $W^\top \in \mathcal{FP}$, and hence $W^\top$ is companion based. But then, since $W^\top$ and $(W^\top)^{-1}$ have no common pole, we can apply item (ii) of Proposition 11.10 with $W^\top$ in place of W to show that W is companion based. On the other hand, it is a simple matter to verify that W does not have a λ-independent fixed point in $\mathbb{C}^2$. Thus, in this case W is companion based and $W \notin \mathcal{FP}$.

Returning to the general case, we can use Theorems 11.14 and 11.16 to show that W is companion based implies that W or $W^\top$ belongs to the class $\mathcal{FP}$. More precisely, when W and W^{-1} do not have a common pole, then W is companion based if and only if W or $W^\top$ has a global fixed point, and when W and W^{-1} have a common pole, then W is companion based if and only if W has a global fixed point. We omit the details.

In conclusion we mention that the case when W and W^{-1} do have a common pole is of special interest in view of the connection with the two machine flow shop problem from combinatorial job scheduling theory to be made in the next chapter (Section 12.4 in particular).

11.5 Complete factorization (companion based)

In this section, combining the results from Section 11.2 with those of Section 10.3 , we derive necessary and sufficient conditions for the existence of complete factorizations of companion based matrix functions.

Theorem 11.17. *Let W be a companion based rational $m \times m$ matrix function, and let n be the McMillan degree of W (assumed to be positive in order to avoid trivialities). Then W admits a complete factorization if and only if there exist an ordering $\alpha_1, \ldots, \alpha_n$ of the poles of W (pole-multiplicities taken into account) and an ordering $\alpha_1^\times, \ldots, \alpha_n^\times$ of the zeros of W (zero-multiplicities taken into account) such that*

$$\alpha_k \neq \alpha_j^\times, \qquad k, j = 1, \ldots, n, \ k < j. \tag{11.46}$$

In fact, given such orderings, there exist complete factorizations of W and W^{-1} of the form

$$\begin{aligned} W(\lambda) &= \left(I_m + \frac{1}{\lambda - \alpha_1} R_1\right) \cdots \left(I_m + \frac{1}{\lambda - \alpha_n} R_n\right), \\ W^{-1}(\lambda) &= \left(I_m - \frac{1}{\lambda - \alpha_n^\times} R_n\right) \cdots \left(I_m - \frac{1}{\lambda - \alpha_1^\times} R_1\right), \end{aligned}$$

where $R_1, \ldots, R_n$ are rank one $m \times m$ matrices.

Proof. Let $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ be a minimal realization of W with A and $A^\times = A - BC$ first companion matrices. The condition on the poles and zeros of W can be rephrased as a requirement on the eigenvalues of A and $A^\times$. In fact it amounts to the existence of orderings $\alpha_1, \ldots, \alpha_n$ of the eigenvalues of A and

$\alpha_1^\times, \ldots, \alpha_n^\times$ of the eigenvalues of $A^\times$ (in both cases algebraic multiplicities taken into account) such that (11.46) holds. In turn, this requirement is equivalent to the condition that the first companion matrices A and $A^\times$ admit simultaneous reduction to complementary triangular forms (see Theorem 11.3). The first part of the theorem is now immediate from Theorem 10.9. The second part follows by combining the details contained in the notes (c) and (d) (see the one but last paragraph preceding Theorem 11.8) with the proof of Theorem 10.5. □

From the above proof, it is clear that Theorem 11.17 can also be formulated in terms of realization (so that one obtains a formulation in the same vein as, for instance, Theorem 10.12). The following reformulation of the first part of Theorem 11.17 is along this tack.

Theorem 11.18. *Let W be a companion based rational $m \times m$ matrix function, and let $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ be a minimal realization of W, so that n is the McMillan degree of W (assumed to be positive in order to avoid trivialities). Then W admits a complete factorization if and only if there exists an ordering $\mu_1, \ldots, \mu_s$ of the (different) elements of $\sigma(A) \cup \sigma(A^\times)$ such that*

$$\sum_{i=1}^{t} m_{A^\times}(\mu_i) \leq 1 + \sum_{i=1}^{t-1} m_A(\mu_i), \qquad t = 1, \ldots, s.$$

Proof. On account of the state space similarity theorem, we may assume that A and $A^\times = A - BC$ are first companions. By Theorem 11.8, the above requirement on (the algebraic multiplicities of) the eigenvalues of A and $A^\times$ then amounts to the condition that A and $A^\times$ admit simultaneous reduction to complementary triangular forms. Again the desired result is immediate from Theorem 10.9. □

We conclude this section with an example illustrating the above theorem. Let W be given by

$$W(\lambda) = \begin{bmatrix} 1 & \dfrac{1}{(\lambda-1)(\lambda-3)^4(\lambda-4)(\lambda-5)^3(\lambda-6)} \\ 0 & \dfrac{(\lambda-2)^2(\lambda-6)}{(\lambda-1)(\lambda-3)(\lambda-5)} \end{bmatrix}.$$

From the material presented in Section 11.4, we see that W is companion based, its pole and zero-polynomial are

$$\begin{aligned} p(\lambda) &= (\lambda-1)(\lambda-3)^4(\lambda-4)(\lambda-5)^3(\lambda-6), \\ p^\times(\lambda) &= (\lambda-2)^2(\lambda-3)^3(\lambda-4)(\lambda-5)^2(\lambda-6)^2, \end{aligned}$$

respectively, and the McMillan degree of W is 10. Also, taking for A and Z the first companion matrices associated with p and $p^\times$, respectively, and appropriately

choosing the matrices B and C, we can produce a minimal realization $W(\lambda) = I + C(\lambda - A)^{-1}B$ of W such that $A^{\times} = Z$. Note that A and Z are precisely the matrices featuring in the example given at the end of Section 11.2. From the last paragraph of that example it is now clear that Theorem 11.18 can be applied to show that W admits a complete factorization. In fact, retracing the steps in the argument leading to (the if part of) Theorem 11.18, one gets complete factorizations of W and W^{-1} where the poles of the elementary factors appear in the order $1, 5, 3, 4, 6$ and $2, 6, 4, 3, 5$, respectively (see (11.29) or the formulas in Theorem 11.17). We shall return to this example in Subsection 11.6.5 below, where the factors will be calculated explicitly using Maple procedures.

11.6 Maple procedures for calculating complete factorizations

In this section Maple procedures are presented to calculate complete factorizations of a proper rational 2×2 matrix function W of the form

$$W(\lambda) = \begin{bmatrix} 1 & w_{12}(\lambda) \\ 0 & w_{22}(\lambda) \end{bmatrix} \quad \text{and} \quad \lim_{\lambda \to \infty} W(\lambda) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{11.47}$$

We know from Lemma 11.15 that W is companion based, its pole-polynomial $p(\lambda)$ is equal to the least common multiple of the denominators of $w_{12}(\lambda)$ and $w_{22}(\lambda)$, and its zero-polynomial $p^{\times}(\lambda)$ is given by $p^{\times}(\lambda) = p(\lambda)w_{22}(\lambda)$. Hence we can apply Theorems 11.17 and 11.18 to check whether W admits a complete factorization, and if this the case, to construct such factorizations. Throughout n is the McMillan degree of W.

In the first part of Subsection 11.6.2 a Maple procedure is provided which calculates the least common multiple polynomial of the denominators of the entries of any rational square matrix function. When applied to $W(\lambda)$, this yields the pole-polynomial $p(\lambda)$. Subsequently, the zero-polynomial $p^{\times}(\lambda)$ is constructed, the poles and zeros of $W(\lambda)$ are calculated using the polynomials $p(\lambda)$ and $p^{\times}(\lambda)$, and the set of different poles and zeros is determined.

The second step is to find an ordering $\alpha_1, \ldots, \alpha_n$ of the poles of W (pole-multiplicities taken into account) and an ordering $\zeta_1, \ldots, \zeta_n$ of the zeros of W (zero-multiplicities taken into account) such that

$$\alpha_k \neq \zeta_j, \qquad k, j = 1, \ldots, n, \ k < j. \tag{11.48}$$

In the second part of Subsection 11.6.2 we shall present a Maple procedure which produces an ordering $\alpha_1, \ldots, \alpha_n$ of the poles of W and an ordering $\zeta_1, \ldots, \zeta_n$ of the zeros of W such that

$$\alpha_k \neq \zeta_j, \qquad k, j = 1, \ldots, n, \ k \leq j - h, \tag{11.49}$$

with the positive integer h as small as possible (cf., formula (11.19)). The procedure given provides first an ordering $\mu_1, \ldots, \mu_s$ of the different elements of $\sigma(A) \cup \sigma(Z)$ such that (11.20) holds, where A and Z are the first companion matrices associated with $p(\lambda)$ and $p^{\times}(\lambda)$, respectively. In fact, all such orderings are obtained.

Next, a Maple procedure is given (see Subsection 11.6.3) to find a transformation matrix T such that the matrix TAT^{-1} is in upper-triangular form and the matrix TZT^{-1} is in lower-triangular form. This procedure is a Maple implementation of formula (11.12) which will return a matrix

$$T = U(\zeta_2, \ldots, \zeta_n; \alpha_1, \ldots, \alpha_{n-1})^{\top},$$

which does the job; see item (c) in the second paragraph preceding Theorem 11.8.

As a fourth step, a procedure (Subsection 11.6.4) is implemented to calculate degree one factors, given a realization of $W(\lambda)$ with the state matrix and associate state matrix in complementary triangular form. In this case the code is a translation into Maple of formula (10.13), and the preceding formulas (10.10)–(10.12), where the identities described by the formulas (10.10)–(10.12) are the result of the triangularization previously calculated. Here, the starting point is the realization described in Part 3 of the proof of Lemma 11.15, which is transformed into the desired form by using the state space similarity given by the matrix T appearing in the previous paragraph.

Finally (see Subsection 11.6.5), we conclude with the example appearing at the end of Section 11.5. Although an ordering of poles and zeros is already given by (11.29), we will ignore this knowledge and use Maple to calculate all orderings (based on multiplicities; see (11.20) with $h = 1$) instead. As we know such orderings satisfy (11.48), and hence the corresponding W admits a complete factorization. We use Maple to calculate such a factorization for two different orderings.

All procedures and calculations in this section are tested under Maple, version 9, [93]. In the text the Maple command lines start with the symbol >. For introductory texts on Maple, see [78], [107]. The Maple worksheet containing all procedures and commands presented in this section is available on request by email from the fourth author (ACM.Ran@few.vu.nl).

11.6.1 Maple environment and procedures

First the Maple environment is defined by loading some Maple packages.

```
> restart; # almost clean start
> with(LinearAlgebra):
> with(MatrixPolynomialAlgebra):
```

11.6.2 Poles, zeros and orderings

The Maple procedure **LCMDenomMatrixPolynom** returns the least common denominator of W (that is, the least common multiple of the denominators of the entries of W) as a monic Maple expression in a Maple name x.

LCMDenomMatrixPolynom	least common denominator of a rational matrix function
Calling sequence	LCMDenomMatrixPolynom(**W**,**x**)
Parameters	W - square rational matrix function x - name (unevaluated)
Output	expression in **x** with leading coefficient 1.

```
> LCMDenomMatrixPolynom:=proc(W,x)
> local k,m,nc,D0,D1,cofmax,mat; mat := convert(W(x),Matrix):
> nc := ColumnDimension(mat): D0:=[]:for k from 1 to nc do
> for m from 1 to nc do D0:=[op(D0),denom(mat[k,m])];
> end do end do;
> D1:=lcm(op(D0)): cofmax:=coeff(D1,x,degree(D1,x)):
> return(expand(D1/cofmax));end;
```

The procedure **GetPolesandZeros** extracts from given pole- and zero-polynomial functions, the poles and zeros, multiplicities included and the set of different poles and zeros. The output is a list of Vectors of Maple type.

GetPolesandZeros	calculate
Calling sequence	GetPolesandZeros(**pf**,**zf**)
Parameters	pf - polynomial function (= pole polynomial) zf - polynomial function (= zero polynomial)
Output	list: first element: Vector of poles (multiplicities included) second element: Vector of zeros (multiplicities included) third element: Vector of different poles and zeros

```
> GetPolesandZeros:=proc(pf,zf)
> local poles, zeros, mu;
> poles:=[solve(pf(x),x)]: print('poles'=poles);
> zeros:=[solve(zf(x),x)]: print('zeros'=zeros);
> mu:=convert([op(op(convert(poles,list)),
> op(convert(zeros,list)))],Vector[row]): print(‘Set of different
> poles and zeros‘=mu); return([convert(poles,Vector[row]),
> convert(zeros,Vector[row]),mu]);
> end proc;
```

The next step is to convert the vectors of poles and zeros of W into vectors of multiplicities. For this purpose we use Maple procedure **GetMultiplicity**.

GetMultiplicity	calculate for each member of a given (second) vector, how many times it is a member of an other (first) vector
Calling sequence	GetMultiplicity(**p,mu**)
Parameters	p - Vector mu - Vector
Output	Vector mP such that $mP_i = \#\{k \mid p_k = mu_i\}$.

```
> GetMultiplicity:=proc(p::Vector,mu::Vector)
> local dimp,dimgv,mP,k,m;
> dimp:=Dimension(p): dimgv:=Dimension(mu):
> mP:=Vector[row](dimgv,0):
> for k from 1 to dimgv do for m from 1 to dimp do
> if evalb(p[m]=mu[k]) then mP[k]:=mP[k]+1; end if:
> end do: end do: return(mP);end;
```

For the conversion of the vectors of poles and zeros of $W(\lambda)$ into vectors of multiplicities the starting point is the Maple Vector *mu* build from the set $\{\mu_1, \ldots, \mu_s\} = \sigma(A) \cup \sigma(Z)$. The Vector p is then either the vector of the poles of $W(\lambda)$ or the vector of zeros of $W(\lambda)$. E.g, the Maple command *muA:=GetMultiplicity(poles,mu)* will return a Maple vector muA such that $muA_t = \#\{k \mid poles_k = mu_t, k = 1, \ldots, n\}$ for $t = 1, \ldots, s$. Analogously, *muZ:=GetMultiplicity(zeros,mu)* will return a Maple vector muZ such that $muZ_t = \#\{k \mid zeros_k = mu_t, k = 1, \ldots, n\}$ for $t = 1, \ldots, s$.

The actual search for a feasible ordering, satisfying condition (11.20), is performed in the procedure **GetAllMOrderings**. In this procedure, starting with $h = 1$, the procedure **GetMOrderingsH** is called with increasing h till an ordering is found satisfying (11.20). Actually, if for some (minimal) h an ordering is found, all orderings satisfying (11.20) are returned.

GetAllMOrderings	calculate all feasible, see condition (11.20), orderings of poles and zeros for minimal h
Calling sequence	GetAllMOrderings(mA,mZ,mu)
Parameters	mA - Vector (of multiplicties of poles) mZ - Vector (of multiplicties of zeros) mu - Vector (of different poles and zeros)
Output	List, with first element is h and the other elements are the output of **GetMOrderingsH**.

```
> GetAllMOrderings:=proc(mA::Vector,mZ::Vector,mu::Vector)
> local h,kordering,orderings, newperm;
> newperm:=true:h:=0:kordering:=0:while (kordering=0) do
> h:=h+1:orderings:=GetMOrderingsH(mA,mZ,mu,h,newperm):
> newperm:=false: kordering:=orderings[1]: end do:
> return(h,orderings);end;
```

The procedure **GetMOrderingsH** needs some further explanation. For a given positive integer h, all permutations of $\{\mu_1, \ldots, \mu_s\} = \sigma(A) \cup \sigma(Z)$ (see the Maple variable *mu*), are tested; this test is performed in the procedure **TestOrderingMAZ**. If the test is successful, then a variable counting the number of admissible permutations (i.e., those permutations that yield a feasible ordering) is increased by one. Subsequently, the vectors of multiplicities of poles and zeros are converted back to vector of poles and zeros with multiplicities included and ordered (with

blocks of equal poles and zeros, respectively) according to the found ordering of $\{\mu_1, \ldots, \mu_s\}$; see procedure **GetOrderedVector**. The resulting vectors are added to an output list.
Since in Maple the calculation of permutations is very time- and cpu-consuming, a boolean variable is added to the arguments list such that only one, initial call to Maple's procedure *permute* is needed; the output of *permute* is assigned to the Maple variable *allperm* which is defined as global.

GetMOrderingsH	calculate all feasible orderings of poles and zeros for given h
Calling sequence	GetMOrderingsH(**muA,muZ,mu,h,newperm**)
Parameters	muA - Vector (of multiplicities of poles) zeros - Vector (of multiplicities of zeros) mu - Vector: different poles and zeros h - positive integer newperm - boolean
Output	If orderings are found, the output is a list, of which the first element is the number of feasible orderings and the second and third elements are list with entries the ordering of poles and zeros respectively. If no ordering is found, the output is just 0.

```
> GetMOrderingsH:=proc(muA,muZ,mu,h,newperm)
> local nmu,nperm,kk,k,perm,orderingP,orderingZ; global allperm;
> nmu:= Dimension(mu): if newperm=true then
> allperm:= combinat[permute](nmu): end if:
> nperm:= combinat[numbperm](nmu):
> kk:=0: for k from 1 to nperm do perm:=allperm[k]:
> if (TestOrderingMAZ(muA,muZ,perm,h)) then kk:=kk+1:
> orderingP[kk]:=GetOrderedVector(muA,mu,convert(perm,list)):
> orderingZ[kk]:=GetOrderedVector(muZ,mu,convert(perm,list)):
> end if: end do: if kk>0 then return(kk,orderingP,orderingZ)
> else return(kk); end if: end proc;
```

Let $\{\mu_1, \ldots, \mu_s\} = \sigma(A) \cup \sigma(Z)$ and let ma and mz be vectors such that ma_t and mz_t, $t = 1, \ldots, s$, denote the multiplicity of μ_t as pole and zero of $W(\lambda)$, respectively, and let $perms$ be a permutation of $(1, \ldots, s)$. Then a call to the Maple procedure *TestOrderingMAZ(ma,mz,perms,h)* will test whether the ordering $perms$ satisfies condition (11.20) for given h. To be specific, the following condition is tested:

$$\sum_{i=1}^{t} mz_{perms(i)} \leq h + \sum_{i=1}^{t-1} ma_{perms(i)}, \quad t = 1, \ldots, s.$$

TestOrderingMAZ	test whether a given ordering satisfies condition (11.20) for a given h.
Calling sequence	TestOrderingMAZ(**ma,mz,perms,h**)
Parameters	ma - Vector of multiplicities of poles mz - Vector of multiplicities of zeros perms - Vector; permutation vector, ordering h - positive integer
Output	boolean: true if ordering for given h satisfies condition (11.20) otherwise false

```
> TestOrderingMAZ:= proc(ma,mz,perms,h)
> local s, bol, t;
> s:=Dimension(mz): bol:=true: t:=0: while (t<s) and bol do
> t:=t+1: bol:=not((add(mz[perms[m1]],m1=1..t)-
> add(ma[perms[m2]],m2=1..(t-1)))>(h)): end do;return(bol);end;
```

GetOrderedVector	calculate a vector V such that, for a given ordering, the multiplicity vector of V equals a given ordered (multiplicity) vector (reverse of **GetMultiplicity**).
Calling sequence	GetOrderedVector(**ma,mu,ordering**)
Parameters	ma - Vector (multiplicity vector) mu - Vector of different elements ordering - List: permutation list
Output	Vector

```
> GetOrderedVector:=proc(ma,mu,ordering)
> local no,n1,tk,nc, k, m, mup, mpp, orderedV;
> mup:=mu[ordering]: mpp:=ma[ordering]: no:=Dimension(mu):
> n1:=add(ma[k],k=1..no): nc:=0: orderedV:=Vector[row](n1,0):
> for k from 1 to no do tk:=mpp[k]: for m from 1 to tk do
> nc:=nc+1: orderedV[nc]:=mup[k]: end do: end do:
> return(orderedV);end proc;
```

11.6.3 Triangularization routines (complete)

This part implements the triangularization of a companion matrices A and Z, given an ordering of poles and zeros of W satisfying condition (11.46). The Maple code is based on the construction exposed in the second paragraph preceding Theorem 11.8. This involves the calculation of the matrix $T{=}U(\zeta_2,\ldots,\zeta_n;\alpha_1,\ldots,\alpha_{n-1})^T$ where $(\zeta_1,\ldots,\zeta_n)$ are ordered zeros and $(\alpha_1,\ldots,\alpha_n)$ ordered poles; see (11.12). The code is split in two: the procedure **Scol(oZ,oA,j)** outputs the jth column of the matrix $U(\zeta_2,\ldots,\zeta_n;\alpha_1,\ldots,\alpha_{n-1})$. The second and main procedure is called **Tcomplete** which calls **Scol** for $k=1,\ldots,n$, and collects the returned vectors in a $n\times n$ matrix.

Tcomplete	calculate transpose of $U(\zeta_2, \ldots, \zeta_n; \alpha_1, \ldots, \alpha_{n-1})$
Calling sequence	Tcomplete(**oZ,oA**)
Parameters	oZ - Vector (of ordered zeros)
	oA - Vector (of ordered poles)
Output	Matrix T such that TAT^{-1} is upper triangular,
	TZT^{-1} is lower-triangular

```
> Tcomplete:=proc(oZ::Vector,oA::Vector)
> local k,nc,S;
> nc:=Dimension(oA): if not (Dimension(oZ)=nc) then
> error "Input vectors should have equal length"; end if:
> S:=Scol(oZ,oA,0): for k from 2 to nc do S:=<S|Scol(oZ,oA,k-1)>:
> end do: return(Transpose(S));end;
```

Scol	calculate one column of $U(\zeta_2, \ldots, \zeta_n; \alpha_1, \ldots, \alpha_{n-1})$
Calling sequence	Scol(**oZ,oA,j**)
Parameters	oZ - Vector (e.g. ordered zeros)
	oA - Vector (e.g. ordered poles)
	j - integer $0 \leq j < Dimension(oA)$
Output	Vector

```
> Scol:=proc(oZ,oA,j)
> local pol, k, nc, vj, x;
> nc:=Dimension(oA);vj:=Vector(nc):pol:=1:
> if (j>0) then for k from 1 to j do pol:=pol*(x-oA[k])
> end do: end if: if ((j+1)<nc) then
> for k from (j+1) to (nc-1) do pol:=pol*(x-oZ[k+1]):
> end do: end if:
> for k from 1 to nc do vj[k]:=coeff(pol,x,k-1);end do:
> return(vj); end;
```

11.6.4 Factorization procedures

We begin with some general remarks. In constructing factorizations into elementary factors for concrete examples (see, e.g., the next subsection) the starting point will be the minimal realization $W(\lambda) = I_2 + C(\lambda I_n - A)^{-1}B$ given in Part 3 of the proof of Lemma 11.15. Thus A is the first companion matrix associated with the pole-polynomial $p(\lambda)$, and B and C are given by (11.41). Note that in this case $A^\times = A - BC = Z$, the first companion matrix associated with the zero-polynomial $p^\times(\lambda)$. In what follows we write

$$Amin = A, \qquad Bmin = B, \qquad Cmin = C, \qquad Amincross = A^\times.$$

We do not use subscripts here because in Maple subscripts play a different role. Furthermore, we shall use the state space transformation T constructed in the

preceding subsection, to produce the matrices

$$Atr = TAminT^{-1}, \qquad Btr = TBmin, \qquad Ctr = CminT^{-1},$$

$$Atrcross = TAmincrossT^{-1}.$$

Thus we have a minimal realization $W(\lambda) = I_2 + Ctr(\lambda I_n - Atr)^{-1}Btr$, where Atr is upper triangular and $Atrcross = Atr - BtrCtr$ is lower triangular.

The Maple procedures to create a factorization of $W(\lambda)$ into elementary factors given in this section calculate the elementary factors from the upper triangular form of Atr and lower triangular form of $Atrcross$, using formula (10.13) and the preceding formulas (10.10)–(10.12). In the present subsection, with some abuse of notation, the label tr will be omitted. Thus we start with a minimal realization $W(\lambda) = I_2 + C(\lambda I_n - A)^{-1}B$ with A in upper triangular form and $A^{\times} = A - BC$ is lower triangular form.

For each $k = 1, \ldots, n$, a factor is of the form

$$I_2 + \frac{1}{\lambda - \alpha_k} R_k$$

where I_2 is the 2×2 identity matrix, R_k is a 2×2 matrix and α_k is kth pole which is equal to the kth main diagonal element of A.

The main procedure is called **MakeFactorization** while, for each $k = 1, \ldots, n$, the R_k-matrix is calculated in the procedure **MakeRmatrix**; this procedure is based on formula (10.13). The output of **MakeFactorization** is a vector of length n with each entry is a realization factor (as matrix function).

As a test facility, the procedure **Factors2Transfer** is written; it returns, for a vector of realization factors, a transfer function equal to the product of those factors. The factors itself should be matrix functions.

MakeFactorization	calculate a factorization if (at least) state space matrices are in complementary triangular form
Calling sequence	MakeFactorization(**A,B,C,x**)
Parameters	A - Matrix (state space matrix in triangular form) B,C - Matrices (input- and output-matrices) x - name (unevaluated Maple name)
Output	Vector, containing all the factors as matrix expressions in x

```
> MakeFactorization:=proc(A,B,C,x)
> local WW, nv, m, Im, k, R, oa;
> m:=RowDimension(C);nv:=Dimension(A):Im:=IdentityMatrix(m):
> WW:=Vector(nv): for k from 1 to nv[1] do R:=MakeRmatrix(B,C,k):
> WW[k]:=
> unapply(map(factor,Im+ScalarMultiply(R,1/(x-A[k,k]))),x):
> end do: return(WW);end;
```

MakeRmatrix	calculate R matrix for kth pole
Calling sequence	MakeRmatrix(**B,C,k**)
Parameters	B,C - Matrices (input- and output-matrices)
	k - integer, index
Output	Matrix (kth R matrix)

```
> MakeRmatrix:=proc(B,C,k)
> local nc, mat, i, j;
> nc:=RowDimension(C); mat:=Matrix(nc,nc,0); for i from 1 to
> nc do for j from 1 to nc do mat[i,j]:=C[i,k]*B[k,j]; end do end
> do: return(mat);end;
```

Factors2Transfer	Calculate from given factorization factors
	the transfer function
Calling sequence	Factors2Transfer(**AllFactors**,**x**)
Parameters	AllFactors - Vector: elements are Matrix functions
	x - name (unevaluated Maple name)
Output	Matrix: rational matrix function

```
> Factors2Transfer := proc(AllFactors,x)
> local Wtest, DimS, k, n, Wdum, ResultW;
> DimS:=ColumnDimension(AllFactors[1](x)):
> n:=Dimension(AllFactors): Wtest:=IdentityMatrix(DimS):
> for k from 1 to n do
> Wtest:=map(simplify,Wtest.AllFactors[k](x)): end do:
> Wdum:=convert(map(factor,map(simplify,evalm(Wtest))),Matrix):
> ResultW:=unapply(Wdum,x): return(ResultW);end proc;
```

11.6.5 Example

The above defined procedures are applied to the example given at the end of Section 11.5:

```
> W:=x-><<1,0>|<1/((x-1)*(x-3)^4*(x-4)*(x-5)^3*(x-6)),
> (x-2)^2*(x-6)/((x-1)*(x-3)*(x-5))>>: `W(lambda)`=W(lambda);
```

$$W(\lambda) = \begin{bmatrix} 1 & \dfrac{1}{(\lambda-1)(\lambda-3)^4(\lambda-4)(\lambda-5)^3(\lambda-6)} \\ 0 & \dfrac{(\lambda-2)^2(\lambda-6)}{(\lambda-1)(\lambda-3)(\lambda-5)} \end{bmatrix} \tag{11.50}$$

We will use Lemma 11.15 to calculate first the least common denominator polynomial q of W. Then the pole-polynomial p is q. Subsequently, again using Lemma 11.15, we calculate the zero-polynomial $p^{\times}$.

```
> q:=unapply(LCMDenomMatrixPolynom(W,x),x):
> ppoles:=q:pzeros:=unapply(simplify(W(x)[2,2]*q(x)),x):
> 'p(lambda)'=sort(collect(ppoles(lambda),lambda),lambda);
> '(p^(x))(lambda)'=sort(collect(pzeros(lambda),lambda),lambda);
```

$$\begin{aligned} p(\lambda) &= \lambda^{10} - 38\,\lambda^9 + 640\,\lambda^8 - 6284\,\lambda^7 + 39778\,\lambda^6 - 169304\,\lambda^5 + \\ &\quad 489456\,\lambda^4 - 945684\,\lambda^3 + 1162485\,\lambda^2 - 814050\,\lambda + 243000 \\ p^{\times}(\lambda) &= \lambda^{10} - 39\,\lambda^9 + 674\,\lambda^8 - 6794\,\lambda^7 + 44217\,\lambda^6 - 194071\,\lambda^5 + \\ &\quad 581556\,\lambda^4 - 1174536\,\lambda^3 + 1529712\,\lambda^2 - 1159920\,\lambda + 388800 \end{aligned}$$

The next step is the calculation of poles and zeros and orderings, satisfying condition (11.20):

```
> res1:=GetPolesandZeros(ppoles,pzeros):
> poles:=res1[1]:zeros:=res1[2]:mu:=res1[3]:
> npoles:=Dimension(poles);nzeros:=Dimension(zeros);
> nmu:=Dimension(mu);
> muA:=GetMultiplicity(poles,mu);muZ:=GetMultiplicity(zeros,mu);
> AllMOrderings:=GetAllMOrderings(muA,muZ,mu):AllMOrderings;
```

$$AllMOrderings = 1,\ 12,\ orderingP,\ orderingZ$$

Hence, we found 12 orderings of poles and zeros which satisfy condition (11.20) for $h = 1$, the first element of *AllMOrderings*. As actual ordering we take the sixth found ordering and use this ordering in the Maple variables *orderedA* and *orderedZ*; this ordering is just equal to the one in (11.29):

```
> orderedA:=AllMOrderings[3][6]:‘alpha‘=orderedA;
> orderedZ:=AllMOrderings[4][6]:‘zeta‘=orderedZ;
```

$$\begin{aligned} \alpha &= [1,5,5,5,3,3,3,3,4,6] \\ \zeta &= [5,5,3,3,3,4,6,6,2,2] \end{aligned}$$

This concludes the search for a feasible ordering with $h = 1$.

To construct a minimal companion based realization of $W(\lambda)$, we start from the minimal realization of $W(\lambda)$ given in Part 3 of the proof of Lemma 11.15. This minimal realization is written (cf., the first paragraph of the preceding subsection) as

$$W(\lambda) = Dmin + Cmin(\lambda I_n - Amin)^{-1}Bmin.$$

Thus *Amin* is the first companion matrix associated with the pole-polynomial p, *Bmin* and *Cmin* are as in (11.41), *Dmin* the (2×2) identity matrix, and *Amincross* is the first companion matrix associated with the zero-polynomial $p^{\times}$. The Maple commands are as follows.

```
> r:=unapply(simplify(W(lambda)[1,2]*ppoles(lambda)),lambda):
> Amin:=Transpose(CompanionMatrix(ppoles(lambda),lambda));
> Bmin:=Matrix(npoles,2):Bmin[npoles,2]:=1:
> Cmin:=Matrix(2,npoles,0): for k from 1 to npoles do
> Cmin[1,k]:=coeff(r(lambda),lambda,k-1):
> Cmin[2,k]:=coeff(pzeros(lambda)-ppoles(lambda),lambda,k-1):
> end do: Dmin:=IdentityMatrix(2,2):
> Amincross:=Transpose(CompanionMatrix(pzeros(lambda),lambda));
```

Note that the standard Maple procedure *CompanionMatrix* gives the second companion matrix.

By Theorem 11.3, *Amin* and *Amincross*, with the given ordering of poles and zeros, allow for simultaneously triangularization in complementary triangular forms. Calling the procedure **Tcomplete** with arguments *orderedZ* and *orderedA*, with $Z = Amincross$ and $A = Amin$ will output a transformation matrix which is needed to bring *Amin* in upper-triangular form. In Maple this matrix carries the name Tr.

```
> Tr:=Tcomplete(orderedZ,orderedA):‘T‘=Tr;
```

$$Tr = \begin{bmatrix} -77760 & 216432 & -262656 & 182376 & -79836 & 22847 & -4274 & 504 & -34 & 1 \\ -15552 & 55728 & -84672 & 72072 & -38028 & 12931 & -2838 & 388 & -30 & 1 \\ -25920 & 89424 & -129888 & 105048 & -52388 & 16765 & -3452 & 442 & -32 & 1 \\ -43200 & 143280 & -198528 & 152200 & -71596 & 21539 & -4162 & 500 & -34 & 1 \\ -72000 & 229200 & -302240 & 219096 & -97028 & 27421 & -4976 & 562 & -36 & 1 \\ -54000 & 176400 & -239880 & 179912 & -82567 & 24181 & -4542 & 530 & -35 & 1 \\ -27000 & 92700 & -133890 & 107621 & -53332 & 16963 & -3474 & 443 & -32 & 1 \\ -13500 & 48600 & -74295 & 63743 & -33979 & 11707 & -2613 & 365 & -29 & 1 \\ -20250 & 69525 & -100980 & 82272 & -41704 & 13698 & -2924 & 392 & -30 & 1 \\ -40500 & 128925 & -172260 & 128904 & -60092 & 18202 & -3596 & 448 & -32 & 1 \end{bmatrix}$$

Next one calculates the inverse of Tr and put *Amin* in upper-triangular form (and *Amincross* in lower-triangular form) and apply corresponding transformations to *Bmin* and *Cmin*; the resulting matrices are named *Atr*, *Btr*, *Ctr* and *Atrcross*.

```
> Trinv:=MatrixInverse(Tr):
> Atr:=Tr.Amin.Trinv:‘A‘=Atr; Btr:=Tr.Bmin:Ctr:=Cmin.Trinv:
> Atrcross:=Tr.(Amincross).Trinv:’A^(x)’=Atrcross;
```

For our example this yields:

$$Atr = \begin{bmatrix} 1 & 0 & 2 & 2 & 0 & -1 & -3 & -3 & 2 & 4 \\ 0 & 5 & 2 & 2 & 0 & -1 & -3 & -3 & 2 & 4 \\ 0 & 0 & 5 & 2 & 0 & -1 & -3 & -3 & 2 & 4 \\ 0 & 0 & 0 & 5 & 0 & -1 & -3 & -3 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & -1 & -3 & -3 & 2 & 4 \\ 0 & 0 & 0 & 0 & 0 & 3 & -3 & -3 & 2 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & -3 & 2 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 2 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 \end{bmatrix}$$

and

$$Atrcross = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & -2 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & -2 & -2 & 3 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & -2 & -2 & 0 & 4 & 0 & 0 & 0 & 0 \\ 4 & 0 & -2 & -2 & 0 & 1 & 6 & 0 & 0 & 0 \\ 4 & 0 & -2 & -2 & 0 & 1 & 3 & 6 & 0 & 0 \\ 4 & 0 & -2 & -2 & 0 & 1 & 3 & 3 & 2 & 0 \\ 4 & 0 & -2 & -2 & 0 & 1 & 3 & 3 & -2 & 2 \end{bmatrix}$$

Finally, with *Amin* and *Amincross* in triangular form, the factorization procedure **MakeFactorization** is called with arguments the Maple variables *Atr*, *Ctr*, *Btr* and name λ. With the given ordering of poles and zeros, the factorization is complete (see Theorem 11.17).

```
> Allfactors:=
> map(simplify,MakeFactorization(Atr,Btr,Ctr,lambda)):
```

and the result is printed on the console:

```
> afactors:=Vector[row](npoles,0):
> for k from 1 to npoles do afactors[k]:=Allfactors[k](lambda):
> end do: print(‘Elementary factors‘=afactors);
```

The elementary factors are:

$$\begin{bmatrix} 1 & -\frac{1}{2400}\,(\lambda-1)^{-1} \\ 0 & \frac{\lambda-5}{\lambda-1} \end{bmatrix}, \quad \begin{bmatrix} 1 & \frac{1}{288}\,(\lambda-5)^{-1} \\ 0 & 1 \end{bmatrix},$$

$$\begin{bmatrix} 1 & -\frac{13}{1350}\,(\lambda-5)^{-1} \\ 0 & \frac{\lambda-3}{\lambda-5} \end{bmatrix}, \quad \begin{bmatrix} 1 & \frac{29}{900}\,(\lambda-5)^{-1} \\ 0 & \frac{\lambda-3}{\lambda-5} \end{bmatrix},$$

$$\begin{bmatrix} 1 & \frac{1}{144}\,(\lambda-3)^{-1} \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & -\frac{14}{675}\,(\lambda-3)^{-1} \\ 0 & \frac{\lambda-4}{\lambda-3} \end{bmatrix},$$

$$\begin{bmatrix} 1 & \frac{4}{225}\,(\lambda-3)^{-1} \\ 0 & \frac{\lambda-6}{\lambda-3} \end{bmatrix}, \quad \begin{bmatrix} 1 & \frac{16}{135}\,(\lambda-3)^{-1} \\ 0 & \frac{\lambda-6}{\lambda-3} \end{bmatrix},$$

$$\begin{bmatrix} 1 & -\frac{1}{6}\,(\lambda-4)^{-1} \\ 0 & \frac{\lambda-2}{\lambda-4} \end{bmatrix}, \quad \begin{bmatrix} 1 & \frac{1}{54}\,(\lambda-6)^{-1} \\ 0 & \frac{\lambda-2}{\lambda-6} \end{bmatrix}.$$

With these factors (ordered from left to right and top to bottom) we have a complete factorization of the rational matrix function W considered at the end of Section 11.5.

To test whether the foregoing calculations are indeed a factorization of the initially given rational matrix function, give the following Maple commands. Note that **Factors2Transfer** will return a rational matrix function which should be equal to (11.50).

```
> Wtest:=Factors2Transfer(Allfactors,lambda): print(‘Product
> elementary factors‘=Wtest(lambda));
```

This concludes the test of the calculations related to the sixth ordering.

It will be convenient to put the triangularization and factorization commands in one procedure.

```
> MakeCompleteFactors:=proc(Amin,Bmin,Cmin,orderP,orderZ,x)
> local Tr,Trinv,Atr,Btr,Ctr,Allfactors,afactors,k,np;
> np:=Dimension(orderP): Tr:=Tcomplete(orderZ,orderP):
> Trinv:=MatrixInverse(Tr): Atr:=Tr.Amin.Trinv:
> Btr:=Tr.Bmin:Ctr:=Cmin.Trinv:
> Allfactors:=map(simplify,MakeFactorization(Atr,Btr,Ctr,x)):
> afactors:=Vector[row](np,0): for k from 1 to np do
> afactors[k]:=Allfactors[k](x): end do:
> print(‘Elementary factors‘=afactors);end proc;
```

The procedure **MakeCompleteFactors** has as arguments the companion pole and zero polynomial based realization matrices A, B, C of W, see (11.41), (in Maple named *Amin*, *Bmin* and *Cmin*), a vector of ordered poles, a vector of ordered zeros and a Maple name x.

Since all orderings have been calculated and were collected in the Maple variable *AllMOrderings*, the next commands will give a factorization. As an example the 5th ordering is taken which differ from the original ordering only in the zero's:

```
> orderP:=AllMOrderings[3][5]:orderZ:=AllMOrderings[4][5]:
> print(‘ordering poles‘=orderP);print(‘ordering zeros‘=orderZ);
> MakeCompleteFactors(Amin,Bmin,Cmin,orderP,orderZ,lambda);
```

The orderings are:

$$\begin{aligned} \textit{ordering poles} &= [1,5,5,5,3,3,3,3,4,6], \\ \textit{ordering zeros} &= [5,5,3,3,3,4,2,2,6,6]. \end{aligned}$$

The elementary factors are (again ordered from left to right and top to bottom):

$$\begin{bmatrix} 1 & -\frac{1}{2400}\,(\lambda-1)^{-1} \\ 0 & \frac{\lambda-5}{\lambda-1} \end{bmatrix}, \quad \begin{bmatrix} 1 & \frac{1}{288}\,(\lambda-5)^{-1} \\ 0 & 1 \end{bmatrix},$$

$$\begin{bmatrix} 1 & -\frac{13}{1350}\,(\lambda-5)^{-1} \\ 0 & \frac{\lambda-3}{\lambda-5} \end{bmatrix}, \quad \begin{bmatrix} 1 & \frac{29}{900}\,(\lambda-5)^{-1} \\ 0 & \frac{\lambda-3}{\lambda-5} \end{bmatrix},$$

$$\begin{bmatrix} 1 & \frac{1}{144}\,(\lambda-3)^{-1} \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & -\frac{14}{675}\,(\lambda-3)^{-1} \\ 0 & \frac{\lambda-4}{\lambda-3} \end{bmatrix},$$

$$\begin{bmatrix} 1 & -\dfrac{4}{675}\,(\lambda-3)^{-1} \\ 0 & \dfrac{\lambda-2}{\lambda-3} \end{bmatrix}, \quad \begin{bmatrix} 1 & \dfrac{32}{6075}\,(\lambda-3)^{-1} \\ 0 & \dfrac{\lambda-2}{\lambda-3} \end{bmatrix},$$

$$\begin{bmatrix} 1 & -\dfrac{151}{12150}\,(\lambda-4)^{-1} \\ 0 & \dfrac{\lambda-6}{\lambda-4} \end{bmatrix}, \quad \begin{bmatrix} 1 & \dfrac{1}{810}\,(\lambda-6)^{-1} \\ 0 & 1 \end{bmatrix}.$$

11.7 Appendix: invariant subspaces of companion matrices

In this section we present detailed information about the lattice of invariant subspaces of $n \times n$ first companion matrices. For a large part the material presented here is standard. It will be used in the next chapter, in particular, it will play an important role in the proof of Theorem 12.2. Whenever there is reason to do so, $n \times n$ matrices are identified in the customary manner with linear operators on $\mathbb{C}^n$.

First we fix some notation. As before, if α is a complex number and A is a square matrix, then $m_A(\alpha)$ will denote the algebraic multiplicity of α as an eigenvalue of A when $\alpha \in \sigma(A)$, and $m_A(\alpha) = 0$ otherwise. For n and k integers, $0 \le k < n$, and $\alpha \in \mathbb{C}$, write $\mathbf{v}_k(\alpha)$ for the vector in $\mathbb{C}^n$ having for its jth component

$$\binom{j-1}{k}\alpha^{j-1-k}$$

when j is among $(k+1), \ldots, n$, and zero otherwise. An alternative way of introducing $\mathbf{v}_k(\alpha)$ is via the expression

$$\mathbf{v}_k(\alpha) = \frac{1}{k!}\frac{d^k}{d\alpha^k}\begin{bmatrix} 1 \\ \alpha \\ \alpha^2 \\ \vdots \\ \alpha^{n-1} \end{bmatrix}, \qquad k = 0, \ldots, n-1.$$

Let A be an $n \times n$ first companion matrix, and let α be an eigenvalue of A. Then $\mathbf{v}_0(\alpha)$ is the unique (up to a nonzero scalar multiple) eigenvector corresponding to A and α. More generally, for $k = 1, \ldots, m_A(\alpha)$, the vectors $\mathbf{v}_0(\alpha), \ldots, \mathbf{v}_{k-1}(\alpha)$ form a basis for $\operatorname{Ker}(A - \alpha I_n)^k$. In fact, $\mathbf{v}_0(\alpha), \ldots, \mathbf{v}_{m_A(\alpha)-1}(\alpha)$ form a Jordan chain for A, that is,

$$A\mathbf{v}_0(\alpha) = \alpha\mathbf{v}_0(\alpha), \quad A\mathbf{v}_k(\alpha) = \alpha\mathbf{v}_k(\alpha) + \mathbf{v}_{k-1}(\alpha), \tag{11.51}$$

where k is in the range 1 up to $m_A(\alpha) - 1$. Using this result, we shall describe the invariant subspaces of A.

Fix an ordering $\mathbf{a}_1, \ldots, \mathbf{a}_s$ of the different eigenvalues of A, write $\mathbf{m}_1, \ldots, \mathbf{m}_s$ for the corresponding algebraic multiplicities, and introduce

$$\mathbf{V} = \left[\begin{array}{cccccccc} \mathbf{v}_0(\mathbf{a}_1) & \cdots & \mathbf{v}_{\mathbf{m}_1-1}(\mathbf{a}_1) & \cdots & \mathbf{v}_0(\mathbf{a}_s) & \cdots & \mathbf{v}_{\mathbf{m}_s-1}(\mathbf{a}_s) \end{array} \right]. \tag{11.52}$$

Then $\mathbf{V}$ is an $n \times n$ matrix (because $\mathbf{m}_1 + \cdots + \mathbf{m}_s = n$) and

$$\det \mathbf{V} = \prod_{k=2}^{s} \prod_{j=1}^{k-1} (\mathbf{a}_k - \mathbf{a}_j)^{\mathbf{m}_j \mathbf{m}_k}, \tag{11.53}$$

where the empty product (appearing when $s = 1$) is read as 1. Since $\mathbf{a}_1, \ldots, \mathbf{a}_s$ are different, the matrix $\mathbf{V}$ is invertible. Also (11.51) gives that $\mathbf{V}A\mathbf{V}^{-1}$ has upper triangular Jordan form. In fact,

$$\mathbf{V}A\mathbf{V}^{-1} = \left[\begin{array}{cccc} J_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & J_s \end{array} \right], \tag{11.54}$$

where, for $j = 1, \ldots, s$, the matrix J_t is the $\mathbf{m}_t \times \mathbf{m}_t$ upper triangular Jordan block with eigenvalue $\mathbf{a}_t$. Since each eigenvalue has geometric multiplicity one, for each eigenvalue there is only one Jordan block, and hence the lattice of invariant subspaces of A is finite. In fact, as the following proposition shows, it consists of $\mathbf{m}_1 \times \cdots \times \mathbf{m}_s$ members.

Proposition 11.19. *Let A be an $n \times n$ first companion matrix, and let $\mathbf{a}_1, \ldots, \mathbf{a}_s$ be an ordering of the different eigenvalues of A with $\mathbf{m}_1, \ldots, \mathbf{m}_s$ being the corresponding algebraic multiplicities. If $m_1, \ldots, m_s$ are non-negative integers not exceeding the numbers $\mathbf{m}_1, \ldots, \mathbf{m}_s$, respectively, then*

$$M = \operatorname{span} \{ \mathbf{v}_0(\mathbf{a}_1), \ldots, \mathbf{v}_{m_1-1}(\mathbf{a}_1), \ldots, \mathbf{v}_0(\mathbf{a}_s), \ldots, \mathbf{v}_{m_s-1}(\mathbf{a}_s) \} \tag{11.55}$$

is an invariant subspace for A. Conversely, if M is an invariant subspace of A, then there exist unique non-negative integers $m_1, \ldots, m_s$, not exceeding $\mathbf{m}_1, \ldots, \mathbf{m}_s$ respectively, such that (11.55) *holds.*

Proof. Let M be an invariant subspace for A. Then $\mathbf{V}^{-1}[M]$ is an invariant subspace for the upper triangular Jordan matrix J appearing in the right-hand side of (11.54). Since the upper triangular Jordan blocks $J_1, \ldots, J_s$ correspond to different eigenvalues, we can decompose $\mathbf{V}^{-1}[M]$ in a unique way as $\mathbf{V}^{-1}[M] = M_1 \dot{+} \cdots \dot{+} M_s$ with M_t an invariant subspace for J_t. Put $m_t = \dim M_t$, so that m_t is a non-negative integer not exceeding $\mathbf{m}_t$. Now J_t is unicellular, i.e., it has only one

complete chain of invariant subspace. In particular, J_t has only one invariant subspace of dimension m_t, and this subspace is the span of the first m_t elements in the standard basis for $\mathbb{C}^{\mathbf{m}_t}$, the space on which the $\mathbf{m}_t \times \mathbf{m}_t$ matrix J_t acts as a linear operator. For M_t viewed as as subspace of $\mathbb{C}^n$, this means

$$M_t = \text{span}\,\{e_{\tilde{m}_1+\cdots+\tilde{m}_{t-1}+1}, \ldots, e_{\tilde{m}_1+\cdots+\tilde{m}_{t-1}+m_t}\},$$

where $e_1, \ldots, e_n$ is the standard basis (consisting of the unit vectors) in $\mathbb{C}^n$. Hence $\mathbf{V}[M_t]$ is spanned by the vectors $\mathbf{v}_0(\mathbf{a}_t), \ldots, \mathbf{v}_{m_t-1}(\mathbf{a}_t)$, and the representation (11.55) follows from $M = \mathbf{V}[M_1] \dotplus \cdots \dotplus \mathbf{V}[M_s]$. So far about existence. Uniqueness is clear from the fact that, if M is given by (11.55) and $\mathbf{V}^{-1}[M]$ is written in the form of a direct sum $M_1 \dotplus \cdots \dotplus M_s$ as above, then necessarily $m_t = \dim M_t$. $\square$

Now let $\alpha_1, \ldots, \alpha_n$ be an ordering of the eigenvalues of A (algebraic multiplicities taken into account). We introduce the *generalized Vandermonde matrix* associated with this ordering as the $n \times n$ matrix $V = V(\alpha_1, \ldots, \alpha_n)$ for which the jth column is the vector $\mathbf{v}_{\nu(j)}(\alpha_j)$ where $\nu(j)$ is the number of times that the eigenvalue α_j appears among its predecessors $\alpha_1, \ldots, \alpha_{j-1}$. Clearly V can be obtained from the matrix $\mathbf{V}$ appearing above by an appropriate permutation of the columns. Therefore, modulo a plus or minus sign, the determinant of V is equal to the product in the right-hand side of (11.53). In particular, V is invertible. Up to a permutation similarity, $V^{-1}AV$ is an upper triangular Jordan matrix. A closer look reveals that $V^{-1}AV$ is upper triangular with diagonal elements $\alpha_1, \ldots, \alpha_n$. The argument, using (11.51), runs as follows. Let $j \in \{1, \ldots, n\}$. If $\nu(j) = 0$, then the jth column of AV is $A\mathbf{v}_0(\alpha_j) = \alpha_j \mathbf{v}_0(\alpha_j)$, and so the jth column of $V^{-1}AV$ has α_j as its jth component and zeros everywhere else. Now assume $\nu(j)$ is positive, hence $j > 1$. Then the jth column of AV is

$$A\mathbf{v}_{\nu(j)}(\alpha_j) = \alpha_j \mathbf{v}_{\nu(j)}(\alpha_j) + \mathbf{v}_{\nu(j)-1}(\alpha_j).$$

It follows that the jth column of $V^{-1}AV$ has α_j on the jth position and zeros everywhere else, except the number 1 on the kth position where

$$k = \max\{l \mid l = 1, \ldots, (j-1),\ \alpha_l = \alpha_j\} < j,$$

and so $\nu(k) = \nu(j) - 1$.

We proceed this review by discussing complete chains of invariant subspaces of a first companion $n \times n$ matrix A. Let $\alpha_1, \ldots, \alpha_n$ be an ordering of the eigenvalues of A (algebraic multiplicities counted) and write M_l for the span of the first l columns of the generalized Vandermonde $V = V(\alpha_1, \ldots, \alpha_n)$. Then, as $V^{-1}AV$ is upper triangular, $\{0\} = M_0 \subset M_1 \subset \cdots \subset M_{n-1} \subset M_n = \mathbb{C}^n$ is a complete chain of invariant subspaces for A. As the next proposition shows, the converse is also true.

Proposition 11.20. *Let A be an $n \times n$ first companion matrix, and let*

$$\{0\} = M_0 \subset M_1 \subset \cdots \subset M_{n-1} \subset M_n = \mathbb{C}^n \tag{11.56}$$

be a complete chain of A-invariant subspaces. Then there exists a unique ordering $\alpha_1, \dots, \alpha_n$ *of the eigenvalues of* A (*algebraic multiplicities counted*) *such that, for* $l = 0, \dots, n$, *the subspace* M_l *is the span of the first* l *columns of the generalized Vandermonde matrix* $V(\alpha_1, \dots, \alpha_n)$.

We shall refer to $V(\alpha_1, \dots, \alpha_n)$ as the *generalized Vandermonde matrix for the chain* (11.56). An alternative formulation of the conclusion of the proposition reads this way: there exists a unique ordering $\alpha_1, \dots, \alpha_n$ of the eigenvalues of A (algebraic multiplicities counted) such that, for $l = 0, \dots, n$,

$$M_l = \operatorname{span}\{\mathbf{v}_{\nu(1)}(\alpha_1), \dots, \mathbf{v}_{\nu(l)}(\alpha_l)\}, \qquad l = 0, \dots, n, \tag{11.57}$$

where $\nu(j)$ is the number of times that the eigenvalue α_j appears among its predecessors $\alpha_1, \dots, \alpha_{j-1}$.

Proof. For the sake of completeness we present the full proof, which will be split into two parts. The first part deals with the uniqueness statement. Throughout we use the notations introduced above.

Part 1. Suppose we have two orderings $\hat{\alpha}_1, \dots, \hat{\alpha}_n$ and $\check{\alpha}_1, \dots, \check{\alpha}_n$ of the eigenvalues of A (algebraic multiplicities counted) such that, for $l = 0, \dots, n$, the subspace spanned by the first l columns of $V(\hat{\alpha}_1, \dots, \hat{\alpha}_n)$ coincides with the subspace spanned by the first l columns of $V(\check{\alpha}_1, \dots, \check{\alpha}_n)$. Thus, for $l = 0, \dots, n$,

$$\operatorname{span}\{\mathbf{v}_{\hat{\nu}(1)}(\hat{\alpha}_1), \dots, \mathbf{v}_{\hat{\nu}(l)}(\hat{\alpha}_l)\} = \operatorname{span}\{\mathbf{v}_{\check{\nu}(1)}(\check{\alpha}_1), \dots, \mathbf{v}_{\check{\nu}(l)}(\check{\alpha}_l)\}, \tag{11.58}$$

where $\hat{\nu}(j)$ is the number of times that the eigenvalue $\hat{\alpha}_j$ appears among its predecessors $\hat{\alpha}_1, \dots, \hat{\alpha}_{j-1}$, and $\check{\nu}(j)$ is the number of times that the eigenvalue $\check{\alpha}_j$ appears among $\check{\alpha}_1, \dots, \check{\alpha}_{j-1}$. For $t = 0, \dots, s$ and $l = 0, \dots, n$, introduce

$$\begin{aligned} \hat{m}_t[l] &= \sharp\{k \mid k = 1, \dots, l,\ \hat{\alpha}_k = \mathbf{a}_t\}, \\ \check{m}_t[l] &= \sharp\{k \mid k = 1, \dots, l,\ \check{\alpha}_k = \mathbf{a}_t\}, \end{aligned}$$

where, as before, the symbol $\sharp$ stands for number of elements. Fix (for the time being) $t \in \{1, \dots, s\}$, $l \in \{0, \dots, n\}$, and consider the $\hat{m}_t[l]$ vectors among

$$\mathbf{v}_{\hat{\nu}(1)}(\hat{\alpha}_1), \dots, \mathbf{v}_{\hat{\nu}(l)}(\hat{\alpha}_l) \tag{11.59}$$

corresponding to the eigenvalue $\mathbf{a}_t$. Writing

$$\{k \mid k = 1, \dots, l,\ \hat{\alpha}_k = \mathbf{a}_t\} = \{k_1, \dots, k_{\hat{m}_t[l]}\},$$

with $k_1 < \cdots < k_{\hat{m}_t[l]}$, these vectors are $\mathbf{v}_{\hat{\nu}(k_1)}(\mathbf{a}_t), \dots, \mathbf{v}_{\hat{\nu}(k_{\hat{m}_t[l]})}(\mathbf{a}_t)$. But $\hat{\nu}(k_j) = j-1$ for $j = 1, \dots, \hat{m}_t[l]$, and we conclude that the vectors under consideration are $\mathbf{v}_0(\mathbf{a}_t), \dots, \mathbf{v}_{\hat{m}_t[l]-1}(\mathbf{a}_t)$. Letting t now run from 1 to s, we see that via a suitable reordering, (11.59) can be transformed into

$$\mathbf{v}_0(\mathbf{a}_1), \dots, \mathbf{v}_{\hat{m}_1[l]-1}(\mathbf{a}_1), \dots, \mathbf{v}_0(\mathbf{a}_s), \dots, \mathbf{v}_{\hat{m}_s[l]-1}(\mathbf{a}_s).$$

So the latter vectors span the subspace in the left-hand side of (11.58). In the same vein we have that

$$\mathbf{v}_0(\mathbf{a}_1), \dots, \mathbf{v}_{\check{m}_1[l]-1}(\mathbf{a}_1), \dots, \mathbf{v}_0(\mathbf{a}_s), \dots, \mathbf{v}_{\check{m}_s[l]-1}(\mathbf{a}_s).$$

span the subspace in the right-hand side of (11.58). However, the left and right-hand side of (11.58) are the same, and it follows that

$$\hat{m}_t[l] = \check{m}_t[l], \qquad t = 1, \dots, s,\ l = 0, \dots, n.$$

From this it is clear that the orderings $\hat{\alpha}_1, \dots, \hat{\alpha}_n$ and $\check{\alpha}_1, \dots, \check{\alpha}_n$ must coincide. Indeed, if $\hat{\alpha}_l = \mathbf{a}_t$, then $\check{m}_t[l] = \hat{m}_t[l] = \hat{m}_t[l-1] + 1 = \check{m}_t[l-1] + 1$, and it follows that $\check{a}_l = \mathbf{a}_t = \hat{a}_l$.

Part 2. We now prove existence. Given the complete chain of A-invariant subspaces as in the proposition, write M_l in the form

$$M_l = \operatorname{span}\{\mathbf{v}_0(\mathbf{a}_1), \dots, \mathbf{v}_{m_1[l]-1}(\mathbf{a}_1), \dots, \mathbf{v}_0(\mathbf{a}_s), \dots, \mathbf{v}_{m_s[l]-1}(a_s)\},$$

where $m_1[l], \dots, m_s[l]$ are non-negative integers not exceeding the algebraic multiplicities $\mathbf{m}_1, \dots, \mathbf{m}_s$, respectively. Note that

$$\sum_{t=1}^{s} m_t[l] \;=\; \dim M_l = l, \qquad l = 0, \dots, n,$$

while, for the extreme values $l = 0$ and $l = n$,

$$m_t[0] = 0, \quad m_t[n] = \mathbf{m}_t, \qquad t = 1, \dots, s.$$

Further it is clear from our considerations concerning (11.55) that

$$m_t[l-1] \;\le\; m_t[l] \;\le\; m_t[l-1] + 1, \qquad t = 1, \dots, s,\ j = 1, \dots, n.$$

Thus the value of $m_t[l] - m_t[l-1]$ is either 0 or 1. Also

$$\sum_{t=1}^{s} \big(m_t[l] - m_t[l-1]\big) \;=\; l - (l-1) = 1, \qquad j = 1, \dots, n.$$

Hence the differences $m_t[l] - m_t[l-1]$ are 0, except for one which is equal to 1.

For $l = 1, \dots, n$, write $\tau(l)$ for the unique integer t among $1, \dots, s$ such that $m_t[l] - m_t[l-1] = 1$ for $t = \tau(l)$. A little later, we shall see that

$$m_t[l] \;=\; \sharp\{k \mid k = 1, \dots, l,\ \tau(k) = t\}, \qquad t = 1, \dots, s,\ l = 1, \dots, n. \tag{11.60}$$

assuming this for the moment, we proceed as follows. Put $\alpha_k = \mathbf{a}_{\tau(k)}$. Then α_k is an eigenvalue of A. For $t = 1, \dots, s$, we have

$$\sharp\{k \mid k = 1, \dots, n,\ \alpha_k = \mathbf{a}_t\} \;=\; \sharp\{k \mid k = 1, \dots, n,\ \tau(k) = t\},$$

and the latter, by (11.60), is equal to $m_t[n]$ which, in turn, is just $\mathbf{m}_t$. Thus $\alpha_1, \ldots, \alpha_n$ is an ordering of the eigenvalues of A, algebraic multiplicities taken into account. As we shall see, the associated generalized Vandermonde matrix $V(\alpha_1, \ldots, \alpha_n)$ has the desired property.

Take l among the integers $1, \ldots, n$, and consider the subspace of $\mathbb{C}^n$ spanned by the first l columns of $V(\alpha_1, \ldots, \alpha_n)$, that is, by the vectors

$$\mathbf{v}_{\nu(1)}(\alpha_1), \ldots, \mathbf{v}_{\nu(l)}(\alpha_l), \tag{11.61}$$

where $\nu(j)$ is the number of times that the eigenvalue α_j appears among its predecessors $\alpha_1, \ldots, \alpha_{j-1}$. Take $t \in \{1, \ldots, s\}$. Then, as we see by another application of (11.60),

$$\sharp\{k \mid k = 1, \ldots, l,\ \alpha_k = \mathbf{a}_t\} \;=\; \sharp\{k \mid k = 1, \ldots, l,\ \tau(k) = t\} = m_t[l],$$

and it follows (see the first paragraph of the proof) that via a suitable change of order (11.61) can be rearranged into

$$\mathbf{v}_0(\mathbf{a}_1), \ldots, \mathbf{v}_{m_1[l]-1}(\mathbf{a}_1), \ldots, \mathbf{v}_0(\mathbf{a}_s), \ldots, \mathbf{v}_{m_s[l]-1}(\mathbf{a}_s).$$

So the latter vectors span the same subspace as those of (11.61). They also span M_l. Thus M_l is given by (11.57), as desired.

We still have to establish (11.60). Recall that $\tau(l)$ is the unique integer t among $1, \ldots, s$ such that $m_t[l] - m_t[l-1] = 1$ for $t = \tau(l)$. Thus, employing the familiar Kronecker delta notation, the property uniquely determining $\tau(l)$ can be expressed as follows:

$$m_t[l] \;=\; m_t[l-1] + \delta_{t,\tau(l)}, \qquad t = 1, \ldots, s,\ l = 1, \ldots, n. \tag{11.62}$$

The proof of (11.60) now goes by finite induction (on l).

Let $t \in \{1, \ldots, n\}$, and take $l = 1$. Then the right-hand side of the identity in (11.60) is equal to the number of integers k in the singleton set $\{1\}$ such that $\tau(k) = t$. If $\tau(1) = t$, this number is 1; if $\tau(1) \neq t$, it is equal to 0. So in this situation ($l = 1$), the right-hand side of the identity in (11.60) equals $\delta_{t,\tau(1)}$. However, by (11.62), together with $m_t[0] = 0$, we have that $m_t[1] = \delta_{t,\tau(1)}$ too. Hence (11.60) is true for $l = 1$.

Turning to the induction step, let $t \in \{1, \ldots, n\}$ and $l \in \{1, \ldots, n-1\}$. Clearly $\sharp\{k \mid k = 1, \ldots, (l+1),\ \tau(k) = t\}$ is equal to

$$\sharp\{k \mid k = 1, \ldots, l,\ \tau(k) = t\} \;+\; \sharp\{k \mid k = l+1,\ \tau(k) = t\}.$$

If $\tau(l+1) = t$, the second term in the latter expression is 1; if $\tau(l+1) \neq t$, it is equal to 0. In other words, the number in question is $\delta_{t,\tau(l+1)}$. Combining this with (11.60), which may be assumed to hold by induction hypothesis, we get

$$\sharp\{k \mid k = 1, \ldots, l+1,\ \tau(k) = t\} \;=\; m_t[l] + \delta_{t,\tau(l+1)}.$$

By (11.62), the right-hand side of this identity is $m_t[l+1]$, and the desired result follows. □

Notes

Propositions 11.1 and 11.2 can be found in somewhat different form in [24]. The latter paper also deals with simultaneous reduction to companion forms of an arbitrary number (instead of just pairs) of matrices. The proof of Proposition 11.1 given here, exhibiting a connection with the Bezout matrix, is from [20]; see also [122]. Section 11.2 is based on Section 3 of [19]. Observations strongly related to Theorem 11.8 can be found in Section 2 of [24], and Section 4 of [20].

The material on companion based matrix functions of Sections 11.3 and 11.4 is inspired by Section 3 in [20]. The approach there is more general in that also companion based matrix functions of size larger than two are considered. The results of Section 11.5 on complete factorization of companion based matrix functions can be traced back to Sections 3 and 6 in [19], and to Section 4 in [20]. In the latter paper (possibly non-complete) minimal factorization of companion based matrix functions is discussed in detail, including (canonical) Wiener-Hopf factorization. It is interesting to note (see Theorem 4.1 in [20]) that the class of companion based matrix functions behaves well under minimal factorization: *If $W = W_1W_2$ is a minimal factorization of a companion based matrix function W, then W_1 and W_2 are companion based too.* The proofs of Theorems 11.17 and 11.18 are constructive as long as the poles and zeros of the companion based matrix function W are known. This fact is illustrated by the Maple procedures presented in Section 11.6 which have been written by Johan F. Kaashoek. The material on invariant subspaces of companion matrices in Section 11.7 is of text book type (cf., Section 2.11, Exercises 21 and 22 in [92]) but not readily available in the specific form needed for Chapter 12. A similar (though less elaborate) exposition can be found in the proof of Lemma 1.1 in [24].

Chapter 12

Quasicomplete Factorization and Job Scheduling

In this chapter a connection is made between the issue of quasicomplete factorization discussed in Section 10.4 and a problem from the theory of combinatorial job scheduling. The problem in question is the so-called two machine flow shop problem (2MFSP for short) where one wants to find optimal schedules for processing jobs on two machines, given certain precedence constraints. It turns out that such problems are in correspondence (one-to-one, essentially) with the companion based rational matrix functions considered in the previous chapter. We show that the number of factors in a quasicomplete factorization of a companion based matrix function is directly related to the minimum makespan (i.e., the time needed for carrying out a optimal schedule) of the associated instance of 2MFSP. Illustrative examples are given. In one of them the (computationally fast) algorithm called Johnson's rule for 2MFSP is used to compute the quasidegree of a companion based function.

The present chapter consists of five sections. The first presents a combinatorial lemma that will be used in the analysis of quasicomplete factorization of companion based matrix functions. The latter topic is the main subject of the second section. In the third section we introduce the two machine flow shop problem and review some of the related results, including Johnson's rule. In the fourth section we establish the relation to quasicomplete factorization of companion based matrix functions. The final section presents Maple procedures to calculate explicitly quasicomplete factorization of a companion based 2×2 matrix function.

12.1 A combinatorial lemma

In the next section we shall consider quasicomplete factorization (into elementary factors) of companion based rational matrix functions. Here we present a combinatorial auxiliary result to be used in that context.

Lemma 12.1. *Let p be positive integer, and let $\hat{a}_1, \ldots, \hat{a}_p$ and $\check{a}_1, \ldots, \check{a}_p$ be two (finite) sequences of elements (not specified at the moment but later to be taken from the complex numbers). Let r be a positive integer not exceeding p, and assume that for $l = r, \ldots, p$ the sequences $\hat{a}_1, \ldots, \hat{a}_l$ and $\check{a}_1, \ldots, \check{a}_{p+r-l}$ have at most $r-1$ entries in common, multiplicities counted. Then there exist permutations $\hat{\pi}$ and $\check{\pi}$ of the set $\{1, \ldots, p\}$ such that*

$$\hat{a}_{\hat{\pi}(\hat{s})} \neq \check{a}_{\check{\pi}(\check{s})}, \qquad \hat{s}, \check{s} = 1, \ldots, p\,;\ \hat{s} + \check{s} \leq p + 2 - r.$$

It is helpful to clarify the hypotheses of the lemma via some notations that will also be used in the proof. For $l = r, \ldots, p$ and $\alpha \in \mathcal{A} = \{\hat{a}_1, \ldots, \hat{a}_p, \check{a}_1, \ldots, \check{a}_p\}$, introduce

$$\begin{aligned} \hat{\nu}_\alpha(l) &= \sharp\{j \mid j = 1, \ldots, l;\ \hat{a}_j = \alpha\}, \\ \check{\nu}_\alpha(l) &= \sharp\{j \mid j = 1, \ldots, l;\ \check{a}_j = \alpha\}, \end{aligned}$$

(where, as in Section 11.1, the symbol $\sharp$ stands for number of elements) and

$$\begin{aligned} \hat{\mu}_\alpha(l) &= \min\{\hat{\nu}_\alpha(l),\, \check{\nu}_\alpha(p + r - l)\}, \\ \check{\mu}_\alpha(l) &= \min\{\check{\nu}_\alpha(l), \hat{\nu}_\alpha(p + r - l)\}. \end{aligned}$$

Note in this context that when l runs through $r, \ldots, p$, then $p + r - l$ runs through $p, \ldots, r$. Obviously $\check{\mu}_\alpha(l) = \hat{\mu}_\alpha(p + r - l)$, and the overlap assumptions of the lemma can be expressed as

$$\sum_{\alpha \in \mathcal{A}} \hat{\mu}_a(l) = \sum_{\alpha \in \mathcal{A}} \check{\mu}_a(p + r - l) < r, \qquad l = r, \ldots, p. \tag{12.1}$$

Clearly there is symmetry here with respect to the two given sequences $\hat{a}_\cdot$ and $\check{a}_\cdot$ (replace l by $p + r - l$). In line with this, the conclusion of the lemma is symmetric in $\hat{a}_\cdot$ and $\check{a}_\cdot$ too.

Proof. As everywhere else in this section where this is convenient, we use the notation introduced above. Take $\alpha \in \mathcal{A}$ and write

$$\begin{aligned} \{j \mid j = 1, \ldots, p;\ \hat{a}_j = \alpha\} &= \{\hat{t}_\alpha[1], \hat{t}_\alpha[2], \ldots, t_\alpha[\hat{\nu}_\alpha(p)]\}, \\ \{j \mid j = 1, \ldots, p;\ \check{a}_j = \alpha\} &= \{\check{t}_\alpha[1], \check{t}_\alpha[2], \ldots, t_\alpha[\check{\nu}_\alpha(p)]\}, \end{aligned}$$

with $\hat{t}_\alpha[1] < \hat{t}_\alpha[2] < \cdots < t_\alpha[\hat{\nu}_\alpha(p)]$ and $\check{t}_\alpha[1] < \check{t}_\alpha[2] < \cdots < t_\alpha[\check{\nu}_\alpha(p)]$. Then, for $l = r, \ldots, p$,

$$\begin{aligned} \{j \mid j = 1, \ldots, l;\ \hat{a}_j = \alpha\} &= \{\hat{t}_\alpha[1], \hat{t}_\alpha[2], \ldots, t_\alpha[\hat{\nu}_\alpha(l)]\}, \\ \{j \mid j = 1, \ldots, l;\ \check{a}_j = \alpha\} &= \{\check{t}_\alpha[1], \check{t}_\alpha[2], \ldots, t_\alpha[\check{\nu}_\alpha(l)]\}. \end{aligned}$$

Now put

$$\begin{aligned}\hat{\mathcal{O}}_\alpha(l) &= \{\hat{t}_\alpha[1], \hat{t}_\alpha[2], \ldots, \hat{t}_\alpha[\hat{\mu}_\alpha(l)]\},\\ \check{\mathcal{O}}_\alpha(l) &= \{\check{t}_\alpha[1], \check{t}_\alpha[2], \ldots, \check{t}_\alpha[\check{\mu}_\alpha(l)]\}.\end{aligned}$$

As $\hat{\mu}_\alpha(l) \le \hat{\nu}_\alpha(l)$ and $\check{\mu}_\alpha(l) \le \check{\nu}_\alpha(l)$, both $\hat{\mathcal{O}}_\alpha(l)$ and $\check{\mathcal{O}}_\alpha(l)$ are subsets of $\{1, \ldots, l\}$.

Taking (disjoint) unions, we obtain

$$\hat{\mathcal{O}}(l) = \bigcup_{\alpha\in\mathcal{A}} \hat{\mathcal{O}}_\alpha(l), \qquad \check{\mathcal{O}}(l) = \bigcup_{\alpha\in\mathcal{A}} \check{\mathcal{O}}_\alpha(l),$$

again both are subsets of $\{1, \ldots, l\}$. We now claim that

$$\hat{a}_{\hat{t}} \ne \check{a}_{\check{t}} \tag{12.2}$$

whenever $\hat{t} \in \{1, \ldots, \hat{l}\} \setminus \hat{\mathcal{O}}(\hat{l})$, $\check{t} \in \{1, \ldots, \check{l}\} \setminus \check{\mathcal{O}}(\check{l})$ with $\hat{l}$ and $\check{l}$ from $\{r, \ldots, p\}$ satisfying $\hat{l} + \check{l} \le p + r$. The proof goes by contradiction. Assume $\hat{a}_{\hat{t}} = \check{a}_{\check{t}} = \alpha$. Then

$$\hat{t} \in \{1, \ldots, \hat{l}\} \setminus \hat{\mathcal{O}}_\alpha(\hat{l}), \qquad \check{t} \in \{1, \ldots, \check{l}\} \setminus \check{\mathcal{O}}_\alpha(\check{l}).$$

Clearly $\hat{t} \in \{j \mid j = 1, \ldots, \hat{l};\ \hat{a}_j = \alpha\} = \{\hat{t}_\alpha[1], \ldots, t_\alpha[\hat{\nu}_\alpha(\hat{l})]\}$. Also

$$\begin{aligned}\{\hat{t}_\alpha[1], \ldots, t_\alpha[\hat{\nu}_\alpha(\hat{l})]\} &= \{\hat{t}_\alpha[1], \ldots, t_\alpha[\hat{\mu}_\alpha(\hat{l})], \ldots, t_\alpha[\hat{\nu}_\alpha(\hat{l})]\}\\ &= \hat{\mathcal{O}}_\alpha(\hat{l}) \cup \{\hat{t}_\alpha[\hat{\mu}_\alpha(\hat{l}) + 1], \ldots, t_\alpha[\hat{\nu}_\alpha(\hat{l})]\},\end{aligned}$$

and, since $\hat{t} \notin \hat{\mathcal{O}}_\alpha(\hat{l})$, it follows that $\hat{\nu}_\alpha(\hat{l}) > \hat{\mu}_\alpha(\hat{l}) = \check{\nu}_\alpha(p + r - \hat{l})$. As $p + r - \hat{l} \ge \check{l}$, we have

$$\begin{aligned}\check{\nu}_\alpha(p + r - \hat{l}) &= \sharp\{j \mid j = 1, \ldots, p + r - \hat{l};\ \check{a}_j = \alpha\}\\ &\ge \sharp\{j \mid j = 1, \ldots, \check{l};\ \check{a}_j = \alpha\} = \check{\nu}_\alpha(\check{l}),\end{aligned}$$

and so $\hat{\nu}_\alpha(\hat{l}) > \check{\nu}_\alpha(\check{l})$. In the same vein (or, if one prefers, by reasons of symmetry), we also have $\check{\nu}_\alpha(\check{l}) > \check{\mu}_\alpha(\check{l}) = \hat{\nu}_\alpha(p + r - \check{l}) \ge \hat{\nu}_\alpha(\hat{l})$, and a contradiction has been obtained which shows that (12.2) does indeed hold.

Next we turn to the construction of the permutation $\hat{\pi}$. Here the overlap assumptions (12.1) come into play. Note that

$$\sharp\,\hat{\mathcal{O}}(l) = \sum_{\alpha\in\mathcal{A}} \sharp\,\hat{\mathcal{O}}_\alpha(l) = \sum_{\alpha\in\mathcal{A}} \hat{\mu}_\alpha(l), \qquad l = r, \ldots, p.$$

Thus (12.1) gives $\sharp\,\hat{\mathcal{O}}(l) < r$. Specializing to $l = r$, we have $\sharp\,\hat{\mathcal{O}}(r) < r$. On the other hand $\hat{\mathcal{O}}(r) \subset \{1, \ldots, r\}$. So $\hat{\mathcal{O}}(r)$ is a proper subset of $\{1, \ldots, r\}$ and we can take $\hat{\pi}(1) \in \{1, \ldots, r\} \setminus \hat{\mathcal{O}}(r)$. In case $r \le p - 1$, we proceed as follows. Clearly

$$\sharp\Big(\{\hat{\pi}(1)\} \cup \hat{\mathcal{O}}(r + 1)\Big) \le 1 + \sharp\,\hat{\mathcal{O}}(r + 1) < 1 + r.$$

Combining this with $\{\hat{\pi}(1)\} \cup \hat{\mathcal{O}}(r+1) \subset \{1,\dots,r+1\}$ leads to the choice of $\hat{\pi}(2)$ satisfying $\hat{\pi}(2) \neq \hat{\pi}(1)$ and $\hat{\pi}(2) \in \{1,\dots,r+1\} \setminus \hat{\mathcal{O}}(r+1)$. When $r \leq p-2$, this procedure can be continued. Indeed, let $s \in \{1,\dots,p-r\}$ and assume that the different integers $\hat{\pi}(1),\dots,\hat{\pi}(s)$ have been chosen in such a way that

$$\hat{\pi}(j) \in \{1,\dots,r+j-1\} \setminus \hat{\mathcal{O}}(r+j-1), \qquad j = 1,\dots,s.$$

Then $\{\hat{\pi}(1),\dots,\hat{\pi}(s)\} \cup \hat{\mathcal{O}}(r+s) \subset \{1,\dots,r+s\}$. The number of elements in the left-hand side of this inclusion is at most $s + \sharp\,\hat{\mathcal{O}}(r+s)$, hence smaller than $s+r$. Therefore it is possible to pick $\hat{\pi}(s+1)$ from $\{1,\dots,r+s\}$ such that $\hat{\pi}(s+1) \neq \hat{\pi}(1),\dots,\hat{\pi}(s)$ and $\hat{\pi}(s+1) \in \{1,\dots,r+s\} \setminus \hat{\mathcal{O}}(r+s)$. In other words, the above expression (displayed) is also valid for $j = s+1$.

In this way (more formally, by finite induction), the existence has been established of an injective function $\hat{\pi} : \{1,\dots,p-r+1\} \to \{1,\dots,p\}$ with

$$\hat{\pi}(j) \in \{1,\dots,r+j-1\} \setminus \hat{\mathcal{O}}(r+j-1), \qquad j = 1,\dots,p-r+1.$$

We complete this function to a permutation of $\{1,\dots,p\}$ by choosing mutually different values $\hat{\pi}(p-r+2),\dots,\hat{\pi}(p)$ from the set

$$\{1,\dots,p\} \setminus \{\pi(1),\dots,\hat{\pi}(p-r+1)\}.$$

Analogously (or, if one prefers, by symmetry) there exists a permutation $\check{\pi}$ of $\{1,\dots,p\}$ satisfying

$$\check{\pi}(j) \in \{1,\dots,r+j-1\} \setminus \check{\mathcal{O}}(r+j-1), \qquad j = 1,\dots,p-r+1.$$

Now suppose $\hat{s}, \check{s} \in \{1,\dots,p\}$ and $\hat{s} + \check{s} \leq p+2-r$. Put

$$\hat{t} = \hat{\pi}(\hat{s}), \qquad \check{t} = \check{\pi}(\check{s}), \qquad \hat{l} = r + \hat{s} - 1, \qquad \check{l} = r + \check{s} - 1.$$

Then $\hat{l}, \check{l} \in \{1,\dots,p\}$ and

$$\hat{l} + \check{l} \;=\; 2r - 2 + \hat{s} + \check{s} \;\leq\; 2r - 2 + p + 2 - r \;=\; p + r.$$

Also $\hat{s}, \check{s} \in \{1,\dots,p+r-1\}$ and so

$$\hat{t} \;=\; \hat{\pi}(\hat{s}) \in \{1,\dots,\hat{l}\} \setminus \hat{\mathcal{O}}(\hat{l}), \qquad \check{t} \;=\; \check{\pi}(\check{s}) \in \{1,\dots,\check{l}\} \setminus \check{\mathcal{O}}(\check{l}).$$

Thus we have the situation considered in the second paragraph of this proof. Therefore (12.2) holds, i.e., $\hat{a}_{\hat{\pi}(\hat{s})} \neq \check{a}_{\check{\pi}(\check{s})}$, as desired. □

We close this section with an example illustrating Lemma 12.1 and its proof.

Take $p = 9$ and let the (finite) sequences $\hat{a}_1, \dots, \hat{a}_9$ and $\check{a}_1, \dots, \check{a}_9$ be given – schematically – by

	1	2	3	4	5	6	7	8	9
$\hat{a}_\cdot$:	♣	♠	♠	♣	✠	♡	♣	♢	✠
$\check{a}_\cdot$:	♡	✠	✠	✠	♡	✠	♣	♢	♠

For $l = 4, \dots, 9$, the sequences $\hat{a}_1, \dots, \hat{a}_l$ and $\check{a}_1, \dots, \check{a}_{13-l}$ have at most 3 entries in common, multiplicities counted. Thus the overlap conditions of the lemma are fulfilled for $r = 4$. It is not hard to verify that

$$\begin{aligned}
\hat{\mathcal{O}}_4 &= \{1, 2\}, & \check{\mathcal{O}}_4 &= \{1, 2, 3\},\\
\hat{\mathcal{O}}_5 &= \{1, 5\}, & \check{\mathcal{O}}_5 &= \{1, 2\},\\
\hat{\mathcal{O}}_6 &= \{1, 5, 6\}, & \check{\mathcal{O}}_6 &= \{1, 2\},\\
\hat{\mathcal{O}}_7 &= \{5, 6\}, & \check{\mathcal{O}}_7 &= \{1, 2, 7\},\\
\hat{\mathcal{O}}_8 &= \{5, 6\}, & \check{\mathcal{O}}_8 &= \{2, 7\},\\
\hat{\mathcal{O}}_9 &= \{5, 6, 9\}, & \check{\mathcal{O}}_9 &= \{7, 9\}.
\end{aligned}$$

One can now construct permutations $\hat{\pi}$ and $\check{\pi}$ of $\{1, \dots, 9\}$ along the lines indicated in the above proof. In the present case different choices can be made. One of the possible outcomes is given – schematically – by

	1	2	3	4	5	6	7	8	9
$\hat{\pi}(.)$:	3	4	2	1	7	8	5	6	9
$\check{\pi}(.)$:	4	3	6	5	1	8	7	9	2

The sequences $\hat{a}_{\hat{\pi}(1)}, \dots, \hat{a}_{\hat{\pi}(9)}$ and $\check{a}_{\check{\pi}(1)}, \dots, \check{a}_{\check{\pi}(9)}$ can now be displayed as

	1	2	3	4	5	6	7	8	9
$\hat{a}_{\hat{\pi}(.)}$:	♠	♣	♠	♣	♣	♢	✠	♡	✠
$\check{a}_{\check{\pi}(.)}$:	✠	✠	✠	♡	♡	♢	♣	♠	✠

A simple check shows that

$$\hat{a}_{\hat{\pi}(\hat{s})} \neq \check{a}_{\check{\pi}(\check{s})}, \qquad \hat{s}, \check{s} = 1, \dots, 9, \;\; \hat{s} + \check{s} \leq 7$$

as required.

12.2 Quasicomplete factorization (companion based)

In this section we consider quasicomplete factorization of companion based functions. For the relevant definitions, see Section 10.4.

Theorem 12.2. *Let W be a companion based rational $m \times m$ matrix function, and let n be the McMillan degree of W (assumed to be positive in order to avoid trivialities). Then the quasidegree $\delta_{\mathrm{q}}(W)$ of W is the smallest integer d larger than or equal to n for which there exist an ordering $\alpha_1, \ldots, \alpha_n$ of the poles of W (pole-multiplicities taken into account) and an ordering $\alpha_1^{\times}, \ldots, \alpha_n^{\times}$ of the zeros of W (zero-multiplicities taken into account) such that*

$$\alpha_k \neq \alpha_j^{\times}, \qquad k, j = 1, \ldots, n, \;\; k < j - (d - n). \tag{12.3}$$

If (12.3) holds for a certain integer d, then so it does when one replaces d by any larger integer. Also (12.3) is vacuously satisfied for $d = 2n - 1$. This is in line with the estimate $\delta_{\mathrm{q}}(W) \leq 2\delta(W) - 1$ of the quasidegree given in (10.30).

The proof of Theorem 12.2 is quite involved and will be split in two parts. The first depends heavily on the combinatorial lemma of the previous section. We begin with a few lemmas.

Lemma 12.3. *Let X_-, X_0, X_+ and X be finite-dimensional Banach spaces, and assume $X = X_- \dot{+} X_0 \dot{+} X_+$. For M a subspace of X, write $M[0] = (M + X_-) \cap X_0$. Then $M[0] = P_0[M \cap (X_- \dot{+} X_0)]$, where P_0 is the projection of X onto X_0 along $X_- \dot{+} X_+$. In addition, assume that $T : X \to X$ is a (bounded) linear operator whose 3×3 operator matrix representation with respect to the given decomposition has the upper triangular form*

$$T = \begin{bmatrix} * & * & * \\ 0 & T_0 & * \\ 0 & 0 & * \end{bmatrix} : X_- \dot{+} X_0 \dot{+} X_+ \to X_- \dot{+} X_0 \dot{+} X_+,$$

with the stars denoting unspecified (possibly nonzero) entries acting between the appropriate spaces, and with $T_0 : X_0 \to X_0$. Then a sufficient condition for $M[0]$ to be invariant for T_0 is that M is invariant for T.

Proof. Take $x \in M[0]$. Then $x \in X_0$, so $P_0 x = x$. Also $x = m + x_-$ for some $m \in M$ and $x_- \in X_-$. Now $m = -x_- + x \in X_- \dot{+} X_0$ and $P_0 m = -P_0 x_- + P_0 x = P_0 x = x$. Thus $M[0] \subset P_0[M \cap (X_- \dot{+} X_0)]$. For the reverse inclusion, assume $y = P_0 m$ with $m \in M \cap (X_- \dot{+} X_0)$. Write $m = x_- + x_0$ with $x_- \in X_-$ and $x_0 \in X_0$. Then $y = P_0 m = P_0 x_0 = x_0 = m - x_- \in M + X_-$. Also $y = P_0 m \in X_0$, and it follows that $P_0[M \cap (X_- \dot{+} X_0)] \subset M[0]$, as desired.

Suppose now that M is an invariant subspace for T. Then, as $X_- \dot{+} X_0$ is T-invariant too, so is $M \cap (X_- \dot{+} X_0)$. This will be used to prove the T_0-invariance of $M[0] = P[M \cap (X_- \dot{+} X_0)]$.

With respect to the decomposition $X = X_- \dot{+} X_0 \dot{+} X_+$, the operators P_0T and P_0TP_0 have the form

$$P_0T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & T_0 & * \\ 0 & 0 & 0 \end{bmatrix}, \qquad P_0TP_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & T_0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

From the first identity we see that $X_- \subset \operatorname{Ker} P_0T$, from the second that $T_0x_0 = P_0TP_0x_0$ for all $x_0 \in X_0$. Take $m_0 \in M[0] \subset X_0$. Then $m_0 = P_0m$ for some $m \in M \cap (X_- \dot{+} X_0)$. Now $T_0m_0 = P_0TP_0m_0 = P_0TP_0m$. Also $m - P_0m \in X_-$ (for $m \in X_- \dot{+} X_0$), so $m - P_0m \in \operatorname{Ker} P_0T$. Hence $T_0m_0 = P_0TP_0m = P_0Tm$. But Tm belongs to $M \cap (X_- \dot{+} X_0)$, and so $T_0m_0 \in P_0[M \cap (X_- \dot{+} X_0)] = M[0]$, as desired. □

With the same notation as above, we also have the following result.

Lemma 12.4. *Let X_-, X_0, X_+ and X be finite-dimensional Banach spaces such that $X = X_- \dot{+} X_0 \dot{+} X_+$, and let M and N be subspaces of X. Then*

$$\dim\big(M[0] \cap N[0]\big) \leq \dim(M \cap N) + \dim X_-, \tag{12.4}$$

$$\dim\left(\frac{X_0}{M[0]+N[0]}\right) \leq \dim\left(\frac{X}{M+N}\right) + \dim X_+. \tag{12.5}$$

Also, if $M \subset N$, then $M[0] \subset N[0]$ and $\dim\big(N[0]/M[0]\big) \leq \dim(N/M)$.

Proof. The argument for (12.4) goes as follows. Choose a basis $m_1^0, \dots, m_s^0$ in $M[0]$. Note here that we may assume that $M[0]$ (as well as $N[0]$) is non-trivial, otherwise (12.4) is evident. For $j = 1, \dots, s$, choose m_j^- in X_- such that $m_j^0 - m_j^- \in M$. Then, clearly, $(m_1^0 - m_1^-), \dots, (m_s^0 - m_s^-) \in M \cap \big(M[0] + X_-\big)$. Suppose $z_1(m_1^0 - m_1^-) + \cdots + z_s(m_s^0 - m_s^-) = 0$, where $z_1, \dots, z_n \in \mathbb{C}$. Then

$$z_1m_1^0 + \cdots + z_sm_s^0 = z_1m_1^- + \cdots + z_sm_s^-.$$

The left-hand side of this identity is in $M[0] \subset X_0$, the right-hand side in X_-. But $X_0 \cap X_- = \{0\}$. So $(z_1m_1^0 + \cdots + z_sm_s^0) = (z_1m_1^- + \cdots + z_sm_s^-) = 0$. Since $m_1^0, \dots, m_s^0$ are linearly independent, it follows that $z_1 = \cdots = z_s = 0$. Thus the elements $(m_1^0 - m_1^-), \dots, (m_s^0 - m_s^-)$ are linearly independent too.

Putting $\widehat{M} = \operatorname{span}\{(m_1^0 - m_1^-), \dots, (m_s^0 - m_s^-)\}$, we obtain a subspace $\widehat{M}$ of X for which $\widehat{M} \subset M \cap \big(M[0]+X_-\big)$ and $\dim \widehat{M} = \dim M[0]$. In the same vein there is a subspace $\widehat{N}$ of X satisfying $\widehat{N} \subset N \cap \big(N[0] + X_-\big)$ and $\dim \widehat{N} = \dim N[0]$. By a standard identity

$$\dim \widehat{M} + \dim \widehat{N} = \dim\big(\widehat{M} \cap \widehat{N}\big) + \dim\big(\widehat{M} + \widehat{N}\big),$$

and it follows that $\dim M[0] + \dim N[0] = \dim\big(\widehat{M} \cap \widehat{N}\big) + \dim\big(\widehat{M} + \widehat{N}\big)$. As $\widehat{M} + \widehat{N} \subset M[0] + N[0] + X_-$, the dimension of $\widehat{M} + \widehat{N}$ does not exceed $\dim\big(M[0] + N[0]\big) + \dim X_-$. Hence

$$\dim M[0] + \dim N[0] \le \dim\big(\widehat{M} \cap \widehat{N}\big) + \dim\big(M[0] + N[0]\big) + \dim X_-.$$

Together with $\dim(M[0] \cap N[0]) = \dim M[0] + \dim N[0] - \dim\big(M[0] + N[0]\big)$, (again the standard identity), this gives

$$\dim\big(M[0] \cap N[0]\big) \le \dim\big(\widehat{M} \cap \widehat{N}\big) + \dim X_-.$$

As the dimension of $\widehat{M} \cap \widehat{N}$ does not exceed that of $M \cap N$ inequality (12.4) follows.

Next, we deal with (12.5). Let P_+ be the projection of X onto X_+ along $X_- \dot{+} X_0$. Choose $m_1, \dots, m_t$ in M such that the vectors $P_+m_1, \dots, P_+m_t$ span the subspace $P_+[M]$. Take $m \in M$. Then there exist $\gamma_1, \dots, \gamma_t \in \mathbb{C}$ such that $P_+m = \gamma_1 P_+m_1 + \cdots + \gamma_t P_+m_t$. Put $y = m - (\gamma_1 m_1 + \cdots + \gamma_t m_t)$. Then $y \in M$ and $y \in \operatorname{Ker} P_+ = X_- \dot{+} X_0$. Now

$$m = (y - P_0y) + P_0y + (\gamma_1 m_1 + \cdots + \gamma_t m_t).$$

The second term P_0y in the right-hand side belongs to $P_0[M \cap (X_- \dot{+} X_0)] = M[0]$ (see Lemma 12.3). For the first term $y - P_0y$ the following holds. On the one hand it belongs to $\operatorname{Ker} P_0 = X_- \dot{+} X_+$, on the other hand it is a member of $(X_- \dot{+} X_0) + X_0 = X_- \dot{+} X_0$. Hence $y - P_0y \in (X_- \dot{+} X_+) \cap (X_- \dot{+} X_0) = X_-$, and we conclude that $m \in X_- + M[0] + \operatorname{span}\{m_1, \dots, m_t\}$. In case $P_+[M]$ is non-trivial, one can take t equal to the dimension of $P_+[M] \subset X_+$ so that $\dim\big(\operatorname{span}\{m_1, \dots, m_t\}\big) \le \dim X_+$. The latter can also be arranged when $P_+[M] = \{0\}$. Just take $t = 1$ and $m_1 = 0$.

Putting $\widetilde{M} = \operatorname{span}\{m_1, \dots, m_t\}$, we obtain a subspace $\widetilde{M}$ of X such that $M \subset X_- + M[0] + \widetilde{M}$ and $\dim \widetilde{M} \le \dim X_+$. In the same vein there is a subspace $\widetilde{N}$ of X satisfying $N \subset X_- + N[0] + \widetilde{N}$ and $\dim \widetilde{N} \le \dim X_+$. Let L be a linear complement of $M + N$ in X. Then $X = L \dot{+} (M + N) = L + X_- + \big(M[0] + N[0]\big) + \widetilde{M} + \widetilde{N}$, and it follows that

$$\dim X \le \dim L + \dim X_- + \dim\big(M[0] + N[0]\big) + 2\dim X_+.$$

Combining this with the identity $\dim X = \dim X_- + \dim X_0 + \dim X_+$, we get the inequality $\dim X_0 - \dim\big(M[0] + N[0]\big) \le \dim L + \dim X_+$, which amounts to the same (12.5).

Finally, assume $M \subset N$. Then evidently $M[0] \subset N[0]$. By standard quotient space arguments, there exist an injective linear mapping

$$\frac{(N + X_-) \cap X_0}{(M + X_-) \cap X_0} \;\to\; \frac{N + X_-}{M + X_-},$$

and a surjective linear mapping

$$\frac{N}{M} \;\to\; \frac{N+X_-}{M+X_-}\,.$$

It follows that

$$\dim \frac{N[0]}{M[0]} \;=\; \dim\left(\frac{(N+X_-)\cap X_0}{(M+X_-)\cap X_0}\right) \;\le\; \dim\left(\frac{N+X_-}{M+X_-}\right) \;\le\; \dim\frac{N}{M},$$

and the proof is complete. □

Let us mention that in the application of Lemma 12.4 given below, the dimension of X_0 is strictly larger than the dimensions of X_- and X_+. Note that in that case the inequalities (12.4) and (12.5) are non-trivial.

First part of the proof of Theorem 12.2. Put $q = \delta_{\mathrm{q}}(W)$. Then $\delta(W) = n \le q \le 2n-1$ and we shall prove that there exist an ordering $\alpha_1, \ldots, \alpha_n$ of the poles of W (pole-multiplicities taken into account) and an ordering $\alpha_1^\times, \ldots, \alpha_n^\times$ of the zeros of W (zero-multiplicities taken into account) such that (12.3) is satisfied with $d = q$. When $q = 2n-1$, there is nothing to prove. So we shall assume that $q \le 2n-2$ (and $n \ge 2$). It is convenient to break up the argument into a number of steps.

Step 1. Since W admits a factorization into q elementary factors, Theorem 10.5 guarantees that W can be written as $W(\lambda) = I_m + C(\lambda I_q - A)^{-1}B$ with A an upper and $A^\times = A - BC$ a lower triangular $q \times q$ matrix. As has been discussed in detail in Section 10.1, these triangularity conditions come down to the existence of a complete chain $M_0 \subset M_1 \subset \cdots \subset M_q$ of A-invariant subspaces, and a complete chain $M_0^\times \subset M_1^\times \subset \cdots \subset M_q^\times$ of $A^\times$-invariant subspaces, such that

$$M_j \dotplus M_{q-j}^\times = \mathbb{C}^q, \qquad j = 0, \ldots, q. \tag{12.6}$$

Here, as everywhere else where this is convenient, matrices are identified in the usual way with operators acting between Euclidean spaces.

Step 2. Put $X = \mathbb{C}^q$. By the material on dilations presented in Section 7.3, we know that X admits a decomposition $X = X_- \dotplus X_0 \dotplus X_+$ such that A, B and C have the form

$$A \;=\; \begin{bmatrix} * & * & * \\ 0 & A_0 & * \\ 0 & 0 & * \end{bmatrix} : X_- \dotplus X_0 \dotplus X_+ \to X_- \dotplus X_0 \dotplus X_+,$$

$$B \;=\; \begin{bmatrix} * \\ B_0 \\ 0 \end{bmatrix} : \mathbb{C}^m \to X_- \dotplus X_0 \dotplus X_+,$$

$$C \;=\; \begin{bmatrix} 0 & C_0 & * \end{bmatrix} : X_- \dotplus X_0 \dotplus X_+ \to \mathbb{C}^m,$$

with the stars denoting unspecified (possibly nonzero) entries acting between the appropriate spaces, and with

$$W(\lambda) = I_m + C_0(\lambda I_{X_0} - A_0)^{-1}B_0 \tag{12.7}$$

being a minimal realization of W, so in particular $\dim X_0 = n$. Given the complete chains of subspaces from Step 1, we now construct comparable chains in the space X_0.

Step 3. For this we employ the material (including the notation) contained in Lemmas 12.3 and 12.4. So we apply the [0]-operation introduced in Lemma 12.3 to the complete chains of invariant subspaces featuring in Step 1. In other words, we consider the subspaces

$$M_j[0] = (M_j + X_-) \cap X_0, \qquad M_j^\times[0] = (M_j^\times + X_-) \cap X_0.$$

Note that $M_j[0]$ is A_0-invariant and $M_j^\times[0]$ is $A_0^\times$-invariant (see Lemma 12.3). Here it should be taken into account that the operator $A^\times = A - BC$ has the representation

$$A^\times = \begin{bmatrix} * & * & * \\ 0 & A_0^\times & * \\ 0 & 0 & * \end{bmatrix} : X_- \dot{+} X_0 \dot{+} X_+ \to X_- \dot{+} X_0 \dot{+} X_+,$$

again with the stars denoting possibly nonzero entries, and with $A_0^\times$ being the associate main operator of the unital system $(A_0, B_0, C_0, \mathbb{C}^m, \mathbb{C}^n)$ underlying (12.7), i.e., $A_0^\times = A_0 - B_0C_0$.

Step 4. For $j = 0, \dots, q-1$, we have

$$M_j[0] \subset M_{j+1}[0], \qquad \dim\left(\frac{M_{j+1}[0]}{M_j[0]}\right) \le \dim\left(\frac{M_{j+1}}{M_j}\right) = 1,$$

(see Lemma 12.4). Put $d_j = \dim M_j[0]$. Then $d_0 = 0$ (for $M_0 = \{0\}$) and $d_q = \dim X_0 = n$ (because $M_q = X$). Further $d_0 \le d_1 \le \cdots \le d_{q-1} \le d_q$ and two consecutive elements in this (finite) sequence differ at most one. Hence $\{d_0, \dots, d_q\} = \{0, \dots, n\}$. For $k = 0, \dots, n$, let

$$j(k) = \min\{j \mid j = 0, \dots, q;\, d_j = k\}, \qquad N_k = M_{j(k)}[0].$$

Then $N_0 \subset N_1 \subset \cdots \subset N_n$ is a complete chain of A_0-invariant subspaces. Introducing

$$j^\times(k) = \min\{j \mid j = 0, \dots, q;\, \dim M_j^\times[0] = k\}, \qquad N_k^\times = M_{j^\times(k)}^\times[0]$$

we obtain a complete chain $N_0^\times \subset N_1^\times \subset \cdots \subset N_n^\times$ of $A_0^\times$-invariant subspaces as well.

Step 5. Let M, N be a pair of subspaces of X, and consider the pair of associated subspaces $M[0], N[0]$ of X_0. In view of the matching conditions (12.6) we are interested in the cases $M \cap N = \{0\}$ and $M + N = X$. Here is what results from the inequalities (12.4) and (12.5) in Lemma 12.4. If $M \cap N = \{0\}$, then

$$\dim(M[0] \cap N[0]) \leq d_-, \tag{12.8}$$

where $d_- = \dim X_-$; if $M + N = X$, then

$$\dim\left(\frac{X_0}{M[0] + N[0]}\right) \leq d_+, \tag{12.9}$$

where $d_+ = \dim X_+$.

In line with the remark immediately following the proof of Lemma 12.4, we note that (12.8) and (12.9) are trivial when $\dim X_0 \leq d_- = \dim X_-$. This, however, is not the case here. Indeed, by assumption $q \leq 2n-2$, where $n = \dim X_0$, hence $\dim X_- = q - n - d_+ \leq n - 2$ and $\dim X_+ = q - n - d_- \leq n - 2$.

Step 6. Consider the inequality

$$\dim\left(N_l \cap N_k^\times\right) \leq d_-. \tag{12.10}$$

Because its left-hand side is bounded above by $\min\{l, k\}$, the inequality is certainly valid when l or k does not exceed d_-. The estimate, however, also holds in non-trivial cases. In fact, (12.10) is satisfied for $l, k = (d_- + 1), \ldots, n$ with $l + k \leq n + d_- - d_+$ (where it should be noted that $2(d_- + 1) = q - n - d_+ + d_- + 2 \leq 2n - 2 - n - d_+ + d_- + 2 = n + d_- - d_+$ and $d_- + 1 = q - n - d_+ + 1 \leq 2n - 2 - n + 1 = n - 1$). The reasoning, employing the matching conditions (12.6), runs this way.

Recall that

$$N_l = M_{j(l)}[0], \qquad N_k^\times = M_{j^\times(k)}^\times[0].$$

Now $M_{j(l)} \dotplus M_{q-j(l)}^\times = \mathbb{C}^q$. If $j(l) + j^\times(k) \leq q$, we have $M_{j(l)} \cap M_{j^\times(k)}^\times \subset M_{j(l)} \cap M_{q-j(l)}^\times = \{0\}$ and (12.10) follows from (12.8). When $j(l) + j^\times(k) \geq q$, we have $M_{j(l)} + M_{j^\times(k)}^\times \supset M_{j(l)} + M_{q-j(l)}^\times = X$, and (12.9) gives

$$n - \dim\left(M_{j(l)}[0] + M_{j^\times(k)}^\times[0]\right) \leq d_+.$$

Thus $\dim(N_l + N_k^\times) \geq n - d_+$. But then

$$\begin{aligned} \dim\left(N_l \cap N_k^\times\right) &= \dim N_l + \dim N_k^\times - \dim\left(N_l + N_k^\times\right) \\ &= l + k - \dim\left(N_l + N_k^\times\right) \\ &\leq l + k - n + d_+. \end{aligned}$$

As the last expression is bounded above by d_- when $l + k \leq n + d_- - d_+ =$, the desired inequality (12.10) follows.

Step 7. Recall that (12.7) is a minimal realization of W. The function W is companion based. Therefore, by the state space isomorphism theorem, we may assume that $X_0 = \mathbb{C}^n$ and that the matrix representations of A_0 and $A_0^\times = A_0 - B_0C_0$ with respect to the standard basis in $\mathbb{C}^n$ are first companions. As was indicated in Proposition 11.20, complete chains of invariant subspaces of first companion matrices can be described with the help of generalized Vandermonde matrices. This is the key for the rest of the argument.

Let $V(a_1, \dots, a_n)$ be the generalized Vandermonde matrix for the complete chain $N_0 \subset N_1 \subset \cdots \subset N_n$ of A_0-invariant subspaces. Thus $a_1, \dots, a_n$ is an appropriate ordering of the eigenvalues of A_0 (algebraic multiplicities counted) and, using the notation of Section 11.7,

$$N_l = \operatorname{span}\{\mathbf{v}_{\nu(1)}(a_1), \dots, \mathbf{v}_{\nu(l)}(a_l)\}, \qquad l = 0, \dots, n,$$

where $\nu(j)$ is the number of times that the eigenvalue a_j appears among its predecessors $a_1, \dots, a_{j-1}$. Similarly, let $V(a_1^\times, \dots, a_n^\times)$ be the generalized Vandermonde matrix for the complete chain $N_0^\times \subset N_1^\times \subset \cdots \subset N_n^\times$ of $A_0^\times$-invariant subspaces. So $a_1^\times, \dots, a_n^\times$ is a suitable ordering of the eigenvalues of $A_0^\times$ (algebraic multiplicities counted) and

$$N_k^\times = \operatorname{span}\{\mathbf{v}_{\nu^\times(1)}(a_1^\times), \dots, \mathbf{v}_{\nu^\times(k)}(a_k^\times)\}, \qquad k = 0, \dots, n,$$

where $\nu^\times(j)$ is the number of times that the eigenvalue $a_j^\times$ appears among the numbers $a_1^\times, \dots, a_{j-1}^\times$.

We are now ready to set things up for the application of the combinatorial lemma of the preceding section. Put $p = n - (d_+ + 1)$ and $r = d_- + 1$. Then $p \le n - 1$ and r is a positive integer not exceeding p (for $d_- + 1 = q - n - d_+ + 1 \le n - d_+ - 1$). For $l = r, \dots, p$ (hence $p + r - l$ in the same range) and $\alpha \in \mathcal{A} = \{a_1, \dots, a_p, a_1^\times, \dots, a_p^\times\}$, introduce

$$\begin{aligned} \nu_\alpha(l) &= \sharp\{j \mid j = 1, \dots, l;\ a_j = \alpha\}, \\ \nu_\alpha^\times(l) &= \sharp\{j \mid j = 1, \dots, l;\ a_j^\times = \alpha\}, \end{aligned}$$

and $\mu_\alpha(l) = \min\{\nu_\alpha(l), \nu_\alpha^\times(p + r - l)\}$. Note that $\sum_{\alpha\in\mathcal{A}} \mu_\alpha(l)$ is the number of common entries in $a_1, \dots, a_l$ and $a_1^\times, \dots, a_{p+r-l}^\times$ (multiplicities counted). Take $\alpha \in \mathcal{A}$. Then $\mathbf{v}_0(\alpha), \dots, \mathbf{v}_{\nu_\alpha(l)}(\alpha)$ appear among the first l columns of $V(a_1, \dots, a_n)$ which span N_l. Similarly, $\mathbf{v}_0(\alpha), \dots, \mathbf{v}_{\nu_\alpha^\times(p+r-l)}(\alpha)$ appear among the first $p+r-l$ columns of $V(a_1^\times, \dots, a_n^\times)$ which span $N_{p+r-l}^\times$. Thus the vectors

$$\mathbf{v}_0(\alpha), \dots, \mathbf{v}_{\mu_\alpha(l)}(\alpha), \qquad \alpha \in \mathcal{A}$$

belong to $N_l \cap N_{p+r-l}^\times$. These vectors, being different columns of $V(a_1, \dots, a_n)$ or, for that matter, of $V(a_1^\times, \dots, a_n^\times)$, are linearly independent. Their total number is $\sum_{\alpha\in\mathcal{A}} \mu_\alpha(l)$, and so

$$\sum_{\alpha\in\mathcal{A}} \mu_\alpha(l) \le \dim\big(N_l \cap N_{p+r-l}^\times\big).$$

The integers l and $p+r-l$ are in the range $r=(d_-+1)$ up to p, with p not exceeding n. Also, $l+(p+r-l)=p+r=n+d_--d_+$. Hence (12.10) holds with $k=p+r-l$ (see Step 6). It follows that $\sum_{\alpha\in\mathcal{A}}\mu_\alpha(l)\le d_-$, and we conclude that the number of common entries in $a_1,\dots,a_l$ and $a_1^\times,\dots,a_{p+r-l}^\times$ (multiplicities counted) is at most $d_-=r-1$.

Now apply Lemma 12.1. This gives two permutations σ and $\sigma^\times$ of $\{1,\dots,p\}$ for which

$$a_{\sigma(s)}\neq a^\times_{\sigma^\times(t)},\qquad s,t=1,\dots,p\,,\ \ s+t\le p+2-r. \tag{12.11}$$

We complete σ to a permutation of $\{1,\dots,n\}$ and do likewise with $\sigma^\times$. By slight abuse of notation, these extended permutations are again denoted by σ and $\sigma^\times$. For $k=1,\dots,n$, put

$$\alpha_k=a_{\sigma(k)},\qquad \alpha_k^\times=\alpha^\times_{\sigma^\times(n+1-k)}.$$

Clearly $\alpha_1,\dots,\alpha_n$ is an ordering of the eigenvalues of A_0 and $\alpha_1^\times,\dots,\alpha_n^\times$ is an ordering of the eigenvalues of $A_0^\times$, algebraic multiplicities taken into account. We claim that (12.3) is satisfied with $d=q$. Let $k,j\in\{1,\dots,n\}$, and assume $k<j+n-q$. Then

$$k+(n+1-j)\ \le\ 2n-q\ =\ n-(d_-+d_+)=p+2-r.$$

As $p+2-r\le p+1$, both k and $n+1-j$ do not exceed p. Taking $s=k$ and $t=n+1-j$ in (12.11), it follows that $\alpha_k\neq\alpha_j^\times$.

In closing we recall the fact that the eigenvalues of A_0 coincide with the poles of W and those of $A_0^\times$ coincide with the zeros of W, in both cases the appropriate multiplicities taken into account (cf., Chapter 8). □

Next, we turn to the second part of the proof of Theorem 12.2. First we establish some auxiliary results.

Proposition 12.5. *Let A be an $n\times n$ first companion matrix and let B be an $n\times m$ matrix. The following statements are equivalent:*

(i) *the pair (A,B) is controllable,*

(ii) *there exists an invertible $n\times n$ matrix T such that $AT=TA$ and, in addition,*

$$\begin{bmatrix}0\\ \vdots\\ 0\\ 1\end{bmatrix}\in \operatorname{Im}T^{-1}B.$$

The latter can be rephrased as: the last column of T is a linear combination of the columns of B.

Proof. The column vector appearing in the displayed formula in (ii) will be denoted by e_n. Since A is a first companion, the $n \times n$ matrix

$$\begin{bmatrix} e_n & Ae_n & \cdots & A^{n-1}e_n \end{bmatrix}$$

is invertible. Hence, given a vector $x \in \mathbb{C}^n$, there exist (unique) complex numbers $p_0, \ldots, p_{n-1}$ (depending on x) such that $\sum_{j=0}^{n-1} p_j A^j e_n = x$. The latter can be rewritten as $p(A)e_n = x$, where p is the scalar polynomial $p(\lambda) = p_0 + \lambda p_1 + \cdots + \lambda^{n-1} p_{n-1}$.

By assumption, the pair (A, B) is controllable, in particular $B \neq 0$. Let $x_1, \ldots, x_r$ be vectors in $\mathbb{C}^n$ that span $\operatorname{Im} B$. With $x_1, \ldots, x_r$, we associate scalar polynomials $p_1, \ldots, p_r$ in the way indicated above. Thus

$$p_j(A)e_n = x_j, \qquad j = 1, \ldots, r.$$

We claim that a common zero α of $p_1, \ldots, p_r$ can not be an eigenvalue of A. Suppose it is. For $j = 1, \ldots, r$, the polynomial p_j is divisible by the linear factor $\lambda - \alpha$, and so $\operatorname{Im} p_j(A) \subset \operatorname{Im}(A - \alpha I_n)$. Hence

$$x_j = p_j(A)e_n \in \operatorname{Im}(A - \alpha I_n), \qquad j = 1, \ldots, r,$$

and, as a consequence, $\operatorname{Im} B \subset \operatorname{Im}(A - \alpha I_n)$. Along with (A, B), the pair $(A - \alpha I_n, B)$ is controllable, and we conclude that $A - \alpha I_n$ has to be invertible. In other words, α is not an eigenvalue of A.

Let α be an eigenvalue of A. Then at least one of the complex numbers $p_1(\alpha), \ldots, p_r(\alpha)$ is nonzero. Hence the set of vectors $(\beta_1, \ldots, \beta_r)$ in $\mathbb{C}^r$ determined by

$$\sum_{k=1}^{r} \beta_k p_k(\alpha) \neq 0$$

is open and dense in $\mathbb{C}^r$. But then the (finite) intersection of these sets over all α in the spectrum of A is (open and) dense too. In particular, this intersection is nonempty. Thus there exist complex numbers $\beta_1, \ldots, \beta_r$ such that the polynomial $q = \sum_{j=1}^{r} \beta_j p_j$ does not vanish on the spectrum of A. Define $T = q(A)$. Then T is invertible and, of course, $AT = TA$. As,

$$Te_n = q(A)e_n = \sum_{j=1}^{r} \beta_j p_j(A)e_n = \sum_{j=1}^{r} \beta_j x_j \in \operatorname{Im} B,$$

the lemma is proved. □

Underlying the definition of quasicomplete factorization and quasidegree is Theorem 10.15. In the proof of that theorem, the spectral assignment theorem (Theorem 6.5.1 in [70]) is used. The next proposition is a specialization of the latter result to first companions.

Proposition 12.6. *Let A be an $n \times n$ first companion matrix, let B be an $n \times m$ matrix, and assume*

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \in \operatorname{Im} B. \tag{12.12}$$

Then, given complex numbers $\gamma_1, \ldots, \gamma_n$ (not necessarily distinct), there exists an $m \times n$ matrix K such that $A + BK$ is again a first companion and $\gamma_1, \ldots, \gamma_n$ are the eigenvalues of $A + BK$ (algebraic multiplicities taken into account).

Note that (12.12), together with the fact that A is a first companion matrix, implies that the pair (A, B) is controllable.

Proof. Write A in the form (11.1), and take $v \in \mathbb{C}^m$ such that $Bv = e_n$. Here e_n stands for the left-hand side of (12.12). Let $c_0, \ldots, c_{n-1}$ be the complex numbers determined by

$$\lambda^n + \sum_{j=0}^{n-1} c_j \lambda^j = (\lambda - \gamma_1) \cdots (\lambda - \gamma_n), \tag{12.13}$$

and introduce the $m \times n$ matrix K via

$$K = \begin{bmatrix} (a_0 - c_0)v & (a_1 - c_1)v & \cdots & (a_{n-1} - c_{n-1})v \end{bmatrix}.$$

Observe that

$$\begin{aligned} A + BK &= A + B\left(v \begin{bmatrix} (a_0 - c_0) & (a_1 - c_1) & \cdots & (a_{n-1} - c_{n-1}) \end{bmatrix}\right) \\ &= A + e_n \begin{bmatrix} (a_0 - c_0) & (a_1 - c_1) & \cdots & (a_{n-1} - c_{n-1}) \end{bmatrix}. \end{aligned}$$

Hence $A + BK$ is a first companion matrix. In fact, $A + BK$ is (11.1) with $a_0, \ldots, a_{n-1}$ replaced by $c_0, \ldots, c_{n-1}$. The characteristic polynomial of $A + BK$ is given by (12.13). Hence the eigenvalues of $A + BK$ are $\gamma_1, \ldots, \gamma_n$ (algebraic multiplicities taken into account). □

After these preparations, we are ready for the second part proof of the proof of Theorem 12.2.

Second part of the proof of Theorem 12.2. Let $d \geq n$ be an integer for which there exist an ordering $\alpha_1, \ldots, \alpha_n$ of the poles of W (pole-multiplicities taken into account) and an ordering $\alpha_1^\times, \ldots, \alpha_n^\times$ of the zeros of W (zero-multiplicities taken into account) such that (12.3) is satisfied, i.e.,

$$\alpha_k \neq \alpha_j^\times, \qquad k, j = 1, \ldots, n, \;\; k < j + n - d.$$

In this second part of the proof of Theorem 12.2, we show that $\delta_{\mathrm{q}}(W) \leq d$.

Note that the case $d = n$ is already covered by Theorem 11.17 (complete factorization, hence $\delta_{\mathrm{q}}(W) = \delta(W) = n$). So we may assume $d \geq n+1$. It may be assumed as well that $d \leq 2n-1$. This is clear from the remark made right after Theorem 12.2.

By assumption, W is companion based. Let $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ be a minimal realization of W such that A and $A^\times = A - BC$ are first companion matrices. Then $\alpha_1, \ldots, \alpha_n$ is an ordering of the eigenvalues of A and $\alpha_1^\times, \ldots, \alpha_n^\times$ is one for the eigenvalues of $A^\times$ (cf., Chapter 8). We shall now appropriately change the realization of W to suit our purpose.

First we shall show that without loss of generality it may be assumed that $\operatorname{Im} B$ contains the vector e_n, where e_n is the column vector in $\mathbb{C}^n$ given by the left-hand side of (12.12). When $A \neq A^\times$, the proof is simple. Indeed, the difference of the first companions A and $A^\times$ then has a nonzero column of which all entries are zero except the last. Hence

$$e_n \in \operatorname{Im}(A - A^\times) = \operatorname{Im} BC \subset \operatorname{Im} B,$$

and so we can even leave the realization as it is. In case $A = A^\times$, the argument is somewhat more involved. Let T be an invertible $n \times n$ matrix, and put $\widetilde{A} = T^{-1}AT$, $\widetilde{B} = T^{-1}B$ and $\widetilde{C} = CT$. Then

$$W(\lambda) = I_m + \widetilde{C}(\lambda I_n - \widetilde{A})^{-1}\widetilde{B}$$

is a minimal realization of W and $\widetilde{A}^\times = T^{-1}A^\times T = T^{-1}AT = \widetilde{A}$. As the pair (A, B) is controllable, we can apply Proposition 12.5 to choose T in such a way that all four matrices $\widetilde{A}$, A, $\widetilde{A}^\times$ and $A^\times$ are the same while, in addition, $e_n \in \operatorname{Im} \widetilde{B}$. Thus what we desire can be reached by replacing B by $T^{-1}B$ and C by CT.

From now on it is assumed that $e_n \in \operatorname{Im} B$, that is, condition (12.12) is satisfied. Let $\gamma_1, \ldots, \gamma_n$ be complex numbers such that $\gamma_1, \ldots, \gamma_{d-n}$ are distinct and outside the spectra of A and $A^\times$. Applying Proposition 12.6 we obtain an $m \times n$ matrix K such that $A + BK$ is a first companion having $\gamma_1, \ldots, \gamma_n$ as its eigenvalues. Introduce the $n \times (d-n)$ matrix

$$\begin{aligned} X &= \begin{bmatrix} \mathbf{v}_0(\gamma_1) & \mathbf{v}_0(\gamma_2) & \ldots & \mathbf{v}_0(\gamma_{d-n}) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & \ldots & 1 \\ \gamma_1 & \gamma_2 & \cdots & \gamma_{d-n} \\ \gamma_1^2 & \gamma_2^2 & \cdots & \gamma_{d-n}^2 \\ \vdots & \vdots & & \vdots \\ \gamma_1^{n-1} & \gamma_2^{n-1} & \cdots & \gamma_{d-n}^{n-1} \end{bmatrix}, \end{aligned} \tag{12.14}$$

and let G be the diagonal $(d-n) \times (d-n)$ matrix with diagonal elements $\gamma_1, \ldots, \gamma_{d-n}$. Then $(A + BK)X = XG$. With F equal to the $m \times (d-n)$ ma-

trix KX, the identity $(A+BK)X = XG$ transforms into the intertwining relation $XG - AX = BF$.

Introduce the matrices

$$\widehat{A} = \begin{bmatrix} A & BF \\ 0 & G \end{bmatrix}, \qquad \widehat{B} = \begin{bmatrix} B \\ 0 \end{bmatrix}, \qquad \widehat{C} = \begin{bmatrix} C & F \end{bmatrix}. \tag{12.15}$$

having dimensions $d \times d$, $d \times m$ and $m \times d$, respectively. Since

$$W(\lambda) = I_m + \widehat{C}(\lambda I_d - \widehat{A})^{-1}\widehat{B}$$

(cf., the material on dilation in Section 7.1), it suffices to prove that $\widehat{A}$ and $\widehat{A}^{\times} = \widehat{A} - \widehat{B}\widehat{C}$ admit simultaneous reduction to complementary triangular forms. Indeed, Theorem 10.5 then gives that W admits a factorization into d elementary factors, so that $\delta_{\mathrm{q}}(W) \leq d$ as desired. The approach we take is via chains of matching subspaces (cf., Section 10.1, in particular, Proposition 10.1; see also the proof of Theorem 10.5).

With the ordering $\alpha_1, \ldots, \alpha_n$ of the eigenvalues of the first companion A, we associate the generalized Vandermonde matrix $V = V(\alpha_1, \ldots, \alpha_n)$ as in Section 11.7. Then V is invertible and $V^{-1}AV$ is upper triangular with diagonal elements $\alpha_1, \ldots, \alpha_n$. Analogously, putting $V^{\times} = V(\alpha_n^{\times}, \ldots, \alpha_1^{\times})$, the matrix $(V^{\times})^{-1}A^{\times}V^{\times}$ is upper triangular with diagonal elements $\alpha_n^{\times}, \ldots, \alpha_1^{\times}$. For clarity, we emphasize that the eigenvalues of $A^{\times}$ have been taken here in the (reversed) order $\alpha_n^{\times}, \ldots, \alpha_1^{\times}$. We will now construct a complete chain of invariant subspaces for $\widehat{A}$, and one for $\widehat{A}^{\times}$ as well. Let us consider $\widehat{A}$ first.

Recall from Section 11.7 that $V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$ with

$$v_j = \mathbf{v}_{\nu(j)}(\alpha_j), \qquad j = 1, \ldots, n,$$

where $\nu(j)$ is the number of times that the eigenvalue α_j appears among its predecessors $\alpha_1, \ldots, \alpha_{j-1}$. For $j = 1, \ldots, d-n$, let e_j be the vector in $\mathbb{C}^{d-n}$ having 1 in the jth position and zeros everywhere else. We now introduce the vectors $a_1, \ldots, a_{d-n}, a_{d-n+1}, \ldots, a_d \in \mathbb{C}^d = \mathbb{C}^n \dotplus \mathbb{C}^{d-n}$ as follows:

$$a_j = \begin{cases} \begin{bmatrix} \mathbf{v}_0(\gamma_j) \\ e_j \end{bmatrix}, & j = 1, \ldots, d-n, \\ \begin{bmatrix} v_{j+n-d} \\ 0 \end{bmatrix}, & j = d-n+1, \ldots, d. \end{cases} \tag{12.16}$$

As is easily seen, these vectors form a basis for $\mathbb{C}^d$. With respect to this basis,

$\widehat{A}$ has upper triangular form. Indeed, for $j = 1, \ldots, d-n$, we have

$$\begin{aligned}\widehat{A}a_j &= \begin{bmatrix} A\mathbf{v}_0(\gamma_j) + BFe_j \\ Ge_j \end{bmatrix} = \begin{bmatrix} A\mathbf{v}_0(\gamma_j) + BKXe_j \\ \gamma_j e_j \end{bmatrix} \\ &= \begin{bmatrix} (A+BK)\mathbf{v}_0(\gamma_j) \\ \gamma_j e_j \end{bmatrix} = \begin{bmatrix} \gamma_j \mathbf{v}_0(\gamma_j) \\ \gamma_j e_j \end{bmatrix} \\ &= \gamma_j a_j.\end{aligned}$$

Also, for $j = d-n+1, \ldots, d$,

$$\begin{aligned}\widehat{A}a_j = \begin{bmatrix} Av_{j+n-d} \\ 0 \end{bmatrix} &\in \operatorname{span}\left\{ \begin{bmatrix} v_1 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} v_{j+n-d} \\ 0 \end{bmatrix} \right\} \\ &= \operatorname{span}\left\{a_{d-n+1}, \ldots, a_j\right\}.\end{aligned}$$

Here we used that $V^{-1}AV$ is upper triangular. Now put

$$\widehat{M}_k = \operatorname{span}\left\{a_1, \ldots, a_k\right\}, \qquad k = 0, \ldots, d. \tag{12.17}$$

Then $\{0\} = \widehat{M}_0 \subset \widehat{M}_1 \subset \widehat{M}_2 \subset \cdots \subset \widehat{M}_{d-1} \subset \widehat{M}_d = \mathbb{C}^d$ is a complete chain of $\widehat{A}$-invariant subspaces.

For $\widehat{A}^\times = \widehat{A} - \widehat{B}\widehat{C}$ the construction is analogous. Note that

$$\widehat{A}^\times = \begin{bmatrix} A & BF \\ 0 & G \end{bmatrix} - \begin{bmatrix} B \\ 0 \end{bmatrix} \begin{bmatrix} C & F \end{bmatrix} = \begin{bmatrix} A^\times & 0 \\ 0 & G \end{bmatrix}.$$

Let $V^\times$ be the $n \times n$ matrix $V^\times = \begin{bmatrix} v_n^\times & v_{n-1}^\times & \cdots & v_1^\times \end{bmatrix}$ with

$$v_j^\times = \mathbf{v}_{\nu^\times(j)}(\alpha_j^\times), \qquad j = 1, \ldots, n,$$

where $\nu^\times(j)$ is the number of times that the eigenvalue $\alpha_j^\times$ appears among the numbers $\alpha_n^\times, \ldots, \alpha_{j+1}^\times$. Furthermore, set

$$a_j^\times = \begin{cases} \begin{bmatrix} 0 \\ e_j \end{bmatrix}, & j = 1, \ldots, d-n, \\ \begin{bmatrix} v_{d+1-j}^\times \\ 0 \end{bmatrix}, & j = d-n+1, \ldots, d. \end{cases} \tag{12.18}$$

Then $a_1^\times, \ldots, a_d^\times$ is a basis for $\mathbb{C}^d$, and with respect to this basis $\widehat{A}^\times$ has upper triangular form. Hence, with

$$\widehat{M}_k^\times = \operatorname{span}\left\{a_1^\times, \ldots, a_k^\times\right\}, \qquad k = 0, \ldots, d, \tag{12.19}$$

we have that $\{0\} = \widehat{M}_0^{\times} \subset \widehat{M}_1^{\times} \subset \widehat{M}_2^{\times} \subset \cdots \subset \widehat{M}_{d-1}^{\times} \subset \widehat{M}_d^{\times} = \mathbb{C}^d$ is a complete chain of $\widehat{A}^{\times}$-invariant subspaces.

We need to prove that

$$\widehat{M}_k \dotplus \widehat{M}_{d-k}^{\times} = \mathbb{C}^d, \qquad k = 1, \ldots, d-1. \tag{12.20}$$

It is convenient to distinguish three cases, depending on the value of k. Recall here that we assumed $n+1 \leq d \leq 2n-1$. From these inequalities we see that $1 \leq d-n < n \leq d-1$.

Case 1. Let $1 \leq k \leq d-n$. In this case $\widehat{M}_k$ is spanned by the k vectors

$$\begin{bmatrix} \mathbf{v}_0(\gamma_1) \\ e_1 \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{v}_0(\gamma_k) \\ e_k \end{bmatrix}, \tag{12.21}$$

and $\widehat{M}_{d-k}^{\times}$ by the $d-k$ vectors

$$\begin{bmatrix} 0 \\ e_1 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ e_{d-n} \end{bmatrix}, \begin{bmatrix} v_n^{\times} \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} v_{k+1}^{\times} \\ 0 \end{bmatrix}. \tag{12.22}$$

Note here that the assumption $k \leq d-n$, taken together with $d-n < n$, implies $d-k > d-n$. The vectors $v_n^{\times}, \ldots, v_{k+1}^{\times}$ are the first $n-k$ columns in the generalized Vandermonde matrix $V^{\times} = V(\alpha_n^{\times}, \ldots, \alpha_1^{\times})$. Also $\gamma_1, \ldots, \gamma_k$ do not appear among the numbers $\alpha_n^{\times}, \ldots, \alpha_{k+1}^{\times}$. Thus

$$\begin{bmatrix} \mathbf{v}_0(\gamma_1) & \cdots & \mathbf{v}_0(\gamma_k) & v_n^{\times} & \cdots & v_{k+1}^{\times} \end{bmatrix}$$

is a matrix of generalized Vandermonde type, and so the n vectors

$$\mathbf{v}_0(\gamma_1), \ldots, \mathbf{v}_0(\gamma_k), v_n^{\times}, \ldots, v_{k+1}^{\times}$$

are linearly independent. As the same is true for $e_1, \ldots, e_{d-n}$, the d vectors given by (12.21) and (12.22) together are linearly independent, and (12.20) is indeed satisfied.

Case 2. Let $d-n < k < n$. In this situation $\widehat{M}_k$ is spanned by the k vectors

$$\begin{bmatrix} \mathbf{v}_0(\gamma_1) \\ e_1 \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{v}_0(\gamma_{d-n}) \\ e_{d-n} \end{bmatrix}, \begin{bmatrix} v_1 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} v_{n+k-d} \\ 0 \end{bmatrix}, \tag{12.23}$$

and $\widehat{M}_{d-k}^{\times}$ by the $d-k$ vectors

$$\begin{bmatrix} 0 \\ e_1 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ e_{d-n} \end{bmatrix}, \begin{bmatrix} v_n^{\times} \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} v_{k+1}^{\times} \\ 0 \end{bmatrix}. \tag{12.24}$$

Note that the assumption $k < n$ is the same as $d - k > d - n$. As before, the vectors $v_n^\times, \ldots, v_{k+1}^\times$ are the first $n - k$ columns in the generalized Vandermonde matrix $V^\times = V(\alpha_n^\times, \ldots, \alpha_1^\times)$. Similarly, the vectors $v_1, \ldots, v_{n+k-d}$ are the first $n + k - d$ columns in $V = V(\alpha_1, \ldots, \alpha_n)$. Also, the condition (12.3) implies that the sets $\{\alpha_{k+1}^\times, \ldots, \alpha_n^\times\}$ and $\{\alpha_1, \ldots, \alpha_{n+k-d}\}$ are disjoint. Finally, the numbers $\gamma_1, \ldots, \gamma_{d-n}$ do not appear among the numbers

$$\alpha_1, \ldots, \alpha_{n+k-d}, \alpha_{k+1}^\times, \ldots, \alpha_n^\times.$$

Thus the matrix

$$\begin{bmatrix} \mathbf{v}_0(\gamma_1) & \cdots & \mathbf{v}_0(\gamma_{d-n}) & v_1 & \cdots & v_{n+k-d} & v_n^\times & \cdots & v_{k+1}^\times \end{bmatrix}$$

is of generalized Vandermonde type, and hence the n vectors

$$\mathbf{v}_0(\gamma_1), \ldots, \mathbf{v}_0(\gamma_{d-n}), v_1, \ldots, v_{n+k-d}, v_n^\times, \ldots, v_{k+1}^\times$$

are linearly independent. As the same is true for $e_1, \ldots, e_{d-n}$, the d vectors given by (12.23) and (12.24) together are linearly independent, and we conclude again that (12.20) holds.

Case 3. Let $n \le k \le d - 1$. Now $\widehat{M}_k$ is spanned by the k vectors

$$\begin{bmatrix} \mathbf{v}_0(\gamma_1) \\ e_1 \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{v}_0(\gamma_{d-n}) \\ e_{d-n} \end{bmatrix}, \begin{bmatrix} v_1 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} v_{n+k-d} \\ 0 \end{bmatrix}, \tag{12.25}$$

and $\widehat{M}_{l-k}^\times$ by the $d - k$ vectors

$$\begin{bmatrix} 0 \\ e_1 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ e_{d-k} \end{bmatrix}. \tag{12.26}$$

Note that in the present case $k > d - n$ and $d - k \le d - n$. To prove that the d vectors given by (12.25) and (12.26) together are linearly independent, it suffices to show that the linear independency condition is satisfied for the vectors

$$\begin{bmatrix} \mathbf{v}_0(\gamma_1) \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{v}_0(\gamma_{d-k}) \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{v}_0(\gamma_{d-k+1}) \\ e_{d-k+1} \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{v}_0(\gamma_{d-n}) \\ e_{d-n} \end{bmatrix},$$

$$\begin{bmatrix} v_1 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} v_{n+k-d} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ e_1 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ e_{d-k} \end{bmatrix}.$$

Now $e_1, \ldots, e_{d-n}$ are linearly independent, and what we have to establish is the linear independence of $\mathbf{v}_0(\gamma_1), \ldots, \mathbf{v}_0(\gamma_{d-k}), v_1, \ldots, v_{n+k-d}$. As $\gamma_{d-k+1}, \ldots, \gamma_{d-n}$ do not appear among $\alpha_1, \ldots, \alpha_{n+k-d}$, the vectors in question form again a matrix of generalized Vandermonde type, and once more we conclude that (12.20) is satisfied. □

Theorem 12.2 can be reformulated a follows.

Theorem 12.7. *Let W be a companion based rational $m \times m$ matrix function and let n be the McMillan degree of W (assumed to be positive in order to avoid trivialities). Furthermore, let $\alpha_1, \ldots, \alpha_n$ be an arbitrary ordering of the poles of W (pole-multiplicities counted), and let $\alpha_1^\times, \ldots, \alpha_n^\times$ be an arbitrary ordering of the zeros of W (zero-multiplicities counted). Then*

$$\delta_{\mathrm{q}}(W) \;=\; n + \min_{\sigma,\tau \in \mathcal{S}(n)} \max\left\{ j-k \;\mid\; k<j,\; \alpha_{\sigma(k)} = \alpha^\times_{\tau(j)} \right\}, \tag{12.27}$$

where $\mathcal{S}(n)$ stands for the collection of all permutations of $\{1, \ldots, n\}$ and $\max \emptyset$ is defined to be zero.

Proof. By Theorem 12.2 there are permutations $\sigma, \tau \in \mathcal{S}(n)$ such that

$$\alpha_{\sigma(k)} \neq \alpha^\times_{\tau(j)}, \qquad k,j = 1, \ldots, n,\;\; k < j - (\delta_{\mathrm{q}} - n),$$

where $\delta_{\mathrm{q}} = \delta_{\mathrm{q}}(W)$. For these permutations we clearly have

$$\max\left\{ j-k \;\mid\; k<j,\; \alpha_{\sigma(k)} = \alpha^\times_{\tau(j)} \right\} \;\leq\; \delta_{\mathrm{q}} - n.$$

Hence the right-hand side in (12.27) does not exceed the left-hand side.

Now conversely. Let $\sigma, \tau \in \mathcal{S}(n)$, and put

$$m_{\sigma,\tau} \;=\; \max\left\{ j-k \;\mid\; k<j,\; \alpha_{\sigma(k)} = \alpha^\times_{\tau(j)} \right\}.$$

If $\alpha_{\sigma(k)} = \alpha^\times_{\tau(j)}$, then either $k < j$ and $j - k \leq m_{\sigma,\tau}$ or $k \geq j$. In the latter situation, we have $j - k \leq m_{\sigma,\tau}$ too, because $m_{\sigma,\tau}$ is non-negative (as $\max \emptyset$ is zero by definition). So $\alpha_{\sigma(k)} = \alpha^\times_{\tau(j)}$ implies $j - k \leq m_{\sigma,\tau}$ which one may also write as

$$\alpha_{\sigma(k)} \neq \alpha^\times_{\tau(j)}, \qquad k,j = 1, \ldots, n,\;\; k < j - m_{\sigma,\tau}.$$

Theorem 12.2 now gives $n + m_{\sigma,\tau} \geq \delta_{\mathrm{q}}(W)$, and it follows that the left-hand side of (12.27) does not exceed the right-hand side. □

Theorems 12.2 and 12.7 are stated in terms of poles and zeros of the given function W. Clearly they can also be formulated in terms of realizations (cf., the remark made after the proof of Theorem 11.17). The following result is phrased along this line. It is a counterpart (in fact, a generalization) of Theorem 11.18.

Theorem 12.8. *Let W be a companion based rational $m \times m$ matrix function, and let $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ be a minimal realization of W, so that n is the McMillan degree of W (assumed to be positive in order to avoid trivialities). Then the quasidegree $\delta_{\mathrm{q}}(W)$ of W is the smallest integer d larger than or equal to n for which there exists an ordering $\mu_1, \ldots, \mu_s$ of the (different) elements of $\sigma(A) \cup \sigma(A^\times)$ such that*

$$\sum_{i=1}^{t} m_{A^\times}(\mu_i) \;\leq\; (d-n) + 1 + \sum_{i=1}^{t-1} m_A(\mu_i), \qquad t = 1, \ldots, s.$$

Proof. From the proof of Theorem 11.8 (take $h = d - n + 1$ there), we see that the above requirement on the (algebraic multiplicities of the) eigenvalues of A and $A^\times$ is equivalent to the existence of orderings of the type mentioned in Theorem 12.2. Recall in this connection that the eigenvalues of A and $A^\times = A - BC$ correspond to the poles and zeros of W, respectively (the appropriate multiplicities taken into account). □

For an integer k, we let $k_+ = \frac{1}{2}(k + |k|)$. In other word, k_+ is the maximum of k and zero.

Theorem 12.9. *Let W be a companion based rational $m \times m$ matrix function, and let $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ be a minimal realization of W, so that n is the McMillan degree of W (assumed to be positive in order to avoid trivialities). Furthermore, let $\mu_1, \ldots, \mu_s$ be an arbitrary ordering of the (different) elements of $\sigma(A) \cup \sigma(A^\times)$. Then*

$$\delta_{\mathrm{q}}(W) \;=\; n + \min_{\sigma \,\in\, \mathcal{S}(s)} \max_{t=1,\ldots,s} \left(-1 + \sum_{i=1}^{t} m_{A^\times}\left(\mu_{\sigma(i)}\right) \;-\; \sum_{i=1}^{t-1} m_A\left(\mu_{\sigma(i)}\right) \right)_+ ,$$

where $\mathcal{S}(s)$ stands for the collection of all permutations of $\{1, \ldots, s\}$.

Proof. The proof is similar to that of Theorem 12.8 but based on Theorem 12.7 instead of Theorem 12.2. □

Theorems 12.7 and 12.9 suggest that calculating the quasidegree is a task of high computational complexity. No matter how this may be in general, for the class of companion based rational matrix functions to which the theorems apply, the computational complexity is actually very low (assuming that its poles and zeros are known). The key to this is a connection with the theory of job scheduling which will be made in the next two sections.

12.3 A review of the two machine flow shop problem

In this section we introduce the two machine flow shop problem and review some related results.

The *two machine flow shop problem* – 2MFSP for short – is concerned with two machines, written $\mathbf{M}_1$ and $\mathbf{M}_2$, and a number of jobs, indexed by the integers $1, \ldots, k$ say. The jobs have to be processed by the two machines. Each job j involves (at most) two operations: a (possible) first operation O_j^1 to be processed on the first machine $\mathbf{M}_1$, and a (possible) second operation O_j^2 to be processed on the second machine $\mathbf{M}_2$. Each machine can be processing at most one operation at the same time. In standard 2MFSP, it is required that for every job j processing O_j^2 on $\mathbf{M}_2$ cannot start until processing O_j^1 on $\mathbf{M}_1$ has been completed. In non-standard versions of 2MFSP other constraints may be imposed.

The processing times of the operations are given and fixed. That of O_j^1 is denoted by s_j, and the processing time of O_j^2 is denoted by t_j. Hence, formally, an instance J of 2MFSP involving k jobs consists of k tuples

$$(s_1, t_1),\ \dots\ , (s_k, t_k) \tag{12.28}$$

specifying the processing times of the operations. Of course these processing times are taken to be non-negative numbers. As already suggested above, we do allow for the possibility that one of the numbers s_j, t_j is zero, meaning that the job indexed j does not require the machine in question ($\mathbf{M}_1$ when $s_j = 0$, $\mathbf{M}_2$ when $t_j = 0$). However, in order to avoid trivialities, we assume that for each j, either s_j or t_j is nonzero (i.e., for each job something has to be done). There is another assumption that we will adopt in the present exposition, namely that the processing times are integers. In practical situations, they will usually be rational numbers which can be made into integers by an appropriate choice of the time unit.

A *schedule* for J is a rule indicating in what order the jobs are carried out on the two machines. This is an informal definition which can be made precise by using two functions (one for $\mathbf{M}_1$ and one for $\mathbf{M}_2$) mapping a time interval into the set $\{1, \dots, k\}$ indexing the collection of jobs. We refrain from burdening the discussion with the details. A schedule is said to be *feasible* if it satisfies the specified constraints. The length of the time interval required to carry out all jobs is called the *makespan* of the schedule. Of course such a makespan is always larger than or equal to the maximum of the numbers

$$s(J) = \sum_{j=1}^{k} s_j, \qquad t(J) = \sum_{j=1}^{k} t_j,$$

where it is assumed that J is given by (12.28). In the versions of 2MFSP considered here (including the standard one), the objective is to find an *optimal schedule*, that is a feasible schedule with smallest possible makespan, the so-called *minimum makespan*.

To give a feel for what is going on, let us first concentrate on the standard version of 2MFSP and indicate some properties of the optimal schedules for this case. The minimum makespan of an instance J of standard 2MFSP will be denoted by $\mu(J)$. Clearly

$$\max\{s(J),\, t(J)\} \ \le\ \mu(J)\ \le\ s(J) + t(J).$$

Also, the hypotheses that all processing times are integers, implies that $\mu(J)$ is an integer too. Indeed, if ϵ is a number strictly between 0 and 1 such that $\mu(J) - \epsilon$ is integer, there would exist a feasible schedule with makespan $\mu(J) - \epsilon$, strictly smaller than $\mu(J)$, which is impossible.

Next, let us turn to some less trivial observations. It is known that each instance J of standard 2MFSP has an optimal *non-preemptive schedule* (see the

textbook [6]). By this we mean that the optimal schedule has the additional property that, once a machine has started processing an operation, it does not start processing another operation until the one it has begun working on has been completed. Additionally it may be assumed that once a machine has been activated, it works uninterrupted until all the operations to be carried out on the machine in question have been completed. This can be achieved by appropriately shifting the jobs on $\mathbf{M}_1$ to the left (i.e., backward in time) and those on $\mathbf{M}_2$ to the right (i.e., forward in time). In this way, $\mathbf{M}_1$ is occupied during the time interval from 0 to $s(J)$, while $\mathbf{M}_2$ is occupied during the time interval from $\mu(J) - t(J)$ to $\mu(J)$ with $s(J)$, $t(J)$ and $\mu(J)$ as above.

A schedule is a *permutation schedule* if it is non-preemptive and for all $i \neq j$ with strictly positive processing times s_i, t_i, s_j and t_j, the operation O_i^2 is processed before the operation O_j^2 on the second machine $\mathbf{M}_2$ if (and only if) the operation O_i^1 is processed before the operation O_j^1 on the first machine $\mathbf{M}_1$. Thus the order of the operations on the first machine is the same as the order of the operations on the second machine. It is known that the optimal schedule can be chosen to be a permutation schedule.

Again these definitions can be formalized by using functions mapping a time interval into the set of integers indexing the collection of jobs, but for our purposes here, it is not necessary to do so. Also we refrain from giving proofs of the observations contained in the preceding paragraph. In fact, using the type of arguments employed in the proof of Theorem 11.8, any schedule can be transformed into a permutation schedule without increasing the makespan.

An optimal permutation schedule for an instance of 2MFSP can be obtained by the application of *Johnson's rule* (see Johnson [82]; also [6]). According to this algorithm, an optimal schedule can be constructed as follows:

Step 1: Introduce $V_1 = \{j \mid s_j < t_j\}$ and $V_2 = \{j \mid s_j \geq t_j\}$.

Step 2: Put the jobs in V_1 in order of increasing processing time (s_j) on $\mathbf{M}_1$, and put the jobs in V_2 in order of decreasing processing time (t_j) on $\mathbf{M}_2$.

Step 3: Process the jobs in V_1 first, and those in V_2 thereafter.

The running time of Johnson's rule is $\mathcal{O}(k \log k)$. Thus 2MFSP belongs to the class of tractable problems that can be solved in polynomial time (cf., Garey and Johnson [44]).

Example. Let J be an instance of 2MPSP involving 6 jobs (so $k = 6$), the tuples specifying the processing times being

$$(s_1, t_1) = (0, 1), \quad (s_2, t_2) = (2, 0), \quad (s_3, t_3) = (3, 4),$$

$$(s_4, t_4) = (1, 1), \quad (s_5, t_5) = (2, 3), \quad (s_6, t_6) = (2, 1).$$

Note that job 1 does not require any action on $\mathbf{M}_1$ and job 2 not on $\mathbf{M}_2$. Further $s(J) = t(J) = 10$ (cf., the standing assumption introduced at the end of the one but last paragraph of this section). Combining Steps 1 and 2 in Johnson's rule, we obtain $V_1 = \{1, 5, 3\}$ and $V_2 = \{4, 6, 2\}$. An optimal permutation schedule is now obtained by processing the jobs in the order $1, 5, 3, 4, 6, 2$ (Step 3); schematically:

$\mathbf{M}_2$:		$*$	1	5	5	5	3	3	3	3	4	6	(2)
$\mathbf{M}_1$:	(1)	5	5	3	3	3	4	6	6	2	2	$*$	
Time :		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	

Here a job number between parenthesis means that for that job no action is required (processing time zero) on the machine in question, and a star indicates that the machine is idle. We conclude that the minimum makespan $\mu(J)$ of this particular instance J of the 2MFSP is equal to 11. The specifics (i.e., the processing times) of this example are inspired by the material contained in the example given at the end of Section 11.2. To facilitate the comparison of the above scheme with (11.29) from the earlier example, the schedule for $\mathbf{M}_2$ has been put on top and that for $\mathbf{M}_1$ at the bottom.

In the above example, the sum of the processing times on the two machines is the same: $s(J) = t(J)$. As far as the minimum makespan is concerned, this equality may be assumed without loss of generality. To see this, consider an instance J of standard 2MFSP, given by (12.28), and assume $s(J)$ and $t(J)$ do not coincide. We now augment J to another instance of 2MFSP by adding a "dummy job" as follows: the job listed (s_{k+1}, t_{k+1}) with

$$s_{k+1} = \sum_{j=1}^{k} (t_j - s_j), \qquad t_{k+1} = 0$$

in case $s(J) < t(J)$, the job listed (s_0, t_0) with

$$s_0 = 0, \qquad t(0) = \sum_{j=1}^{k} (s_j - t_j)$$

in case $s(J) > t(J)$. The instance J_{ext} of 2MFSP obtained this way meets the desired condition $s(J_{\text{ext}}) = t(J_{\text{ext}})$ and is essentially identical to J, satisfying $\mu(J) = \mu(J_{\text{ext}})$ in particular.

So far about the standard 2MFSP. As was already indicated, there are variants of 2MFSP. One of them, actually very closely related to the standard version, is especially appropriate for making the connection with the phenomenon of quasicomplete factorization discussed earlier in this chapter. It is what we shall call the Reduced two machine flow shop problem which we will describe in a moment. But first let us put things in a somewhat wider framework.

The non-standard versions of the two machine flow shop problem come about by relaxation of the predecessor constraints. Thus it is allowed that the processing of O_j^2 already starts before that of O_j^1 has been completed, resulting in an *infeasibility* $\max\{0, F(O_j^1) - S(O_j^2)\}$, where $F(O_j^1)$ denotes the finish time of operation O_j^1 on $\mathbf{M}_1$ and $S(O_j^2)$ stands for the start time of O_j^2 on $\mathbf{M}_2$. Here, of course, only those jobs are taken into account for which the processing times on both machines are positive. The infeasibilities introduced in this way should now be minimized in some prescribed sense. For standard 2MFSP, the requirement is that they are all zero, but there is a variety of other possibilities (see [22] and the references given therein).

Staying still very close to the standard version of 2MFSP, one can require that the infeasiblities do not surpass a given threshold τ. In other words, instead of the standard predecessor restriction we have the relaxed constraint: for each job j in the given instance of 2MFSP, processing O_j^2 on $\mathbf{M}_2$ cannot start until τ time units before processing O_j^1 on $\mathbf{M}_1$ has been completed. An optimal schedule is then obtained by taking one for standard 2MFSP and shifting the jobs on $\mathbf{M}_2$ backwards over a time interval of length $\min\{\tau, \mu(J) - t(J)\}$, resulting in a minimum makespan $\max\{s(J), t(J), \mu(J) - \tau\}$. As a consequence, an optimal schedule can be obtained via Johnson's rule of low computational complexity. Note that for this variant of 2MFSP, the instances J and J_{ext} (see the paragraph directly following the example) are again essentially identical, their minimum makespans coinciding in particular. Thus, from now on, we adopt as a standing assumption that the sum of the processing times on the two machines is the same: $s(J) = t(J)$.

We now specialize to the situation pertinent to the connection with quasicomplete factorization, the case $\tau = 1$ where it is required that non of the infeasibilities exceeds 1. This version of 2MFSP will be named *reduced two machine flow shop problem* – abbreviated $2\text{MFSP}_{\text{red}}$. We shall denote the minimum makespan of an instance J of $2\text{MFSP}_{\text{red}}$ by $\mu_{\text{red}}(J)$ and call it the *reduced minimal makespan* of J. For later use, we explicitly record that

$$\mu_{\text{red}}(J) \;=\; \max\{v(J),\, \mu(J) - 1\}, \tag{12.29}$$

where $v(J) = s(J) = t(J)$. Hence $v(J) \leq \mu_{\text{red}}(J) \leq 2v(J) - 1$. Again it may be assumed that once a machine has been activated, it works uninterrupted until all the operations to be carried out on it have been completed. In that case $\mathbf{M}_1$ is occupied during the time interval from 0 to $v(J)$, while $\mathbf{M}_2$ is occupied during the time interval from $\mu_{\text{red}}(J) - v(J)$ to $\mu_{\text{red}}(J)$. Since the processing times are integers, both $v(J)$ and $\mu(J)$ are integers as well. Also $\mu(J) \geq v(J)$. Thus $\mu_{\text{red}}(J) = v(J)$ when $\mu(J) = v(J)$, and $\mu_{\text{red}}(J) = \mu(J) - 1$ when $\mu(J) \neq v(J)$. It is also worthwhile to recall from the previous paragraph that $2\text{MFSP}_{\text{red}}$ has the same (low!) computational complexity as standard 2MFSP.

Finally, for the example presented above, now considered as an instance of $2\text{MFSP}_{\text{red}}$, we have the optimal permutation schedule

$\mathbf{M}_2$:		1	5	5	5	3	3	3	3	4	6	(2)
$\mathbf{M}_1$:	(1)	5	5	3	3	3	4	6	6	2	2	
Time :		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	

corroborating the identity (12.29): $\mu_{\text{red}}(J) = \mu(J) - 1 = 10$. As before, a job number between parenthesis means that for that job no action is required (processing time zero) on the machine in question.

12.4 Quasicomplete factorization and the 2MSFP

This section is devoted to the connection of the two machine flow shop problem (2MFSP) with quasicomplete factorization. The discussion will draw heavily upon the material on companion based rational matrix functions presented in Sections 11.4 and 12.2.

First we indicate how a companion based matrix function can be associated with an instance of 2MFSP and vice versa. Let W be a companion based $m \times m$ matrix function, of positive McMillan degree to avoid trivialities, and let J be an instance of 2MFSP, determined by (12.28) and satisfying the standing assumptions formulated above. Thus, for $j = 1, \ldots, k$, the processing times s_j and t_j are non-negative integers, not both vanishing, and the sum of the processing times on the two machines $\mathbf{M}_1$ and $\mathbf{M}_2$ is the same. As before, these coinciding sums will be denoted by $v(J)$, so,

$$v(J) = \sum_{j=1}^{k} s_j = \sum_{j=1}^{k} t_j. \tag{12.30}$$

We say that the companion based function W and the instance J of 2MFSP are *associated* if the pole-polynomial p of W and the zero-polynomial $p^{\times}$ of W can be written in the form

$$p(\lambda) = (\lambda - \beta_1)^{t_1}(\lambda - \beta_2)^{t_2} \cdots (\lambda - \beta_k)^{t_k}, \tag{12.31}$$

$$p^{\times}(\lambda) = (\lambda - \beta_1)^{s_1}(\lambda - \beta_2)^{s_2} \cdots (\lambda - \beta_k)^{s_k}, \tag{12.32}$$

with $\beta_1, \ldots, \beta_k$ different complex numbers. Note that in this definition all three standing assumptions have a role. First, the processing times s_j and t_j must be integers in view of (12.31) and (12.32). Second, the pole-polynomial p and the zero-polynomial $p^{\times}$ need to have the same degree, and this is guaranteed by (12.30). Indeed, the common degree of p and $p^{\times}$ is $v(J) = s(J) = t(J)$ and coincides with the McMillan degree $\delta(W)$ of W. Third, the number k, i.e., the number of jobs in J, is also the number of different elements is the union of the set of poles of W

and the set of zeros of W. Here it is used that for each job in J at least one of the processing times is nonzero.

For a given companion based matrix function W (of positive McMillan degree) there exists an instance J of 2MFSP such that W and J are associated. To see this, write the pole and zero-polynomial of W in the form (12.31) and (12.32), respectively, and take for J the instance of 2MFSP given by (12.28). This instance of 2MFSP is uniquely determined by W up to the ordering of the jobs in (12.28), an irrelevant feature from the point of view of job scheduling. Conversely, if J is an instance of 2MFSP with k jobs as in the preceding paragraph, then there do exist companion based matrix functions W such that W and J are associated. This can be seen as follows. First, choose k different complex numbers $\beta_1, \ldots, \beta_k$ (for example $\beta_j = j,\ j = 1, \ldots, k$). Next, introduce the polynomials $p(\lambda) = (\lambda - \beta_1)^{t_1}(\lambda - \beta_2)^{t_2} \cdots (\lambda - \beta_k)^{t_k}$ and $p^\times(\lambda) = (\lambda - \beta_1)^{s_1}(\lambda - \beta_2)^{s_2} \cdots (\lambda - \beta_k)^{s_k}$. Finally, define the 2×2 rational matrix function W by

$$W(\lambda) = \begin{bmatrix} 1 & \dfrac{1}{p(\lambda)} \\ 0 & \dfrac{p^\times(\lambda)}{p(\lambda)} \end{bmatrix}. \tag{12.33}$$

Then W is a companion based with pole-polynomial p and zero-polynomial $p^\times$ (see Section 11.4). Hence W and J are associated.

The function W in (12.33) is completely determined by the given instance J of 2MFSP. There are, however, more possibilities to produce a companion based function associated with J. For example, if T is any invertible 2×2 matrix, then $T^{-1}WT$ and J are associated as well. Still other possibilities are provided by the material in Section 11.4. In any case, if J is an instance of 2MFSP (satisfying our standing assumptions), there exist several companion based matrix functions W such that W and J are associated. However, all these functions have basically the same factorization properties. So, from a factorization point of view, the differences between them are irrelevant and in this relaxed sense, we have uniqueness here as well.

After these preparations, we come to the result we have been aiming at.

Theorem 12.10. *Let W be a companion based rational matrix function of positive McMillan degree, let J be an instance of the two machine flow shop problem, and suppose W and J are associated. Then*

$$\delta_{\mathrm{q}}(W) = \mu_{\mathrm{red}}(J). \tag{12.34}$$

In other words, the quasidegree of W is precisely equal to the reduced minimum makespan of J where the latter is viewed as an instance of $2\mathrm{MFSP}_{\mathrm{red}}$.

In terms of the standard minimum makespan $\mu(J)$, the conclusion of the theorem reads as

$$\delta_{\mathrm{q}}(W) \ = \ \max\{\delta(W),\, \mu(J) - 1\}.$$

This is clear from (12.29) and the identity $v(J) = \delta(W)$. Thus Theorem 12.10 says that either W admits a complete factorization (namely when $\mu(J) \leq v(J) + 1 = \delta(W) + 1$), or the function W has a non-minimal quasicomplete factorization involving $\delta_{\rm q}(W) = \mu(J) - 1$ elementary factors (namely when $\mu(J) > v(J) + 1 = \delta(W) + 1$).

Proof. We begin by fixing notation. Write $n = \delta(W)$ and $d = \mu_{\rm red}(J)$. Then $n = v(J)$ too and $n \leq d \leq 2n-1$. Also, let p be the pole-polynomial of W and let $p^\times$ be the zero-polynomial of W. Then p and $p^\times$ have the same (positive) degree n. Finally, let $\beta_1, \ldots, \beta_k$ be as in the paragraphs above where we discussed the association of companion based functions and instances of 2MFSP, so (12.31) and (12.32) hold.

Consider an optimal schedule for the given instance J of $2\text{MFSP}_{\rm red}$, so one with makespan $d = \mu_{\rm red}(J)$, and assume (without loss of generality) that $\mathbf{M}_1$ is occupied during the time interval from 0 to n, while $\mathbf{M}_2$ is occupied during the time interval from $d-n$ to d. Also assume that the schedule is non-preemptive (or even a permutation schedule, if one desires). Then, in particular, for $l = 1, \ldots, n$, the machine $\mathbf{M}_1$ is working on a single job during the time interval from $l-1$ to l, say the one indexed by $j_1(l) \in \{1, \ldots, k\}$. Set $\alpha_l^\times = \beta_{j_1(l)}$. In this way, we obtain an ordering $\alpha_1^\times, \ldots, \alpha_n^\times$ of the zeros of W, zero-multiplicities counted. Similarly, put $\alpha_l = \beta_{j_2(l)}$ where $j_2(l)$ is the integer among $1, \ldots, k$ uniquely determined by the requirement that $\mathbf{M}_2$ is processing the job indexed by $j_2(l)$ during the time interval from $d-n+l-1$ to $d-n+l$. Then $\alpha_1, \ldots, \alpha_n$ is an ordering of the poles of W, pole-multiplicities counted.

Suppose now that $\alpha_i = \alpha_l^\times$ for some l and i in $\{1, \ldots, n\}$. Then $\mathbf{M}_1$ is busy with the job indexed $j_1(l)$ during the time interval from $l-1$ to l. Also $\mathbf{M}_2$ is working on the job indexed $j_2(i)$ during the time interval from $d-n+i-1$ to $d-n+i$. But $\beta_{j_2(i)} = \alpha_i = \alpha_l^\times = \beta_{j_1(l)}$, and so $j_2(i) = j_1(l)$. Thus the two jobs in question are the same, indexed by $j = j_2(i) = j_1(l)$. Hence, by the predecessor constraints imposed in the case of $2\text{MFSP}_{\rm red}$,

$$d - n + i - 1 \;\geq\; S(O_j^2) \;\geq\; F(O_j^1) - 1 \;\geq l - 1,$$

where, as before, $S(O_j^2)$ denotes the start time of operation O_j^2 on $\mathbf{M}_2$ and $F(O_j^1)$ stands for the finish time of O_j^1 on $\mathbf{M}_1$. The conclusion is that $\alpha_i = \alpha_l^\times$ implies $i \geq l - (d-n)$ and Theorem 12.2 gives $\delta_{\rm q}(W) \leq d = \mu_{\rm red}(J)$.

It remains to establish the converse inequality. Write $\delta_{\rm q} = \delta_{\rm q}(W)$. Again on the basis of Theorem 12.2, we know that there exist an ordering $\alpha_1, \ldots, \alpha_n$ of the poles of W (pole-multiplicities counted) and an ordering $\alpha_1^\times, \ldots, \alpha_n^\times$ of the zeros of W (zero-multiplicities counted) such that

$$\alpha_i \neq \alpha_l^\times, \qquad i, l = 1, \ldots, n,\;\; i < l - (\delta_{\rm q} - n)\,.$$

In the next paragraph these orderings will be used to produce a feasible schedule for J, viewed as an instance of $2\text{MFSP}_{\rm red}$.

For $l = 1, \ldots, n$, there exist unique integers $j(l)$ and $j^{\times}(l)$ in the set $\{1, \ldots, k\}$ such that $\alpha_l = \beta_{j(l)}$ and $\alpha_l^{\times} = \beta_{j^{\times}(l)}$. We now stipulate that machine $\mathbf{M}_1$ processes the job indexed $j^{\times}(l)$ during the time interval from $l-1$ to l, and that $\mathbf{M}_2$ works on the job indexed $j(l)$ during the time interval from $\delta_{\mathrm{q}} - n + l - 1$ to $\delta_{\mathrm{q}} - n + l$. The schedule obtained this way satisfies the predecessor constraints imposed in the case of $2\mathrm{MFSP}_{\mathrm{red}}$. To see this, consider the job from J indexed by $j \in \{1, \ldots, k\}$, and assume that s_j and t_j are both positive. So β_j is both a zero and a pole of W. Hence there are i and l in $\{1, \ldots, n\}$ such that $\beta_j = \alpha_i = \alpha_l^{\times}$. But then $j = j(i) = j^{\times}(l)$, and so the job indexed by j is processed on machine $\mathbf{M}_1$ during the time interval from $l-1$ to l and on $\mathbf{M}_2$ during the time interval from $\delta_{\mathrm{q}} - n + i - 1$ to $\delta_{\mathrm{q}} - n + i$. As $\alpha_i = \alpha_l^{\times}$, we have $i \geq l - (\delta_{\mathrm{q}} - n)$. Taking for i and l the smallest and largest possible value, respectively, it follows that

$$S(O_j^2) \,=\, \delta_{\mathrm{q}} - n + i - 1 \,\geq\, l - 1 \,=\, F(O_j^1) - 1,$$

as required in the case of $2\mathrm{MFSP}_{\mathrm{red}}$. The feasible schedule thus obtained has makespan δ_{q}, and so $\mu_{\mathrm{red}}(J) \leq \delta_{\mathrm{q}} = \delta_{\mathrm{q}}(W)$ as desired. □

Elaborating on the second part of the proof, we note that the feasible schedule constructed there need not be non-preemptive. However, by first reordering $\alpha_1, \ldots, \alpha_n$ and $\alpha_1^{\times}, \ldots, \alpha_n^{\times}$ along the lines indicated in the proof of Theorem 11.8, one can see to it that the schedule becomes not only non-preemptive, but even a permutation schedule.

We conclude this section with three examples. The first illustrates Theorem 12.10.

Example. Let J be an instance of 2MPSP involving 5 jobs, the tuples specifying the processing times being

$$(s_1, t_1) = (2, 2), \quad (s_2, t_2) = (3, 4), \quad (s_3, t_3) = (1, 0),$$

$$(s_4, t_4) = (5, 4), \quad (s_5, t_5) = (1, 2),$$

so that $s(J) = t(J) = 12$ and the standing assumption introduced at the end of the all but last paragraph in the previous section section is satisfied. Choose five distinct complex numbers β_1, β_2, β_3, β_4 and β_5, and introduce

$$W(\lambda) = \begin{bmatrix} 1 & \dfrac{1}{(\lambda - \beta_1)^2(\lambda - \beta_2)^4(\lambda - \beta_4)^4(\lambda - \beta_5)^2} \\ 0 & \dfrac{(\lambda - \beta_3)(\lambda - \beta_4)}{(\lambda - \beta_2)(\lambda - \beta 5)} \end{bmatrix}.$$

Then W is companion based, the pole and zero-polynomial of W are

$$\begin{aligned} p(\lambda) &= (\lambda-\beta_1)^2(\lambda-\beta_2)^4(\lambda-\beta_4)^4(\lambda-\beta_5)^2 \\ &= (\lambda-\beta_1)^2(\lambda-\beta_2)^4(\lambda-\beta_3)^0(\lambda-\beta_4)^4(\lambda-\beta_5)^2, \end{aligned}$$

$$\begin{aligned} p^{\times}(\lambda) &= (\lambda-\beta_1)^2(\lambda-\beta_2)^3(\lambda-\beta_3)(\lambda-\beta_4)^5(\lambda-\beta_5) \\ &= (\lambda-\beta_1)^2(\lambda-\beta_2)^3(\lambda-\beta_3)^1(\lambda-\beta_4)^5(\lambda-\beta_5)^1, \end{aligned}$$

respectively, and $\delta(W) = 12$. Clearly W and J are associated. Thus, by Theorem 12.10, the quasidegree $\delta_{\mathrm{q}}(W)$ of W is equal to the reduced minimum makespan $\mu_{\mathrm{red}}(J)$ of J. Now $\mu_{\mathrm{red}}(J)$ is the maximum of $\delta(W)$ and $\mu(J)-1$, where $\mu(J)$ is the minimum makespan of J viewed as an instance of standard 2MFSP. So we need to determine $\mu(J)$.

Combining Steps 1 and 2 in Johnson's rule described in Section 12.3, we obtain $V_1 = \{j \mid s_j < t_j\} = \{5,2\}$ and $V_2 = \{j \mid s_j \geq t_j\} = \{4,1,3\}$. An optimal permutation schedule is now obtained by processing the jobs in the order $5,2,4,1,3$, (Step 3); schematically (with the schedule for $\mathbf{M}_2$ on top and that for $\mathbf{M}_1$ at the bottom, to keep in line with an earlier example):

$\mathbf{M}_2$:	$*$	$*$	$*$	5	5	2	2	2	2	4	4	4	4	1	1
$\mathbf{M}_1$:	5	2	2	2	4	4	4	4	4	1	1	3	$*$	$*$	$*$
Time :	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**

where a star indicates that the machine is idle. Note that job 3 (with $t_3 = 0$) does not require any action on machine $\mathbf{M}_2$.

We conclude that the (standard) minimum makespan $\mu(J)$ of J is equal to 15. Recall that $\delta_{\mathrm{q}}(W) = \max\{\delta(W), \mu(J)-1\}$. In the present situation, we have $\delta(W) = 12$ and $\mu(J) = 15$. Hence $\delta_{\mathrm{q}}(W) = 14$. In particular $\delta_{\mathrm{q}}(W) > \delta(W)$, so W does not admit a complete factorization. It is worth stressing that these conclusions have been reached by the application of Johnson's rule.

Our second example demonstrates that the general estimate

$$\delta_{\mathrm{q}}(W) \leq 2\delta(W) - 1$$

appearing in the inequalities (10.30), is sharp in the sense that for every positive value of the McMillan degree $\delta(W)$ equality can occur (cf., the examples in Section 10.4, the first one in particular).

Example. Let n be a positive integer, and consider the 2×2 rational matrix function W given by

$$W(\lambda) = \begin{bmatrix} 1 & \dfrac{1}{\lambda^n} \\ 0 & 1 \end{bmatrix}.$$

This function, which has McMillan degree n, also features in the example given at the end of Section 9.1. It was proved there that W does not admit any non-trivial minimal factorization, so $\delta_{\mathrm{q}}(W) > \delta(W) = n$ whenever $n \geq 2$. Note that W is of the form (12.33) with $p(\lambda) = p^{\times}(\lambda) = \lambda^n$. In particular W is companion based. Let J be the instance of 2MFSP consisting of just one job with processing time n on both machines. Then evidently $\mu_{\mathrm{red}}(J) = 2n-1$, and it follows from Theorem 12.10 that $\delta_{\mathrm{q}}(W) = 2n-1$ too.

For the case $n = 2$, hence $\delta_{\mathrm{q}}(W) = 3$, the following quasicomplete factorization

$$\begin{bmatrix} 1 & \dfrac{1}{\lambda^2} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \dfrac{1}{\lambda-1} \\ 0 & \dfrac{\lambda}{\lambda-1} \end{bmatrix} \begin{bmatrix} 1 & -\dfrac{1}{\lambda} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{\lambda-1}{\lambda} \end{bmatrix}$$

has been obtained in Section 10.4. For $n = 3$, hence $\delta_{\mathrm{q}}(W) = 5$, we have

$$\begin{bmatrix} 1 & \dfrac{1}{\lambda^3} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \dfrac{1}{\lambda-1} \\ 0 & \dfrac{\lambda}{\lambda-1} \end{bmatrix} \begin{bmatrix} 1 & \dfrac{2}{2\lambda+1} \\ 0 & \dfrac{2\lambda}{2\lambda+1} \end{bmatrix} \begin{bmatrix} 1 & -\dfrac{2}{\lambda} \\ 0 & 1 \end{bmatrix} \times$$

$$\times \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{\lambda-1}{\lambda} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{2\lambda+1}{2\lambda} \end{bmatrix}$$

as a quasicomplete factorization.

Note that these explicit factorizations are in accordance with Proposition 10.8. Indeed, for the case $n = 2$ we have

$$\alpha_1 = 1, \qquad \alpha_2 = 0, \qquad \alpha_3 = 0,$$

$$\alpha_1^{\times} = 0, \qquad \alpha_2^{\times} = 0, \qquad \alpha_3^{\times} = 1.$$

and for the case $n = 3$ we have

$$\alpha_1 = 1, \qquad \alpha_2 = -1/2, \qquad \alpha_3 = 0, \qquad \alpha_4 = 0, \qquad \alpha_5 = 0,$$

$$\alpha_1^{\times} = 0, \qquad \alpha_2^{\times} = 0, \qquad \alpha_3^{\times} = 0, \qquad \alpha_4^{\times} = 1, \qquad \alpha_5^{\times} = -1/2.$$

In the above quasicomplete factorizations, poles and zeros occur that are not present in the given function that is factorized. On the one hand, this differs from the case of complete (more generally, minimal) factorization. On the other hand, the phenomenon is completely in line with the proof of Theorem 10.15, where the overall possibility of factorization into elementary factors was established. The question is: can one do without such additional new poles and zeros? We shall now

see that the answer is generally negative and that a counterexample is provided by the function W from the previous example with $n = 2$.

Example. Consider the 2×2 rational matrix function W given by

$$W(\lambda) = \begin{bmatrix} 1 & \dfrac{1}{\lambda^2} \\ 0 & 1 \end{bmatrix}.$$

This function has the origin as its only pole and zero, both with multiplicity two. Let

$$W(\lambda) = \left(I + \frac{1}{\lambda - \alpha_1} R_1\right)\left(I + \frac{1}{\lambda - \alpha_2} R_2\right)\left(I + \frac{1}{\lambda - \alpha_3} R_3\right)$$

be factorization of W involving three rank one 2×2 matrices R_1, R_2 and R_3 and three poles α_1, α_2 and α_3 in the right-hand side. Then

$$W^{-1}(\lambda) = \left(I - \frac{1}{\lambda - \alpha_3^\times} R_3\right)\left(I - \frac{1}{\lambda - \alpha_2^\times} R_2\right)\left(I - \frac{1}{\lambda - \alpha_1^\times} R_1\right)$$

is a factorization of W^{-1} and the poles in the right-hand side are the complex numbers $\alpha_j^\times = \alpha_j - \operatorname{trace} R_j,\ j = 1, 2, 3$. The claim is that one cannot have

$$\alpha_1 = \alpha_2 = \alpha_3 = 0 \qquad \text{or} \quad \alpha_1^\times = \alpha_2^\times = \alpha_3^\times = 0.$$

We shall prove this by reductio ad absurdum.

Suppose one of the collection of identities in question holds. Then, in fact, both of them are satisfied. Indeed, specializing the conclusion of Proposition 10.8 by taking $\alpha = 0$ yields

$$\sharp\{j \mid \alpha_j = 0\} - 2 \;=\; \sharp\{j \mid \alpha_j^\times = 0\} - 2 \;\geq\; 0,$$

so, in particular, $\sharp\{j \mid \alpha_j = 0\} = \sharp\{j \mid \alpha_j^\times = 0\}$. Now $\alpha_j = \alpha_j^\times$ if and only if $\operatorname{trace} R_j = 0$. Thus we need to establish the impossibility of a factorization of the form

$$\begin{bmatrix} 1 & \dfrac{1}{\lambda^2} \\ 0 & 1 \end{bmatrix} \;=\; \left(I + \frac{1}{\lambda} R_1\right)\left(I + \frac{1}{\lambda} R_2\right)\left(I + \frac{1}{\lambda} R_2\right) \tag{12.35}$$

in which R_1, R_2 and R_3 are rank one matrices having zero trace. The latter means that R_j has the form

$$R_j = \begin{bmatrix} c_j & b_j \\ a_j & -c_j \end{bmatrix}, \tag{12.36}$$

where a_j, b_j and c_j do not simultaneously vanish and $c_j^2 + a_j b_j = 0$. Substituting (12.36) into (12.35), the right-hand side of the latter becomes

$$\begin{bmatrix} 1+\dfrac{c_1}{\lambda} & \dfrac{b_1}{\lambda} \\ \dfrac{a_1}{\lambda} & 1-\dfrac{c_1}{\lambda} \end{bmatrix} \begin{bmatrix} 1+\dfrac{c_2}{\lambda} & \dfrac{b_2}{\lambda} \\ \dfrac{a_2}{\lambda} & 1-\dfrac{c_2}{\lambda} \end{bmatrix} \begin{bmatrix} 1+\dfrac{c_3}{\lambda} & \dfrac{b_3}{\lambda} \\ \dfrac{a_3}{\lambda} & 1-\dfrac{c_3}{\lambda} \end{bmatrix},$$

and this can be written as

$$\begin{bmatrix} 1+w_{11}(\lambda) & w_{12}(\lambda) \\ w_{21}(\lambda) & 1+w_{22}(\lambda) \end{bmatrix},$$

with $w_{11}(\lambda)$, $w_{12}(\lambda)$, $w_{21}(\lambda)$, and $w_{22}(\lambda)$ given by

$$\begin{aligned} w_{11}(\lambda) &= \frac{1}{\lambda}(c_1+c_2+c_3) + \frac{1}{\lambda^2}(b_1a_2 + b_1a_3 + b_2a_3 + c_1c_2 + c_1c_3 + c_2c_3) \\ &\quad + \frac{1}{\lambda^3}(b_1a_2c_3 - b_1c_2a_3 + c_1b_2a_3 + c_1c_2c_3), \end{aligned}$$

$$\begin{aligned} w_{12}(\lambda) &= \frac{1}{\lambda}(b_1+b_2+b_3) + \frac{1}{\lambda^2}(c_1b_3 + c_2b_3 + c_1b_2 - b_1c_2 - b_1c_3 - b_2c_3) \\ &\quad + \frac{1}{\lambda^3}(b_1a_2b_3 + b_1c_2c_3 + c_1c_2b_3 - c_1b_2c_3), \end{aligned}$$

$$\begin{aligned} w_{21}(\lambda) &= \frac{1}{\lambda}(a_1+a_2+a_3) + \frac{1}{\lambda^2}(a_1c_2 + a_1c_3 + a_2c_3 - c_1a_2 - c_1a_3 - c_2a_3) \\ &\quad + \frac{1}{\lambda^3}(a_1c_2c_3 + a_1b_2a_3 + c_1c_2a_3 - c_1a_2c_3), \end{aligned}$$

$$\begin{aligned} w_{22}(\lambda) &= \frac{-1}{\lambda}(c_1+c_2+c_3) + \frac{1}{\lambda^2}(a_1b_2 + a_1b_3 + a_2b_3 + c_1c_2 + c_1c_3 + c_2c_3) \\ &\quad + \frac{1}{\lambda^3}(a_1c_2b_3 - a_1b_2c_3 - c_1a_2b_3 - c_1c_2c_3). \end{aligned}$$

Inspection of the coefficients of $1/\lambda$ yields

$$a_1 + a_2 + a_3 = 0, \qquad b_1 + b_2 + b_3 = 0, \qquad c_1 + c_2 + c_3 = 0,$$

and it follows that

$$\begin{aligned} 2c_1c_2 &= c_3^2 - c_1^2 - c_2^2 \\ &= -a_3b_3 + a_1b_1 + a_2b_2 \\ &= -(a_1+a_2)(b_1+b_2) + a_1b_1 + a_2b_2 \\ &= -(a_1b_2 + b_1a_2). \end{aligned}$$

By straightforward computations, the expressions for the functions $w_{ij}(\lambda)$, can now be simplified to

$$\begin{aligned} w_{11}(\lambda) &= \frac{1}{\lambda^2}(b_1a_2 + c_1c_2), \quad w_{12}(\lambda) = \frac{1}{\lambda^2}(c_1b_2 - b_1c_2), \\ w_{21}(\lambda) &= \frac{1}{\lambda^2}(a_1c_2 - c_1a_2), \qquad w_{22}(\lambda) = \frac{1}{\lambda^2}(a_1b_2 + c_1c_2), \end{aligned}$$

(where we note that in the simplification process the coefficients of the powers of $1/\lambda^3$ are becoming zero). Again by comparison of coefficients, this time of $1/\lambda^2$, we find the identities

$$b_1a_2 + c_1c_2 = 0, \qquad c_1b_2 - b_1c_2 = 1, \qquad a_1b_2 + c_1c_2 = 0$$

(and, of course, also $a_1c_2 - c_1a_2 = 0$ but that one does not play a role in the further derivation). It follows that

$$\begin{aligned} 1 &= c_1b_2 - b_1c_2 = c_1(c_1b_2 - b_1c_2)b_2 - b_1(c_1b_2 - b_1c_2)c_2 \\ &= (c_1^2b_2 - b_1c_1c_2)b_2 - b_1(c_1b_2c_2 - b_1c_2^2) \\ &= -(a_1b_1b_2 + b_1c_1c_2)b_2 - b_1(c_1b_2c_2 + b_1a_2b_2) \\ &= -b_1(a_1b_2 + c_1c_2)b_2 - b_1(c_1c_2 + b_1a_2)b_2 = 0, \end{aligned}$$

which is an obvious contradiction.

12.5 Maple procedures for quasicomplete factorizations

This section gives Maple procedures to calculate quasicomplete factorizations of a proper rational 2×2 matrix function W of the form

$$W(\lambda) = \begin{bmatrix} 1 & w_{12}(\lambda) \\ 0 & w_{22}(\lambda) \end{bmatrix} \quad \text{and} \quad W(\infty) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{12.37}$$

From Section 10.4 we know that W always admits quasicomplete factorizations. The fact that W is companion based (see Lemma 11.15) allows us to use the method described in Section 12.2 to get such a factorization. The topic of this section is the implementation of the method of Section 12.2 in Maple procedures. Throughout n is the McMillan degree of W.

We assume the reader to be familiar with the contents of Section 11.6. There we have given Maple procedures to get the pole-polynomial and the zero-polynomial of W, its poles and zeros itself, and procedures to get orderings of the poles and zeros for minimal h, $h > 0$, satisfying condition (11.20). Since the McMillan degree of W is equal to n, the quasidegree $\delta_q(W)$ is given by $\delta_q(W) = n+h-1$, where h is smallest positive integer satisfying (11.20); see Theorems 12.2 and 12.9.

Hence from the procedure **GetAllMorderings** in Subsection 11.6.2, especially its first output element, we can calculate the quasidegree of W.

Again, as in Section 11.6, we denote by A the first companion matrix corresponding to the pole-polynomial $p(\lambda)$ of W, and by $Z(= A^{\times})$ the first companion matrix corresponding to the zero-polynomial $p^{\times}(\lambda)$ of W. Thus it remains to provide Maple procedures for the construction of the matrices $\widehat{A}$ and $\widehat{A}^{\times}$ appearing in the second part of the proof of Theorem 12.2, and the triangularization of those matrices in complementary triangular form. From that point on we can apply the Maple factorization procedure **MakeFactorization** presented in Subsection 11.6.4 to get a quasicomplete factorization of W.

The Maple procedures in this section fall apart in three. First, the matrices $\widehat{A}$ and $\widehat{A}^{\times} = \widehat{A} - \widehat{B}\widehat{C}$ are constructed; see formula (12.15). The starting point is a companion based rational matrix function W as in Theorem 12.2. The Maple procedure **QCmatrices** presented in Subsection 12.5.2 follows directly the second part of the proof of Theorem 12.2. This means that a matrix K as in Proposition 12.6 and a matrix X as in formula (12.14) are calculated, and subsequently, $\widehat{A}$, $\widehat{B}$ and $\widehat{C}$ are constructed according to formula (12.15).

The second step is the construction in Maple of the vectors

$$a_1, \ldots, a_{d-n}, a_{d-n+1}, \ldots, a_d,$$

where d is the quasidegree, in (12.16). These vectors are in Maple collected in one matrix denoted by TA; see Maple procedure **MakeBasisA**. Analogously, the vectors $a_1^{\times}, \ldots, a_{d-n}^{\times}$, $a_{d-n+1}^{\times}, \ldots, a_d^{\times}$ in (12.18) are calculated and collected in a Maple matrix denoted by TZ; see procedure **MakeBasisZ**. As linear transformations, the matrices TA and TZ will bring $\widehat{A}$ and $\widehat{A}^{\times}$ in upper-triangular form; see Subsection 12.5.3.

In the third step a procedure is implemented to extract from the previously calculated matrices TA and TZ the matrix S such that $\widehat{A}$ and $\widehat{A}^{\times}$ can be brought in complementary triangular form. The procedure is based on the proof of Proposition 10.1 (c); see Subsection 12.5.4.

Finally, in Subsection 12.5.5 the use of the procedures is elucidated by an example of quasicomplete factorization of a rational matrix function defined in Maple symbols.

As in Section 11.6 all procedures and calculations in this section are tested under Maple, version 9, [93] and, as usual, the Maple command lines start with the symbol >. Also, the Maple worksheet containing all procedures and commands presented in the present section is available on request by email from the fourth author (ACM.Ran@few.vu.nl).

12.5.1 Maple environment

```
> restart;# almost clean start
> with(LinearAlgebra):
> with(MatrixPolynomialAlgebra):
```

Note: To run an example as included in this section, or a newly defined one, one needs to activate all the poles and zeros Maple procedures of Section 11.6.2 and the Maple factorization procedures of Section 11.6.4.

12.5.2 Triangularization routines (quasicomplete)

The starting point for this subsection is a companion based rational matrix function W as in Theorem 12.2. We assume that we have a realization

$$W(\lambda) = D + C(\lambda I_n - A)^{-1}B,$$

where D is the m-dimensional identity matrix and A and $A^\times$ are first companion $n \times n$ matrices. The aim is to construct $\widehat{A}$, $\widehat{B}$ and $\widehat{C}$ as in (12.15). This task is performed by the Maple procedure **QCmatrices**.

The first four arguments of **QCmatrices** are the companion based realization matrices, in Maple named *Amin*, *Bmin*, *Cmin* and *Dmin*; see Section 11.6. The fifth argument is the quasidegree (in Maple denoted by *dqw*). The sixth argument *gammav* needs some attention: it should be a vector of length n with the first $\delta_q(W)-n$ entries distinct and outside the spectra of A and $A^\times$; see the paragraph containing (12.14). Let *mu* be the Maple vector of which the entries are the different elements of the spectra of A and $A^\times$. Here, for the sake of simplicity, we shall take *gammav* so that all its entries are distinct and unequal any element of *mu*. Always, we can take the entries of *gammav* to be Maple symbols $\gamma_1, \ldots, \gamma_n$ different from the elements of *mu*. In case one wants *gammav* to be a vector of complex numbers, the construction of the vector *gammav* given here proceeds as follows. First, define max_{mu} to be the maximum over the real part of the elements of *mu* that are complex numbers (and set $max_{mu} = 0$ if *mu* consists of Maple symbols only). Next *gammav* is then defined as *gammav[k]*$=k + max_{mu}$, $k = 1, \ldots, n$; see procedure **MakeGamma**. Of course, *gammav* should be calculated before any call is made to **QCmatrices** but it is also used (as an argument) to routines which follow up **QCmatrices** and it should not be altered in between.

The procedure **QCmatrices** heavily makes use of generalized Vandermonde matrices which are calculated from a separate procedure. If the quasidegree $\delta_q(W)$ equals the McMillan degree $\delta(W)$, the procedure does nothing and returns just the matrices given as arguments of **QCmatrices**. This allows us to use this procedure and the following procedures also for the case of a complete case factorization, i.e., when $\delta_q(W) = \delta(W) = n$.

Finally, the procedure **QCmatrices** calls the procedure **MakeKmatrix** with second argument a vector v; this vector v should be such that

$$Bmin\, v = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

see Proposition 12.6. Because of the special structure of *Bmin*, see (11.41), one may take $v = [0, 1]$, which is done in **QCmatrices**.

QCmatrices	calculate matrices $\widehat{A}$, $\widehat{B}$, $\widehat{C}$ (and $\widehat{D}$)
Calling sequence	QCmatrices(**Amin,Bmin,Cmin,Dmin,dqw,gammav)**
Parameters	Amin,Bmin,Cmin,Dmin - Realization matrices with Amin companion form dqw - scalar: quasidegree gammav - Vector
Output	list of Matrices Ahat, Bhat, Chat, Dhat.
Note	If dqw = Column Dimension(Amin) = McMillan degree, then the returned matrices are just the matrices (Amin,Bmin,Cmin,Dmin).

```
> QCmatrices:=proc(Amin,Bmin,Cmin,Dmin,dqw,gammav)
> local v, k, Kmat, Xmat, Fmat, Gmat, Ahat, Bhat, Chat, Dhat,
> LastRowA, na;
> if not (IsCompanionForm(Amin)) then error "First
> input matrix should be in companion form but get %1",Amin; end
> if; na:=Dimension(Amin): if (dqw>na[2]) then
> LastRowA:=Row(Amin,na[1]): v:=Vector([0,1]);
> Kmat:=MakeKmatrix(LastRowA,v,gammav):
> Xmat:=MakeXmatrix(gammav,dqw,na[2]):
> Gmat:=DiagonalMatrix(gammav[1..(dqw-na[2])],(dqw-na[2]),
> (dqw-na[2])): Fmat:=Kmat.Xmat:
> Ahat:=<<Amin,ZeroMatrix(dqw-na[2],na[2])>|<Bmin.Fmat,Gmat>>:
> Bhat:=<Bmin,ZeroMatrix(dqw-na[2],ColumnDimension(Bmin))>:
> Chat:=<<Cmin>|<Fmat>>: else Ahat:=Amin: Bhat:=Bmin: Chat:=Cmin:
> end if: Dhat:=Dmin: return(Ahat,Bhat,Chat,Dhat); end proc;
```

Secondary routines used in QCmatrices

The procedure **MakeGamma** constructs a vector with entries outside the spectra of A and $A^{\times}$.

MakeGamma	calculate a vector of length dw = $\delta(W)$
Calling sequence	MakeGamma(**dw,mu,symbols**)
Parameters	dw: scalar (McMillan degree $\delta(W)$) mu: Vector of different poles and zeros symbols: boolean (true or false). If symbols is true then the output vector is a Maple Vector of symbols $(\gamma_1, \ldots, \gamma_{dw})$. Otherwise, the output vector is a Maple Vector with numerical entries only.
Output	Vector(**dw**)

```
> MakeGamma:=proc(dw,mu,symbols)
> local nmu,mm,k,numval,gammav;
> if symbols then gammav:=Vector(dw,symbol=gamma): else
> gammav:=Vector(dw):nmu:=Dimension(mu):k:=0: numval:=[0]:
> for k from 1 to nmu  do if (type(mu[k],complex)) then
> numval:=[op(numval),Re(mu[k])]: end if: end do:
> mm:=max(op(numval)): for k from 1 to dw do gammav[k]:=mm+k:
> end do: end if: return(gammav);end proc;
```

The procedure **IsCompanionMatrix** outputs true if the input matrix is in companion form, otherwise false.

IsCompanionMatrix	test whether a (square) matrix is of companion form
Calling sequence	IsCompanionForm(**A**)
Parameters	A - Matrix
Output	true, if **A** has companion form, else false

```
> IsCompanionForm:=proc(A)
> local na, zm, tm, k, bol;
> na:=Dimension(A): zm:=Matrix(na[1]-1,na[2],0):
> for k from 1 to (na[1]-1) do zm[k,k+1]:=1: end do:
> tm:=SubMatrix(A,[1..na[1]-1],[1..na[2]]):
> bol:=Equal(map(simplify,(tm-zm)),ZeroMatrix(na[1]-1,na[2])):
> return(bol);end proc;
```

The procedure **GenVandermondeMatrix** returns a generalized Vandermonde matrix according to the definition in the paragraph after the proof of Proposition 11.19.

GenVandermondeMatrix	calculate the generalized Vandermonde matrix
Calling sequence	GenVandermondeMatrix(**r**,**av**)
Parameters	r - scalar (row dimension returned matrix) av - Vector of algebraic values
Output	Matrix(**r**,c) M with c is dimension of **av**; $M_{j,k} = \binom{j-1}{k} av_k^{j-1-k}$, $j = (k+1), \ldots, \mathbf{r}$, zero otherwise

```
> GenVandermondeMatrix:=proc(r,av::Vector)
> local GVM,m,v,j,c;
> c:=Dimension(av):v:=0:GVM:=GenVandermondeVector(r,av[1],v):
> for j from 2 to c do v:=0:for m from 1 to (j-1) do
> if (av[m]=av[j]) then v:=v+1: end if: end do:
> GVM:=<GVM|GenVandermondeVector(r,av[j],v)>:end do:
> return(GVM);end proc;
```

The actual calculation of the columns of a generalized Vandermonde matrix is done by the procedure **GenVandermondeVector**.

GenVandermondeVector	calculate a column (vector) of the generalized Vandermonde matrix
Calling sequence	GenVandermondeVector(**r,a,k**)
Parameters	r - scalar (dimension (length) output vector) a - scalar value k - integer (column index)
Output	Vector(**r**) V with V is kth column of a generalized Vandermonde matrix

```
> GenVandermondeVector:=proc(r,a,k)
> local colgvm,j;
> colgvm:=Vector(r,0): for j from 1 to r do if (j>k) then
> colgvm[j]:=binomial(j-1,k)*a^(j-1-k):end if: end do:
> return(colgvm); end proc;
```

The next two procedures will output a matrix K with the properties described in Proposition 12.6 and a matrix X defined as in formula (12.14).

MakeKmatrix	calculate matrix K as in Proposition 12.6
Calling sequence	MakeKmatrix(**a,v,gammav**)
Parameters	a - Vector: final row of companion matrix v - Vector: solution of $B.v = e_n$; see Proposition 12.6 gammav - Vector: output of **MakeGamma**
Output	Matrix K

```
> MakeKmatrix:= proc(a,v,gammav)
> local nc,k,Kmat,pol;
> nc:=Dimension(a):pol:=1: for k from 1 to (nc) do
> pol:=pol*(lambda-gammav[k]):end do: pol:=expand(pol):
> Kmat:=ScalarMultiply(v,-a[1]-coeff(expand(pol),lambda,0)):
> for k from 2 to (nc) do Kmat:=
> <Kmat|ScalarMultiply(v,-a[k]-coeff(expand(pol),lambda,k-1))>:
> end do:return(Kmat);end proc;
```

MakeXmatrix	calculate matrix X, see formula (12.14)
Calling sequence	MakeXmatrix(**gammav,d,n**)
Parameters	gammav - Vector: output of **MakeGamma** d - scalar ($d > n$) n - scalar
Output	Matrix $X(n, d-n)$.

```
> MakeXmatrix:=proc(gammav,d,n)
> local gd,k,m,X;
> X:=GenVandermondeVector(n,gammav[1],0):
> for k from 2 to (d-n) do
> X:=<X|GenVandermondeVector(n,gammav[k],0)>: end do:
> return(convert(X,Matrix)); end proc;
```

12.5.3 Transformations into upper triangular form

The two procedures **MakeBasisA** and **MakeBasisZ** (where Z refers to $A^{\times}$) construct bases such that $\widehat{A}$ and $\widehat{A}^{\times}$ are in upper-triangular form with respect to those bases.

MakeBasisA	calculate a basis (transformation matrix) as in the paragraph containing formula (12.16)
Calling sequence	MakeBasisA(**orderedA,dqw,gammav**)
Parameters	orderedA - Vector of ordered poles dqw - scalar: quasi degree gammav - Vector: output of **MakeGamma**
Output	Matrix(**dqw,dqw**). If **dqw**=Dimension(**orderedA**)=McMillan degree then this matrix is just the generalized Vandermonde matrix with vector **orderedA**.

```
> MakeBasisA:=proc(orderedA,dqw,gammav)
> local BasisMat,id,zerod,n,j,jj,GVM,npoles,k;
> npoles:=Dimension(orderedA):
> GVM:=GenVandermondeMatrix(npoles,orderedA):
> n:=Dimension(GVM)[1]: if dqw>n then id:=IdentityMatrix(dqw-n):
> zerod:=Vector(dqw-n,0):
> BasisMat:=<GenVandermondeVector(n,gammav[1],0),Column(id,1)>:
> for j from 2 to (dqw-n) do BasisMat:=
> <BasisMat|<GenVandermondeVector(n,gammav[j],0),Column(id,j)>>:
> end do: for j from (dqw-n+1) to dqw do jj:=j+n-dqw:
> BasisMat:=<BasisMat|<Column(GVM,jj),zerod>>:end do: else
> BasisMat:=GVM: end if: return(BasisMat);end proc;
```

MakeBasisZ	calculate a basis (transformation matrix) as in the paragraph containing formula (12.12.17)
Calling sequence	MakeBasisZ(**orderedZ,dqw,gammav**)
Parameters	orderedZ - Vector of ordered zeros dqw - scalar : quasi degree gammav - Vector: output of **MakeGamma**
Output	Matrix(**dqw,dqw**). If **dqw**=Dimension(**orderedZ**)= McMillan degree then this matrix is just the generalized Vandermonde matrix with vector **orderedZ** in reverse order.

```
> MakeBasisZ:=proc(orderedZ,dqw,gammav)
> local BasisMat,id,zerod,n,j,jj,nzeros,ReorderedZ,GVM,k;
> nzeros:=Dimension(orderedZ):
> ReorderedZ:=ReverseOrder(orderedZ):
> GVM:=GenVandermondeMatrix(nzeros,ReorderedZ):
> n:=Dimension(GVM)[1]: if dqw>n then
> id:=IdentityMatrix(dqw-n):zerod:=Vector(n,0):
```

```
> BasisMat:=<zerod,Column(id,1)>: for j from 2 to (dqw-n) do
> BasisMat:=<BasisMat|<zerod,Column(id,j)>>: end do:
> zerod:=Vector(dqw-n,0): for j from (dqw-n+1) to dqw do
> jj:=dqw+1-j:jj:=j-dqw+n:
> BasisMat:=<BasisMat|<Column(GVM,jj),zerod>>: end do: else
> BasisMat:=GVM:end if; return(BasisMat);end proc;
```

The procedure **ReverseOrder** reverts the order of elements of a vector.

ReverseOrder	calculate vector with elements in reverse order of a given vector
Calling sequence	ReverseOrder(**v**)
Parameters	v - Vector or list
Output	Vector rv such that $rv_k = v_{n-k+1}$, $k = 1, \cdots, n$, with n is dimension of **v**.

```
> ReverseOrder:=proc(v)
> local rv,n,k;
> n:=Dimension(v): rv:=Vector(n): for k from 1 to n do
> rv[k]:=v[n-k+1]: end do: return(rv);end proc;
```

12.5.4 Transformation into complementary triangular forms

In Maple we use the name TA for the output of **MakeBasisA**; it is a transformation which brings $\widehat{A}$ in upper-triangular form. Similarly, TZ is the output of **MakeBasisZ**; it is a the transformation which brings $\widehat{A}^{\times}$ in upper-triangular form. Then the procedure **UpperLowerTransformation** extracts from TA and TZ a transformation which allows for a simultaneous reduction to upper- and lower-triangularization of $\widehat{A}$ and $\widehat{A}^{\times}$, respectively; see Proposition 10.1 (c).

UpperLowerTransformation	calculate a matrix S which brings $\widehat{A}$ and $\widehat{A}^{\times}$ in upper- and lower triangular form (see Section 10.1)
Calling sequence	UpperLowerTransformation(**TA,TZ,dqw**)
Parameters	TA - Matrix: output of MakeBasisA TZ - Matrix: output of MakeBasisZ dqw - quasidegree
Ouput	Matrix(**dqw,dqw**)

```
> UpperLowerTransformation:=proc(TA,TZ,dqw)
> local S1,S,k;
> S:=op(IntersectionBasis([[Column(TA,1)],
> [Column(TZ,[1..dqw])]])):
> for k from 2 to dqw do
> S1:=op(IntersectionBasis([[Column(TA,[1..k])],
> [Column(TZ,[1..dqw-k+1])]])): S:=<S|S1>:end do:
> return(map(simplify,S));end proc;
```

12.5.5 An example: symbolic and quasicomplete

The next lines define a rational 2×2 matrix function W of the form (12.37) in symbolic variables (unevaluated names). We start with two polynomials $q(\lambda)$ and $q^{\times}(\lambda)$ of degree n; in this example $n = 3$. We take the degree of the numerator of w_{12} to be equal to $n - 2$.

```
> n:=3:
> p:='p':r:='r':q:='q':
> sw1:=<<1,0>|<'s'(lambda)/ 'q(lambda)', '(q^(x))(lambda)'/
> 'q(lambda)'>>: 'W(lambda)'=sw1;
> ppol:=proc(x,p,nn) local r,k; r:=1: for k from 1 to (nn) do r:=
> r*(x-p[k-1]): end do: return(r); end proc:
> p:='p': p:=array(0..(n-1)):
> pp:=x->ppol(x,p,n): 'q(lambda)'=pp(lambda);
> z:='z': z:=array(0..(n-1)):
> zz:=x->ppol(x,z,n): 'q^(x)'(lambda) = zz(lambda);
> r:='r': r:=array(0..(n-1)):
> rs:=x->ppol(x,r,n-2): 's(lambda)'=rs(lambda);
```

$$
\begin{aligned}
W(\lambda) &= \begin{bmatrix} 1 & \dfrac{s(\lambda)}{q(\lambda)} \\ 0 & \dfrac{q^{\times}(\lambda)}{q(\lambda)} \end{bmatrix} \\
q(\lambda) &= (\lambda - p_0)(\lambda - p_1)(\lambda - p_2) \\
q^{\times}(\lambda) &= (\lambda - z_0)(\lambda - z_1)(\lambda - z_2) \\
s(\lambda) &= \lambda - r_0.
\end{aligned}
$$

In this example, we assume that the polynomial q has two equal zeros ($p_1 = p_0$) and that all zeros of the polynomial $q^{\times}$ are equal to the zeros of q, that is $z_i = p_i,\ i = 0, 1, 2$. If one wants to change or leave out any condition, one has to rerun the foregoing Maple lines and change (or comment out) the next two lines. Note that Maple will consider variables with different, unevaluated names as different.

```
> p:='p': z:='z': r:='r': p[1]:=p[0];
> z[0]:=p[0]; z[1]:=p[1]; z[2]:=p[2];
```

Next the rational function is defined:

```
> W:=x-><<1,0>|<rs(x)/pp(x),zz(x)/pp(x)>>:
> 'W(lambda)'=W(lambda);
```

$$
W(\lambda) = \begin{bmatrix} 1 & \dfrac{\lambda - r_0}{(\lambda - p_0)^2 (\lambda - p_2)} \\ 0 & 1 \end{bmatrix}. \tag{12.38}
$$

Now we will start our example showing how to use the previously defined Maple procedures, with W as defined in (12.38) (and neglecting any knowledge about how W is constructed). As before, most of the actual output is not shown. The Maple variable $DimS$ is used throughout this subsection.

```
> DimS:=ColumnDimension(W(x));
```

Next we get the pole-polynomial and zero-polynomial of W in (12.38), and its poles and zeros itself, just as in Section 11.6. So we will give the following Maple lines without any comment.

```
> q:=unapply(LCMDenomMatrixPolynom(W,x),x):
> ppoles:=q:pzeros:=unapply(simplify(W(x)[2,2]*q(x)),x):
> 'p(lambda)'=sort(collect(ppoles(lambda),lambda),lambda);
> '(p^(x))(lambda)'=sort(collect(pzeros(lambda),lambda),lambda);
```

$$\begin{aligned} p(\lambda) &= \lambda^3 + (-2\,p_0 - p_2)\,\lambda^2 + \left({p_0}^2 + 2\,p_0 p_2\right)\lambda - {p_0}^2 p_2 \\ p^{\times}(\lambda) &= \lambda^3 + (-2\,p_0 - p_2)\,\lambda^2 + \left({p_0}^2 + 2\,p_0 p_2\right)\lambda - {p_0}^2 p_2 \end{aligned}$$

```
> res1:=GetPolesandZeros(ppoles,pzeros):
> poles := res1[1]: zeros:=res1[2]:
> mu:=res1[3]:
> npoles:=Dimension(poles);nzeros:=Dimension(zeros);
> nmu:=Dimension(mu);
```

The calculation of companion based realization matrices results from a Maple implementation of Lemma 11.15; see again Section 11.6. Again, the realization matrices are named in Maple *Amin, Bmin, Cmin* and *Dmin*.

```
> r:=unapply(simplify(W(x)[1,2]*ppoles(x)),x):r(lambda):
> Amin:=Transpose(CompanionMatrix(ppoles(x),x)):
> Bmin:=Matrix(npoles,DimS,0):Bmin[npoles,DimS]:=1:
> Cmin:= Matrix(DimS,npoles):
> for k from 1 to npoles do
> Cmin[1,k]:= coeff(r(lambda),lambda,k-1):
> Cmin[2,k] := coeff(pzeros(lambda)-ppoles(lambda),lambda,k-1):
> end do: Dmin:=IdentityMatrix(DimS,DimS):
> Amincross:=Transpose(CompanionMatrix(pzeros(x),x)):
> 'A'=Amin;
> `transpose(B)`=Transpose(Bmin);
> `C`=Cmin;
```

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ {p_0}^2 p_2 & -{p_0}^2 - 2p_0 p_2 & 2p_0 + p_2 \end{bmatrix}$$

$$transpose\,(B) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} -r_0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The next step, see again Section 11.6, is getting feasible orderings of poles and zeros.

```
> muA:=GetMultiplicity(poles,mu);
> muZ:=GetMultiplicity(zeros,mu);
> ResultOrdering:=GetAllMOrderings(muA,muZ,mu);
> h:=ResultOrdering[1];
> ‘number of
> orderings‘=ResultOrdering[2];
> orderedA:=ResultOrdering[3][1];
> orderedZ:=ResultOrdering[4][1];
```

$$\begin{aligned} h &= 2 \\ number\ of\ orderings &= 2 \\ orderedA &= [p_0, p_0, p_2] \\ orderedZ &= [p_0, p_0, p_2] \end{aligned}$$

The Maple variables *orderedA* and *orderedZ* are the used orderings of poles and zeros of W, respectively. In this case, we have taken the first found ordering. If one would like to have results on a different ordering, one should change the Maple variables *orderedA* and *orderedZ*, e.g., *orderedA:=ResultOrdering[3][2]* and *orderedZ:=ResultOrdering[4][2]*, and re-run the worksheet from this point on.

The Maple variable *dw* is the McMillan degree $\delta(W)$ and is equal to 3. The quasidegree $\delta_q(W)$ is denoted in Maple as *dqw*. The value of *dqw* is 4 since $h = 2$.

```
> dw := npoles: dqw:=h-1+npoles:
> ’delta[q](W)’ = dqw;
> print(’delta[q](W)-delta(W)’=dqw-dw);
```

$$\begin{aligned} \delta_q(W) &= 4 \\ \delta_q(W) - \delta(W) &= 1 \end{aligned}$$

Triangularization: quasicomplete case
First the matrices $\widehat{A}$, $\widehat{B}$, $\widehat{C}$ and $\widehat{D}$ are made by calling **QCmatrices** with arguments the previously calculated matrices *Amin*,*Bmin*, *Cmin*, *Dmin* and the quasidegree value *dqw*. In Maple the output matrices are named *Ahat, Bhat, Chat* and *Dhat*. The matrix $\widehat{A}^{\times}$ is named *Ahatcross*.

```
> gammav:=MakeGamma(dw,mu,true):
```

In this case, *gammav* is a vector with elements $\gamma_1, \gamma_2, \gamma_3$.

```
> Allhat:=QCmatrices(Amin,Bmin,Cmin,Dmin,dqw,gammav):
> Ahat := map(simplify,Allhat[1]): Bhat :=
> Allhat[2]: Chat := map(simplify,Allhat[3]): Dhat := Allhat[4]:
> Ahatcross:=Ahat-Bhat.Chat:
```

For typographical reasons we show only the transpose of $\widehat{A}$:

$$\widehat{A}^{\top} = \begin{bmatrix} 0 & 0 & {p_0}^2 p_2 & 0 \\ 1 & 0 & -{p_0}^2 - 2\,p_0 p_2 & 0 \\ 0 & 1 & 2\,p_0 + p_2 & 0 \\ 0 & 0 & -{p_0}^2 p_2 + \gamma_1 {p_0}^2 + 2\,\gamma_1 p_0 p_2 - 2\,{\gamma_1}^2 p_0 - {\gamma_1}^2 p_2 + {\gamma_1}^3 & \gamma_1 \end{bmatrix}.$$

Next, we calculate the matrices TA and TZ which bring $\widehat{A}$ and $\widehat{A}^{\times}$ in upper-triangular form.

```
> TA:=MakeBasisA(orderedA,dqw,gammav):
> TZ:=MakeBasisZ(orderedZ,dqw,gammav):
```

Finally, the matrix S which will bring $\widehat{A}$ and $\widehat{A}^{\times}$ in complementary triangular form, is constructed. Applying S to to previously calculated matrices *Ahat* etc., will result in matrices *Atr, Atrcross, Btr, Ctr* (and *Dtr*) with *Atr* and *Atrcross* indeed in complementary triangular form.

```
> S:=UpperLowerTransformation(TA,TZ,dqw);
> Sinv:=MatrixInverse(S):
> Atr:=convert(map(simplify,Sinv.Ahat.S),Matrix):
> Atrcross:=convert(map(simplify,Sinv.Ahatcross.S),Matrix):
> Btr:=convert(Sinv.Bhat,Matrix);
> Ctr:=convert(map(simplify,Chat.S),Matrix):
> Dtr:=Dhat;
```

$$\widehat{A} = \begin{bmatrix} \gamma_1 & 0 & \alpha_1 & \alpha_2 \\ 0 & p_0 & \alpha_3 & \alpha_4 \\ 0 & 0 & p_0 & \alpha_5 \\ 0 & 0 & 0 & p_2 \end{bmatrix},$$

where

$$\begin{aligned} \alpha_1 &= {p_2}^2 - \gamma_1 p_2 - p_0 p_2 + \gamma_1 p_0, \\ \alpha_2 &= \frac{-{p_0}^2 p_2 + {p_0}^2 \gamma_1 + 2\,p_0 \gamma_1 p_2 - 2\,{\gamma_1}^2 p_0 - {\gamma_1}^2 p_2 + {\gamma_1}^3}{p_0 - p_2}, \\ \alpha_3 &= -\frac{{p_2}^2 - \gamma_1 p_2 - p_0 p_2 + \gamma_1 p_0}{-p_0 + \gamma_1}, \\ \alpha_4 &= -\frac{p_0 p_2 - \gamma_1 p_0 - \gamma_1 p_2 + {\gamma_1}^2}{p_0 - p_2}, \\ \alpha_5 &= \frac{{p_0}^2 + {\gamma_1}^2 - 2\,\gamma_1 p_0}{p_0 - p_2}, \end{aligned}$$

and

$$\widehat{A}^{\times} = \begin{bmatrix} p_0 & 0 & 0 & 0 \\ 1 & p_0 & 0 & 0 \\ \beta_1 & 0 & p_2 & 0 \\ \beta_2 & 0 & \beta_3 & \gamma_1 \end{bmatrix}$$

with

$$\begin{aligned} \beta_1 &= -\frac{-p_0+\gamma_1}{-p_2+\gamma_1}, \\ \beta_2 &= \frac{p_0-p_2}{-p_0+\gamma_1}, \\ \beta_3 &= \frac{-{p_0}^2p_2+2\,p_0{p_2}^2-{p_2}^3+{p_0}^2\gamma_1-2\,p_0\gamma_1p_2+{p_2}^2\gamma_1}{\left(-p_0+\gamma_1\right)^2}. \end{aligned}$$

Having obtained a realization (*Atr,Btr,Ctr,Dtr*) with *Atr* and *Atrcross* in complementary triangular form, the elementary factors follow from a call to **MakeFactorization**; see Subsection 11.6.4.

```
> Allfactors:=
> map(simplify,MakeFactorization(Atr,Btr,Ctr,lambda)):
```

Finally, the factors in the factorization can be shown:

```
> afactors := Vector[row](dqw,0): for k
> from 1 to dqw do afactors[k]:=Allfactors[k](lambda): end do:
> print(‘Elementary factors‘=afactors);
```

The elementary factors (with ordering left to right and top to bottom) are:

$$\begin{bmatrix} 1 & \dfrac{-r_0+\gamma_1}{(\lambda-\gamma_1)(p_0-\gamma_1)(p_2-\gamma_1)} \\ 0 & \dfrac{\lambda-p_0}{\lambda-\gamma_1} \end{bmatrix}, \quad \begin{bmatrix} 1 & \dfrac{-r_0+p_0}{(\lambda-p_0)(p_0-p_2)(p_0-\gamma_1)} \\ 0 & 1 \end{bmatrix},$$

$$\begin{bmatrix} 1 & \dfrac{r_0-p_2}{(\lambda-p_0)(p_2-\gamma_1)(p_0-p_2)} \\ 0 & \dfrac{\lambda-p_2}{\lambda-p_0} \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{\lambda-\gamma_1}{\lambda-p_2} \end{bmatrix}.$$

One may apply a final test:

```
> Wtest:=Factors2Transfer(Allfactors,lambda):
```

and show that *Wtest* is equal to *W*.

12.5.6 Concluding remarks

The implementation of the method of Section 12.2 in Maple, as done in the foregoing Sections 12.5 and 11.6, has two main features.

First of all, it allows for getting (quasi-)complete factorizations of a proper rational 2×2 matrix function W as in formula (12.37), completely defined in symbolic names. Secondly, one can calculate all feasible orderings of the set of the different elements of $\sigma(A) \cup \sigma(Z)$ such that (11.20) holds, where A and Z are the first companion matrices associated with the pole polynomial and zero-polynomial of W, respectively. From there on, one can calculate the corresponding orderings of poles and zeros of W and subsequently, the corresponding factorizations.

The advantages referred to above have also their drawbacks. We have already mentioned in Section 11.6 that in Maple the calculation of all permutations (i.e. orderings) is very much time- and cpu-consuming for sets with more than, say, 8 elements. To overcome this problem, one could get one ordering by applying Johnson's rule, see Section 12.3. Note that Johnson's rule is of order $k \log(k)$, where k is the number of different elements of $\sigma(A) \cup \sigma(Z)$, while our procedure is at least of order $k!$. On the other hand, as soon as Johnson's rule has been used to produce a desired ordering, the Maple procedures given in this section can be used to calculate corresponding quasicomplete factorizations. We also note that this Maple implementation of Johnson's rule in producing one valid ordering can be used in the same manner in the case of complete factorizations, $h = 1$, cf., Section 11.6.

For the sake of completeness, a Maple implementation of Johnson's rule is provided here. The calling sequence of the Maple procedure **JohnsonRule** is just the same as the procedure **GetAllMOrderings** in Section 11.6. Although not part of the specific Johnson algorithm, the first element of the Maple output will be the value of h as used in condition (11.20). Moreover, to be completely in line with the procedure **GetAllMOrderings** which has as second output argument, the number of found orderings, we add also in the output of the procedure **JohnsonRule** as second argument the value 1 since in contrast with our implementation, this procedure will give only one ordering.

JohnsonRule	calculate ordering of poles and zeros
Calling sequence	JohnsonRule(mA,mZ,mu)
Parameters	mA - Vector (of multiplicties of poles) mZ - Vector (of multiplicties of zeros) mu - Vector (of different poles and zeros)
Output	List, with first element is h, see condition (11.20) and with second argument, the number 1 third element, a list with the ordering of poles and the ordering of zeros

```
> JohnsonRule:=proc(mA,mZ,mu)
> local V1,V2,VZ1,VA2,S,lv1,nm,k,h;
> nm:=Dimension(mu):VZ1:=[]:VA2:=[]:V1:=[]:V2:=[]:
> for k from 1 to nm do if (mZ[k]<mA[k]) then
> V1:=[op(V1),k]:VZ1:=[op(VZ1),mZ[k]] :
> else V2:=[op(V2),k]:VA2:=[op(VA2),-1*mA[k]]: end if:end do:
> lv1:=ord(VZ1)[2]:S:=[]:for k from 1 to nops(V1) do
> S:=[op(S),V1[lv1[k]]]:end do: lv1:=ord(VA2)[2]:
> for k from 1 to nops(V2) do
> S:=[op(S),V2[lv1[k]]]:end do: h:=1: while (h<nm) and
> not (TestOrderingMAZ(mA,mZ,S,h)) do h:=h+1: end do:
> return(h,1,[GetOrderedVector(muA,mu,S),
> GetOrderedVector(muZ,mu,S)]);
> end proc;
```

The procedure **JohnsonRule** uses the following sorting procedure **ord**.

ord	sort a list in increasing order
Calling sequence	ord(x)
Parameters	x - Maple list
Output	List, with first element is the sorted list and second element, the reordered index positions, the permutation vector

```
> ord:= proc(x) local i,s;
> s:=sort([seq([x[i],i],i=1..nops(x))],(a,b)->evalb(a[1]<b[1]));
> [map(x->x[1],s),map(x->x[2],s)] end:
```

A second point of consideration is the use of the built in Maple procedure *IntersectionBasis* in our procedure **UpperLowerTransformation**. In case of symbolic names for the poles and zeros, this procedure is again very slow. For instance, a problem with McMillan degree n is 6, and quasidegree $\delta_q(W) = 8$ it takes about 3 minutes to calculate for all three found orderings their minimal factorizations.

Notes

For the largest part the first four sections in this chapter are based on and an elaboration of [23]. The Maple procedures presented in Section 12.5 were made by Johan F. Kaashoek. As we have mentioned the running time of Johnson's rule is $\mathcal{O}(k \log k)$ for a 2MFSP with k jobs. The analogous problem with three or more machines is NP-hard.

Part IV

Stability of Factorization and of Invariant Subspaces

Numerical computations of the factors in a factorization lead in a natural way to the problem of stability of factors under small perturbations of the initial matrix function. The entire present part is devoted to this problem. The state space approach to factorization allows one to deal with the problem of stable factors in terms of stability of invariant subspaces of matrices or operators. It turns out that in general the factors of a minimal factorization of a rational matrix function are unstable. Only in some special cases, including the case of canonical factorization, we have stability of the factors. A full description of these stable cases is given. This part consists of three chapters (13–15).

Chapter 13 has partly a preparatory character. Some illustrative examples are given, and the theory of distances between subspaces is reviewed. The stability of the factors in a canonical factorization is proved. Applications to transfer functions and to Riccati equations are included. Chapter 14 is the main chapter of this part. The notion of a stable invariant subspace is introduced, and all stable invariant subspaces of a matrix are described. The stronger notion of Lipschitz stability of subspaces is studied separately. For a matrix it is shown that the Lipschitz stable invariant subspaces coincide with the spectral subspaces. On the basis of these theorems a full description is given of all minimal factorizations of finite-dimensional systems with stable and Lipschitz stable factors. Applications are given for factorizations of rational matrix function and matrix polynomials. The results are specified further for Riccati equations. Chapter 15 contains the study of factorization and stability of the factors in the real case. The results are based on the study of the stability of real invariant subspaces.

Chapter 13

Stability of Spectral Divisors

In numerical computations of minimal factors of a given transfer function questions concerning the conditioning of the factors turn up naturally. According to the division theory developed in the previous chapters, all minimal factorizations may be obtained in an explicit way in terms of supporting projections of minimal systems. This fact allows one to reduce questions concerning the conditioning of minimal factorizations to questions concerning the stability of divisors of a system. In the present chapter we study the matter of stability of spectral divisors mainly. In this case the investigation can be carried out for finite- as well as for infinite-dimensional state spaces. The invariant subspace method employed in this chapter will also be used to prove that "spectral" solutions of an operator Riccati equation are stable. The case of minimal non-spectral factorizations will be considered in the next chapter.

13.1 Examples and first results for the finite-dimensional case

The property of having non-trivial minimal factorizations is ill-conditioned. For example it may happen that a transfer function admits non-trivial minimal factorizations while after a small perturbation the perturbed function has no such factorizations. On the other hand it may also happen that the perturbed function admits non-trivial minimal factorizations while the original function does not have this property. To see this we consider the following examples. Let

$$W_\varepsilon(\lambda) = \begin{bmatrix} 1 + \dfrac{1}{\lambda} & \dfrac{\varepsilon}{\lambda^2} \\ 0 & 1 + \dfrac{1}{\lambda} \end{bmatrix}. \tag{13.1}$$

For each ε this matrix function is the transfer function of the unital minimal system $\Theta_\varepsilon = (A_\varepsilon, I, I; \mathbb{C}^2, \mathbb{C}^2)$, where I is the identity on $\mathbb{C}^2$ and

$$A_\varepsilon = \begin{bmatrix} 0 & \varepsilon \\ 0 & 0 \end{bmatrix}.$$

Note that the associate main operator $A_\varepsilon^\times$ of Θ_ε is given by $A_\varepsilon^\times = A_\varepsilon - I$. To find a non-trivial minimal factorization of the function (13.1), we have to find non-trivial divisors of the system Θ_ε (cf., Theorem 9.3), i.e., we must look for non-trivial subspaces M and $M^\times$ of $\mathbb{C}^2$, invariant under A_ε and $A_\varepsilon - I$, respectively, such that

$$M \dot{+} M^\times = \mathbb{C}^2.$$

Note that A_ε and $A_\varepsilon - I$ have the same invariant subspaces, and for $\varepsilon \neq 0$ there is only one such space of dimension one, namely the first coordinate space. It follows that for $\varepsilon \neq 0$ the function (13.1) has no non-trivial minimal factorizations. For $\varepsilon = 0$ we have

$$W_0(\lambda) = \begin{bmatrix} 1 + \dfrac{1}{\lambda} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 + \dfrac{1}{\lambda} \end{bmatrix}$$

and this factorization is minimal, because the McMillan degree of $W_0(\lambda)$ is equal to 2 and the McMillan degree of each of the factors is one.

Next consider the function

$$W_\varepsilon(\lambda) = \begin{bmatrix} 1 & \dfrac{1}{\lambda^2 - \varepsilon^2} \\ 0 & 1 \end{bmatrix}.$$

Put

$$A_\varepsilon = \begin{bmatrix} \varepsilon & 1 \\ 0 & -\varepsilon \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Then W_ε is the transfer function of the unital system $\Theta_\varepsilon = (A_\varepsilon, B, C; \mathbb{C}^2, \mathbb{C}^2)$. As Θ_ε is minimal, the McMillan degree of W_ε is equal to 2. For $\varepsilon \neq 0$ we have the following factorization

$$W_\varepsilon(\lambda) = \begin{bmatrix} 1 & \dfrac{1}{2\varepsilon(\lambda - \varepsilon)} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \dfrac{-1}{2\varepsilon(\lambda + \varepsilon)} \\ 0 & 1 \end{bmatrix}.$$

By comparing the McMillan degrees of the factors with the McMillan degree of W_ε, we see that this factorization is minimal. On the other hand, as has been

established at the end of Section 9.1, the function

$$W_0(\lambda) = \begin{bmatrix} 1 & \dfrac{1}{\lambda^2} \\ 0 & 1 \end{bmatrix}.$$

does not admit any non-trivial minimal factorization.

Although the first example proves that in general minimal factorizations are not stable, the next theorem shows that in an important case the possibility to factorize in a minimal way is stable under small perturbations. This theorem will appear as a corollary to the main stability theorem to be proved in this chapter.

Theorem 13.1. *Consider the minimal realization*

$$W_0(\lambda) = I_m + C_0(\lambda I_n - A_0)^{-1}B_0,$$

and assume that W_0 admits a (minimal) factorization

$$W_0 = W_{01}W_{02}, \qquad W_{0j}(\lambda) = I_m + C_{0j}(\lambda I_{n_j} - A_{0j})^{-1}B_{0j},$$

where $n = n_1 + n_2$ and the factors W_{01} and W_{02} have neither common zeros nor common poles. Then, given $\varepsilon > 0$, there exists $\omega > 0$ with the following property. If A, B and C are matrices of appropriate sizes, with

$$\|A - A_0\| + \|B - B_0\| + \|C - C_0\| < \omega, \tag{13.2}$$

then the realization $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ is minimal and W admits a (minimal) factorization:

$$W(\lambda) = W_1W_2, \qquad W_j(\lambda) = I_m + C_j(\lambda I_{n_j} - A_j)^{-1}B_j,$$

such that the factors W_1 and W_2 have no common zeros and no common poles and

$$\|A_{0j} - A_j\| < \varepsilon, \qquad \|B_{0j} - B_j\| < \varepsilon, \qquad \|C_{0j} - C_j\| < \varepsilon$$

for $j = 1, 2$.

Later we shall avoid the ε/ω-language and give more explicit formulas for the relation between the quantity in the left-hand side of (13.2) and the perturbation of the factors (see Theorem 13.7). In Section 13.4 it will also be shown that the factors change analytically whenever the operators appearing in the minimal realization of the original function do so (see Theorem 13.8).

The results referred to above will appear as corollaries to infinite-dimensional stability theorems for certain divisors of systems, which deal mainly with the case of spectral factorization (see Section 13.3). In the next chapter the case of stable non-spectral minimal factorizations will be completely described (see Theorem 14.9).

The next section is of preliminary nature; there we describe the relation between angular operators and the minimal and maximal opening between subspaces. In Section 13.5 we employ the method of Section 13.3 to prove stability for certain solutions of the Riccati equation.

13.2 Opening between subspaces and angular operators

From the description of the factors of a system in terms of angular operators (see Theorem 5.5) it is clear that for our purposes it is important to know how the angular operator changes when the operators in the system are perturbed a little. For this reason we study the properties of angular operators in terms of the minimal and maximal opening between certain subspaces.

Let M_1 and M_2 be closed subspaces of the Banach space X. The number

$$\eta(M_1, M_2) \;=\; \inf\{\|x+y\| \mid x \in M_1,\, y \in M_2,\, \max(\|x\|, \|y\|) = 1\}$$

will be called the *minimal opening* between M_1 and M_2. Note that always $0 \leq \eta \leq 1$ except when both M_1 and M_2 are the zero space in which case $\eta = \infty$. It is well known (see [71], Lemma 1) that $\eta(M_1, M_2) > 0$ if and only if $M_1 \cap M_2 = \{0\}$ and $M_1 \dotplus M_2$ is closed. If Π is a projection of the space X, then

$$\max\{\|\Pi\|, \|I-\Pi\|\} \;\leq\; \frac{1}{\eta(\operatorname{Im}\Pi, \operatorname{Ker}\Pi)}. \tag{13.3}$$

To see this, note that for each $z \in X$ we have

$$\|z\| \;=\; \|\Pi z + (I-\Pi)z\| \;\geq\; \eta(\operatorname{Im}\Pi, \operatorname{Ker}\Pi) \cdot \max(\|\Pi z\|, \|(I-\Pi)z\|).$$

Sometimes it will be convenient to describe $\eta(M_1, M_2)$ in terms of the *minimal angle* $\varphi_{\min}$ between M_1 and M_2. By definition (cf., [71]) this quantity is given by the following formulas:

$$0 \;\leq\; \varphi_{\min} \leq \frac{\pi}{2}, \qquad \sin\varphi_{\min} \;=\; \eta(M_1, M_2).$$

Now let us assume that M_1 and M_2 are closed subspaces of a Hilbert space H with inner product $< .,. >$, and let Q_1 and Q_2 be the orthogonal projections of H onto M_1 and M_2, respectively. Note that

$$\inf\{\|x+y\| \mid y \in M_2\} \;=\; \|x - Q_2 x\|, \qquad x \in M_1.$$

It follows that

$$\eta(M_1, M_2) \;=\; \min\left\{ \inf_{0 \neq x \in M_1} \frac{\|x - Q_2 x\|}{\|x\|},\; \inf_{0 \neq y \in M_2} \frac{\|y - Q_1 y\|}{\|y\|} \right\}.$$

If both M_1 and M_2 are non-trivial, then the two infima in the right-hand side of

the previous identity are equal. This follows from

$$\begin{aligned}\inf_{0\neq x\in M_1}\left(\frac{\|x-Q_2x\|}{\|x\|}\right)^2 &= \inf_{0\neq x\in M_1}\frac{\|x\|^2-\|Q_2x\|^2}{\|x\|^2} = 1-\sup_{0\neq x\in M_1}\frac{\|Q_2x\|^2}{\|x\|^2}\\ &= 1-\sup_{\substack{x\in M_1\\ x\neq 0}}\sup_{\substack{y\in M_2\\ y\neq 0}}\frac{|<x,y>|^2}{\|x\|^2\|y\|^2}\\ &= 1-\sup_{\substack{y\in M_2\\ y\neq 0}}\sup_{\substack{x\in M_1\\ x\neq 0}}\frac{|<x,y>|^2}{\|x\|^2\|y\|^2}\\ &= 1-\sup_{0\neq y\in M_2}\frac{\|Q_1y\|^2}{\|y\|^2} = \inf_{0\neq y\in M_2}\left(\frac{\|y-Q_1y\|}{\|y\|}\right)^2.\end{aligned}$$

From the previous equalities it also follows that

$$1-\eta(M_1,M_2)^2 = \sup_{0\neq x\in M_1}\frac{\|Q_2x\|^2}{\|x\|^2} = \sup_{0\neq y\in M_2}\frac{\|Q_1y\|^2}{\|y\|^2}, \tag{13.4}$$

provided both M_1 and M_2 contain nonzero elements.

Returning to the Banach space case, put

$$\rho(M_1,M_2) = \sup_{0\neq x\in M_1}\inf_{y\in M_2}\frac{\|x-y\|}{\|x\|}.$$

If $M_1=\{0\}$, then $\rho(M_1,M_2)=0$ by definition. When P and Q are projections of X, then for $x\in\operatorname{Im}P$ and $y\in\operatorname{Im}Q$ we have

$$\inf_{y\in\operatorname{Im}Q}\|x-y\| \le \|Px-Qx\| \le \|P-Q\|\cdot\|x\|,$$

and thus $\rho(\operatorname{Im}P,\operatorname{Im}Q)\le\|P-Q\|$. The number

$$\operatorname{gap}(M_1,M_2) = \max\{\rho(M_1,M_2),\,\rho(M_2,M_1)\}$$

is the so-called *gap* (or *maximal opening*) between the subspaces M_1 and M_2. There exists an extensive literature on this concept, see, e.g., [86] and the references given there. From what we remarked above we see that the gap has the following property: if P and Q are projections of X, then $\operatorname{gap}(\operatorname{Im}P,\operatorname{Im}Q)\le\|P-Q\|$.

In the Hilbert space case we actually have $\operatorname{gap}(\operatorname{Im}P,\operatorname{Im}Q) = \|P-Q\|$, provided that the projections P and Q are orthogonal. Furthermore

$$\rho(M_2,M_1^\perp) = \sqrt{1-\eta(M_1,M_2)^2} = \cos\varphi_{\min} \tag{13.5}$$

whenever $M_1\neq\{0\}$. To see this, note that for $M_2\neq\{0\}$

$$\rho(M_2,M_1^\perp) = \sup_{0\neq y\in M_2}\frac{\|y-(I-Q_1)y\|}{\|y\|} = \sup_{0\neq y\in M_2}\frac{\|Q_1y\|}{\|y\|},$$

where Q_1 is the orthogonal projection onto M_1. But then we can use (13.4) to get formula (13.5). If $M_2 = \{0\}$, then (13.5) holds trivially.

The next lemma is well known, but explicit references are difficult to give. For this reason it will be presented with full proof.

Lemma 13.2. *Let Π_0, Π and Π_1 be projections of the Banach space X, and assume that $\operatorname{Ker}\Pi_0 = \operatorname{Ker}\Pi = \operatorname{Ker}\Pi_1$. Let R and R_1 be the angular operator relative to Π_0 of the angular subspaces $\operatorname{Im}\Pi$ and $\operatorname{Im}\Pi_1$, respectively. The following statements hold true:*

(i) $\eta(\operatorname{Ker}\Pi_0, \operatorname{Im}\Pi_0) \cdot \rho(\operatorname{Im}\Pi_1, \operatorname{Im}\Pi) \le \|R_1 - R\|$;

(ii) *if $\rho(\operatorname{Im}\Pi_1, \operatorname{Im}\Pi) < \eta(\operatorname{Ker}\Pi, \operatorname{Im}\Pi)$, then*

$$\|R_1 - R\| \le \frac{\rho(\operatorname{Im}\Pi_1, \operatorname{Im}\Pi)(1 + \|R\|)}{\eta(\operatorname{Ker}\Pi, \operatorname{Im}\Pi) - \rho(\operatorname{Im}\Pi_1, \operatorname{Im}\Pi)}.$$

In particular, if $\rho(\operatorname{Im}\Pi_1, \operatorname{Im}\Pi_0) < \eta(\operatorname{Ker}\Pi_0, \operatorname{Im}\Pi_0)$, then

$$\|R_1\| \le \frac{\rho(\operatorname{Im}\Pi_1, \operatorname{Im}\Pi_0)}{\eta(\operatorname{Ker}\Pi_0, \operatorname{Im}\Pi_0) - \rho(\operatorname{Im}\Pi_1, \operatorname{Im}\Pi_0)}. \tag{13.6}$$

Finally, if X is a Hilbert space and Π_0 is an orthogonal projection, then $\|R_1\| = \operatorname{ctg}\varphi_{\min}$, where $\varphi_{\min}$ is the minimal angle between $\operatorname{Ker}\Pi_0$ and $\operatorname{Im}\Pi_1$.

Proof. First we present the proof of the second part of the lemma. We begin with formula (13.6). Put $\rho_0 = \rho(\operatorname{Im}\Pi_1, \operatorname{Im}\Pi_0)$ and $\eta_0 = \eta(\operatorname{Ker}\Pi_0, \operatorname{Im}\Pi_0)$. Recall (cf., Proposition 5.1) that

$$R_1 = (\Pi_1 - \Pi_0)|_{\operatorname{Im}\Pi_0}. \tag{13.7}$$

For $x \in \operatorname{Im}\Pi_1$ and $z \in \operatorname{Im}\Pi_0$ we have

$$\|(\Pi_1 - \Pi_0)x\| = \|(I - \Pi_0)x\| = \|(I - \Pi_0)(x - z)\| \le \|I - \Pi_0\| \cdot \|x - z\|.$$

Taking the infimum over all $z \in \operatorname{Im}\Pi_0$ and using inequality (13.3), one sees that

$$\|(\Pi_1 - \Pi_0)x\| \le \frac{\rho_0}{\eta_0}\|x\|, \qquad x \in \operatorname{Im}\Pi_1. \tag{13.8}$$

Now recall that $R_1 y + y \in \operatorname{Im}\Pi_1$ for each $y \in \operatorname{Im}\Pi_0$. As $R_1 y \in \operatorname{Ker}\Pi_0 = \operatorname{Ker}\Pi_1$, we see from (13.7) that

$$(\Pi_1 - \Pi_0)(R_1 y + y) = R_1 y.$$

So, using (13.8), we obtain

$$\|R_1 y\| \le \frac{\rho_0}{\eta_0}\|R_1 y + y\|, \qquad y \in \operatorname{Im}\Pi_0.$$

It follows that $(1-\rho_0\eta_0^{-1})\|R_1y\| \le \rho_0\eta_0^{-1}\|y\|$ for each $y \in \operatorname{Im}\Pi_0$, which proves the inequality (13.6).

Next, assume that X is a Hilbert space, and that Π_0 is an orthogonal projection. If $\operatorname{Ker}\Pi_0 = \{0\}$, then $R_1 = 0$ and $\varphi_{\min} = \frac{\pi}{2}$, and hence, in that case, we certainly have $\|R_1\| = \operatorname{ctg}\varphi_{\min}$. So we assume that $\operatorname{Ker}\Pi_0 \neq \{0\}$. Then, by (13.4),

$$\cos^2\varphi_{\min} \;=\; 1-\eta(\operatorname{Ker}\Pi_0, \operatorname{Im}\Pi_1)^2 \;=\; \sup_{0\neq x\in\operatorname{Im}\Pi_1}\left(\frac{\|(I-\Pi_0)x\|}{\|x\|}\right)^2.$$

Given $x \in \operatorname{Im}\Pi_1$, there exists $y \in \operatorname{Im}\Pi_0$ such that $x = R_1y+y$. As $(I-\Pi_0)x = R_1y$, this implies that

$$\begin{aligned}\cos^2\varphi_{\min} &= \sup_{0\neq y\in\operatorname{Im}\Pi_0}\frac{\|R_1y\|^2}{\|R_1y+y\|^2}\\ &= \sup_{0\neq y\in\operatorname{Im}\Pi_0}\frac{\|R_1y\|^2}{\|y\|^2+\|R_1y\|^2} \;=\; \frac{\|R_1\|^2}{1+\|R_1\|^2}.\end{aligned}$$

Hence, $\|R_1\| = \operatorname{ctg}\varphi_{\min}$, and we have proved the second part of the theorem.

Next we establish (i). Take an arbitrary $y \in \operatorname{Im}\Pi_1$. Then $y = R_1x + x$ for some $x \in \operatorname{Im}\Pi_0$. Note that $Rx + x \in \operatorname{Im}\Pi$. So

$$\inf_{z\in\operatorname{Im}\Pi}\|y-z\| \;\le\; \|y-(Rx+x)\| \le \|R_1-R\|\cdot\|x\|.$$

Then $\|y\| = \|R_1x+x\| \ge \eta_0\|x\|$, where $\eta_0 = \eta(\operatorname{Ker}\Pi_0, \operatorname{Im}\Pi_0)$. It follows that $\eta_0 d(y, \operatorname{Im}\Pi) \le \|R_1-R_2\|\cdot\|y\|$. This proves (i).

Finally, we turn to statement (ii). Recall that

$$R_1 = (\Pi_1-\Pi_0)|_{\operatorname{Im}\Pi_0}, \qquad R = (\Pi-\Pi_0)|_{\operatorname{Im}\Pi_0}.$$

So, $(R_1-R)x = (\Pi_1-\Pi)x$ for each $x \in \operatorname{Im}\Pi_0$. Let $\widetilde{R}$ be the angular operator of $\operatorname{Im}\Pi_1$ with respect to Π. Note that $\widetilde{R}y = (\Pi_1-\Pi)y$ for all $y \in \operatorname{Im}\Pi$. Take $x \in \operatorname{Im}\Pi_0$. As $\operatorname{Im}(I-\Pi) = \operatorname{Ker}\Pi = \operatorname{Ker}\Pi_1$, we have $(\Pi_1-\Pi)x = (\Pi_1-\Pi)\Pi x = \widetilde{R}\Pi x$. Now

$$\|\Pi x\| \;\le\; \|(\Pi-\Pi_0)x + \|\Pi_0 x\| \;\le\; (\|R\|+1)\,\|x\|.$$

It follows that

$$\|(R_1-R)x\| \;\le\; \|\widetilde{R}\|\,\big(\|R\|+1\big)\|x\|. \tag{13.9}$$

As $\rho(\operatorname{Im}\Pi_1, \operatorname{Im}\Pi) < \eta = \eta(\operatorname{Ker}\Pi, \operatorname{Im}\Pi)$, we can use formula (13.6) for Π instead of Π_0 to show that

$$\|\widetilde{R}\| \;\le\; \frac{\rho(\operatorname{Im}\Pi_1, \operatorname{Im}\Pi)}{\eta-\rho(\operatorname{Im}\Pi_1, \operatorname{Im}\Pi)}.$$

Substituting this in (13.9) gives the desired inequality. □

The following lemma will be useful in the next section.

Lemma 13.3. *Let P, $P^\times$, Q and $Q^\times$ be projections of the Banach space X, and put $\alpha_0 = \frac{1}{6}\eta(\operatorname{Im} P, \operatorname{Im} P^\times)(\|P^\times\| + 1)^{-1}$. Assume $X = \operatorname{Im} P \dotplus \operatorname{Im} P^\times$ and*

$$\|P - Q\| + \|P^\times - Q^\times\| < \alpha_0. \tag{13.10}$$

Then $X = \operatorname{Im} Q \dotplus \operatorname{Im} Q^\times$ and there is an invertible operator $S : X \to X$ such that

(i) $S[\operatorname{Im} Q] = \operatorname{Im} P, \quad S[\operatorname{Im} Q^\times] = \operatorname{Im} P^\times$,
(ii) $\max\{\|S - I\|, \|S^{-1} - I\|\} \le \beta\big(\|P - Q\| + \|P^\times - Q^\times\|\big)$,

where $\beta = 2\big(\alpha_0\eta(\operatorname{Im} P, \operatorname{Im} P^\times)\big)^{-1}$.

Proof. Recall that

$$\operatorname{gap}(\operatorname{Im} P, \operatorname{Im} Q) \le \|P - Q\|, \qquad \operatorname{gap}(\operatorname{Im} P^\times, \operatorname{Im} Q^\times) \le \|P^\times - Q^\times\|.$$

Thus condition (13.10) implies that

$$2\operatorname{gap}(\operatorname{Im} P, \operatorname{Im} Q) + 2\operatorname{gap}(\operatorname{Im} P^\times, \operatorname{Im} Q^\times) < \eta(\operatorname{Im} P, \operatorname{Im} P^\times).$$

But then we may apply [71], Theorem 2 to show that $X = \operatorname{Im} Q \dotplus \operatorname{Im} Q^\times$.

Note that (13.10) implies that $\|P - Q\| < 1/4$. Hence $S_1 = I + P - Q$ is invertible, and we can write $S_1^{-1} = I + V$ with $\|V\| \le \frac{4}{3}\|P - Q\| < \frac{1}{3}$. As $I - P + Q$ is invertible too, we have

$$\operatorname{Im} P = P(I - P + Q)X = PQX = (I + P - Q)QX = S_1(\operatorname{Im} Q). \tag{13.11}$$

Moreover,

$$\begin{aligned} S_1Q^\times S_1^{-1} - P^\times &= (I + P - Q)Q^\times(I + V) - P^\times \\ &= Q^\times + (P - Q)Q^\times + Q^\times V + (P - Q)Q^\times V - P^\times \\ &= Q^\times - P^\times + (P - Q)(Q^\times - P^\times) + (P - Q)P^\times + \\ &\quad + (Q^\times - P^\times)V + P^\times V + (P - Q)(Q^\times - P^\times)V + \\ &\quad + (P - Q)P^\times V. \end{aligned}$$

So $\|S_1Q^\times S_1^{-1} - P^\times\| \le 3\|Q^\times - P^\times\| + 3\|P - Q\| \cdot \|P^\times\|$. But then

$$\begin{aligned} \rho(\operatorname{Im} S_1Q^\times S_1^{-1}, \operatorname{Im} P^\times) &\le \|S_1Q^\times S_1^{-1} - P^\times\| \\ &\le 3\big(\|P - Q\| + \|P^\times - Q^\times\|\big)\big(\|P^\times\| + 1\big) \\ &\le \frac{1}{2}\eta(\operatorname{Im} P, \operatorname{Im} P^\times). \end{aligned}$$

Let Π_0 be the projection of X along $\operatorname{Im} P$ onto $\operatorname{Im} P^\times$, and let Π be the projection of X along $\operatorname{Im} Q$ onto $\operatorname{Im} Q^\times$. Put $\widetilde{\Pi} = S_1\Pi S_1^{-1}$. Then $\widetilde{\Pi}$ is a projection of X, and by (13.11) we have $\operatorname{Ker}\widetilde{\Pi} = \operatorname{Ker}\Pi_0$. Furthermore, $\operatorname{Im}\widetilde{\Pi} = \operatorname{Im} S_1Q^\times S_1^{-1}$,

and so we have

$$\rho(\operatorname{Im}\widetilde{\Pi}, \operatorname{Im}\Pi_0) \leq \frac{1}{2}\eta(\operatorname{Ker}\Pi_0, \operatorname{Im}\Pi_0).$$

Hence, if R denotes the angular operator of $\operatorname{Im}\widetilde{\Pi}$ with respect to Π_0, then because of formula (13.6) in Lemma 13.2, we get

$$\|R\| \leq \frac{1}{\alpha_0}(\|P-Q\| + \|P^\times - Q^\times\|). \tag{13.12}$$

Next, put $S_2 = I - R\Pi_0$, and set $S = S_2S_1$. Clearly, S_2 is invertible, in fact, $S_2^{-1} = I + R\Pi_0$. It follows that S is invertible too. From the properties of the angular operator one easily sees that with this choice of S statement (i) holds true. It remains to prove (ii).

To prove (ii) we simplify our notation. Put $d = \|P-Q\| + \|P^\times - Q^\times\|$, and let $\eta = \eta(\operatorname{Im}P, \operatorname{Im}P^\times)$. From $S = (I - R\Pi_0)(I + P - Q)$ and the fact that $\|P-Q\| < \frac{1}{4}$ one deduces that $\|S - I\| \leq \|P-Q\| + \frac{5}{4}\|R\| \cdot \|\Pi_0\|$. For $\|R\|$ an upper bound is given by (13.12), and from (13.3) we know that $\|\Pi_0\| \leq \eta^{-1}$. It follows that

$$\|S - I\| \leq d + \frac{5}{4}d(\alpha_0\eta)^{-1}. \tag{13.13}$$

Finally, we consider S^{-1}. Recall that $S_1^{-1} = I + V$ with $\|V\| \leq \frac{4}{3}\|P-Q\| \leq \frac{1}{3}$. Hence

$$\begin{aligned}\|S^{-1} - I\| &\leq \|V\| + \|V\| \cdot \|\Pi_0\| \cdot \|R\| + \|R\| \cdot \|\Pi_0\| \\ &\leq \frac{4}{3}\|P-Q\| + \frac{4}{3}\|R\| \cdot \|\Pi_0\| \\ &\leq \frac{4}{3}d + \frac{4}{3}d(\alpha_0\eta)^{-1}.\end{aligned}$$

Using the fact that $\alpha_0\eta \leq \frac{1}{6}$, it is easy to derive statement (ii) from (13.13) and the previous inequality. □

13.3 Stability of spectral divisors of systems

To state the main theorem of this section we need the following definition. If $\Theta = (A, B, C; X, Y)$ and $\Theta_0 = (A_0, B_0, C_0; X, Y)$ are two systems, then the *distance* between Θ and Θ_0 is defined to be

$$\|\Theta - \Theta_0\| = \|A - A_0\| + \|B - B_0\| + \|C - C_0\|.$$

In particular, we set $\|\Theta\| = \|A\| + \|B\| + \|C\|$. If $W(\lambda)$ and $W_0(\lambda)$ are the transfer functions of Θ and Θ_0, respectively, then

$$\|W(\lambda) - W_0(\lambda)\| \leq \frac{\|\Theta - \Theta_0\| \cdot \|\Theta\| \cdot \|\Theta_0\|}{\|A\| \cdot \|A_0\|},$$

provided $|\lambda| > 2\max\{\|A\|, \|A_0\|\}$.

Theorem 13.4. *Let $\Theta_0 = (A_0, B_0, C_0; X, Y)$ be a system with a supporting projection Π_0, and put $A_0^\times = A_0 - B_0C_0$. Assume that*

$$\operatorname{Ker}\Pi_0 = \operatorname{Im}P(A_0;\Gamma), \qquad \operatorname{Im}\Pi_0 = \operatorname{Im}P(A_0^\times;\Gamma^\times),$$

where Γ and $\Gamma^\times$ are Cauchy contours which split the spectra of A_0 and $A_0^\times$, respectively. Then there exist positive constants α, β_1 and β_2 such that the following holds. If $\Theta = (A, B, C; X, Y)$ is a system such that $\|\Theta - \Theta_0\| < \alpha$, then Γ splits the spectrum of A, $\Gamma^\times$ splits the spectrum of $A^\times = A - BC$,

$$X \;=\; \operatorname{Im}P(A;\Gamma) \dotplus \operatorname{Im}P(A^\times;\Gamma^\times),$$

the projection Π of X along $\operatorname{Im}P(A;\Gamma)$ onto $\operatorname{Im}P(A^\times;\Gamma^\times)$ is a supporting projection for Θ, and there exists a similarity transformation S such that

$$\|S - I\| \;\le\; \beta_1\|\Theta - \Theta_0\|,$$

$\Pi_0 = S\Pi S^{-1}$, and the projection Π_0 is a supporting projection for the system $\widetilde{\Theta} = (SAS^{-1}, SB, CS^{-1}; X, Y)$ while for the corresponding factors we have

(i) $\|\operatorname{pr}_{I-\Pi_0}(\Theta_0) - \operatorname{pr}_{I-\Pi_0}(\widetilde{\Theta})\| \;\le\; \beta_2\|\Theta - \Theta_0\|$,

(ii) $\|\operatorname{pr}_{\Pi_0}(\Theta_0) - \operatorname{pr}_{\Pi_0}\widetilde{\Theta})\| \;\le\; \beta_2\|\Theta - \Theta_0\|$.

Furthermore, if Θ_0 is minimal and the spaces X and Y are finite-dimensional, then α can be chosen such that Θ is minimal whenever $\|\Theta - \Theta_0\| < \alpha$.

From the proof of the theorem it will become clear that in the first part of the theorem we may take for the constant α the following quantity:

$$\alpha \;=\; \frac{1}{1 + \|\Theta_0\|}\min\left\{1,\; \frac{1}{2\gamma},\; \frac{\alpha_0\pi}{2\gamma^2\ell}\right\},$$

where ℓ is the maximum of the lengths of the curves Γ and $\Gamma^\times$,

$$\gamma = \max\left\{\max_{\lambda\in\Gamma}\|(\lambda - A_0)^{-1}\|,\; \max_{\lambda\in\Gamma^\times}\|(\lambda - A_0^\times)^{-1}\|\right\},$$

and $\alpha_0 = \frac{1}{6}\eta(\operatorname{Ker}\Pi_0, \operatorname{Im}\Pi_0)(\|P(A_0^\times;\Gamma)\| + 1)^{-1}$. Furthermore, we may take

$$\begin{aligned}
\beta_1 &= 4\big(1 + \|\Theta_0\|\big)\,\gamma^2\ell\left(\pi\alpha_0\eta(\operatorname{Ker}\Pi_0, \operatorname{Im}\Pi_0)\right)^{-1},\\
\beta_2 &= \frac{9}{\eta(\operatorname{Ker}\Pi_0, \operatorname{Im}\Pi_0)^3}\left(1 + \frac{2\gamma^2\ell}{\pi\alpha_0}\,\|\Theta_0\|\,\big(1 + \|\Theta_0\|\big)\right).
\end{aligned}$$

To prove Theorem 13.4 we first establish the following auxiliary result.

Theorem 13.5. *Let $\Theta_0 = (A_0, B_0, C_0; X, Y)$ be a system with supporting projection Π_0, and assume that*

$$\operatorname{Ker}\Pi_0 = \operatorname{Im} P, \qquad \operatorname{Im}\Pi_0 = \operatorname{Im} P^\times,$$

where P and $P^\times$ are given projections of X. Put

$$\alpha_0 \;=\; \frac{1}{6}\eta(\operatorname{Im} P, \operatorname{Im} P^\times)\big(\|P^\times\| + 1\big)^{-1}.$$

Let $\Theta = (A, B, C; X, Y)$ be another system, and let Q and $Q^\times$ be projections of X such that

$$A[\operatorname{Im} Q] \subset \operatorname{Im} Q, \qquad A^\times[\operatorname{Im} Q^\times] \subset \operatorname{Im} Q^\times, \tag{13.14}$$

$$\|P - Q\| + \|P^\times - Q^\times\| \;<\; \alpha_0. \tag{13.15}$$

Then $X = \operatorname{Im} Q \dotplus \operatorname{Im} Q^\times$. Moreover there exists an invertible operator $S : X \to X$ such that $S^{-1}\Pi_0 S$ is the projection Π of X onto $\operatorname{Im} Q^\times$ along $\operatorname{Im} Q$, the projection Π_0 is a supporting projection for the system $\widetilde{\Theta} = (SAS^{-1}, SB, CS^{-1}; X, Y)$, while for the corresponding factors we have

$$\max\big\{\|\mathrm{pr}_{\,I-\Pi_0}(\Theta_0) - \mathrm{pr}_{\,I-\Pi_0}(\widetilde{\Theta})\|,\, \|\mathrm{pr}_{\,\Pi_0}(\Theta) - \mathrm{pr}_{\,\Pi_0}(\widetilde{\Theta})\|\big\} \;\le \tag{13.16}$$

$$\le \quad \frac{9}{\eta(\operatorname{Im} P, \operatorname{Im} P^\times)^3}\left(\|\Theta - \Theta_0\| + \frac{1}{\alpha_0}\|\Theta_0\| \cdot \big(\|P - Q\| + \|P^\times - Q^\times\|\big)\right).$$

Proof. From Lemma 13.3 we know that $X = \operatorname{Im} Q \dotplus \operatorname{Im} Q^\times$. Let Π be the projection of X along $\operatorname{Im} Q$ onto $\operatorname{Im} Q^\times$. Then (13.14) implies that Π is a supporting projection for Θ. Take S as in Lemma 13.3. Then we see from statement (i) in Lemma 13.3 that $S\Pi S^{-1} = \Pi_0$. But then it is clear that Π_0 is a supporting projection for $\widetilde{\Theta}$.

Let Θ_{01} and $\widetilde{\Theta}_1$ be the left factors of Θ_0 and $\widetilde{\Theta}$ associated with Π_0, and let Θ_{02} and $\widetilde{\Theta}_2$ be the corresponding right factors. From the definition of the factors (see Section 2.4) it is clear that

$$\|\Theta_{01} - \widetilde{\Theta}_1\| \;\le\; \|I - \Pi_0\|\left(\|A_0 - \widetilde{A}\| + \|B_0 - \widetilde{B}\| + \|C_0 - \widetilde{C}\|\right).$$

It follows that $\|\Theta_{01} - \widetilde{\Theta}_1\| \le \|I - \Pi_0\| \cdot \|\Theta_0 - \widetilde{\Theta}\|$. Similarly, $\|\Theta_{02} - \widetilde{\Theta}_2\| \le \|\Pi_0\| \cdot \|\Theta_0 - \widetilde{\Theta}\|$. Using (13.3) we obtain

$$\max_{i=1,2}\|\Theta_{0i} - \widetilde{\Theta}_i\| \le \frac{\|\Theta_0 - \widetilde{\Theta}\|}{\eta(\operatorname{Im} P, \operatorname{Im} P^\times)}. \tag{13.17}$$

As $\|\Theta_0 - \widetilde{\Theta}\| \le \|\Theta_0 - \Theta\| + \|\Theta - \widetilde{\Theta}\|$, it remains to compute a suitable upper bound for $\|\Theta - \widetilde{\Theta}\|$.

Put $S = I + V$ and $S^{-1} = I + W$. Note that

$$\begin{aligned}\|\Theta - \widetilde{\Theta}\| &= \|A - SAS^{-1}\| + \|B - SB\| + \|C - CS^{-1}\| \\ &\leq \|A\| \cdot \big(\|V\| + \|W\| + \|V\| \cdot \|W\|\big) + \|B\| \cdot \|V\| + \|C\| \cdot \|W\|.\end{aligned}$$

By Lemma 13.3(ii) we have $\max\{\|V\|, \|W\|\} \leq 2d\big(\alpha_0\eta\big)^{-1}$, where $d = \|P - Q\| + \|P^\times - Q^\times\|$ and $\eta = \eta(\operatorname{Im} P, \operatorname{Im} P^\times)$. It follows that

$$\|\Theta - \widetilde{\Theta}\| \leq \frac{4d}{\alpha_0\eta}\left(1 + \frac{d}{\alpha_0\eta}\right)\|\Theta\|. \tag{13.18}$$

Since $d\alpha_0^{-1} < 1$ and $\eta \leq 1$, we can use (13.18) to show that

$$\begin{aligned}\|\Theta_0 - \widetilde{\Theta}\| &\leq \|\Theta_0 - \Theta\| + \frac{8d}{\alpha_0\eta^2}\|\Theta\| \\ &\leq \|\Theta_0 - \Theta\| + \frac{8d}{\alpha_0\eta^2}\|\Theta - \Theta_0\| + \frac{8d}{\alpha_0\eta^2}\|\Theta_0\| \\ &\leq \frac{9}{\eta^2}\|\Theta - \Theta_0\| + \frac{8d}{\alpha_0\eta^2}\|\Theta_0\| \\ &\leq \frac{9}{\eta^2}\left(\|\Theta - \Theta_0\| + \frac{d}{\alpha_0}\|\Theta_0\|\right).\end{aligned}$$

By using this in (13.17) we obtain the desired inequality (13.16). □

Proof of Theorem 13.4. Take γ, ℓ, α_0 and α as in the first paragraph after Theorem 13.4, and take $\|\Theta - \Theta_0\| < \alpha$. In particular, we have $\|\Theta - \Theta_0\| < 1$. Note that

$$\begin{aligned}\|A^\times - A_0^\times\| &\leq \|A - A_0\| + \|B - B_0\| \cdot \|C - C_0\| + \\ &\qquad + \|B_0\| \cdot \|C - C_0\| + \|C_0\| \cdot \|B - B_0\| \\ &\leq \|\Theta - \Theta_0\| \cdot \big(1 + \|\Theta_0\|\big).\end{aligned}$$

It follows that

$$\max\{\|A - A_0\|, \|A^\times - A_0^\times\|\} \leq \|\Theta - \Theta_0\| \cdot \big(1 + \|\Theta_0\|\big) = \frac{\nu}{2\gamma}, \tag{13.19}$$

where $0 \leq \nu < 1$. Using elementary spectral theory, we may conclude from (13.19) that the curves Γ and $\Gamma^\times$ split the spectra of A and $A^\times$, respectively, while in addition

$$\begin{aligned}\|(\lambda - A)^{-1} - (\lambda - A_0)^{-1}\| &\leq 2\gamma^2\|\Theta - \Theta_0\| \cdot \big(1 + \|\Theta_0\|\big), \qquad \lambda \in \Gamma, \\ \|(\lambda - A^\times)^{-1} - (\lambda - A_0^\times)^{-1}\| &\leq 2\gamma^2\|\Theta - \Theta_0\| \cdot \big(1 + \|\Theta_0\|\big), \qquad \lambda \in \Gamma^\times.\end{aligned}$$

Hence for the corresponding Riesz projections we have,

$$\|P(A;\Gamma) - P(A_0;\Gamma)\| + \|P(A^\times;\Gamma^\times) - P(A_0^\times;\Gamma^\times)\| \le \tag{13.20}$$

$$\le 2\frac{\gamma^2\ell}{\pi}\|\Theta - \Theta_0\|\big(1 + \|\Theta_0\|\big) < \alpha_0.$$

So, for $P = P(A_0;\Gamma)$, $P^\times = P(A_0^\times, \Gamma^\times)$, $Q = P(A;\Gamma)$ and $Q^\times = P(A^\times;\Gamma^\times)$, the coinciding conditions (13.10) and (13.15) are satisfied. Hence we can apply Lemma 13.3 and Theorem 13.5 to the four projections P, $P^\times$, Q and $Q^\times$.

It follows that $X = \operatorname{Im} P(A;\Gamma) \dot{+} \operatorname{Im} P(A^\times;\Gamma^\times)$. Further, if Π is the projection of X along $\operatorname{Im} P(A;\Gamma)$ onto $\operatorname{Im} P(A^\times;\Gamma^\times)$, then Π is a supporting projection for the system Θ. Also there exists a similarity transformation S such that $\Pi_0 = S\Pi S^{-1}$ and Π_0 is a supporting projection for the system

$$\widetilde{\Theta} = (SAS^{-1}, SB, CS^{-1}; X, Y).$$

Finally, by virtue of Lemma 13.3(ii) and formulas (13.16) and (13.20), we have $\|S - I\| \le \beta_1\|\Theta - \Theta_0\|$ and

$$\max\big\{\|\mathrm{pr}_{I-\Pi_0}(\Theta_0) - \mathrm{pr}_{I-\Pi_0}(\widetilde{\Theta}_0)\|,\ \|\mathrm{pr}_{\Pi_0}(\Theta_0) - \mathrm{pr}_{\Pi_0}(\widetilde{\Theta}_0)\|\big\} \le$$

$$\le \beta_2\|\Theta - \Theta_0\|,$$

where β_1 and β_2 are as in the paragraph after Theorem 13.4.

Now suppose that Θ_0 is minimal, and that X and Y are finite-dimensional. The minimality of Θ_0 and the finite dimensionality of X imply that for some k the operator $\operatorname{col}(C_0A_0^j)_{j=0}^k$ is injective and the operator $\operatorname{row}(A_0^jB_0)_{j=0}^k$ is surjective. As Y is finite-dimensional too, it follows that for

$$\|\Theta - \Theta_0\| = \|A - A_0\| + \|B - B_0\| + \|C - C_0\|$$

sufficiently small the operator $\operatorname{col}(CA^j)_{j=0}^k$ will be injective and the operator $\operatorname{row}(A^jB)_{j=0}^k$ will be surjective. This implies that Θ will be minimal whenever $\|\Theta - \Theta_0\|$ is sufficiently small. This completes the proof of Theorem 13.4. □

Theorem 13.6. *Let $\Theta_\varepsilon = (A_\varepsilon, B_\varepsilon, C_\varepsilon; X, Y)$ be a system, and assume that the operators A_ε, B_ε, and C_ε depend analytically on ε in a neighborhood of $\varepsilon = 0$. Put $A_0^\times = A_0 - B_0C_0$, and let Π_0 be a supporting projection of Θ_0. Assume that*

$$\operatorname{Ker}\Pi_0 = \operatorname{Im} P(A_0;\Gamma), \qquad \operatorname{Im} P_0 = \operatorname{Im} P(A_0^\times;\Gamma^\times),$$

where Γ and $\Gamma^\times$ are Cauchy contours that split the spectra of A_0 and $A_0^\times$, respectively. Then for $|\varepsilon|$ sufficiently small, there exists a similarity transformation S_ε, depending analytically on ε, such that $S_0 = I$ and the projection Π_0 is a supporting projection for the system

$$\widetilde{\Theta}_\varepsilon = (S_\varepsilon A_\varepsilon S_\varepsilon^{-1}, S_\varepsilon B_\varepsilon, C_\varepsilon S_\varepsilon^{-1}; X, Y).$$

In particular, if

$$\begin{aligned} \mathrm{pr}_{I-\Pi_0}(\widetilde{\Theta}_\varepsilon) &= (\widetilde{A}_{1\varepsilon}, \widetilde{B}_{1\varepsilon}, \widetilde{C}_{1\varepsilon}; \operatorname{Ker}\Pi_0, Y), \\ \mathrm{pr}_{\Pi_0}(\widetilde{\Theta}_\varepsilon) &= (\widetilde{A}_{2\varepsilon}, \widetilde{B}_{2\varepsilon}, \widetilde{C}_{2\varepsilon}; \operatorname{Im}\Pi_0, Y), \end{aligned}$$

the operators $\widetilde{A}_{1\varepsilon}, \widetilde{A}_{2\varepsilon}, \widetilde{B}_{1\varepsilon}, \widetilde{B}_{2\varepsilon}, \widetilde{C}_{1\varepsilon}$ *and* $\widetilde{C}_{2\varepsilon}$ *depend analytically on* ε.

Proof. We know already that for $|\varepsilon|$ sufficiently small the Cauchy contours Γ and $\Gamma^\times$ split the spectra of A_ε and $A_\varepsilon^\times$, respectively. Put

$$P_\varepsilon = P(A_\varepsilon; \Gamma), \qquad P_\varepsilon^\times = P(A_\varepsilon^\times, \Gamma^\times).$$

From the Cauchy integral formulas for the Riesz projections P_ε and $P_\varepsilon^\times$ it follows that P_ε and $P_\varepsilon^\times$ depend analytically on ε.

Now we proceed as in the proof of Lemma 13.3. Put $S_{1\varepsilon} = I + P_0 - P_\varepsilon$. Then $S_{1\varepsilon}$ depends analytically on ε, the operator $S_{10} = I$, and hence $S_{1\varepsilon}$ is invertible for $|\varepsilon|$ sufficiently small.

Let Π_ε be the projection of X along $\operatorname{Im} P_\varepsilon$ onto $\operatorname{Im} P_\varepsilon^\times$. As both P_ε and $P_\varepsilon^\times$ are analytic functions of ε, the same is true for Π_ε (cf., [105]). It follows that $\widetilde{\Pi}_\varepsilon = S_{1\varepsilon}\Pi_\varepsilon S_{1\varepsilon}^{-1}$ is analytic in ε also. Note that $\widetilde{\Pi}_0 = \Pi_0$.

Next we consider the angular operator R_ε of $\operatorname{Im}\widetilde{\Pi}_\varepsilon$ with respect to Π_0. Recall (see Section 5.1) that

$$R_\varepsilon = (\widetilde{\Pi}_\varepsilon - \Pi_0)|_{\operatorname{Im}\Pi_0}.$$

It follows that R_ε depends analytically on ε and R_0 is the zero operator. So the operator $S_{2\varepsilon} = I - R_\varepsilon\Pi_0$ is analytic in ε and $S_{20} = I$. In particular, we see that $S_{2\varepsilon}$ is invertible for $|\varepsilon|$ sufficiently small. Now put $S_\varepsilon = S_{2\varepsilon}S_{1\varepsilon}$. Then for $|\varepsilon|$ sufficiently small S_ε has all desired properties. □

13.4 Applications to transfer functions

In this section we shall prove Theorem 13.1. We begin with its infinite-dimensional analogue. Throughout this section X and Y are Banach spaces.

Theorem 13.7. *Consider the transfer function*

$$W_0(\lambda) = I_Y + C_0(\lambda I_X - A_0)^{-1}B_0, \tag{13.21}$$

and assume that W_0 *admits a factorization*

$$W_0 = W_{01}W_{02}, \qquad W_{0j}(\lambda) = I_Y + C_{0j}(\lambda I_{X_j} - A_{0j})^{-1}B_{0j},$$

such that (*with* $A_{0j}^\times = A_{0j} - B_{0j}C_{0j}$ *as usual*)

$$\sigma(A_{01}) \cap \sigma(A_{02}) = \emptyset, \qquad \sigma(A_{01}^\times) \cap \sigma(A_{02}^\times) = \emptyset, \tag{13.22}$$

while, in addition, the system $\Theta_0 = (A_0, B_0, C_0; X, Y)$ *is similar to the product* $\Theta_{01}\Theta_{02}$*, where* $\Theta_{0j} = (A_{0j}, B_{0j}, C_{0j}; X_j, Y)$*. Then there exist positive constants* α_0 *and* β_0 *such that the following holds. If* A, B *and* C *are matrices of appropriate sizes, with*

$$\|A - A_0\| + \|B - B_0\| + \|C - C_0\| < \alpha_0, \tag{13.23}$$

then the transfer function $W(\lambda) = I_Y + C(\lambda I_X - A)^{-1}B$ *admits a factorization*

$$W = W_1 W_2, \qquad W_j(\lambda) = I_Y + C_j(\lambda I_{X_j} - A_j)^{-1}B_j, \tag{13.24}$$

such that (with $A_j^\times = A_j - B_j C_j$ *as usual)*

$$\sigma(A_1) \cap \sigma(A_2) = \emptyset, \qquad \sigma(A_1^\times) \cap \sigma(A_2^\times) = \emptyset, \tag{13.25}$$

and, for $j = 1, 2$*,*

$$\|A_j - A_{0j}\| + \|Bj - B_{0j}\| + \|C_j - C_{0j}\| \leq \tag{13.26}$$

$$\leq \beta_0(\|A - A_0\| + \|B - B_0\| + \|C - C_0\|). \tag{13.27}$$

Proof. Let $T : X \to X_1 \dot{+} X_2$ be a system similarity between Θ_0 and $\Theta_{01}\Theta_{02}$. Assume (13.23) holds, and put

$$\overline{\Theta} = (TAT^{-1}, TB, CT^{-1}; X_1 \dot{+} X_2, Y).$$

Note that for the system distance $\|\Theta_{01}\Theta_{02} - \overline{\Theta}\|$, we have

$$\begin{aligned} \|\Theta_{01}\Theta_{02} - \overline{\Theta}\| &= \|(TA_0T^{-1} - TAT^{-1}\| + \|TB_0 - TB\| + \\ &\qquad + \|C_0T^{-1} - CT^{-1}\| \\ &\leq (\|A - A_0\| + \|B - B_0\| + \|C - C_0\|) \cdot \\ &\qquad \cdot\big(\|T\| \cdot \|T^{-1}\| + \|T\| + \|T^{-1}\|\big). \end{aligned}$$

Relative to the direct sum $X_1 \dot{+} X_2$ the main operator of the system $\overline{\Theta}_0 = \Theta_{01}\Theta_{02}$ and the associated main operator (respectively) have the following form

$$\overline{A}_0 = \begin{bmatrix} A_{01} & * \\ 0 & A_{02} \end{bmatrix}, \qquad \overline{A}_0^\times = \begin{bmatrix} A_{01}^\times & 0 \\ * & A_{02}^\times \end{bmatrix}.$$

Put a Cauchy contour Γ around $\sigma(A_{01})$ that separates the spectrum $\sigma(A_{01})$ from $\sigma(A_{02})$. Similarly, put a Cauchy contour $\Gamma^\times$ around $\sigma(A_{02}^\times)$ such that $\Gamma^\times$ separates $\sigma(A_{02}^\times)$ from $\sigma(A_{01}^\times)$. Then we can apply Lemma 5.9 to show that

$$X_1 = \operatorname{Im} P(\overline{A}_0; \Gamma), \qquad X_2 = \operatorname{Im} P(\overline{A}_0^\times; \Gamma^\times).$$

It follows that we may apply Theorem 13.4 to the system $\overline{\Theta}_0 = \Theta_{01}\Theta_{02}$.

Let α and β_2 be the positive numbers that according to Theorem 13.4 correspond to the system $\overline{\Theta}_0$. Put

$$\alpha_0 = \alpha\big(\|T\| \cdot \|T^{-1}\| + \|T\| + \|T^{-1}\|\big)^{-1}.$$

Now assume that (13.23) holds. Then $\|\overline{\Theta}_0 - \overline{\Theta}\| < \alpha$. So by Theorem 13.4 there exists a similarity transformation S such that for the system

$$\widetilde{\Theta} = (STAT^{-1}S^{-1}, STB, CT^{-1}S^{-1}; X_1 \dot{+} X_2, Y)$$

the projection Π_0 of $X_1 \dot{+} X_2$ along X_1 onto X_2 is a supporting projection. This shows that W admits a factorization of the form (13.24). Moreover we know that

$$\begin{aligned} \|\mathrm{pr}_{I-\Pi_0}(\overline{\Theta}_0) - \mathrm{pr}_{I-\Pi_0}(\widetilde{\Theta})\| &\leq \beta_2\|\overline{\Theta}_0 - \overline{\Theta}\|, \\ \|\mathrm{pr}_{\Pi_0}(\overline{\Theta}) - \mathrm{pr}_{\Pi_0}(\widetilde{\Theta})\| &\leq \beta_2\|\overline{\Theta}_0 - \overline{\Theta}\|. \end{aligned}$$

But this is the same as

$$\|A_{0i} - A_i\| + \|B_{0i} - B_i\| + \|C_{0i} - C_i\| \leq \beta_2\|\overline{\Theta}_0 - \overline{\Theta}\|, \qquad i = 1, 2.$$

So, if we take

$$\beta_0 = \beta_2\big(\|T\| \cdot \|T^{-1}\| + \|T\| + \|T^{-1}\|\big),$$

then (13.26) holds true.

Let $\overline{A}$ be the main operator of $\overline{\Theta}$, and let $\overline{A}^{\times}$ be the main operator of the associated system $\overline{\Theta}^{\times}$. As $\|\overline{\Theta}_0 - \overline{\Theta}\| < \alpha$, we can apply Theorem 13.4 to show that the curves Γ and $\Gamma^{\times}$ split the spectra of $\overline{A}$ and $\overline{A}^{\times}$, respectively, and

$$X_1 \dot{+} X_2 = \mathrm{Im}\, P(\overline{A}; \Gamma) \dot{+} \mathrm{Im}\, P(\overline{A}^{\times}; \Gamma^{\times}).$$

Let Π be the projection of $X_1 \dot{+} X_2$ along $\mathrm{Im}\, P(\overline{A}; \Gamma)$ onto $\mathrm{Im}\, P(\overline{A}^{\times}; \Gamma^{\times})$. Then $\Pi_0 = S\Pi S^{-1}$. It follows that $\sigma(A_1)$ is inside the contour Γ and $\sigma(A_2)$ is outside the contour Γ. Similarly, $\sigma(A_2^{\times})$ is inside $\Gamma^{\times}$ and $\sigma(A_1^{\times})$ is outside $\Gamma^{\times}$. In particular, we see that (13.25) holds true. This completes the proof of the theorem. □

To prove Theorem 13.1, we shall show that Theorem 13.1 appears as a corollary of Theorem 13.7. To do this, let us assume that X and Y are finite-dimensional. Further, let us assume that the realization (13.21) is minimal. Applying the last paragraph of Theorem 13.4, we see that in Theorem 13.7 the positive number α_0 may be chosen such that (13.23) implies that the realization $W(\lambda) = I_Y + C(\lambda I_X - A)^{-1}B$ is also minimal. Next we observe that the assumption in Theorem 13.7 that Θ_0 is similar to the product $\Theta_{01}\Theta_{02}$ may be replaced by

$$\dim X = \dim X_1 + \dim X_2, \tag{13.28}$$

because we have assumed that Θ_0 is minimal. Moreover, again because of minimality, the condition (13.22) is equivalent to the requirement that the factors W_{01} and W_{02} have no common zeros and no common poles, and, similarly, (13.25) is equivalent to the statement that the factors W_1 and W_2 have no common zeros and no common poles. By virtue of (13.28), the minimality of the realizations of W_0 and W implies that $W_0(\lambda) = W_{01}(\lambda)W_{02}\lambda$ and $W(\lambda) = W_1(\lambda)W_2\lambda$ are minimal factorizations (cf., Section 9.1). Using the above remarks it is simple to obtain Theorem 13.1 as a corollary of Theorem 13.7.

Using Theorem 13.6 in the same way as Theorem 13.4 has been used in the proof of Theorem 13.7, one can see that the following analytic version of Theorem 13.7 holds true.

Theorem 13.8. *Consider the transfer function*

$$W_\varepsilon(\lambda) = I_Y + C_\varepsilon(\lambda I_X - A_\varepsilon)^{-1}B_\varepsilon,$$

with the operators A_ε, B_ε and C_ε depending analytically on ε in a neighborhood of $\varepsilon = 0$. Assume that W_0 admits a factorization

$$W_0 = W_{01}W_{02}, \qquad W_{0j}(\lambda) = I_Y + C_{0j}(\lambda I_{X_j} - A_{0j})^{-1}B_{0j},$$

such that (with $A_{0j}^\times = A_{0j} - B_{0j}C_{0j}$ as usual)

$$\sigma(A_{01}) \cap \sigma(A_{02}) = \emptyset, \qquad \sigma(A_{01}^\times) \cap \sigma(A_{02}^\times) = \emptyset,$$

while, in addition, the system $\Theta_0 = (A_0, B_0, C_0; X, Y)$ is similar to the product $\Theta_{01}\Theta_{02}$, where $\Theta_{0j} = (A_{0j}, B_{0j}, C_{0j}; X_j, Y)$. Then for $|\varepsilon|$ sufficiently small the transfer function W_ε admits a factorization,

$$W_\varepsilon = W_{1\varepsilon}W_{2\varepsilon}, \qquad W_{j\varepsilon}(\lambda) = I_Y + C_j^\varepsilon(\lambda X_{I_\varepsilon} - A_j^\varepsilon)^{-1}B_j^\varepsilon,$$

such that (with $(A_{0j}^\varepsilon)^\times = A_{0j}^\varepsilon - B_{0j}^\varepsilon C_{0j}^\varepsilon$ as usual)

$$\sigma(A_{01}^\varepsilon) \cap \sigma(A_{02}^\varepsilon) = \emptyset, \qquad \sigma\left((A_{01}^\varepsilon)^\times\right) \cap \sigma\left((A_{02}^\varepsilon)^\times\right) = \emptyset,$$

the operators $A_1^\varepsilon, A_2^\varepsilon, B_1^\varepsilon, B_2^\varepsilon, C_1^\varepsilon$ and C_2^ε depend analytically on ε, and for $\varepsilon = 0$ they are equal to $A_{01}, A_{02}, B_{01}, B_{02}, C_{01}$ and C_{02}, respectively.

13.5 Applications to Riccati equations

In this section we show that the method of Sections 5.4 and 13.3 can also be used to prove stability theorems for certain solutions of the Riccati equation. Throughout this section X_1 and X_2 are Banach spaces, and we use the symbol $\mathcal{L}(X_j, X_i)$ to denote the space of all bounded linear operators from X_j into X_i.

Theorem 13.9. *For $i, j = 1, 2$, let $T_{ij} \in \mathcal{L}(X_j, X_i)$, and let $R \in \mathcal{L}(X_2, X_1)$ be a solution of*

$$RT_{21}R + RT_{22} - T_{11}R - T_{12} = 0. \tag{13.29}$$

Assume $\sigma(T_{11} - RT_{21})$ and $\sigma(T_{22} + T_{21}R)$ are disjoint, and let Γ be a Cauchy contour with $\sigma(T_{22} + T_{21}R)$ in the inner domain of Γ and $\sigma(T_{11} - RT_{21})$ in the outer domain. Then there exist positive constants α and β such that the following holds. If $S_{ij} \in \mathcal{L}(X_j, X_i)$, and

$$\|S_{ij} - T_{ij}\| \le \alpha, \qquad i, j = 1, 2, \tag{13.30}$$

then the equation

$$QS_{21}Q + QS_{22} - S_{11}Q - S_{12} = 0 \tag{13.31}$$

has a solution $Q \in \mathcal{L}(X_2, X_1)$ such that $\sigma(S_{22} + S_{21}Q)$ lies in the inner domain of Γ, the set $\sigma(S_{11} - QS_{21})$ lies in the outer domain of Γ, and

$$\|R - Q\| \le \beta \max_{i,j=1,2} \|T_{ij} - S_{ij}\|. \tag{13.32}$$

Proof. Consider the operators

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}, \qquad S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

on $X = X_1 \dotplus X_2$. Assume that X is endowed with the norm $\|(x_1, x_2)\| = \|x_1\| + \|x_2\|$. Then

$$\|T - S\| \le \max_{i,j=1,2} \|T_{ij} - S_{ij}\|. \tag{13.33}$$

As the Riccati equation (13.29) has a solution R such that $\sigma(T_{11} - RT_{21})$ and $\sigma(T_{22} + T_{21}R)$ do not intersect, we know from Proposition 5.10 that the space

$$N_R = \{(Rz, z) \mid z \in X_2\}$$

is a spectral subspace for T. In fact, if Γ is as in the statement of the theorem, then Γ splits the spectrum of T and $N_R = \operatorname{Im} P(T; \Gamma)$.

Let ℓ be the length of Γ, and put $\gamma = \max_{\lambda \in \Gamma} \|(\lambda - T)^{-1}\|$. Take S such that $\|T - S\| < (2\gamma)^{-1}$. By elementary spectral theory this implies that Γ splits the spectrum of S and

$$\|(\lambda - T)^{-1} - (\lambda - S)^{-1}\| \le 2\gamma^2 \|S - T\|, \qquad \lambda \in \Gamma.$$

But then $\|P(T; \Gamma) - P(S; \Gamma)\| \le \dfrac{\gamma^2 \ell}{\pi} \|S - T\|$.

As $X = X_1 \dotplus N_R$, the number $\eta(X_1, N_R)$ is positive. Put

$$\alpha = \min \left\{ \frac{1}{4\gamma}, \frac{\pi}{4\gamma^2 \ell} \eta(X_1, N_R) \right\},$$

and assume that (13.30) holds true. By (13.33) this implies that $\|T-S\| < 2\alpha \leq (2\gamma)^{-1}$, and we can apply the result of the previous paragraph to show that

$$\|P(T;\Gamma) - P(S;\Gamma)\| \leq \frac{1}{2}\eta(X_1, N_R).$$

In particular we see that

$$\operatorname{gap}(N_R, \operatorname{Im} P(S;\Gamma)) \leq \frac{1}{2}\eta(X_1, N_R). \tag{13.34}$$

By [71], Theorem 2 this implies that $X = X_1 \dotplus \operatorname{Im} P(S;\Gamma)$. It follows that there exists $Q \in \mathcal{L}(X_2, X_1)$ such that

$$N_Q = \{Qz + z \big| z \in X_2\} = \operatorname{Im} P(S;\Gamma).$$

By Proposition 5.10, this operator Q is a solution of equation (13.31), the spectrum $\sigma(S_{22}+S_{21}Q)$ is in the inner domain of Γ and $\sigma(S_{11}-QS_{21})$ is in the outer domain of Γ.

According to (13.34), we have $\operatorname{gap}(N_R, N_Q) \leq \frac{1}{2}\eta(X_1, N_R)$. So we can apply Lemma 13.2(ii) to show that

$$\|R-Q\| \leq \frac{2(1+\|R\|)}{\eta(X_1, N_R)} \operatorname{gap}(N_R, N_Q). \tag{13.35}$$

But

$$\begin{aligned} \operatorname{gap}(N_R, N_Q) &\leq \|P(T;\Gamma) - P(S;\Gamma)\| \leq \frac{\gamma^2\ell}{\pi}\|T-S\| \\ &\leq 2\frac{\gamma^2\ell}{\pi} \max_{i,j=1,2} \|T_{ij} - S_{ij}\|. \end{aligned} \tag{13.36}$$

Put

$$\beta = 4\big(1+\|R\|\big)\frac{\gamma^2\ell}{\pi\eta(X_1, N_R)}.$$

Then we see from (13.35) and (13.36) that (13.32) holds true. This completes the proof of the theorem. □

Using arguments similar to the ones employed in the proof of Theorem 13.6, one can see that the following analytic analogue of the previous theorem holds true.

Theorem 13.10. *For $i, j = 1, 2$, let $T_{ij}(\varepsilon) : X_j \to X_i$ be bounded linear operators depending analytically on ε in a neighborhood of $\varepsilon = 0$. Let $R \in \mathcal{L}(X_2, X_1)$ be a solution of*

$$RT_{21}(0)R + RT_{22}(0) - T_{11}(0)R - T_{12}(0)R = 0,$$

and assume that $\sigma\big(T_{11}(0) - RT_{21}(0)\big)$ *and* $\sigma\big(T_{22}(0) + T_{21}(0)R\big)$ *are disjoint. Then for* $|\varepsilon|$ *sufficiently small, there exists* $R(\varepsilon) \in \mathcal{L}(X_1, X_2)$*, depending analytically on* ε*, such that* $R(0) = R$*,*

$$R(\varepsilon)T_{21}(\varepsilon)R(\varepsilon) + R(\varepsilon)T_{22}(\varepsilon) - T_{11}(\varepsilon)R(\varepsilon) - T_{12}(\varepsilon) = 0,$$

and

$$\sigma\big(T_{11}(\varepsilon) - R(\varepsilon)T_{21}(\varepsilon)\big) \cap \sigma(T_{22}\big(\varepsilon) + T_{21}(\varepsilon)R(\varepsilon)\big) = \emptyset.$$

Notes

The material in this chapter is taken from Chapter VII in [14]. The notion of a gap between subspaces has been introduced and developed in [89]. It was developed further and used in Fredholm theory in [58] and [85]. As our main sources for topological properties of subspaces we used [71] and [86]. For Euclidean spaces they can also been found in Chapter 13 of [70].

Chapter 14

Stability of Divisors

In this chapter we shall prove that there exist stable factorizations which are not spectral factorizations. In fact, for the finite-dimensional case we shall give a complete description of all possible stable minimal factorizations. It will also be shown that stability amounts to the same as the property of being isolated provided the underlying field is complex (which will be the case in this chapter).

14.1 Stable invariant subspaces

In the previous chapter we have implicitly been dealing with invariant subspaces which have a certain stability property. In this section we shall investigate this matter explicitly.

Let T be a bounded linear operator on a Banach space X. A closed T-invariant subspace N of X is called *stable* if given $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. If S is a bounded linear operator on X and $\|S - T\| < \delta$, then S has a closed invariant subspace M such that $\operatorname{gap}(M, N) < \varepsilon$. The property of being a stable invariant subspace is similarity invariant in the following sense. Let E be an invertible operator on X, and introduce $\widetilde{T} = E^{-1}TE$, $\widetilde{N} = E^{-1}[N]$. Then $\widetilde{N}$ is a stable invariant subspace for $\widetilde{T}$ if (and only if) N is a stable invariant subspace for T. The argument is straightforward and involves the condition number $\|E^{-1}\| \cdot \|E\|$ of E.

If N is the image of a Riesz projection corresponding to T, then N is clearly a stable invariant subspace for T. In general, not every stable T-variant subspace is of this form. For the finite-dimensional case we shall give a complete description.

Let A be a $k \times k$ matrix. As usual we identify A with its canonical action on $\mathbb{C}^k$. The generalized eigenspace $\operatorname{Ker}(\lambda_0 - A)^k$ of A corresponding to the eigenvalue λ_0 will be denoted by $N(\lambda_0)$.

Theorem 14.1. *Let $\lambda_1, \dots, \lambda_r$ be the different eigenvalues of the $k \times k$ matrix A. A subspace N of $\mathbb{C}^k$ is a stable A-invariant subspace if and only if $N = N_1 \dot{+} \cdots \dot{+} N_r$, where for each j the space N_j is an arbitrary A-invariant subspace of $N(\lambda_j)$ whenever $\dim \operatorname{Ker}(\lambda_j - A) = 1$, while otherwise $N_j = \{0\}$ or $N_j = N(\lambda_j)$.*

The proof of Theorem 14.1 will be based on a series of lemmas and an auxiliary theorem which is of some interest in itself. To state the latter theorem we recall the following notion. Given a $k \times k$ matrix A, a chain

$$M_0 \subset M_1 \subset \cdots \subset M_{k-1} \subset M_k$$

of A-invariant subspaces in $\mathbb{C}^k$, written in shorthand as $\{M_j\}$, is said to be *complete* if $\dim M_j = j$ for $j = 0, \dots, k$.

Theorem 14.2. *Given $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds true. If B is a $k \times k$ matrix with $\|B - A\| < \delta$ and $\{M_j\}$ is a complete chain of B-invariant subspaces, then there exists a complete chain $\{N_i\}$ of A-invariant subspaces such that $\operatorname{gap}(N_j, M_j) < \varepsilon$ for $j = 1, \dots, k-1$.*

In general, the chain $\{N_j\}$ for A will depend on the choice of B. To see this, consider

$$A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad B_\nu^- = \begin{bmatrix} 0 & 0 \\ \nu & 0 \end{bmatrix}, \qquad B_\nu^+ = \begin{bmatrix} 0 & \nu \\ 0 & 0 \end{bmatrix},$$

where $\nu \in \mathbb{C}$. For $\nu \neq 0$ the unique one-dimensional invariant subspace of B_ν^- is $\{0\} \dot{+} \mathbb{C}$, while the only one-dimensional invariant subspace for B_ν^+ is $\mathbb{C} \dot{+} \{0\}$.

Proof. Assume that the conclusion of the theorem is not correct. Then there exists $\varepsilon > 0$ with the property that for every positive integer m there exists a $k \times k$ matrix B_m satisfying $\|B_m - A\| < 1/m$ and a complete chain $\{M_{mj}\}$ of B_m-invariant subspaces such that for every complete chain $\{N_j\}$ of A-invariant subspaces we have

$$\max_{1 \le j \le k-1} \operatorname{gap}(N_j, M_{mj}) \ge \varepsilon, \qquad m = 1, 2, \dots . \tag{14.1}$$

Denote by P_{mj} the orthogonal projection of $\mathbb{C}^k$ onto M_{mj}. Since $\mathbb{C}^k$ is finite-dimensional and all P_{mj} are in the unit ball of $L(\mathbb{C}^k, \mathbb{C}^k)$, there exist a subsequence $m_1, m_2, \dots$ of the sequence of positive integers and operators $P_1, \dots, P_{k-1}$ on $\mathbb{C}^k$ such that

$$\lim_{i \to \infty} P_{m_i j} = P_j, \qquad j = 0, \dots, k.$$

Note that $P_1, \dots, P_{k-1}$ are orthogonal projections and that $N_j = \operatorname{Im} P_j$ has dimension j. By passing to the limits it follows from $B_m P_{mj} = P_{mj} B_m P_{mj}$ that $AP_j = P_j A P_j$. Hence N_j is A-invariant. Since $P_{mj} = P_{m,j+1} P_{mj}$ we have $P_j = P_{j+1} P_j$, and thus $N_j \subset N_{j+1}$. It follows that N_j is a complete chain of A-invariant subspaces. Finally, $\operatorname{gap}(N_j, M_{m_i j}) = \|P_j - P_{m_i j}\| \to 0$ for $i \to \infty$. But this contradicts (14.1), and the proof is complete. □

Corollary 14.3. *If A has only one eigenvalue, λ_0 say, and if* $\operatorname{Ker}(\lambda_0 - A)$ *is one-dimensional, then each invariant subspace of A is stable.*

Proof. The conditions on A are equivalent to the requirement that for each $j = 0, \dots, k$, the operator A has only one j-dimensional invariant subspace and the non-trivial invariant subspaces form a complete chain. So we may apply the previous theorem to get the desired result. □

Lemma 14.4. *If A has only one eigenvalue, λ_0 say, and if* $\operatorname{Ker}(\lambda_0 - A)$ *has dimension at least two, then the only stable A-invariant subspaces are* $\{0\}$ *and* $\mathbb{C}^k$.

Proof. Let $J = \operatorname{diag}(J_1, \dots, J_s)$ be a Jordan matrix for A. Here J_i is a simple Jordan block with λ_0 on the main diagonal and of size κ_i, say. As $\dim \operatorname{Ker}(\lambda_0 - A) \geq 2$ we have $s \geq 2$. By similarity, it suffices to prove that J has no non-trivial stable invariant subspace.

Let $e_1, \dots, e_k$ be the standard basis for $\mathbb{C}^k$. Define on $\mathbb{C}^k$ the operator T_ε by setting $T_\varepsilon e_i = \varepsilon e_{i-1}$ if $i = \kappa_1 + \cdots + \kappa_j + 1$, $j = 1, \dots, s-1$, and $T_\varepsilon e_i = 0$ otherwise. Put $B_\varepsilon = J + T_\varepsilon$. Then $B_\varepsilon \to J$ as $\varepsilon \to 0$. For $\varepsilon \neq 0$ the operator B_ε has exactly one j-dimensional invariant subspace namely, $N_j = \operatorname{sp}\{e_1, \dots, e_j\}$. Here $j = 0, \dots, k$. It follows that N_j is the only candidate for a stable J-invariant subspace of dimension j.

Now consider $\widetilde{J} = \operatorname{diag}(J_s, \dots, J_1)$. Repeating the argument of the previous paragraph for $\widetilde{J}$ instead of J, we see that N_j is the only candidate for a stable $\widetilde{J}$-invariant subspace of dimension j. But $J = S\widetilde{J}S^{-1}$, where S is the similarity transformation that reverses the order of the blocks in J. It follows that SN_j is the only candidate for a stable J-invariant subspace of dimension j. However, as $s \geq 2$, we have $SN_j \neq N_j$ for $j = 1, \dots, k-1$, and the proof is complete. □

Corollary 14.3 and Lemma 14.4 together prove Theorem 14.1 for the case when A has one eigenvalue only. The next two lemmas will show that the general version of the theorem may be proved by reduction to the case of a single eigenvalue.

In the remainder of this section X will be a complex Banach space and T will be a bounded linear operator on X.

Lemma 14.5. *Let Γ be a Cauchy contour splitting the spectrum of T, let T_0 be the restriction of T to* $\operatorname{Im} P(T;\Gamma)$, *and let N be a closed subspace of* $\operatorname{Im} P(T;\Gamma)$. *Then N is a stable invariant subspace for T if and only if N is a stable invariant subspace for T_0.*

Proof. Suppose N is a stable invariant subspace for T_0, but not for T. Then one can find $\varepsilon > 0$ such that for every positive integer m there exists $S_m \in \mathcal{L}(X)$ satisfying

$$\|S_m - T\| < \frac{1}{m}, \tag{14.2}$$

while in addition

$$\text{gap}\,(N, M) \geq \varepsilon, \qquad M \in \Omega_m. \tag{14.3}$$

Here Ω_m denotes the collection of all closed invariant subspaces of S_m. From (14.2) it is clear that $S_m \to T$. By assumption Γ splits the spectrum of T. Thus, for m sufficiently large, the contour Γ will split the spectrum of S_m as well. Moreover, $P(S_m, \Gamma) \to P(T; \Gamma)$, and hence $\text{Im}\, P(S_m; \Gamma)$ tends to $\text{Im}\, P(T; \Gamma)$ in the gap topology. But then, for m sufficiently large,

$$\text{Ker}\, P(T; \Gamma) \dot{+} \text{Im}\, P(S_m; \Gamma) = X$$

(cf, [71], Theorem 2).

Let R_m be the angular operator of $\text{Im}\, P(S_m; \Gamma)$ with respect to $P(T; \Gamma)$. Here, as in the sequel, m is supposed to be sufficiently large. Recall that $P(S_m; \Gamma) \to P(T; \Gamma)$. Thus we have $R_m \to 0$. Put

$$E_m = \begin{bmatrix} I & R_m \\ 0 & I \end{bmatrix},$$

where the matrix representation corresponds to the decomposition

$$X = \text{Ker}\, P(T; \Gamma) \dot{+} \text{Im}\, P(T; \Gamma). \tag{14.4}$$

Then E_m is invertible with inverse

$$E_m^{-1} = \begin{bmatrix} I & -R_m \\ 0 & I \end{bmatrix}.$$

Furthermore, $E_m[\text{Im}\, P(T; \Gamma)] = \text{Im}\, P(S_m; \Gamma)$ and $E_m \to I$ when $m \to \infty$.

Put $T_m = E_m^{-1} S_m E_m$. Then $T_m \text{Im}\, P(T; \Gamma) \subset \text{Im}\, P(T; \Gamma)$ and $T_m \to T$. Let T_{m0} be the restriction of T_m to $\text{Im}\, P(T; \Gamma)$. Then $T_{m0} \to T_0$. As N is a stable invariant subspace for T_0, there exists a sequence $N_1, N_2, \ldots$ of closed subspaces of $\text{Im}\, P(T; \Gamma)$ such that each N_m is T_{m0}-invariant and $\text{gap}\,(N_m, N) \to 0$. Note that N_m is also T_m-invariant.

Now put $M_m = E_m N_m$. Then M_m is a closed invariant subspace for S_m. Thus $M_m \in \Omega_m$. Since $E_m \to I$ if $m \to \infty$, one can easily deduce that $\text{gap}\,(M_m, N_m) \to 0$. Together with $\text{gap}\,(N_m, N) \to 0$ this gives $\text{gap}\,(M_m, N) \to 0$, which contradicts (14.3).

Next assume that N is a stable invariant subspace for T, but not for T_0. Then one can find $\varepsilon > 0$ such that for every positive integer m there exists a bounded linear operator S_{m0} on $\text{Im}\, P(T; \Gamma)$ satisfying

$$\|S_{m0} - T_0\| < \frac{1}{m}, \tag{14.5}$$

while in addition

$$\operatorname{gap}(N, M) \geq \varepsilon, \qquad M \in \Omega_{m0}. \tag{14.6}$$

Here Ω_{m0} denotes the collection of all closed invariant subspaces of S_{m0}. Let T_1 be the restriction of T to $\operatorname{Ker} P(T;\Gamma)$ and write

$$S_m = \begin{bmatrix} T_1 & 0 \\ 0 & S_{m0} \end{bmatrix},$$

where the matrix representation corresponds to the decomposition given in (14.4). From the inequality (14.5) it is clear that $S_m \to T$. Hence, as N is a stable invariant subspace for T, there exists a sequence $N_1, N_2, \ldots$ of closed subspaces of X such that N_m is S_m-invariant and $\operatorname{gap}(N_m, N) \to 0$. Put $M_m = P(T;\Gamma)N_m$. Since $P(T;\Gamma)$ commutes with S_m, we have that M_m is an invariant subspace for S_{m0}. As N is a closed subspace of $\operatorname{Im} P(T;\Gamma)$, the minimal opening $\eta = \eta(N, \operatorname{Ker} P(T;\Gamma))$ is strictly positive. From Lemma 2 in [71], we know that $\operatorname{gap}(N_m, N) \to 0$ implies that $\eta(N_m, \operatorname{Ker} P(T;\Gamma)) \geq \frac{1}{2}\eta > 0$. It follows that $N_m + \operatorname{Ker} P(T;\Gamma)$ is closed. But then M_m is also closed by Lemma IV.2.9 in [75]. Hence M_m is a closed invariant subspace for S_{m0}. In other words $M_m \in \Omega_{m0}$. We shall now prove that $\operatorname{gap}(M_m, N) \to 0$, thus obtaining a contradiction to (14.6).

Take $y \in M_m$ with $\|y\| \leq 1$. Then $y = P(T;\Gamma)x$ for some $x \in M_m$. As

$$\begin{aligned} \|y\| = \|P(T;\Gamma)x\| &\geq \inf\{\|x - u\| \mid u \in \operatorname{Ker} P(T;\Gamma)\} \\ &\geq \eta(N_m, \operatorname{Ker} P(T;\Gamma)) \cdot \|x\|, \end{aligned}$$

we see that $\|y\| \geq \frac{1}{2}\eta\|x\|$ for m sufficiently large. Using this it is not difficult to deduce that

$$\operatorname{gap}(M_m, N) \leq \left(1 + \frac{2}{\eta}\right) \|P(T;\Gamma)\| \cdot \operatorname{gap}(N_m, N)$$

for m sufficiently large. We conclude that $\operatorname{gap}(N_m, N) \to 0$ when $m \to \infty$, and the proof is complete. □

Lemma 14.6. *Let N be a complemented invariant subspace for T, and assume that the Cauchy contour Γ splits the spectrum of T and the spectrum of the restriction operator $T|_N$. If N is stable for T, then $P(T;\Gamma)N$ is a stable closed invariant subspace for the restriction T_0 of T to $\operatorname{Im} P(T;\Gamma)$.*

Proof. It is clear that $M = P(T;\Gamma)N$ is T_0-invariant. For each $\lambda \in \Gamma$ we have $(\lambda - T|_N)^{-1} = (\lambda - T)^{-1}|_N$. This implies that

$$M = P(T;\Gamma)N = \operatorname{Im} P(T|_N;\Gamma) \subset N,$$

and it follows that M is closed.

Assume that M is not stable for T_0. Then M is neither stable for T by Lemma 14.5. Hence there exist $\varepsilon > 0$ and a sequence $S_1, S_2, \ldots$ such that

$$\operatorname{gap}(L, M) \geq \varepsilon, \qquad L \in \Omega_m, \tag{14.7}$$

where Ω_m denotes the set of all closed invariant subspaces of S_m, while moreover $S_m \to T$ for $m \to \infty$.

As N is stable for T, one can find a sequence $N_1, N_2, \ldots$ of closed subspaces such that $S_m N_m \subset N_m$ and $\operatorname{gap}(N_m, N) \to 0$. Also, since Γ splits the spectrum of T and $S_m \to T$, the contour Γ will split the spectrum of S_m for m sufficiently large. But then, without loss of generality, we may assume that Γ splits the spectrum of each S_m. Again using $S_m \to T$, it follows that $P(S_m; \Gamma) \to P(T; \Gamma)$.

Let Z be a closed complement of N in X, that is, $X = Z \dot{+} N$. Because $\operatorname{gap}(N_m, N) \to 0$, we have $X = Z \dot{+} N_m$ for m sufficiently large. So, without loss of generality, we may assume that $X = Z \dot{+} N_m$ for each m. Let R_m be the angular operator of N_m with respect to the projection of X along Z onto N, and put

$$E_m = \begin{bmatrix} I & R_m \\ 0 & I \end{bmatrix},$$

where the matrix representation corresponds to the decomposition $X = Z \dot{+} N$. Note that $T_m = E_m^{-1} S_m E_m$ leaves invariant N. Since $R_m \to 0$, we have $E_m \to I$, and so $T_m \to T$.

By assumption Γ splits the spectrum of $T|_N$. As $T_m \to T$ and N is invariant under T_m, the contour Γ will split the spectrum of $T_m|_N$ as well, provided m is sufficiently large. But then we may assume that this happens for all m. Also, we have

$$\lim_{m \to \infty} P(T_m|_N; \Gamma) \;=\; P(T|_N; \Gamma).$$

Hence $M_m = \operatorname{Im} P(T_m|_N; \Gamma) \to \operatorname{Im} P(T|_N; \Gamma) = M$ in the gap topology.

Now consider $L_m = E_m M_m$. Then L_m is a closed S_m-invariant subspace of X. In other words, $L_m \in \Omega_m$. From $E_m \to I$ it follows that $\operatorname{gap}(L_m, M_m) \to 0$. The latter, together with $\operatorname{gap}(M_m, M) \to 0$, implies that $\operatorname{gap}(L_m, M) \to 0$. So we arrive at a contradiction to (14.7) and the proof is complete. □

Proof of Theorem 14.1. Suppose N is a stable invariant subspace for A. Put $N_j = P_j N$, where P_j is the Riesz projection corresponding to A and λ_j. Then $N = N_1 \dot{+} \cdots \dot{+} N_r$. By Lemma 14.6 the space N_j is a stable invariant subspace for the restriction A_j of A to $N(\lambda_j)$. But A_j has one eigenvalue only, namely λ_j. So we may apply Lemma 14.4 to prove that N_j has the desired form.

Conversely, assume that each N_j has the desired form, and let us prove that $N = N_1 \dot{+} \cdots \dot{+} N_r$ is a stable invariant subspace for A. By Corollary 14.3, the space N_j is a stable invariant subspace for the restriction A_j of A to $\operatorname{Im} P_j$. Hence we may apply Lemma 14.5 to show that each N_j is a stable invariant subspace for A. But then the same is true for the direct sum $N = N_1 \dot{+} \cdots \dot{+} N_r$. □

For shortness sake, the proofs of Lemmas 14.5 and 14.6 were given by reductio ad absurdum. It is of some practical interest to note that they could have been given in a more constructive way.

The next theorem indicates the way in which Theorem 14.1 will be applied in the context of minimal factorization theory.

Theorem 14.7. *Let X_1 and X_2 be finite-dimensional spaces, and let*

$$A = \begin{bmatrix} A_1 & A_0 \\ 0 & A_2 \end{bmatrix}$$

be a linear operator acting on $X = X_1 \dot{+} X_2$. Then X_1 is a stable invariant subspace for A if and only if each common eigenvalue of A_1 and A_2 is an eigenvalue of A of geometric multiplicity one.

Proof. It is clear that X_1 is an invariant subspace for A. We know from Theorem 14.1 that X_1 is stable if and only if for each Riesz projection P of A corresponding to an eigenvalue λ_0 with $\dim \operatorname{Ker}(\lambda_0 - A) \geq 2$, we have $PX_1 = \{0\}$ or $PX_1 = \operatorname{Im} P$.

Let P be a Riesz projection of A corresponding to an arbitrary complex number λ_0. Also, for $i = 1, 2$, let P_i be the Riesz projection associated with A_i and λ_0. Then P has the form

$$P = \begin{bmatrix} P_1 & P_1 Q_1 + Q_2 P_2 \\ 0 & P_2 \end{bmatrix},$$

where Q_1 and Q_2 are certain linear operators acting from X_2 into X_1 (cf., the proof of Theorem 8.19). It follows that $\{0\} \neq PX_1 \neq \operatorname{Im} P$ if and only if λ_0 is a common eigenvalue of A_1 and A_2. This proves the theorem. □

14.2 Lipschitz stable invariant subspaces

In this section we consider a different concept of stability. Let T be a bounded linear operator on a Banach space X. A closed T-invariant subspace N of X is called *Lipschitz stable* if there exist $\delta > 0$ and $K > 0$ such that the following statement holds true. If S is a bounded linear operator on X and $\|S - T\| < \delta$, then S has a closed invariant subspace M such that $\operatorname{gap}(M, N) \leq K\|S - T\|$. Clearly, a Lipschitz stable invariant subspace is also a stable one.

If N is the image of a Riesz projection corresponding to T, then N is a Lipschitz stable invariant subspace for T. To see this, we argue as follows. Let Γ be a closed positively oriented Jordan curve not intersecting the spectrum of T such that $N = \operatorname{Im} P(T; \Gamma)$. For δ small enough and S a bounded linear operator on X such that $\|S - T\| < \delta$ also S will have no spectrum on Γ. Thus the Riesz

projection $P(S;\Gamma)$ is well defined too. Put $M = \operatorname{Im} P(S;\Gamma)$. From the material on the gap presented in Section 13.2 we recall that $\operatorname{gap}(M,N) \leq \|P(S;\Gamma) - P(T;\Gamma)\|$, and so

$$\begin{aligned} \operatorname{gap}(M,N) &\leq \|\frac{1}{2\pi i}\int_\Gamma (\lambda I - S)^{-1} - (\lambda I - T)^{-1}\, d\lambda\| \\ &= \|\frac{1}{2\pi i}\int_\Gamma (\lambda I - S)^{-1}(T-S)(\lambda I - T)^{-1}\, d\lambda\|. \end{aligned}$$

Now let C be such that $\max_{\lambda\in\Gamma} \|(\lambda I - T)^{-1}\| < C$. Such a C exists as $(\lambda I - T)^{-1}$ is continuous on Γ and Γ is compact. Take δ small enough so that $\delta C < 1$. Since $\lambda - S = (\lambda - T)\big(I - (\lambda - T)^{-1}(S-T)\big)$, and as $\|(\lambda - T)^{-1}(S-T)\| < \delta C < 1$ for $\lambda \in \Gamma$, we see that $\|(\lambda I - S)^{-1}\| < C(1-\delta C)^{-1},\ \lambda \in \Gamma$. Hence

$$\operatorname{gap}(M,N) \leq \frac{1}{2\pi}\ell(\Gamma)\,\frac{C^2}{1-\delta C}\,\|S-T\|,$$

where $\ell(\Gamma)$ denotes the length of Γ. Thus the spectral subspace N is a Lipschitz stable T-invariant subspace.

Not every stable T-variant subspace is Lipschitz stable. In fact, for the finite-dimensional case we shall show that the Lipschitz stable subspaces are precisely the images of Riesz projections.

Theorem 14.8. *Let X be a finite-dimensional Hilbert space, let T be a linear operator on X, and let N be a T-invariant subspace. Then N is Lipschitz stable if and only if it is the image of a Riesz projection for T.*

Proof. The arguments above show that a spectral subspace is Lipschitz stable. Hence we only need to show the converse. Let N be a Lipschitz stable T-invariant subspace. Since N is stable, we know (see Theorem 14.1) that for every eigenvalue λ of T with $\dim \operatorname{Ker}(\lambda - T) \geq 2$ either N contains the spectral subspace of T corresponding to λ, or N has zero intersection with that spectral subspace.

As in the proof of Lemma 14.6 one shows that for every eigenvalue λ of T the subspace $\operatorname{Im} P_\lambda(T)N$ is Lipschitz stable for the restriction of T to $\operatorname{Im} P_\lambda(T)$. Here $P_\lambda(T)$ is the Riesz projection of T corresponding to the eigenvalue λ. Recall that the spectral subspace $\operatorname{Im} P_\lambda(T)$ of T corresponding to an eigenvalue λ is given by $\operatorname{Ker}(\lambda - T)^n$, where n is the dimension of X. Also note that N is Lipschitz stable for T if and only if $S^{-1}N$ is Lipschitz stable for $S^{-1}TS$.

Now consider an eigenvalue λ of T with $\dim \operatorname{Ker}(\lambda - T) = 1$, and assume that $N \cap \operatorname{Ker}(\lambda - T)^n \neq \{0\}$. We have to show that $\operatorname{Ker}(\lambda - T)^n \subset N$. Assume this is not the case. Let $x_1, \ldots, x_p$ be a Jordan chain for T corresponding to the eigenvalue λ such that $x_1, \ldots, x_p$ form a basis for the spectral subspace $\operatorname{Ker}(\lambda - T)^n$. By our assumption and the fact that N is a stable invariant subspace, we see from the previous section that $N \cap \operatorname{Ker}(\lambda - T)^n = \operatorname{span}\{x_1, \ldots, x_j\}$ for some $j < p$. By the arguments of the previous paragraph we may assume that $X = \operatorname{span}\{x_1, \ldots, x_p\}$,

and that T is in Jordan normal form. More precisely, we may assume that $X = \mathbb{C}^p$ and $N = \text{span}\,\{e_1, \ldots, e_j\}$, where e_i is the ith unit vector in $\mathbb{C}^p$, while $T = \lambda I_p + J$ where J is a single Jordan block of order p with 0 on the main diagonal, i.e,

$$J = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & 0 & 1 \\ 0 & \cdots & \cdots & 0 & 0 \end{bmatrix}.$$

Now consider the perturbation $T(\varepsilon)$ obtained from T by changing the zero in the lower left-hand corner to $\varepsilon > 0$, that is,

$$T(\varepsilon) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & & \ddots & \lambda & 1 \\ \varepsilon & 0 & \cdots & 0 & \lambda \end{bmatrix}.$$

It is easily checked that the eigenvalues of $T(\varepsilon)$ are the pth roots of ε, i.e., they are given by $\varepsilon^{1/p}\exp(\ell\frac{2i\pi}{p})$ for $\ell = 1, \ldots, p$. The eigenvector of $T(\varepsilon)$ corresponding to $\lambda_\ell = \varepsilon^{1/p}\exp(\ell\frac{2i\pi}{p})$ is given by

$$y_\ell = \begin{bmatrix} 1 \\ \lambda_\ell \\ \lambda_\ell^2 \\ \vdots \\ \lambda_\ell^{p-1} \end{bmatrix}.$$

Thus, any j-dimensional invariant subspace of $T(\varepsilon)$ is spanned by j of these vectors. Let M be any one of them. Then M is spanned by, say, $y_{\ell_1}, \ldots, y_{\ell_j}$. Denote by P the orthogonal projection onto N, and by Q the orthogonal projection onto M. Let y_k be any one of the eigenvectors spanning M. Then

$$\text{gap}\,(N, M) \;=\; \|P - Q\| \;\geq\; \frac{1}{\|y_k\|}\|Py_k - Qy_k\| = \sqrt{\frac{\sum_{i=j}^{p-1} |\lambda_k^i|^2}{\sum_{i=0}^{p-1} |\lambda_k^i|^2}}\,.$$

Since $|\lambda_k| = \varepsilon^{1/p}$, we see that for ε sufficiently small

$$\operatorname{gap}(N, M) \geq \frac{1}{2}\sqrt[p]{\varepsilon^j}.$$

On the other hand, $\|T - T(\varepsilon)\| = \varepsilon$. From this it becomes clear that for $j = 1, \ldots, n-1$ the space N cannot be Lipschitz stable. □

14.3 Stable minimal factorizations of rational matrix functions

Throughout this section W_0, W_{01} and W_{02} are proper rational $m \times m$ matrix functions with value I_m at infinity. We assume that $W_0 = W_{01}W_{02}$ and that this factorization is minimal. In view of Theorems 13.1 and 13.7 the following definition is natural. Let

$$W_0(\lambda) = I_m + C_0(\lambda I_n - A_0)^{-1}B_0, \tag{14.8}$$

$$W_{0j}(\lambda) = I_m + C_{0j}(\lambda I_{n_j} - A_{0j})^{-1}B_{0j}, \qquad j = 1, 2, \tag{14.9}$$

be minimal realizations of W_0, W_{01} and W_{02}. The factorization $W_0 = W_{01}W_{02}$ is called *stable* if for each $\varepsilon > 0$ there exists $\omega > 0$ such that $\|A - A_0\| + \|B - B_0\| + \|C - C_0\| < \omega$ implies that the realization

$$W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B \tag{14.10}$$

is minimal and W admits a minimal factorization $W = W_1W_2$,

$$W_j(\lambda) = I_m + C_j(\lambda I_{n_j} - A_j)^{-1}B_j, \qquad j = 1, 2, \tag{14.11}$$

with the extra property that $\|A_j - A_{0j}\| + \|B_j - B_{0j}\| + \|C_j - C_{0j}\| < \varepsilon$ for $j = 1, 2$.

We make a few comments. The fact that the realization of W_0 in (14.8) is minimal, implies that the realization of W in (14.10) will also be minimal whenever the quantity $\|A - A_0\| + \|B - B_0\| + \|C - C_0\|$ is sufficiently small (regardless of ε). Next, note that the minimality of the factorization $W = W_1W_2$ with W_i given by (14.10) is clear from the minimality of the realization (14.10) and the identity $n = n_1 + n_2$ holding for the state space dimensions. Finally, since in the finite-dimensional case all minimal realizations of a given transfer function are mutually similar, the above definition does not depend on the particular choice of the minimal realizations (14.8) and (14.9).

From Theorem 13.1 we see that a sufficient condition for the factorization $W_0 = W_{01}W_{02}$ to be stable is that W_{01} and W_{02} have no common poles and no common zeros. The next theorem characterizes stability of minimal factorization in terms of spectral data.

Theorem 14.9. *Suppose $W_0 = W_{01}W_{02}$ is a minimal factorization. This factorization is stable if and only if each common pole (respectively zero) of W_{01} and W_{02} is a pole (respectively zero) of W_0 of geometric multiplicity one.*

In connection with this result (and a number of theorems below) we recall from Sections 8.2 and 8.4 (see Corollary 8.10 and the last sentence of the first paragraph of Section 8.4) that a pole (zero) λ_0 of W_0 has geometric multiplicity one if and only if the order of λ_0 as a pole of W_0 (W_0^{-1}) is equal to the local degree $\delta(W_0;\lambda_0)$ ($\delta(W_0^{-1};\lambda_0)$).

The proof of Theorem 14.9 will be given in a number of steps. Recall that there is a one-one correspondence between minimal factorizations and supporting projections of minimal realizations (see Theorem 9.3). Therefore we begin by characterizing stability of minimal factorizations in terms of supporting projections. This leads to the notion of a stable supporting projection.

Let Π_0 be a supporting projection for the system $\Theta_0 = (A_0, B_0, C_0; X, Y)$. We call Π_0 *stable* if, given $\varepsilon > 0$, there exists $\omega > 0$ such that the following is true. If $\Theta = (A, B, C; X, Y)$ is a system satisfying $\|\Theta - \Theta_0\| < \omega$, then Θ has a supporting projection Π such that $\|\Pi - \Pi_0\| < \varepsilon$.

Lemma 14.10. *Let $W_0(\lambda) = I_m + C_0(\lambda I_n - A_0)^{-1}B_0$ be a minimal realization, and let Π_0 be the supporting projection for the system $\Theta_0 = (A_0, B_0, C_0; \mathbb{C}^n, \mathbb{C}^m)$ corresponding to the minimal factorization $W_0 = W_{01}W_{02}$. This factorization is stable if and only if Π_0 is stable.*

Proof. We know already that for $\|A - A_0\| + \|B - B_0\| + \|C - C_0\|$ sufficiently small the realization $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ will be minimal. So, if Π_0 is stable, we can apply Theorem 13.5 to show that the factorization $W_0 = W_{01}W_{02}$ is stable too.

Conversely, let the factorization $W_0 = W_{01}W_{02}$ be a stable factorization, and assume Π_0 is not stable. Then there exist $\varepsilon > 0$ and a sequence $\Theta_1, \Theta_2, \ldots$ of systems such that $\|\Theta_k - \Theta_0\| \to 0$ for $k \to \infty$ and $\|\Pi - \Pi_0\| \geq \varepsilon$ for each supporting projection Π of Θ_k. Since Θ_0 is minimal and $\|\Theta_k - \Theta_0\| \to 0$, we may assume that Θ_k is minimal for all k. Also we may assume that for each k the transfer function $W_k = W_{\Theta_k}$ admits a minimal factorization $W_k = W_{k1}W_{k2}$,

$$W_{kj}(\lambda) = I_m + C_{kj}(\lambda I_{n_j} - A_{kj})^{-1}B_{kj}, \qquad j = 1, 2,$$

such that for $j = 1, 2$ we have

$$A_{kj} \to A_{0j}, \qquad B_{kj} \to B_{0j}, \qquad C_{kj} \to C_{0j}, \qquad (k \to \infty). \tag{14.12}$$

Here $I_m + C_{0j}(\lambda I_{n_j} - A_{0j})^{-1}B_{0j}$ is a minimal realization of W_{0j}.

Let Π_k be the supporting projection for Θ_k corresponding to the minimal factorization $W_k = W_{k1}W_{k2}$. Write $\Theta_{ki} = (A_{ki}, B_{ki}, C_{ki}; \mathbb{C}^{n_j}, \mathbb{C}^m)$. Then $\Theta_{k1}\Theta_{k2}$ and Θ_k are similar, say with system similarity $S_k : \mathbb{C}^{n_i} \oplus \mathbb{C}^{n_2} \to \mathbb{C}^n$. For $k = 0, 1, 2, \ldots$, we have $\Pi_k = S_k P S_k^{-1}$, where P is the projection of $\mathbb{C}^{n_1} \oplus \mathbb{C}^{n_2}$ along $\mathbb{C}^{n_1}$ onto $\mathbb{C}^{n_2}$. From Theorem 7.7 we know how S_k can be described explicitly. This description, together with (14.12) and $\|\Theta_k - \Theta_0\| \to 0$, gives $S_k \to S_0$. So $\Pi_k \to \Pi_0$, which contradicts the fact that $\|\Pi_k - \Pi_0\| \geq \varepsilon$ for all n. We conclude that Π_0 must be stable. □

Next we make the connection with stable invariant subspaces.

Lemma 14.11. *Let $\Theta_0 = (A_0, B_0, C_0; X, Y)$ be a given system, and let Π_0 be a supporting projection for this system. Then Π_0 is stable if and only if* $\operatorname{Ker}\Pi_0$ *and* $\operatorname{Im}\Pi_0$ *are stable invariant subspaces for A_0 and $A_0^\times = A_0 - B_0C_0$, respectively.*

Proof. Let $\operatorname{Ker}\Pi_0$ and $\operatorname{Im}\Pi_0$ be stable invariant subspaces for A_0 and $A_0^\times$, respectively. Assume Π_0 is not stable. Then there exist $\varepsilon > 0$ and a sequence $\Theta_1, \Theta_2, \dots$ of systems such that $\|\Theta_k - \Theta_0\| \to 0$ and $\|\Pi - \Pi_0\| \geq \varepsilon$ for every supporting projection Π of Θ_k. Write $\Theta_k = (A_k, B_k, C_k; X, Y)$. Then clearly

$$A_k \to A_0, \qquad A_k^\times = A_k - B_kC_k \to A_0 - B_0C_0 = A_0^\times, \qquad (k \to \infty).$$

But then our hypothesis ensures the existence of two sequences $M_1, M_2, \dots$ and $M_1^\times, M_2^\times, \dots$ of closed subspaces of X such that

$$A_k[M_k] \subset M_k, \qquad A_k^\times[M_k^\times] \subset M_k^\times, \qquad k = 1, 2, \dots,$$

while in addition

$$\operatorname{gap}(M_k, \operatorname{Ker}\Pi_0) \to 0, \qquad \operatorname{gap}(M_k^\times, \operatorname{Im}\Pi_0) \to 0, \qquad (k \to \infty). \tag{14.13}$$

By [71], Theorem 2 we may assume that $X = M_k \dot{+} M_k^\times$ for all k. Let Π_k be the projection of X along M_k onto $M_k^\times$. Then Π_k is a supporting projection for Θ_k. Moreover, it follows from (14.13) that $\Pi_k \to \Pi_0$. This contradicts the fact that $\|\Pi_k - \Pi_0\| \geq \varepsilon$ for all k. So Π_0 must be stable.

Now conversely, let Π_0 be a stable supporting projection for Θ_0 and assume $\operatorname{Ker}\Pi_0$ is not stable for A_0. Then there exist $\varepsilon > 0$ and a sequence $A_1, A_2, \dots$ of bounded linear operators on X such that $A_k \to A_0$ and $\operatorname{gap}(M, \operatorname{Ker}\Pi_0) \geq \varepsilon$ for each closed invariant subspace M of A_k. Put $\Theta_k = (A_k, B_0, C_0; X, Y)$. Then $\|\Theta_k - \Theta_0\| \to 0$. So we can find a sequence $\Pi_1, \Pi_2, \dots$ of projections such that Π_k is a supporting projection for Θ_k and $\Pi_k \to \Pi_0$ when $k \to \infty$. Hence $\operatorname{Ker}\Pi_k$ is a closed invariant subspace for A_k and $\operatorname{gap}(\operatorname{Ker}\Pi_k, \operatorname{Ker}\Pi_0) \to 0$. But this conflicts with $\operatorname{gap}(\operatorname{Ker}\Pi_k, \operatorname{Ker}\Pi_0) \geq \varepsilon$, $k = 1, 2, \dots$. So $\operatorname{Ker}\Pi_0$ must be stable for A_0. Likewise $\operatorname{Im}\Pi_0$ is a stable invariant subspace for $A_0^\times$. □

We now come to the proof of Theorem 14.9. Recall that W_0, W_{01} and W_{02} are proper rational $m \times m$ matrix functions that are analytic at ∞ with value I_m. Moreover $W_0 = W_{01}W_{02}$, and this factorization are minimal.

Proof of Theorem 14.9. Let $W_0(\lambda) = I_m + C_0(\lambda I_n - A_0)^{-1}B_0$ be a minimal realization for W_0, and let Π_0 be the supporting projection for the system $\Theta_0 = (A_0, B_0, C_0; \mathbb{C}^n, \mathbb{C}^m)$ corresponding to the minimal factorization $W_0 = W_{01}W_{02}$. From Lemma 14.10 we know that this factorization is stable if and only if Π_0 is stable.

With respect to the decomposition $\mathbb{C}^n = \operatorname{Ker}\Pi_0 \dot{+} \operatorname{Im}\Pi_0$, we write

$$A_0 = \begin{bmatrix} A_1 & * \\ 0 & A_2 \end{bmatrix}.$$

Applying Theorem 14.7 we see that $\operatorname{Ker}\Pi_0$ is a stable invariant subspace for A_0 if and only if each common eigenvalue of A_1 and A_2 is an eigenvalue of A_0 of geometric multiplicity one. But then Lemma 9.2 gives that $\operatorname{Ker}\Pi_0$ is stable for A_0 if and only if each common eigenvalue of A_1 and A_2 is a pole of W_0 of geometric multiplicity one. Observe now that A_1 and A_2 are the main operators in the systems $\operatorname{pr}_{\Pi_0}(\Theta)$ and $\operatorname{pr}_{I_n-\Pi_0}(\Theta)$, respectively. Since these systems are minimal, we have that $\sigma(A_1)$ and $\sigma(A_2)$ coincide with the sets of poles of W_{01} and W_{02}, respectively. Hence $\operatorname{Ker}\Pi_0$ is stable for A_0 if and only if each common pole of W_{01} and W_{02} is a pole of W_0 of geometric multiplicity one. Likewise $\operatorname{Im}\Pi_0$ is stable for $A_0^\times$ if and only if each common zero of W_{01} and W_{01} is a zero of W_0 of geometric multiplicity one. The desired result is now immediate from Lemma 14.11. □

In the remainder of this section we deal with Lipschitz stability. As before, W_0, W_{01} and W_{02} are proper rational $m \times m$ matrix functions with value I_m at infinity, and we assume that $W_0 = W_{01}W_{02}$ is a minimal factorization. Let

$$W_0(\lambda) = I_m + C_0(\lambda I_n - A_0)^{-1}B_0,$$

$$W_{0j}(\lambda) = I_m + C_{0j}(\lambda I_{n_j} - A_{0j})^{-1}B_{0j}, \qquad j = 1, 2,$$

be minimal realizations of W_0, W_{01} and W_{02}. The factorization $W_0 = W_{01}W_{02}$ is called *Lipschitz stable* if there are positive constants ω and K such that the inequality $\|A - A_0\| + \|B - B_0\| + \|C - C_0\| < \omega$ implies that the realization

$$W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$$

is minimal and W admits a minimal factorization $W = W_1W_2$,

$$W_j(\lambda) = I_m + C_j(\lambda I_{n_j} - A_j)^{-1}B_j, \qquad j = 1, 2,$$

with the extra property that

$$\|A_j - A_{0j}\| + \|B_j - B_{0j}\| + \|C_j - C_{0j}\| \\ \leq K\big(\|A - A_0\| + \|B - B_0\| + \|C - C_0\|\big), \qquad j = 1, 2.$$

The comments that have been made in the second paragraph of this section (about the definition given in the beginning of that section) apply here too. In particular, the definition of Lipschitz stability given above does not depend on the particular choice of the minimal realizations for W_0, W_{01} and W_{02}. Note that we already encountered a Lipschitz stable factorization in Theorem 13.7 above, without using the term there (cf., Theorem 14.12 below).

Given the above realization of W_0, one has that the minimal factorization $W_0 = W_{01}W_{02}$ is Lipschitz stable if and only if for the corresponding supporting

projection Π_0, the kernel $\operatorname{Ker}\Pi_0$ is a Lipschitz stable A_0-invariant subspace and the range $\operatorname{Im}\Pi_0$ is a Lipschitz stable $A_0^\times$-invariant subspace. Here, as usual, we have $A_0^\times = A_0 - B_0C_0$. This is analogous to the situation for stable invariant subspaces (cf., Lemma 14.11). The results of Section 14.2 now imply that the factorization $W_0 = W_{01}W_{02}$ is Lipschitz stable if and only if both $\operatorname{Ker}\Pi_0$ and $\operatorname{Im}\Pi_0$ are images of Riesz projections for A_0 and $A_0^\times$, respectively. This leads to the following result.

Theorem 14.12. *The minimal factorization $W_0 = W_{01}W_{02}$ is Lipschitz stable if and only if W_{01} and W_{02} have no common poles and no common zeros.*

14.4 Stable complete factorizations

Let W_0 be a rational $m \times m$ matrix function with minimal realization

$$W_0(\lambda) = I_m + C_0(\lambda I_n - A_0)^{-1}B_0. \tag{14.14}$$

Suppose

$$W_0(\lambda) = \left(I_m + \frac{1}{\lambda - \alpha_{01}}R_{01}\right)\cdots\left(I_m + \frac{1}{\lambda - \alpha_{0n}}R_{0n}\right) \tag{14.15}$$

is a complete factorization of W_0. We say that this complete factorization (14.16) is *stable* if for all $\varepsilon > 0$ there exists $\omega > 0$ such that

$$\|A - A_0\| + \|B - B_0\| + \|C - C_0\| < \omega$$

implies that the realization $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ is minimal and W admits a complete factorization

$$W(\lambda) = \left(I_m + \frac{1}{\lambda - \alpha_1}R_1\right)\cdots\left(I_m + \frac{1}{\lambda - \alpha_n}R_n\right) \tag{14.16}$$

with the extra property that

$$|\alpha_j - \alpha_{0j}| + \|R_j - R_{0j}\| < \varepsilon, \qquad j = 1, \ldots, n. \tag{14.17}$$

Note that this definition is not completely analogous to the one of a stable minimal factorization involving just two factors as given in Section 14.3. To mimic that one, we should write the complete factorizations (14.15) and (14.16) in the form

$$W_0(\lambda) = \left(I_m + \frac{1}{\lambda - \alpha_{01}}c_{01}b_{01}^\top\right)\cdots\left(I_m + \frac{1}{\lambda - \alpha_{0n}}c_{0n}b_{0n}^\top\right), \tag{14.18}$$

$$W(\lambda) = \left(I_m + \frac{1}{\lambda - \alpha_1}c_1b_1^\top\right)\cdots\left(I_m + \frac{1}{\lambda - \alpha_n}c_nb_n^\top\right), \tag{14.19}$$

with b_{0j}, c_{0j}, b_j, c_j nonzero vectors in $\mathbb{C}^n$, and the estimates (14.17) as

$$|\alpha_j - \alpha_{0j}| + \|b_j - b_{0j}\| + \|c_j - c_{0j}\| < \varepsilon, \qquad j = 1, \ldots, n.$$

A routine argument shows that the two possibilities amount to the same.

Theorem 14.13. *Let W_0 be a proper rational $m \times m$ matrix function with $W_0(\infty) = I_m$. A necessary condition for W_0 to admit a stable complete factorization is that the poles and zeros of W_0 all have geometric multiplicity one. In that case there are only finitely many complete factorizations of W_0 and these are all stable.*

Note that the theorem does not guarantee that W_0 admits a complete factorization. The number of these factorizations might be zero.

Recall from Section 8.2 that a complex number λ_0 is a pole of W_0 of geometric multiplicity one if and only if $\dim \operatorname{Pol}(W; \lambda_0) = 1$. Similarly (see Section 8.1), λ_0 is a zero of W_0 of geometric multiplicity one if and only if $\dim \operatorname{Ker}(W; \lambda_0) = 1$.

Now assume that W_0 is given by the minimal realization (14.14). Then, using the final statement of Corollary 8.22 we see that $\dim \operatorname{Ker}(\lambda_0 - A_0) = \dim \operatorname{Pol}(W; \lambda_0)$. It follows that λ_0 is a pole of W_0 of geometric multiplicity one if and only if λ_0 is an eigenvalue of geometric multiplicity one of A_0. Applying this result to W_0^{-1} and using Lemma 8.8, we obtain that λ_0 is a zero of W_0 of geometric multiplicity one if and only λ_0 is an eigenvalue of geometric multiplicity one of the associate state matrix $A_0^\times$. Therefore the poles and zeros of W_0 all have geometric multiplicity one if and only if both A_0 and $A_0^\times$ are nonderogatory. Hence an equivalent way to formulate Theorem 14.13 is as follows.

Theorem 14.14. *Let W_0 be a proper rational $m \times m$ matrix function with minimal realization $W_0(\lambda) = I_m + C_0(\lambda I_n - A_0)^{-1} B_0$. A necessary condition for W_0 to admit a stable complete factorization is that both A_0 and $A_0^\times = A_0 - B_0 C_0$ are nonderogatory. In that case there are only finitely many complete factorizations of W_0 and these are all stable.*

Companion matrices are nonderogatory. Thus we have the following immediate consequence of Theorem 14.14.

Corollary 14.15. *Let W_0 be a companion based rational $m \times m$ matrix function. Then there are only finitely many complete factorizations of W_0 and these are all stable.*

Again Theorem 14.14 and Corollary 14.15 do not guarantee that W_0 admit a complete factorization.

From Sections 10.1–10.3 we know that complete factorizations are closely related to complete chains of invariant subspaces. The proof of Theorem 14.13 is therefore similar to that of Theorem 14.9 provided one has an analogue of Theorem 14.1, where the single invariant subspace featuring there is replaced by a complete chain of invariant subspaces. Thus our task here is to analyze stability of complete chains of invariant subspaces. First we give the formal definition.

Let A be an $n \times n$ matrix, whenever convenient identified with its canonical action on $\mathbb{C}^n$, considered here as a Hilbert space. Suppose $\mathcal{M} = \{M_l\}_{l=0}^n$ is a complete chain of A-invariant subspaces, i.e.,

$$\{0\} = M_0 \subset M_1 \subset \cdots \subset M_{n-1} \subset M_n = \mathbb{C}^n$$

and, for $l = 1, \ldots, n-1$, $\dim M_l = l$ and $A[M_l] \subset M_l$. We call $\mathcal{M}$ *stable* (for A) if given $\varepsilon > 0$, there exists $\delta > 0$ such that the following is true: if B is a an $n \times n$ matrix and $\|B - A\| < \delta$, then there exists a complete chain $\mathcal{N} = \{N_l\}_{l=0}^k$ of B-invariant subspaces for which the *gap between* $\mathcal{M}$ *and* $\mathcal{N}$ defined by

$$\mathrm{GAP}\,(\mathcal{M}, \mathcal{N}) = \max_{l=1,\ldots,n-1} \mathrm{gap}\,(N_l, M_l) \tag{14.20}$$

does not exceed ε. Note that the values 0 and n for l do not play a role in (14.20) because all chains start with $\{0\}$ and end with $\mathbb{C}^n$.

Theorem 14.16. *The matrix A has a stable complete chain of invariant subspaces if and only if A is nonderogatory. In that case, A has a finite number of complete chains of invariant subspaces and all these chains are stable.*

Proof. Suppose A has a stable complete chain $\{M_l\}_{l=0}^k$ of invariant subspaces. From the definition given above it is clear that $M_0, \ldots, M_k$ are then stable invariant subspaces for A. Assume that A is derogatory, and let λ_0 be an eigenvalue of A with $\dim \mathrm{Ker}\,(\lambda_0 - A) \geq 2$. Write P for the spectral projection corresponding to λ_0 and A_0 for the restriction of A (viewed as an operator) to $\mathrm{Im}P$. According to Lemma 14.6, the subspaces $PM_0, \ldots, PM_k$ are stable for A_0. Hence, on account of Lemma 14.4, these subspaces are either trivial or have dimension at least two. However, as $\{M_l\}_{l=0}^k$ is a complete chain of A-invariant subspaces, there must be at least one j for which the dimension of PM_j is one (cf., first part of the proof of Theorem 12.2, Step 4). Contradiction. The conclusion is that A is nonderogatory.

Next assume that A is nonderogatory and let $\mathcal{M} = \{M_l\}_{l=0}^k$ be a complete chain of A-invariant subspaces. By Theorem 14.1 each A-invariant subspace is stable. In particular, all members M_l of $\mathcal{M}$ are stable. But then $\mathcal{M}$ is a stable complete chain of invariant subspaces for A by an observation made on page 464 of the book [70], in a comment connected to Theorem 15.6.1 in the same book [70], which is concerned with the more general notion of a stable lattice.

To complete the proof, we recall that if the matrix A is nonderogatory, it has only a finite number of invariant subspaces and hence the collection of all complete chains of A-invariant subspaces is a finite set. In fact, if the nonderogatory matrix A has s different eigenvalues with algebraic multiplicities $m_1, \ldots, m_s$, the number of complete chains of A-invariant subspaces is

$$\frac{(m_1 + m_2 + \cdots + m_s)!}{m_1! \times m_2! \times \cdots \times m_s!}.$$

This follows from Proposition 11.19 by virtue of the well-known fact that the nonderogatory matrices are precisely those that are similar to first companions. □

Proof of Theorems 14.13 *and* 14.14. We shall focus on Theorem 14.14 which is nothing else than a reformulation of Theorem 14.13.

Let $W_0(\lambda) = I_m + C_0(\lambda - A_0)^{-1}B_0$ be a minimal realization of W_0. As we know, there is a one-to-one correspondence between minimal factorizations of W_0 and direct sum decompositions $\mathbb{C}^n = M \dot{+} M^\times$ where the subspaces M and $M^\times$ are invariant for A and $A^\times$, respectively. From Lemmas 14.10 and 14.11 we see that a minimal factorization is stable if and only if the corresponding subspaces M and $M^\times$ are stable for A and $A^\times$, respectively. This fact has a straightforward analogue for complete factorizations. Indeed, a complete factorization of W_0 corresponds with two complete chains of subspaces: a chain

$$\{0\} = M_0 \subset M_1 \subset \cdots \subset M_{n-1} \subset M_n = \mathbb{C}^n \tag{14.21}$$

of A-invariant subspaces, and a chain

$$\{0\} = M_0^\times \subset M_1^\times \subset \cdots \subset M_{n-1}^\times \subset M_n^\times = \mathbb{C}^n \tag{14.22}$$

of $A^\times$-invariant subspaces, such that for $j = 1, \ldots, n-1$ the subspaces M_j and $M_{n-j}^\times$ match in the sense that $M_j \dot{+} M_{n-j}^\times = \mathbb{C}^n$. Now the complete factorization in question is stable if and only if (14.21) is a stable complete chain for A and (14.22) is a stable complete chain for $A^\times$. In view of this, Theorem 14.14 is an immediate consequence of Theorem 14.16. □

Next, we consider Lipschitz stability of complete chains of invariant subspaces. Let $\mathcal{M} = \{M_l\}_{l=0}^k$ be a complete chain of A-invariant subspaces. This chain is called *Lipschitz stable* if there are positive constants δ and K such that the following holds: if B is a $k \times k$ matrix with $\|A - B\| < \delta$, then B has a complete chain of invariant subspaces $\mathcal{N} = \{N_l\}_{l=0}^k$ with $\mathrm{GAP}\,(\mathcal{M}, \mathcal{N}) \leq K\|A - B\|$.

Theorem 14.17. *The $k \times k$ matrix A has a Lipschitz stable complete chain of invariant subspaces if and only if A has k distinct eigenvalues. In that case all complete chains of invariant subspaces are Lipschitz stable.*

Proof. Suppose that A has a Lipschitz stable complete chain of invariant subspaces. From the above definition it follows that each of the invariant subspaces M_l is Lipschitz stable, and hence must be a spectral subspace (see Theorem 14.8). Then it is easily seen that A must have k distinct eigenvalues.

Conversely, assume that A has k distinct eigenvalues. Then, for δ small enough, if $\|A - B\| < \delta$ also B has k distinct eigenvalues. Now selecting a complete chain of invariant subspaces of such a matrix is equivalent to choosing an ordering of the eigenvalues. Suppose that the eigenvalues $\lambda_1, \ldots, \lambda_k$ of A are ordered so that for the corresponding unit eigenvectors x_l, $l = 1, \ldots, k$ we have $M_l = \mathrm{span}\,\{x_1, \ldots, x_l\}$. Let $\mu_1, \ldots \mu_k$ be the eigenvalues of B ordered so that $|\lambda_l - \mu_l|$ is small, and let y_l, $l = 1, \ldots, k$ be the corresponding eigenvalues. Consider $N_l = \mathrm{span}\,\{y_1, \ldots, y_l\}$. Then, letting B tend to A and using the fact that all M_l's are Lipschitz stable we see that the chain $\mathcal{M}$ is Lipschitz stable. □

Next, we define Lipschitz stability of complete factorizations. Suppose W_0 is a rational $m \times m$ matrix function with minimal realization (14.14). The complete factorization (14.15) of W_0 is called *Lipschitz stable* if there are positive constants ω and K such that $\|A-A_0\|+\|B-B_0\|+\|C-C_0\| < \omega$ implies that the realization $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ is minimal and W admits a complete factorization (14.16) with the property that

$$|\alpha_j - \alpha_{0j}| + \|R_j - R_{0j}\| \leq K\big(\|A - A_0\| + \|B - B_0\| + \|C - C_0\|\big),$$

where j is allowed to take the values $1, \ldots, n$. An equivalent definition using (14.18) and (14.19) is, of course, again possible.

Theorem 14.18. *Let W_0 be a proper rational $m \times m$ matrix function with $W_0(\infty) = I_m$. A necessary condition for W_0 to admit a Lipschitz stable complete factorization is that the poles and zeros of W_0 all have geometric and algebraic multiplicity one. In that case all complete factorizations are Lipschitz stable.*

Proof. Like in the proof of Theorem 14.14 we see that a complete factorization is Lipschitz stable if and only if the corresponding complete chains of invariant subspaces for A_0 and $A_0^\times$ are Lipschitz stable. The theorem is then a direct consequence of Theorem 14.17. □

14.5 Stable factorizations of monic matrix polynomials

Throughout this section m will be a fixed positive integer. Given a positive integer ℓ, we denote the set of all monic $m \times m$ matrix polynomials of degree ℓ by $\mathcal{MP}_\ell$. If L_1 and L_2 are in $\mathcal{MP}_\ell$, say

$$L_j(\lambda) = \lambda^\ell I + \sum_{i=0}^{\ell-1} \lambda^j A_{ji}, \qquad j = 1, 2,$$

we put

$$\|L_1 - L_2\| = \sum_{i=0}^{\ell-1} \|A_{1i} - A_{2i}\|.$$

This defines a metric on $\mathcal{MP}_\ell$.

Suppose L_0, L_{01} and L_{02} are monic $m \times m$ matrix polynomials of degree p, q and r, respectively. So $L_0 \in \mathcal{MP}_p$, $L_{01} \in \mathcal{MP}_q$ and $L_{02} \in \mathcal{MP}_r$. Assume $L_0 = L_{02}L_{01}$. We say that this factorization is *stable* if, given $\varepsilon > 0$, there exists $\delta > 0$ with the following property. If $L \in \mathcal{MP}_p$ and $\|L - L_0\| < \delta$, then L admits a factorization $L = L_2L_1$ with $L_1 \in \mathcal{MP}_q$, $L_2 \in \mathcal{MP}_r$ and

$$\|L_j - L_{0j}\| < \varepsilon, \qquad j = 1, 2.$$

The aim of this section is to characterize stability of a factorization in terms of spectral data. We begin by making the connection with stable invariant subspaces.

This will be done via the notion of a supporting subspace, here always taken with respect to first companion systems (see Section 3.4). For briefness sake we shall simply speak about supporting subspaces (of the first companion operator) of the given monic matrix polynomial L_0. Recall that there is a one-one correspondence between the supporting subspaces of L_0 and the factorizations of L_0 into monic operator polynomials.

Lemma 14.19. *Let L_0, L_{01} and L_{02} be monic $m \times m$ matrix polynomials, and assume $L_0 = L_{02}L_{01}$. This factorization is stable if and only if the corresponding supporting subspace is stable for the first companion operator of L_0.*

Proof. It is possible to give a rather quick proof based on [68], Theorem 3. We prefer however to present a more direct argument.

As before, we write p for the degree of L_0 and q for that of L_{01}. Further, the first companion operator of L_0 is indicated by C_0, the supporting subspace of L_0 corresponding to the factorization $L_0 = L_{02}L_{01}$ by M_0.

Suppose the factorization is stable. In order to prove that M_0 is a stable invariant subspace for C_0 we consider a sequence $C_1, C_2, \ldots$ of operators converging to C_0. Using the Kronecker delta notation, put

$$Q = \operatorname{row}(\delta_{j1}I)_{j=1}^{p}, \qquad S_n = \operatorname{col}(QC_n^j)_{j=0}^{p-1}, \qquad n = 0, 1, 2, \ldots .$$

Then $S_1, S_2, \ldots$ converges to S_0 which is equal to the identity operator on $\mathbb{C}^{mp}$. So, passing if necessary to a subsequence, we may assume that S_n is invertible for all n. Write $S_n^{-1} = \operatorname{row}(U_{ni})_{i=1}^{p}$. Then

$$U_{ni} \to \operatorname{col}(\delta_{ji}I)_{j=1}^{p}, \qquad i = 1, \ldots, p. \tag{14.23}$$

A straightforward calculation shows that $S_nC_nS_n^{-1}$ is the first companion operator associated with the monic operator polynomial

$$L_n(\lambda) = \lambda^p I - \sum_{i=1}^{p} \lambda^{i-1} QC_n^p U_{ni}.$$

From (14.23) and the fact that $C_n \to C_0$ it follows that $\|L_n - L_0\| \to 0$. But then we may assume that for all n the polynomial L_n admits a factorization $L_n = L_{n2}L_{n1}$ with $L_{n1} \in \mathcal{MP}_q$, $L_{n2} \in \mathcal{MP}_r$, $r = p - q$, and

$$\|L_{nj} - L_{0j}\| \to 0, \qquad j = 1, 2.$$

Let M_n be the supporting subspace corresponding to the factorization $L_n = L_{n2}L_{n1}$. We shall show that $M_n \to M_0$ in the gap topology. In order to do this we describe M_n as follows. Let D_n be the first companion operator of L_{n1}. Then M_n is the image of the operator

$$\operatorname{col}(QD_n^i)_{i=0}^{p-1} : \mathbb{C}^{kr} \to \mathbb{C}^{kp}$$

(see Section 3.4). Define P to be the projection from $\mathbb{C}^{mp} = \mathbb{C}^{mr} \dot{+} \mathbb{C}^{m(p-r)}$ onto $\mathbb{C}^{mr}$ given by $P = \begin{bmatrix} I & 0 \end{bmatrix}$. Since P is surjective, we have that $M_n = \operatorname{Im} P_n$, where $P_n = \big(\operatorname{col}(QD_n^i)_{i=0}^{p-1}\big)P$ has the form

$$P_n = \begin{bmatrix} I & 0 \\ F_n & 0 \end{bmatrix} : \mathbb{C}^{mr} \dot{+} \mathbb{C}^{m(p-r)} \to \mathbb{C}^{mr} \dot{+} \mathbb{C}^{m(p-r)}.$$

Observe that P_n is a projection. Now $\|L_{n1} - L_{01}\| \to 0$ implies that $F_n \to F_0$. Hence $P_n \to P_0$. But $\operatorname{gap}(M_n, M_0) = \operatorname{gap}(\operatorname{Im} P_k, \operatorname{Im} P_0) \leq \|P_n - P_0\|$, and so $\operatorname{gap}(M_n, M_0) \to 0$.

Put $V_n = S_n^{-1} M_n$. Then V_n is an invariant subspace for C_n. Moreover, it follows from $S_n \to I$ that $\operatorname{gap}(V_n, M_n) \to 0$. But then $\operatorname{gap}(V_n, M) \to 0$, and the first part of the proof is complete.

Next assume that M_0 is a stable invariant subspace of C_0, and let $L_1, L_2, \ldots$ be a sequence in $\mathcal{MP}_p$ converging to L_0. Denote the first companion operator of L_n by C_n. Then $C_n \to C_0$, and hence there exists a C_n-invariant subspace M_n of $\mathbb{C}^{mp}$ such that $\operatorname{gap}(M_n, M_0) \to 0$. Recall now that $\mathbb{C}^{mp} = M_0 \dot{+} N_q$, where

$$N_q = \left\{ x = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} \in \mathbb{C}^{mp} \,\middle|\, x_j \in \mathbb{C}^m,\ x_1 = \cdots = x_q = 0 \right\}. \tag{14.24}$$

So, passing if necessary to a subsequence, we may assume that

$$\mathbb{C}^{mp} = M_n \dot{+} N_q, \qquad n = 0, 1, 2, \ldots . \tag{14.25}$$

This means that M_n is a supporting subspace for L_n. Let $L_n = L_{n2} L_{n1}$ be the corresponding factorization. We need to show that $\|L_{n1} - L_{01}\| \to 0$ and $\|L_{n2} - L_{02}\| \to 0$.

With respect to the decomposition (14.24) we write

$$C_n = \begin{bmatrix} C_{n1} & C_{n0} \\ 0 & C_{n2} \end{bmatrix}, \qquad Q_n = \begin{bmatrix} Q_{n1} & Q_{n2} \end{bmatrix}.$$

The polynomial L_{n1} can be explicitly expressed in terms of C_{n1} and Q_{n1} (cf., Section 3.4). A complication here is that the decomposition (14.25) depends on n. This difficulty however can be easily overcome by the usual angular operator argument. From the expression for L_{n1} one then sees that $\|L_{n1} - L_{01}\| \to 0$. In the same way one shows that $\|L_{n2} - L_{02}\| \to 0$, and the proof is complete. □

Recall that a complex number λ_0 is an eigenvalue of the matrix polynomial L if $L(\lambda_0)$ is not invertible. In that case $\operatorname{Ker} L(\lambda_0)$ is non-trivial and its dimension is the geometric multiplicity of λ_0 as an eigenvalue of L. This number is also equal to the geometric multiplicity of λ_0 as an eigenvalue of the first companion operator of L.

Theorem 14.20. *Let L_0, L_{01} and L_{02} be monic $m \times m$ matrix polynomials, and assume $L_0 = L_{02}L_{01}$. This factorization is stable if and only if each common eigenvalue of L_{01} and L_{02} is an eigenvalue of L_0 of geometric multiplicity one.*

Proof. Let M_0 be the supporting subspace of L_0 corresponding to the factorization $L_0 = L_{02}L_{01}$. From Lemma 14.19 we know that this factorization is stable if and only if M_0 is a stable invariant subspace for the first companion operator C_0 of L_0. Let p be the degree of L_0, let q be the degree of L_{01}, and let N_q be as in (14.24). Then $\mathbb{C}^{mp} = M_0 \dot{+} N_q$. With respect to this decomposition we write

$$C_0 = \begin{bmatrix} C_{01} & C_{00} \\ 0 & C_{02} \end{bmatrix}.$$

Then it is known (cf., Section 3.4 and the one but last paragraph of Section 4.3) that a complex number is an eigenvalue of C_{0i} if and only if it is an eigenvalue of $L_{0i},\ i = 1, 2$. The desired result is now obtained by applying Theorem 14.7. □

Next, we discuss Lipschitz stability. Let $L_0 = L_{02}L_{01}$ be a factorization of the monic matrix polynomial L_0 into monic factors, as above in the definition of stable factorization given at the start of this section. We shall say that this factorization is *Lipschitz stable* if there are positive constant K and δ such that $L \in \mathcal{M}_p$ with $\|L - L_0\| < \delta$ admits a factorization $L = L_2L_1$ with $L_1 \in \mathcal{M}_q, L_2 \in \mathcal{M}_r$ and $\|L_i - L_{0i}\| \leq K\|L - L_0\|$. One can prove (see Theorem 17.3.1 in [70]) that the factorization is Lipschitz stable if and only if the corresponding supporting subspace is Lipschitz stable. This gives the following theorem.

Theorem 14.21. *The factorization $L_0 = L_{02}L_{01}$ is Lipschitz stable if and only if L_{01} and L_{02} have no common eigenvalues.*

14.6 Stable solutions of the operator Riccati equation

Consider the operator Riccati equation

$$XT_{21}X + XT_{22} - T_{11}X - T_{12} = 0. \tag{14.26}$$

Here $T_{ij} \in \mathcal{L}(Y_j, Y_i),\ i, j = 1, 2$, where Y_1 and Y_2 are assumed to be finite-dimensional Banach spaces. A solution $R : Y_2 \to Y_1$ of (14.26) is said to be *stable* if for each $\varepsilon > 0$ there exists $\delta > 0$ such that the following is true: if $S_{ij} \in \mathcal{L}(Y_j, Y_i),\ i, j = 1, 2$, and $\max_{i,j=1,2} \|S_{ij} - T_{ij}\| < \delta$, then the Riccati equation

$$XS_{21}X + XS_{22} - S_{11}X - S_{12} = 0$$

has a solution $Q : Y_2 \to Y_1$ for which $\|Q - R\| < \varepsilon$.

Theorem 14.22. *A solution R of the operator Riccati equation* (14.26) *is stable if and only if each common eigenvalue of $T_{11}-RT_{21}$ and $T_{22}+T_{21}R$ is an eigenvalue of geometric multiplicity one of the operator*

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} : Y_1 \dot{+} Y_2 \to Y_1 \dot{+} Y_2.$$

Proof. Let R be an operator from Y_2 into Y_1. Put $N = \{Rz + z \mid z \in Y_2\}$. Then $Y_1 \dot{+} N = Y_1 \dot{+} Y_2$ and R is the angular operator for N with respect to the projection of $Y_1 \dot{+} Y_2$ along Y_1 onto Y_2. By Proposition 5.4, the hypothesis that R is a solution of (14.26) is equivalent to the assumption that N is an invariant subspace for T. It is not difficult to prove that R is a stable solution of (14.26) if and only if N is a stable invariant subspace for T. The latter is the case if and only if Y_2 is a stable invariant subspace for the operator given by the right-hand side of (5.7). The desired result is now an immediate consequence of Theorem 14.7. □

A solution R of (14.26) is called *Lipschitz stable* if there are positive constants K and δ such that $\max_{i,j=1,2} \|S_{ij} - T_{ij}\| < \delta$ implies that the Riccati equation

$$XS_{21}X + XS_{22} - S_{11}X - S_{12} = 0$$

has a solution Q with $\|Q-R\| \leq K(\max_{i,j=1,2} \|S_{ij}-T_{ij}\|)$. The proof of Theorem 14.22 shows that R is Lipschitz stable if and only if the subspace $N = \{Rz + z \mid z \in Y_2\}$ is a Lipschitz stable invariant subspace for T. This observation is the main ingredient of the proof of the following theorem. Compare also Proposition 5.10.

Theorem 14.23. *A solution R of the Riccati equation* (14.26) *is Lipschitz stable if and only if $T_{11} - RT_{21}$ and $T_{22} + T_{21}R$ have no eigenvalues in common.*

14.7 Stability of stable factorizations

Let X be a finite-dimensional Banach space, and let T be a bounded linear operator on X. If N is a stable invariant subspace for T then, by definition for each $\varepsilon > 0$ there exists $\delta > 0$ such that $\|S-T\| < \delta$ implies that S has an invariant subspace M with $\text{gap}\,(M, N) < \varepsilon$. On the basis of Theorem 14.7 one can prove that for an appropriate choice of δ the space M may always be chosen to be stable for S. This is the contents of the next theorem.

Theorem 14.24. *Let N be a stable invariant subspace for a linear operator T acting on a finite-dimensional space X. Then, given $\varepsilon > 0$, there exists $\delta > 0$ such that $\|S - T\| < \delta$ implies that S has a stable invariant subspace M such that* $\text{gap}\,(M, N) < \varepsilon$.

Proof. Suppose not. Then there exist $\varepsilon > 0$ and a sequence $S_1, S_2, \ldots$ of linear operators on X converging to T such that for $k = 1, 2, \ldots$

$$\text{gap}\,(M, N) \geq \varepsilon, \qquad M \in \Omega_k.$$

Here Ω_k denotes the collection of all stable invariant subspaces for S_k. Since N is stable for T and $S_k \to T$ there exists a sequence $N_1, N_2, \ldots$ of subspaces of X with $S_k[N_k] \subset N_k$ and $\text{gap}\,(N_k, N) \to 0$. For k sufficiently large, we have $\text{gap}\,(N_k, N) < \varepsilon$, and hence $N_k \notin \Omega_k$. So, passing if necessary to a subsequence, we may assume that for all k the S_k-invariant subspace N_k is not stable.

Let Z be an algebraic complement of N in X. Since N_k converges to N in the gap topology, we may assume that $Z \dot{+} N_k = Z \dot{+} N = X$ for all k. Let R_k be the angular operator of N_k with respect to the projection of X onto N along Z. Then $R_k \to 0$. Write

$$E_k = \begin{bmatrix} I & R_k \\ 0 & I \end{bmatrix},$$

where the matrix representation is taken relative to the decomposition $X = Z \dot{+} N$. Then E_k is invertible, $E_k[N] = N_k$ and $E_k \to I$. Put $T_k = E_k^{-1} S_k E_k$. Obviously, $T_k \to T$ and $T_k[N] \subset N$. Note that N is not stable for T_k.

With respect to the decomposition $X = N \dot{+} Z$, we write

$$T = \begin{bmatrix} U & V \\ 0 & W \end{bmatrix}, \qquad T_k = \begin{bmatrix} U_k & V_k \\ 0 & W_k \end{bmatrix}.$$

Then $U_k \to U$ and $W_k \to W$. Since N is not stable for T_k, Theorem 14.7 ensures the existence of a common eigenvalue λ_k of U_k and W_k such that

$$\dim \text{Ker}\,(\lambda_k I - T_k) \geq 2, \qquad k = 1, 2, \ldots . \tag{14.27}$$

Now $|\lambda_k| \leq \|U_k\|$ and $U_k \to U$. Hence, the sequence $\lambda_1, \lambda_2, \ldots$ is bounded. Passing, if necessary, to a subsequence, we may assume that $\lambda_k \to \lambda_0$ for some $\lambda_0 \in \mathbb{C}$. But then

$$\lambda_k I - U_k \to \lambda_0 I - U, \qquad \lambda_k I - W_k \to \lambda_0 I - W, \qquad (k \to \infty).$$

It follows that λ_0 is a common eigenvalue of U and W. Again applying Theorem 14.7, we see that λ_0 is an eigenvalue of T of geometric multiplicity one. But this cannot be true in view of (14.27) and the fact that for $k \to \infty$ we have $\lambda_k I - T_k \to \lambda_0 I - T$. This can be proved by using a standard rank argument. □

With the help of Theorem 14.24 one can sharpen Theorem 14.9 as follows.

Theorem 14.25. *Suppose $W_0 = W_{01} W_{02}$ is a stable minimal factorization involving proper rational $m \times m$ matrix functions that have the value I_m at infinity. Let*

$$W_0(\lambda) = I_m + C_0(\lambda I_n - A_0)^{-1} B_0,$$

$$W_{0j}(\lambda) = I_m + C_{0j}(\lambda I_{n_j} - A_{0j})^{-1} B_{0j}, \qquad j = 1, 2,$$

be minimal realizations for W_0, W_{01} and W_{02}. Then for each $\varepsilon > 0$ there exists $\omega > 0$ with the following property. If

$$\|A - A_0\| + \|B - B_0\| + \|C - C_0\| < \omega,$$

then $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ is a minimal realization and W admits a stable minimal factorization $W = W_1 W_2$,

$$W_j(\lambda) = I_m + C_j(\lambda I_{n_j} - A_j)^{-1}B_j, \qquad j = 1, 2,$$

with the extra property that $\|A_j - A_{0j}\| + \|B_j - B_{0j}\| + \|C_j - C_{0j}\| < \varepsilon$.

Note that each common pole (zero) of W_1 and W_2 is a pole (zero) of W of geometric multiplicity one. So Theorem 14.25 extends Theorem 13.1. Similar refinements can be formulated for Theorems 14.20 and 14.22. For the exact formulation, see [13], Theorems 4.2 and 4.3. The arguments are again based on Theorem 14.24.

Theorem 14.25 has also a counterpart for complete factorizations.

Theorem 14.26. *Let W_0 be a rational $m \times m$ matrix function given by the minimal realization $W_0(\lambda) = I_m + C_0(\lambda I_n - A_0)^{-1}B_0$, and suppose*

$$W_0(\lambda) = \left(I_m + \frac{1}{\lambda - \alpha_{01}}R_{01}\right) \cdots \left(I_m + \frac{1}{\lambda - \alpha_{0n}}R_{0n}\right) \tag{14.28}$$

is a stable complete factorization of W_0. Then for each $\varepsilon > 0$ there exists $\omega > 0$ with the following property. If

$$\|A - A_0\| + \|B - B_0\| + \|C - C_0\| < \omega, \tag{14.29}$$

then $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ is a minimal realization and W admits a stable complete factorization

$$W(\lambda) = \left(I_m + \frac{1}{\lambda - \alpha_1}R_1\right) \cdots \left(I_m + \frac{1}{\lambda - \alpha_n}R_n\right) \tag{14.30}$$

with the extra property that

$$|\alpha_j - \alpha_{0j}| + \|R_j - R_{0j}\| < \varepsilon, \qquad j = 1, \ldots, n. \tag{14.31}$$

Proof. Let ε be a positive number. As (14.28) is a stable complete factorization of W_0, there exists $\omega > 0$ such that (14.29) implies that $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ is a minimal realization and W admits a complete factorization (14.30) which satisfies (14.31). According to Theorem 14.14, both the matrices A_0 and $A_0^\times = A_0 - B_0C_0$ are nonderogatory, i.e., all their eigenspaces are one-dimensional. As is easily seen via a simple rank argument, this is a property that is retained under small perturbations. Thus if ω is taken sufficiently small, then (14.29) implies that A and $A^\times = A - BC$ are nonderogatory too. But then the complete factorization (14.30) is stable by Theorem 14.14. □

The proof of Theorem 14.26 can also be based on the analogue of Theorem 14.24 for stable complete chains of invariant subspaces. This result is interesting in its own right. Employing matrix terminology (just as in Section 14.4), it reads as follows.

Theorem 14.27. *Suppose the $n \times n$ matrix A has a stable complete chain of invariant subspaces $\mathcal{M}$. Then, given $\varepsilon > 0$, there exists $\delta > 0$ with the following property. If B is an $n \times n$ matrix with $\|B - A\| < \delta$, then B has a stable stable complete chain of invariant subspaces $\mathcal{M}$ for which $\mathrm{GAP}(\mathcal{M}, \mathcal{N}) < \varepsilon$, where $\mathrm{GAP}(\mathcal{M}, \mathcal{N})$ is as defined in* (14.20).

Proof. Let ε be a positive number. As $\mathcal{M}$ is a stable complete chain of invariant subspaces for A, there exists $\delta > 0$ such that $\|B - A\| < \delta$ implies that B has a complete chain $\mathcal{N}$ of invariant subspaces such that $\mathrm{GAP}(\mathcal{M}, \mathcal{N}) < \varepsilon$. According to Theorem 14.16, the matrix A is nonderogatory, i.e., all its eigenspaces are one-dimensional. As mentioned earlier, this property is retained under small perturbations. Thus if δ is taken sufficiently small, $\|B - A\| < \delta$ implies that B is nonderogatory too. But then the complete chain $\mathcal{N}$ referred to above is stable for B by Theorem 14.16. □

14.8 Isolated factorizations and related topics

In the previous sections of this chapter we studied invariant subspaces, factorizations and solutions of the Riccati equation from the point of view of stability. In the present section, we deal with another property, namely that of being isolated which, apart from bearing a certain resemblance to stability, turns out to be equivalent to stability. For reasons of systematic presentation, the material has been divided in four subsections.

14.8.1 Isolated invariant subspaces

Let T be a bounded linear operator on a complex Banach space X. A closed invariant subspace N of T is called *isolated* if there exists $\varepsilon > 0$ such that the following holds. If M is a closed invariant subspace of T and $\mathrm{gap}(M, N) < \varepsilon$, then $M = N$. In the same way as for stability, the property of being an isolated invariant subspace is similarity invariant in the following sense. Let E be an invertible operator on X, and introduce $\widetilde{T} = E^{-1}TE$, $\widetilde{N} = E^{-1}[N]$. Then $\widetilde{N}$ is an isolated invariant subspace for $\widetilde{T}$ if (and only if) N is an isolated invariant subspace for T. The argument, which involves the condition number $\|E^{-1}\| \cdot \|E\|$ of E, is straightforward.

In the remainder of this subsection we will restrict ourselves to the finite-dimensional case. Whenever convenient, matrices will be considered as operators. Recall that the generalized eigenspace $\mathrm{Ker}\,(\lambda_0 - A)^k$ of a $k \times k$ matrix A corresponding to the eigenvalue λ_0 is denoted by $N(\lambda_0)$.

Theorem 14.28. *Let $\lambda_1, \ldots, \lambda_r$ be the different eigenvalues of the $k \times k$ matrix A. A subspace N of $\mathbb{C}^k$ is an isolated A-invariant subspace if and only if $N = N_1 \dot{+} \cdots \dot{+} N_r$, where for each j the space N_j is an arbitrary A-invariant subspace of $N(\lambda_j)$ whenever $\dim \operatorname{Ker}(\lambda_j - A) = 1$, while otherwise $N_j = \{0\}$ or $N_j = N(\lambda_j)$.*

Proof. First we deal with the case when A has only one eigenvalue, without loss of generality taken to be zero. So $r = 1$, $\lambda_1 = 0$ and $N(\lambda_1) = \mathbb{C}^k$. There are two different situations that have to be dealt with: $\dim \operatorname{Ker} A = 1$ (nonderogatory A) and $\dim \operatorname{Ker} A \geq 2$ (derogatory A). The first is trivial because there are only a finite number of A-invariant subspaces then. So assume $\dim \operatorname{Ker} A \geq 2$. Obviously the trivial subspaces $\{0\}$ and $\mathbb{C}^k$ are isolated invariant subspaces for A. What we need to show is that the non-trivial A-invariant subspaces are not isolated.

Let N be a non-trivial A-invariant subspace. In the case when $N = \operatorname{Ker} A^p$ for some positive integer p we argue as follows. Let

$$\{x_{jk}\}_{j=1,\ k=0}^{q\ \ j} \tag{14.32}$$

be a basis of $\mathbb{C}^k$ such that the corresponding matrix representation of A has Jordan form. In other words, for $j = 1, \ldots, q$, we have

$$Ax_{j0} = 0, \qquad Ax_{jk} = x_{j,k-1}, \qquad k = 1, \ldots, r_j. \tag{14.33}$$

For convenience we assume that $r_1 \geq r_2 \geq \cdots \geq r_q$. Observe that $\operatorname{Ker} A^p$ is the span of

$$\{x_{jk}\}_{j=1,\ k=0}^{q\ \ r_j \wedge (p-1)}. \tag{14.34}$$

Here $r_j \wedge (p-1)$ is the minimum of r_j and $p-1$. We claim that $r_1 \geq p$. Indeed, for if not $N = \operatorname{Ker} A^p$ would be all of $\mathbb{C}^k$. For $\varepsilon \neq 0$, let N_ε be the span of

$$\{x_{jk}\}_{j=1,\ k=0}^{q-1\ \ r_j \wedge (p-1)} \cup \{x_{qk}\}_{k=0}^{[r_q \wedge (p-1)]-1} \cup \{x_{q,r_q \wedge (p-1)} + \varepsilon x_{1p}\},$$

where the middle term in the union is absent when $r_q \wedge (p-1) = 0$. Since $q = \dim \operatorname{Ker} A \geq 2$, we have that N_ε is an invariant subspace of T. Moreover, $\operatorname{gap}(N_\varepsilon, N) \to 0$ when $\varepsilon \to 0$. As all N_ε are different from N, it follows that N is not isolated.

Next assume that N is not of the form $\operatorname{Ker} A^m$. Since $\operatorname{Ker} A^m = \mathbb{C}^k$ for m sufficiently large and $N \neq \mathbb{C}^k$, there exists a unique non-negative integer p such that

$$\operatorname{Ker} A^p \subset N, \qquad \operatorname{Ker} A^{p+1} \not\subset N.$$

Consider the restriction A_0 of A to N. The spectrum of A_0 consists of zero only. Let (14.32) now denote a basis of N such that the corresponding matrix representation for A_0 has Jordan form. This means that (14.32) is a basis of N for which (14.33) holds. Again we assume that $r_1 \geq r_2 \geq \cdots \geq r_q$. Now $\operatorname{Ker} A^p = \operatorname{Ker} A_0^p$ is the span of (14.34). Since $N \neq \operatorname{Ker} A^p$, it follows that $r_1 \geq p$. Choose $u \in \operatorname{Ker} A^{p+1} \setminus N$, and put

$$u_k = A^{p-k}u, \qquad k = 0, \ldots, p.$$

Then clearly

$$Au_0 = 0, \qquad Au_k = u_{k-1}, \qquad k = 1, \dots, p.$$

Moreover, $u_p = u \notin N$. For $\varepsilon \neq 0$, we now define N_ε to be the span of

$$\{x_{jk}\}_{j=2,k=0}^{q \quad r_j} \cup \{x_{1k}\}_{k=0}^{r_1-p-1} \cup \{x_{1,r_1-p+k} + \varepsilon u_k\}_{k=0}^{p},$$

where the middle term in the union is absent when $p = r_1$. Then N_ε is well defined for $r_1 \geq p$. Observe that N_ε is A-invariant and $\operatorname{gap}(N_\varepsilon, N) \to 0$ when $\varepsilon \to 0$. Since all N_ε are different from N, it follows that N is not isolated, as desired.

We now drop the restriction that A has only one eigenvalue, so r is allowed to be larger than one. Write $X_1, \dots, X_r$ for the generalized eigenspaces corresponding to the different eigenvalues $\lambda_1, \dots, \lambda_r$ of A. Thus $X_j = N(\lambda_j)$ for $j = 1, \dots, r$. Then $\mathbb{C}^k = X_1 \dot{+} \cdots \dot{+} X_r$, and relative to this decomposition A has the (diagonal) form

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & A_r \end{bmatrix} : X_1 \dot{+} \cdots \dot{+} X_s \to X_1 \dot{+} \cdots \dot{+} X_s,$$

with $\sigma(A_j) = \{\lambda_j\}$ for all j.

Let N be an isolated invariant subspace for A. For $j = 1, \dots, r$, put $N_j = N \cap X_j$. Then N_j is an A-invariant subspace of X_j and $N = N_1 \dot{+} \cdots \dot{+} N_r$. We need to prove that N_j is either $\{0\}$ or X_j whenever $\dim \operatorname{Ker}(\lambda_j - A) \geq 2$. Suppose the latter is the case and $\{0\} \neq N_j \neq X_j$. Then N_j is a non-trivial A_j-invariant subspace where A_j is derogatory and has only one eigenvalue. Hence N_j is not isolated for A_j. But this immediately implies that, contrary to our assumption, N cannot be isolated for A.

Next, assume $N = N_1 \dot{+} \cdots \dot{+} N_r$, where for each j the space N_j is an arbitrary A-invariant subspace of X_j whenever $\dim \operatorname{Ker}(\lambda_j - A) = 1$, while otherwise $N_j = \{0\}$ or $N_j = X_j$. Then, by the preliminary observations made above about the single eigenvalue case, $N_1, \dots, N_r$ are isolated invariant subspaces for $A_1, \dots, A_r$, respectively. Suppose now that N is not isolated for A. This means that there is a sequence of A-invariant subspaces $N^{(1)}, N^{(2)}, \dots$, all different from N, such that $\operatorname{gap}(N^{(n)}, N) \to 0$ when $n \to \infty$. Put $N_j^{(n)} = N^{(n)} \cap X_j$. Then $N_j^{(n)}$ is an invariant subspace for A_j. Also $N^{(n)} = N_1^{(n)} \dot{+} \cdots \dot{+} N_r^{(n)}$ and

$$\lim_{n \to \infty} \operatorname{gap}\left(N_j^{(n)}, N_j\right) = 0, \qquad j = 1, \dots, r.$$

As N_j is isolated for A_j, we may conclude that $N_j^{(n)} = N_j$ for n sufficiently large, depending on j. Now j takes only a finite number of values. It follows that there exists n_0 such that

$$N_j^{(n)} = N_j, \qquad j = 1, \dots, r, \ n = n_0, n_0 + 1, \dots .$$

But then $N^{(n)} = N$ for n larger than or equal to n_0 . This yields a contradiction, and we conclude that N is an isolated A-invariant subspace. □

Comparing Theorem 14.28 to the characterization of stability given in Theorem 14.1, one obtains the conclusion that an invariant subspace is isolated if and only if it is stable. This result is typical for the case of spaces over $\mathbb{C}$. In fact, as we shall see in the next chapter (Section 15.3), when the underlying scalar field is $\mathbb{R}$ instead of $\mathbb{C}$, then stable invariant subspaces are isolated, but the converse is no longer true.

Returning to the situation where the underlying field is $\mathbb{C}$, we note that Theorem 14.7 remains true when the word stable is replaced by isolated. The argument is the same as that for Theorem 14.7 where, of course, one has to read stable instead of isolated and with the reference to Theorem 14.1 replaced by one to Theorem 14.28.

We close this subsection with one additional remark. From Theorem 14.28 one immediately has the following two observations. If $\sigma(T)$ consists of exactly one eigenvalue of geometric multiplicity one, then each invariant subspace of T is isolated; if $\sigma(T)$ consists of exactly one eigenvalue of geometric multiplicity at least two, then no non-trivial invariant subspace of T is isolated. This can be used to give a quick elementary proof of [40], Theorem 9.

14.8.2 Isolated chains of invariant subspaces

Next we turn to chains of subspaces. A complete chain $\mathcal{N}$ of invariant subspaces for the $n \times n$ matrix A is said to be *isolated* (for A) if there exists $\varepsilon > 0$ with the following property. If $\mathcal{M}$ is a complete chain of A-invariant subspaces and $\mathrm{GAP}\,(\mathcal{M},\mathcal{N}) < \varepsilon$, then $\mathcal{M} = \mathcal{N}$. In analogy to what was observed above about isolated subspaces, a complete chain of A-invariant subspaces is isolated if and only if it is stable. This is immediate from the following result combined with Theorem 14.16.

Theorem 14.29. *The matrix A has an isolated complete chain of A-invariant subspaces if and only if A is nonderogatory. In that case, A has a finite number of complete chains of invariant subspaces and all these chains are isolated.*

Proof. Suppose A is nonderogatory. Then, as we already have seen in the last paragraph of the proof of Theorem 14.16, the (nonempty) collection of all complete chains of A-invariant subspaces is finite, and it is obvious that each complete chain of A-invariant subspaces is isolated.

In the remainder of this proof it is assumed that A is derogatory. The aim is to show that there are no isolated complete chains of A-invariant subspaces. First we deal with the case when A has only one eigenvalue, without loss of generality taken to be zero.

Let $\mathcal{M} = \{M_l\}_{l=0}^n$ be a complete chain of A-invariant subspaces, and choose a basis $u_1, \ldots, u_n$ of $\mathbb{C}^n$ such that

$$M_l = \operatorname{span}\{u_1, \ldots, u_l\}, \qquad l = 0, \ldots, n,$$

where following standard practice $\operatorname{span}\{\emptyset\} = \{0\}$. Let $U = [u_1\, u_2 \cdots u_n]$ be the $n \times n$ matrix whose lth column is u_l. Then U is invertible and $U^{-1}AU$ is upper triangular. Clearly $U^{-1}AU$ has the eigenvalues of A on the diagonal. Now (as assumed for the time being) A has zero as its only eigenvalue. Thus the diagonal of $U^{-1}AU$ features only zeros. But then $U^{-1}AU$ is strictly upper triangular and

$$Au_l \in \operatorname{span}\{u_1, \ldots, u_{l-1}\}, \qquad l = 1, \ldots, n. \tag{14.35}$$

Clearly, $U^{-1}AU$ has the zero vector in $\mathbb{C}^n$ as its first column. If the other columns in $U^{-1}AU$ were linearly independent, the rank of $U^{-1}AU$ would be $n-1$ contradicting the fact that A is derogatory. Indeed, the latter means that $\dim \operatorname{Ker} A$ is a least 2 so the (coinciding) ranks of A and $U^{-1}AU$ are at most $n-2$. Choose p among the integers $1, \ldots, n-1$ such that the $(p+1)$th column in $U^{-1}AU$ is a linear combination of the columns of $U^{-1}AU$ in the positions 1 up to (and including) p. Note that one can take $p = 1$ if and only if $U^{-1}AU$ has the zero vector in $\mathbb{C}^n$ not only as its first, but also as its second column. For the specific value $l = p+1$, the expression (14.35) can now be sharpened into

$$Au_{p+1} \in \operatorname{span}\{u_1, \ldots, u_{p-1}\}. \tag{14.36}$$

For $k = 1, 2, \ldots$, put $v_l = u_l$ for $l = 1, \ldots, n,\ l \neq p$ and (with slight abuse of notation because the dependance on k is suppressed)

$$v_p = u_p + \frac{1}{k} u_{p+1}. \tag{14.37}$$

Then, for $l = 1, \ldots, n,\ l \neq p, p+1$,

$$Av_l = Au_l \in \operatorname{span}\{u_1, \ldots, u_{l-1}\} = \operatorname{span}\{v_1, \ldots, v_{l-1}\}.$$

This is evident from (14.35) and the definition of $v_1, \ldots, v_n$. Further,

$$Av_p = Au_p + \frac{1}{k} Au_{p+1} \in \operatorname{span}\{u_1, \ldots, u_{p-1}\} = \operatorname{span}\{v_1, \ldots, v_{p-1}\}.$$

Here we used not only (14.35) but also (14.36). Finally, based on (14.36),

$$Av_{p+1} = Au_{p+1} \in \operatorname{span}\{u_1, \ldots, u_{p-1}\} = \operatorname{span}\{v_1, \ldots, v_{p-1}\}.$$

We conclude that that the subspaces $\operatorname{span}\{v_1, \ldots, v_l\}$ form a complete chain of A-invariant subspaces. We shall denote this chain, which via (14.37) depends on k, by $\mathcal{N}^{(k)} = \{N_l^{(k)}\}_{l=0}^n$. For $l = 0, \ldots, n,\ l \neq p$, we have

$$N_l^{(k)} = \operatorname{span}\{v_1, \ldots, v_l\} = \operatorname{span}\{u_1, \ldots, u_l\} = M_l,$$

and so $\mathrm{GAP}\,(\mathcal{N}^{(k)},\mathcal{M}) = \mathrm{gap}\,(N_p^{(k)},M_p)$. Now

$$\begin{aligned} N_p^{(k)} &= \mathrm{span}\,\{u_1,\dots,u_{p-1},u_p+\frac{1}{n}u_{p+1}\},\\ M_p &= \mathrm{span}\,\{u_1,\dots,u_{p-1},u_p\}, \end{aligned}$$

and it follows that $\lim_{k\to\infty}\mathrm{GAP}\,(\mathcal{N}^{(n)},\mathcal{M}) = \lim_{k\to\infty}\mathrm{gap}\,(N_p^{(k)},M_p) = 0$. For all k, the subspaces $N_p^{(k)}$ and M_p are different, hence the chains $\mathcal{N}^{(k)}$ and $\mathcal{M}$ are different too. We conclude that the complete chain $\mathcal{M}$ of A-invariant subspaces is not isolated.

We now drop the restriction that A has only one eigenvalue. Write $\alpha_1,\dots,\alpha_s$ for the different eigenvalues of A and $P_1,\dots,P_s$ for the corresponding spectral projections. For $j=1,\dots,s$, put $X^{(j)} = \mathrm{Im}\,P_j$. Then $X^{(1)},\dots,X^{(s)}$ are the generalized eigenspaces corresponding to the different eigenvalues $\alpha_1,\dots,\alpha_s$ of A. Hence $\mathbb{C}^k = X^{(1)}\dot{+}\cdots\dot{+}X^{(s)}$ and with respect to this decomposition A has the (diagonal) form

$$A = \begin{bmatrix} A^{(1)} & 0 & \cdots & 0\\ 0 & \ddots & & \vdots\\ \vdots & & \ddots & 0\\ 0 & \cdots & 0 & A^{(s)} \end{bmatrix} : X^{(1)}\dot{+}\cdots\dot{+}X^{(s)} \to X^{(1)}\dot{+}\cdots\dot{+}X^{(s)},$$

with $\sigma\left(A^{(j)}\right) = \{\alpha_j\}$ for all j and at least one of the diagonal entries $A^{(1)},\dots,A^{(s)}$ derogatory, say $A^{(1)}$ (without loss of generality).

Let $\mathcal{M} = \{M_l\}_{l=0}^n$ be a complete chain of A-invariant subspaces. Taking the intersection of the subspaces M_l with the generalized eigenspace X_j we obtain a complete chain $\mathcal{M}^{(j)} = \{M_l^{(j)}\}_{l=0}^{n_j}$ of $A^{(j)}$-invariant subspaces. Here $\dim X_j = n_j$. The subspaces constituting the chain $\mathcal{M}$ can now be written in the form

$$M_l = M_{\nu_l(1)}^{(1)}\dot{+}M_{\nu_l(2)}^{(2)}\dot{+}\cdots\dot{+}M_{\nu_l(s)}^{(s)}, \qquad l=0,\dots,n, \tag{14.38}$$

with $\nu_l(j)$ among $0,\dots,n_j$, and this representation is unique. The nonnegative integers $\nu_l(1),\dots,\nu_l(s)$, being the dimensions of $M_{\nu_l(1)}^{(1)},\dots,M_{\nu_l(s)}^{(1)}$, add up to the dimension l of M_l. Also $\nu_l(j) = \nu_{l-1}(j)$ for all $j=0,\dots,s$ except one, written κ_l, for which $\nu_{(l)}(\kappa_l) = \nu_{l-1}(\kappa_l)+1$. Here $l=1,\dots,n$.

As $A^{(1)}$ is derogatory and has only one eigenvalue, the complete chain of $A^{(1)}$-invariant subspaces $\mathcal{M}^{(1)}$ is not isolated. This means that there exist complete chains of $A^{(1)}$-invariant subspaces $\mathcal{N}_1^{(1)},\mathcal{N}_2^{(1)},\dots$, all different from $\mathcal{M}^{(1)}$, such that $\lim_{k\to\infty}\mathrm{GAP}\,(\mathcal{N}_n^{(1)},\mathcal{M}^{(1)}) = 0$. Write $\mathcal{N}_k^{(1)} = \{N_{kl}^{(1)}\}_{l=0}^{n_1}$ and introduce

$$N_{kl} = N_{k,\nu_l(1)}^{(1)}\dot{+}N_{k,\nu_l(2)}^{(2)}\dot{+}\cdots\dot{+}N_{k,\nu_l(s)}^{(s)}, \qquad l=0,\dots,n,$$

with $\nu_l(1), \ldots, \nu_l(s)$ as in (14.38). Then $\mathcal{N}_k = \{N_{kl}\}_{l=0}^n$ is a complete chain of A-invariant subspaces and

$$\lim_{k\to\infty} \mathrm{GAP}\,(\mathcal{N}_k, \mathcal{M}) = 0.$$

Now $\mathcal{N}_1, \mathcal{N}_2, \ldots$ are all different from $\mathcal{M}$, and we may conclude that $\mathcal{M}$ is not isolated. □

14.8.3 Isolated factorizations

Next we consider factorizations. Let W_0 be a proper rational $m\times m$ matrix function with $W_0(\infty) = I_m$, and let $W_0 = W_{01}W_{02}$ be a minimal factorization. Furthermore, let

$$W_{0j}(\lambda) = I_m + C_{0j}(\lambda I_{n_j} - A_{0j})^{-1}B_{0j}, \qquad j = 1, 2, \tag{14.39}$$

be minimal realizations of W_{01} and W_{02}. The factorization $W_0 = W_{01}W_{02}$ is called *isolated* if the following condition (IF) is fulfilled:

(IF) There exists $\varepsilon > 0$ such that if $W_0 = W_1W_2$, while W_1 and W_2 admit minimal realizations

$$W_j(\lambda) = I_m + C_j(\lambda I_{n_j} - A_j)^{-1}B_j, \qquad j = 1, 2,$$

satisfying

$$\|A_j - A_{0j}\| + \|B_j - B_{0j}\| + \|C_j - C_{0j}\| < \varepsilon, \qquad j = 1, 2,$$

then $W_1 = W_{01}$ and $W_2 = W_{02}$.

As for stable minimal factorization, the definition of isolated minimal factorization does not depend on the particular choice of the minimal realization in (14.39). Indeed, consider another pair of minimal realizations

$$W_{0j}(\lambda) = I_m + \widetilde{C}_{0j}(\lambda I_{n_j} - \widetilde{A}_{0j})^{-1}\widetilde{B}_{0j}, \qquad j = 1, 2, \tag{14.40}$$

and assume condition (IF) is fulfilled. We have to show that this condition is also fulfilled when the minimal realizations in (14.39) are replaced by the minimal realizations in (14.40). As a first step, note that by the state space isomorphism theorem there exist invertible matrices S_1 and S_2 (of appropriate size) such that

$$A_{0j} = S_j\widetilde{A}_{0j}S_j^{-1}, \qquad B_{0j} = S_j\widetilde{B}_{0j}, \qquad C_{0j} = \widetilde{C}_{0j}S_j^{-1}, \qquad j = 1, 2.$$

Put $\widetilde{\varepsilon} = \varepsilon/\alpha$, where

$$\alpha = \max\big\{\|S_1\|\cdot\|S_1^{-1}\|,\ \|S_1\|,\ \|S_1^{-1}\|,\ \|S_2\|\cdot\|S_2^{-1}\|,\ \|S_2\|,\ \|S_2^{-1}\|\big\}.$$

We claim that with the minimal realizations in (14.39) being replaced by the minimal realizations in (14.40), condition (IF) is fulfilled with $\tilde{\varepsilon}$ in place of ε.

To see this, let $W_0 = \widetilde{W}_1\widetilde{W}_2$, where $\widetilde{W}_1$ and $\widetilde{W}_2$ are given by the minimal realizations

$$\widetilde{W}_j(\lambda) = I_m + \widetilde{C}_j(\lambda I_{n_j} - \widetilde{A}_j)^{-1}\widetilde{B}_j, \qquad j = 1, 2,$$

and assume that

$$\|\widetilde{A}_j - \widetilde{A}_{0j}\| + \|\widetilde{B}_j - \widetilde{B}_{0j}\| + \|\widetilde{C}_j - \widetilde{C}_{0j}\| < \widetilde{\varepsilon}, \qquad j = 1, 2.$$

Using the invertible matrices S_1 and S_2 we introduce the matrices:

$$A_j = S_j\widetilde{A}_jS_j^{-1}, \qquad B_j = S_j\widetilde{B}_j, \qquad C_j = \widetilde{C}_jS_j^{-1}, \qquad j = 1, 2.$$

Then $\widetilde{W}_j(\lambda) = I_m + C_j(\lambda I_{n_j} - A_j)^{-1}B_j$, $j = 1, 2$, and these realizations are minimal. Furthermore, we have

$$\begin{aligned}
&\|A_j - A_{0j}\| + \|B_j - B_{0j}\| + \|C_j - C_{0j}\| \\
&= \|S_j\widetilde{A}_jS_j^{-1} - S_j\widetilde{A}_{0j}S_j^{-1}\| + \|S_j\widetilde{B}_j - S_j\widetilde{B}_{0j}\| + \|\widetilde{C}_jS_j^{-1} - \widetilde{C}_{0j}S_j^{-1}\| \\
&\le \|S_j\| \cdot \|S_j^{-1}\| \cdot \|\widetilde{A}_j - \widetilde{A}_{0j}\| + \|S_j\| \cdot \|\widetilde{B}_j - \widetilde{B}_{0j}\| + \|S_j^{-1}\| \cdot \|\widetilde{C}_j - \widetilde{C}_{0j}\| \\
&\le \alpha\big(\|\widetilde{A}_j - \widetilde{A}_{0j}\| + \|\widetilde{B}_j - \widetilde{B}_{0j}\| + \|\widetilde{C}_j - \widetilde{C}_{0j}\|\big) \\
&< \alpha\,\widetilde{\varepsilon} = \varepsilon.
\end{aligned}$$

Since (IF) is fulfilled, we conclude that $\widetilde{W}_1 = W_{01}$ and $\widetilde{W}_1 = W_{01}$, as desired.

Theorem 14.9 remains true with the word stable replaced by isolated.

Theorem 14.30. *Suppose $W_0 = W_{01}W_{02}$ is a minimal factorization. This factorization is isolated if and only if each common pole (zero) of W_{01} and W_{02} is a pole (zero) of W_0 of geometric multiplicity one.*

Thus a minimal factorization of a rational matrix function is isolated if and only if it is stable. This is different when the underlying scalar field is $\mathbb{R}$ instead of $\mathbb{C}$. In that situation stable minimal factorizations are isolated, but the converse is no longer true; see the next chapter (Section 15.4). The proof of Theorem 14.30 is along similar lines as that of Theorem 14.9 but uses Theorem 14.7 with the word stable replaced by isolated (cf., the last paragraph of Subsection 14.8.1).

The next topic to be treated in this subsection is isolated complete factorizations. Let W_0 be a rational $m \times m$ matrix function with $W_0(\infty) = I_m$, and let

$$W_0(\lambda) = \left(I_m + \frac{1}{\lambda - \alpha_{01}}R_{01}\right) \cdots \left(I_m + \frac{1}{\lambda - \alpha_{0n}}R_{0n}\right) \tag{14.41}$$

be a complete factorization of W_0. We say that this complete factorization is *isolated* if there exists $\varepsilon > 0$ such that the following holds. If

$$W(\lambda) = \left(I_m + \frac{1}{\lambda - \alpha_1}R_1\right) \cdots \left(I_m + \frac{1}{\lambda - \alpha_n}R_n\right) \tag{14.42}$$

is a complete factorization of W_0 and

$$|\alpha_j - \alpha_{0j}| + \|R_j - R_{0j}\| < \varepsilon, \qquad j = 1, \ldots, n, \tag{14.43}$$

then (14.41) and (14.42) coincide, i.e., $a_j = \alpha_{0j}$ and $R_j = R_{0j}$, $j = 1, \ldots, n$. Again (see the second paragraph in Section 14.4) this definition is not completely analogous to the one given of an isolated minimal factorization involving just two factors. It can, however, also be given along these lines using representations of the form (14.18) and (14.19).

As might be expected by now, Theorems 14.13 and 14.14, remain true when the word stable is replaced by isolated, and this is true for Corollary 14.15 as well. The precise statements are covered by the following results.

Theorem 14.31. *Let W_0 be a rational $m \times m$ matrix function with $W_0(\infty) = I_m$. A necessary condition for W_0 to admit an isolated complete factorization is that the poles and zeros of W_0 all have geometric multiplicity one. In that case there are only finitely many complete factorizations of W_0 and these are all isolated.*

Theorem 14.32. *Let W_0 be a rational $m \times m$ matrix function with minimal realization*

$$W_0(\lambda) = I_m + C_0(\lambda I_n - A_0)^{-1}B_0.$$

A necessary condition for W_0 to admit an isolated complete factorization is that both A_0 and $A_0^\times = A_0 - B_0C_0$ are nonderogatory. In that case there are only finitely many complete factorizations of W_0 and these are all isolated.

Corollary 14.33. *Let W_0 be a companion based rational $m \times m$ matrix function. Then there are only finitely many complete factorizations of W_0 and these are all isolated.*

Note that these results do not guarantee that W_0 admits a complete factorization. The number of these factorizations might be zero. Observe also that for complete factorizations, the stable and the isolated ones coincide.

As in Section 14.5, let $L_0 = L_{02}L_{01}$ where L_0, L_{01} and L_{02} are monic $m \times m$ matrix polynomials of degree p, q and r, respectively, so $L_0 \in \mathcal{MP}_p$, $L_{01} \in \mathcal{MP}_q$ and $L_{02} \in \mathcal{MP}_r$. We say that the factorization $L_0 = L_{02}L_{01}$ is *isolated* if there exists $\varepsilon > 0$ with the following property. If $L_0 = L_2L_1$ with $L_1 \in \mathcal{MP}_q$ and $L_2 \in \mathcal{MP}_r$, and

$$\|L_j - L_{0j}\| < \varepsilon, \qquad j = 1, 2,$$

then $L_1 = L_{01}$ and $L_2 = L_{02}$. Theorem 14.20 remains true with the word stable replaced by isolated.

Theorem 14.34. *Let L_0, L_{01} and L_{02} be monic $m \times m$ matrix polynomials, and assume $L_0 = L_{02}L_{01}$. This factorization is isolated if and only if each common eigenvalue of L_{01} and L_{02} is an eigenvalue of L_0 of geometric multiplicity one.*

Thus a minimal factorization of monic matrix polynomials is stable if and only if it is isolated. In the case when the underlying scalar field is $\mathbb{R}$ instead of $\mathbb{C}$, things are different. Stable minimal factorizations are then always isolated, but the converse is no longer true. For details, we refer to the next chapter.

14.8.4 Isolated solutions of the Riccati equation

A solution R of a given operator Riccati equation is called *isolated* if there exists $\varepsilon > 0$ such that the following holds. If Q is also a solution of the Riccati equation in question and $\|Q - R\| < \varepsilon$, then $Q = R$. Theorem 14.22 remains true with the word stable replaced by isolated.

Theorem 14.35. *Consider the operator Riccati equation*

$$XT_{21}X + XT_{22} - T_{11}X - T_{12} = 0,$$

with $T_{ij} \in \mathcal{L}(Y_j, Y_i)$, $i, j = 1, 2$, where Y_1 and Y_2 are finite-dimensional Banach spaces. A solution R of this equation is isolated if and only if each common eigenvalue of $T_{11} - RT_{21}$ and $T_{22} + T_{21}R$ is an eigenvalue of geometric multiplicity one of the operator

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} : Y_1 \dot{+} Y_2 \to Y_1 \dot{+} Y_2.$$

Thus as far as solutions of the operator Riccati equation is concerned, the stable and the isolated ones are the same. In the case when the underlying scalar field is $\mathbb{R}$ instead of $\mathbb{C}$, the situation is different. Then stable solutions are isolated, but the converse is no longer true. For details, we refer once more to the next chapter.

Notes

This chapter covers the material from Chapter VIII in [14], with Sections 14.2, 14.4 and a large part of 14.8 as substantial novel additions. For further information about stability of invariant spaces we refer to [70]; see also Chapter S4 in [69]. The notion of Lipschitz stability (Section 14.2) was introduced in [83], where one also can find Theorem 14.8 and its proof. Theorem 14.21 is a particular case of Theorem 17.3.3 in [70]; Theorem 14.23 is Theorem 17.9.3 in [70]. The fact, mentioned in Section 14.8, that in the finite-dimensional case an invariant subspace is stable if and only if it is isolated has been proved in [34] too (see also [37]). The main theorem in [34] contains the characterization given in Theorem 14.1. For

related material, see [40]. Analytic properties of invariant subspaces depending on a complex parameter, also with applications to factorizations of rational matrix functions and to quadratic matrix equations can be found in Part Four of [70]. Stability of chains (more generally, of lattices) of invariant subspaces has been considered in [70], Section 15.6. The results in Section 14.4 on (Lipschitz) stability of complete factorizations are new. Also new are the results in Section 14.8 on isolated complete chains of invariant subspaces and isolated complete factorizations.

As a further development we mention the notion of α-stability of invariant subspaces which originated from [70], Exercise 16.7. An A-invariant subspace M is called α-stable if there exist positive constants δ and K such that $\|A - B\| < \delta$ implies the existence of a B-invariant subspace N with $\text{gap}(M, N) \leq K\|A - B\|^{\alpha}$. It follows that the notion of α-stability is weaker than Lipschitz stability and stronger than usual stability. A full characterization of α-stable invariant subspaces was first given in [98]. The related concept of strong α-stability was introduced and studied in [100]. Other related notions of stability and applications can be found in [99] and the references cited therein. The paper [99], which has survey character, also points the way to the literature on stability of invariant subspaces of matrices with symmetry properties in indefinite inner product spaces. The latter can be applied to study stability of symmetric factorizations for rational matrix functions $W(\lambda)$ that have selfadjoint values for real values of λ.

For the connections with computational aspects, we refer to [17], where among other things rough estimates are given for the number of computations involved in the construction of a minimal factorization of a transfer function.

Chapter 15

Factorization of Real Matrix Functions

In this chapter we review the factorization theory for the case of real matrix functions with respect to real divisors. As in the complex case the minimal factorizations are completely determined by the supporting projections of a given realization, but in this case one has the additional requirement that all linear transformations must be representable by matrices with real entries. Due to the difference between the real and complex Jordan canonical form the structure of the stable real minimal factorizations is somewhat more complicated than in the complex case. This phenomenon is also reflected by the fact that for real matrixes there is a difference between the stable and isolated invariant subspaces.

15.1 Real matrix functions

We begin with some notation and terminology. Let $x = (x_1, \ldots, x_n)^\top$ be a vector in $\mathbb{C}^n$. Then $\overline{x} = (\overline{x}_1, \ldots, \overline{x}_n)^\top$ is called the *conjugate* of x. We say that x is *real* if $x = \overline{x}$. So the real vectors in $\mathbb{C}^n$ are just the elements of $\mathbb{R}^n$.

Let M be a subspace of $\mathbb{C}^n$. Then, by definition, $\overline{M} = \{\overline{x} \mid x \in M\}$. Note that $\overline{M}$ is also a subspace of $\mathbb{C}^n$. We call M *self-conjugate* if $M = \overline{M}$. This notion will be used in Sections 15.2 and 15.3. It is easy to see that M is self-conjugate if and only if there exists a subspace N (uniquely determined by M) of the real vector space $\mathbb{R}^n$ such that $M = \{x + iy \mid x,\ y \in N\}$.

Suppose $A = [a_{jk}]_{j=1,k=1}^{n\ \ m}$ is a complex matrix. By the *conjugate* $\overline{A}$ of A, we mean the matrix

$$\overline{A} = [\overline{a}_{jk}]_{j=1,k=1}^{n\ \ m}.$$

The matrix A is called *real* if $A = \overline{A}$. In other words, A is real if and only if all its entries are real numbers. Now specify bases $e_1, \ldots, e_m$ of $\mathbb{C}^m$ and $f_1, \ldots, f_n$ of

$\mathbb{C}^n$ consisting of real vectors. Then the matrix A defines a linear operator from $\mathbb{C}^m$ into $\mathbb{C}^n$. Note that A is a real matrix if and only if this operator maps real vectors in $\mathbb{C}^m$ into real vectors in $\mathbb{C}^n$.

Let W be a rational $m \times m$ matrix function. We say that W is *real* if $W(\lambda)$ is a real matrix for all real λ in the domain of W (i.e., not being a pole of W). A realization

$$W(\lambda) = D + C(\lambda I_n - A)^{-1}B \tag{15.1}$$

is called a (*minimal*) *real realization* of W if (it is minimal in the sense of Section 7.1 and) A, B, C and D are real matrices. Clearly, if W admits a real realization, then W is a real matrix function. The converse of this is also true; in fact, one can always make a minimal real realization (cf., [116], Lemma 1).

Theorem 15.1. *Let W be a proper rational $m \times m$ matrix function. Assume W is real. Then W admits a minimal real realization.*

Proof. Let n be the McMillan degree of W. Then W admits a minimal realization of the form (15.1), where A, B, C and D are complex matrices of appropriate sizes. Define the rational $m \times m$ matrix function $\overline{W}$ by $\overline{W}(\lambda) = \overline{W(\overline{\lambda})}$. Then clearly $\overline{W}(\lambda) = \overline{D} + \overline{C}(\lambda - \overline{A})^{-1}\overline{B}$ is a minimal realization for $\overline{W}$. For all real λ in the domain of W, we have $\overline{W}(\lambda) = W(\lambda)$. It follows that $\overline{W} = W$, and hence $W(\lambda) = \overline{D} + \overline{C}(\lambda - \overline{A})^{-1}\overline{B}$ is a minimal realization for W. So the systems $(A, B, C, D; \mathbb{C}^n, \mathbb{C}^m)$ and $(\overline{A}, \overline{B}, \overline{C}, \overline{D}; \mathbb{C}^n, \mathbb{C}^m)$ are similar. In particular $D = \overline{D}$, thus D is a real matrix.

Let U be an invertible complex matrix such that

$$U^{-1}AU = \overline{A}, \qquad U^{-1}B = \overline{B}, \qquad CU = \overline{C}. \tag{15.2}$$

Put $\Omega = \operatorname{col}(CA^{j-1})_{j=1}^n$. Then $\overline{\Omega} = \operatorname{col}(\overline{C}\,\overline{A}^{j-1})_{j=1}^n$, and so $\Omega U = \overline{\Omega}$. Due to the minimality, the matrix Ω has rank n. Now we construct a special left inverse $\Omega^{(-1)}$ of Ω and an invertible $n \times n$ matrix S such that

$$\Omega^{(-1)}\overline{\Omega} = S^{-1}\overline{S}.$$

Write $\Omega = [\omega_{ij}]_{i=1,j=1}^{kn\ \ n}$. Choose $1 \le i_1 < i_2 < \cdots < i_n \le kn$ such that

$$S = [\omega_{i_\alpha \beta}]_{\alpha,\beta=1}^n$$

is invertible. Define $\Omega^{(-1)}$ to be the $n \times kn$ matrix all of whose columns are zero except those with index $i_1, \ldots, i_n$, while together the latter form the inverse of S. Then Ω and S have the desired properties, and hence

$$U = \Omega^{(-1)}\overline{\Omega} = S^{-1}\overline{S}.$$

Using this in (15.2) we get

$$SAS^{-1} = \overline{SAS^{-1}}, \qquad SB = \overline{SB}, \qquad CS^{-1} = \overline{CS^{-1}}.$$

Thus SAS^{-1}, SB, and CS^{-1} are real matrices. But then

$$W(\lambda) = D + CS^{-1}(\lambda I_n - SAS^{-1})^{-1}SB$$

is clearly a minimal real realization for W. This completes the proof. □

Let W be a rational $m \times m$ matrix function, and write

$$W(\lambda) = [w_{ij}(\lambda)]_{i,j=1}^{m}.$$

If the functions w_{ij} may be written as quotients of (scalar) polynomials having real coefficients, then obviously W is real. The converse is also true. For the special case when W is proper, this is an easy consequence of Theorem 15.1. For arbitrary real rational $m \times m$ matrix functions, not necessarily proper, the result follows by applying a suitable Möbius transformation mapping the extended real line onto itself.

Next we study real factorizations of rational matrix functions. Let W, W_1 and W_2 be rational $m \times m$ functions, and suppose that $W = W_1W_2$. We say that this factorization is a (*minimal*) *real factorization* if (it is a minimal factorization and) the factors W_1 and W_2 are real. We shall characterize minimal real factorizations in terms of supporting projections. For convenience we restrict ourselves to the case where the functions W, W_1 and W_2 are proper and have the value I_m at infinity.

Theorem 15.2. *Suppose $W(\lambda) = I_m + C(\lambda I_n - A)^{-1}B$ is a minimal real realization. Let Π be a supporting projection of the system $\Theta = (A, B, C; \mathbb{C}^n, \mathbb{C}^m)$, and let $W = W_1W_2$ be the corresponding (minimal) factorization of W. This factorization is real if and only if Π is a real matrix.*

Proof. One checks without difficulty that $\overline{\Pi}$ is also a supporting projection of the system Θ. The corresponding (minimal) factorization is $\overline{W} = \overline{W}_1\overline{W}_2$, where

$$\overline{W_j}(\lambda) = \overline{W_j(\overline{\lambda})}, \qquad j = 1, 2.$$

The desired result is now immediate from Theorem 9.3. □

Let us remark that it may happen that W has plenty of minimal factorizations with non-real factors, but no minimal real factorization. To give an example, let W be the real rational 3×3 matrix function given by

$$W(\lambda) = \begin{bmatrix} \dfrac{\lambda^3 + 2\lambda^2 + 1}{\lambda(\lambda^2+1)} & \dfrac{3\lambda^2+1}{\lambda(\lambda^2+1)} \\ \\ \dfrac{-2\lambda^2 - 1}{\lambda(\lambda^2+1)} & \dfrac{\lambda^3 - 2\lambda^2 - 1}{\lambda(\lambda^2+1)} \end{bmatrix},$$

and introduce

$$A = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & -1 \end{bmatrix}.$$

Then $W(\lambda) = I_2 + C(\lambda I_3 - A)^{-1}B$ is a minimal real realization for W. Observe that

$$A^\times = A - BC = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}.$$

The non-trivial invariant subspaces of A, considered as an operator on $\mathbb{R}^3$ are $\mathbb{R}\dot{+}\mathbb{R}\dot{+}\{0\}$ and $\{0\}\dot{+}\{0\}\dot{+}\mathbb{R}$. The non-trivial invariant subspaces of $A^\times$, considered as an operator on $\mathbb{R}^3$, are $\mathbb{R}\dot{+}\{0\}\dot{+}\{0\}$ and $\{0\}\dot{+}\mathbb{R}\dot{+}\mathbb{R}$. From this it is clear that the system $\Theta = (A, B, C; \mathbb{C}^3, \mathbb{C}^2)$ has no real supporting projections. But then the function W does not admit any minimal real factorization. In particular W has no minimal factorization involving real elementary factors. This is in contrast with the situation where complex factorizations are considered. Indeed, as the function W has simple poles, it can be written as a product of three complex elementary factors (Section 10.2). In fact, one such factorization into degree one factors is given by

$$W(\lambda) = \begin{bmatrix} \dfrac{\lambda+1}{\lambda} & \dfrac{1-i}{\lambda} \\ \dfrac{-1}{\lambda} & \dfrac{\lambda-\lambda+i}{\lambda} \end{bmatrix} \begin{bmatrix} 1 & \dfrac{1+2i}{\lambda+i} \\ 0 & \dfrac{\lambda-i}{\lambda+i} \end{bmatrix} \begin{bmatrix} \dfrac{\lambda-1+i}{\lambda-1} & \dfrac{1-i}{\lambda-1} \\ \dfrac{-1}{\lambda-1} & \dfrac{\lambda-1}{\lambda-i} \end{bmatrix}.$$

Note that from the above example it follows that Theorems 10.10 and 10.11 do not have real counterparts.

15.2 Real monic matrix polynomials

Throughout this section L will be a monic $m \times m$ matrix polynomial. We say that L is *real* if $L(\lambda)$ is a real matrix for all $\lambda \in \mathbb{R}$. An equivalent requirement is that all coefficients of L are real matrices.

Let l be the degree of L. If there exists a monic system Θ,

$$\Theta = (T, R, Q, 0; \mathbb{C}^{ml}, \mathbb{C}^m),$$

such that the transfer function of Θ is L^{-1} and T, R and Q are real matrices, then clearly L is real. The converse is also true. To see this, just take the first companion system (3.3) corresponding to L. This characterization of real monic matrix polynomial could also have been obtained from Theorem 15.1.

Assume that $L = L_2L_1$, where L_1 and L_2 are monic $m \times m$ matrix polynomials. We say that the factorization $L = L_2L_1$ is *real* if the factors L_1 and L_2 (and therefore also L) are real. The next theorem is the analogue of Theorem 15.2.

Theorem 15.3. *Suppose that $\Theta = (T, R, Q, 0; \mathbb{C}^{ml}, \mathbb{C}^m)$ is a monic system such that the transfer function of Θ is L^{-1}, and let T, R and Q be real matrices. Let M be a supporting subspace for Θ, and let $L = L_2L_1$ be the corresponding factorization of L. This factorization is real if and only if M is self-conjugate.*

Proof. Write $\overline{M} = \{\overline{x} \mid x \in M\}$. Then $\overline{M}$ is also a supporting subspace for Θ and the corresponding factorization of L is $L = \overline{L}_2\overline{L}_1$, where

$$\overline{L}_j(\lambda) = \overline{L_j(\overline{\lambda})}, \qquad j = 1, 2.$$

This implies the desired result. □

15.3 Stable and isolated invariant subspaces

In this section we study *stable invariant subspaces* and *isolated invariant subspaces* of operators acting on finite-dimensional real spaces. We refrain from giving the explicit definition of these notions because they are formally the same as those presented in Sections 14.1 and 14.8. In the one but last paragraph of Subsection 14.8.1 we mentioned that in the complex case each stable invariant subspace is isolated and conversely. When the underlying scalar field is the real line, this is no longer true.

We shall begin our investigation by considering some simple special cases. But first we introduce some notation and terminology.

Let E be a real Banach space. The complexification of E will be denoted by E^c. As a set, E^c consists of all (ordered) pairs (x, y) with x and y in E. Instead of (x, y) we shall write $x + iy$. If $\eta = x + iy$ belongs to E^c, then $\overline{\eta} = x - iy$ is called the *conjugate* of η. We call $\eta = x + iy$ *real* if $\eta = \overline{\eta}$ or, equivalently, $y = 0$. The real vectors are identified with those of E in the usual way.

If N is a subspace of E, then N^c is a subspace of E^c. Let M be a subspace of E^c. Then $\overline{M} = \{\overline{\eta} \mid \eta \in M\}$ is also a subspace of E^c. We call M *self-conjugate* if $M = \overline{M}$. Observe that M is self-conjugate if and only if there exists a subspace N of E such that $M = N^c$.

Suppose T is a (bounded) linear operator from E into F. Here E and F are real Banach spaces. Define $T^c : E^c \to F^c$ by $T^c(x + iy) = Tx + iTy$. Then T^c is a (bounded) linear operator which is called the *complexification* of T. For an arbitrary (bounded) linear operator $S : E^c \to F^c$, we define the *conjugate* $\overline{S} : E^c \to F^c$ by $\overline{S}(\eta) = \overline{S(\overline{\eta})}$. Observe that $\overline{S}$ is a (bounded) linear operator. We call S *real* if $S = \overline{S}$. One checks without difficulty that S is real if and only if $S = T^c$ for some (bounded) linear operator $T : E \to F$. Also, S is real if and only if S maps real vectors in E into real vectors in F.

Assume now that E and F are finite-dimensional real spaces with bases $e_1, \ldots, e_m$ and $f_1, \ldots, f_n$, respectively. Note that $e_1, \ldots, e_m$ and $f_1, \ldots, f_n$ form bases of E^c and F^c, respectively. With respect to these bases a linear operator $S : E^c \to F^c$ can be represented by a matrix, say

$$S = [s_{jk}]_{j=1,k=1}^{n \quad m}.$$

A straightforward calculation shows that $\overline{S}$ is then given by

$$\overline{S} = [\overline{s}_{jk}]_{j=1,k=1}^{n \quad m}.$$

Thus S is real if and only if all entries s_{jk} in the matrix representation for S are real. So, after specification of bases consisting of real vectors, real operators between complexifications of finite-dimensional real spaces can be identified with real matrices.

Let T be a linear operator acting on a finite-dimensional real space E. The *spectrum* of T is by definition the spectrum of T^c. It is denoted by $\sigma(T)$. Since the characteristic polynomial of T^c has real coefficients, the spectrum of T is symmetric with respect to the real line. The points of $\sigma(T)$ are called the *eigenvalues* of T. If λ_0 is a real eigenvalue of T, the geometric and algebraic multiplicity of λ_0 are equal to $\dim \operatorname{Ker}(\lambda_0 - T)$ and $\dim \operatorname{Ker}(\lambda_0 - T)^n$, respectively. Here $n = \dim E$.

Lemma 15.4. *Suppose* $\dim E$ *is odd and* $\sigma(T)$ *consists of exactly one real eigenvalue of geometric multiplicity one. Then each invariant subspace of* T *is both stable and isolated.*

Proof. The hypothesis on T implies that T is unicellular. Hence each invariant subspace of T is isolated.

Let N be an invariant subspace of T. Put $k = \dim N$. Since $\dim E$ is odd each operator S on E has an invariant subspace of dimension k. To see this, observe that $\sigma(S)$ contains at least one real point and use the real Jordan normal form for S (cf.,[88], 36.2). The proof that N is stable is now similar to that of Theorem 14.2 (see also the proof of Corollary 14.3). □

Lemma 15.5. *Suppose* $\dim E$ *is even and* $\sigma(T)$ *consists of exactly one real eigenvalue of geometric multiplicity one. Then the even-dimensional invariant subspaces of* T *are stable and the odd-dimensional invariant subspaces of* T *are not stable. All invariant subspaces of* T *are isolated.*

Proof. The last statement of the theorem is clear from the fact that T is unicellular.

Let N be an invariant subspace of T, and put $k = \dim N$. Assume k is even. Then each operator S on E has an invariant subspace of dimension k. This follows from the hypothesis that $\dim E$ is even and the real Jordan normal form of S. Using the same method as in the proof of Lemma 15.4 we can now show that N is stable.

Next assume that k is odd. In order to prove that N is not stable, we may suppose that $\sigma(T) = \{0\}$. With respect to a suitable basis for E, the matrix representation of T has the upper triangular nilpotent Jordan form

$$T = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & 0 & 1 \\ 0 & \cdots & \cdots & 0 & 0 \end{bmatrix}.$$

Now let ε be positive a positive number, and let T_ε be the matrix which one obtains if for even j from 2 to n, the $(j, j-1)$th entry in T is set to be $-\varepsilon^2$ while all other entries remain unchanged. Here n is the order of T, and according to our hypotheses this number is even. Clearly $T_\varepsilon \to T$ as $\varepsilon \downarrow 0$. One checks without difficulty that $\sigma(T_\varepsilon) = \{i\varepsilon, -i\varepsilon\}$. So $\sigma(T_\varepsilon) \cap \mathbb{R} = \emptyset$. Recalling that a (scalar) polynomial with real coefficients and odd degree has at least one real zero, we see that T_ε has no invariant subspaces of odd dimension. This completes the proof. □

From Lemma 15.5 it is already clear that not every isolated invariant subspace will be stable.

Lemma 15.6. *Suppose $\sigma(T)$ consists of exactly one real eigenvalue of geometric multiplicity at least two. Then T has neither stable nor isolated non-trivial invariant subspaces.*

Proof. Let N be a non-trivial invariant subspace of T. The proof that N is not stable is almost verbatim the same as that of Lemma 14.4. To establish that N is also not isolated, use the argument presented in the second and third paragraph of the proof of Theorem 14.28 with $\mathbb{C}^k$ replaced by E and A by T. □

From Lemmas 15.4–15.6, we obtain the following two observations. If $\sigma(T)$ consists of exactly one real eigenvalue of geometric multiplicity one, then each invariant subspace of T is isolated. Also, if $\sigma(T)$ consists of exactly one real eigenvalue of geometric multiplicity at least two, then no non-trivial invariant subspace of T is isolated. In Subsection 14.8.1 we have seen that in the complex case the same conclusions hold even when the single eigenvalue is non-real.

Lemma 15.7. *Suppose $\sigma(T)$ consists of two non-real eigenvalues of geometric multiplicity one. Then each invariant subspace of T is both stable and isolated.*

Proof. First of all, note that T is unicellular. Hence each invariant subspace of T is isolated. Next observe that all invariant subspaces of T are even-dimensional. In particular the dimension of E is even. The rest of the argument is now similar to that presented in the second paragraph of the proof of Lemma 15.5. □

Lemma 15.8. *Suppose $\sigma(T)$ consists of two non-real eigenvalues of geometric multiplicity at least two. Then T has neither stable nor isolated non-trivial invariant subspaces.*

Proof. Let N be a non-trivial invariant subspace of T. The proof that N is not stable is analogous to that of Lemma 14.4. In order to prove that N is also not isolated, we argue as follows.

Consider $N^c = \{x+iy \mid x, y \in N\}$. Observe that N^c is a non-trivial invariant subspace of T^c. The spectrum of T^c consists of two non-real eigenvalues of geometric multiplicity at least two. Denote these eigenvalues by $\alpha + i\beta$ and $\alpha - i\beta$, and let N_+, respectively N_-, be the generalized eigenspace corresponding to T^c and $\alpha + i\beta$, respectively $\alpha - i\beta$. The non-trivial stable invariant subspaces of T^c are N_+ and N_-. In the complex case, however, the notion of a stable and that of an isolated invariant subspace coincide. So the only non-trivial isolated invariant subspaces of T^c are N_+ and N_-. Now $N_- = \{\overline{\eta} \mid \eta \in N_+\}$ and $E^c = N_+ \dot{+} N_-$. From this it is clear that $N_- \neq N^c \neq N_+$. It follows that the T^c-invariant subspace N^c is not isolated.

Let $M_1, M_2, \ldots$ be a sequence of T^c-invariant subspace all different from N^c, such that $\text{gap}\,(M_k, N^c) \to 0$. For $k = 1, 2 \ldots$, we put

$$M_k^+ = M_k \cap N_+, \qquad M_k^- = M_k \cap N_-.$$

Then $M_k = M_k^+ \dot{+} M_k^-$, $\text{gap}\,(M_k^+, N^c \cap N_+) \to 0$ and $\text{gap}\,(M_k^-, N^c \cap N_-) \to 0$ for $k \to \infty$. From $M_k \neq N^c$ and

$$N^c = [N^c \cap N_+] \dot{+} [N^c \cap N_-],$$

we see that either $M_k^+ \neq N^c \cap N_+$ or $M_k^- \neq N^c \cap N_-$. Assume, for instance, that $M_k^+ \neq N^c \cap N_+$ for infinitely many k. Then, by passing to a subsequence, we may assume that $M_k^+ \neq N^c \cap N_+$ for all k. Put

$$L_k = M_k^+ \dot{+} \{\overline{\eta} \mid \eta \in M_k^+\}.$$

Then L_k is T^c-invariant. Moreover, L_k is self-conjugate and hence we have $L_k = N_k^c = \{x + iy \mid x, y \in N_k\}$ for some T-invariant subspace N_k of E.

Observe that

$$N_k^c = [N_k^c \cap N_+] \dot{+} [N_k^c \cap N_-],$$

where $N_k^c \cap N_+ = M_k^+$ and $N_k^c \cap N_- = \{\overline{\eta} \mid \eta \in M_k^+\}$. So

$$\text{gap}\,(N_k^c \cap N_+, N^c \cap N_+) \to 0, \qquad (k \to \infty),$$

$$\text{gap}\,(N_k^c \cap N_-, N^c \cap N_-) \to 0, \qquad (k \to \infty).$$

It follows that $\text{gap}\,(N_k^c, N^c) \to 0$. But then $\text{gap}\,(N_k, N) \to 0$ too. Since $N_k^c \cap N_+ = M_k^+ \neq N^c \cap N_+$, we have that $N_k \neq N$ for all k. We conclude that N is not isolated, and the proof is complete. □

In order to deal with an arbitrary linear operator T on a finite-dimensional real space E, we introduce some more notation and terminology. Let λ_0 be a real eigenvalue of T. Recall that the algebraic multiplicity of λ_0 is equal to $\dim \operatorname{Ker}(\lambda_0 - T)^n$, where $n = \dim E$. The space $\operatorname{Ker}(\lambda_0 - T)^n$ can be described as follows. Consider the spectral projection of T^c corresponding to λ_0. It is easy to see that this spectral projection is a real operator. Hence it is of the form $P(T;\lambda_0)^c$ for some projection $P(T;\lambda_0)$ of E. We call $P(T;\lambda_0)$ the *spectral projection* of T corresponding to λ_0. The image of $P(T;\lambda_0)$ is the space $\operatorname{Ker}(\lambda_0 - T)^n$; it is called the *generalized eigenspace* corresponding to T and λ_0.

Next let $\alpha + i\beta$ be a non-real eigenvalue of T. Then $\alpha - i\beta$ is an eigenvalue of T too, and the geometric (algebraic) multiplicities of $\alpha + i\beta$ and $\alpha - i\beta$ are the same. The spectral projections of T^c corresponding to $\alpha + i\beta$ and $\alpha - i\beta$ are non-real. However, their sum is real. In other words, there exists a projection $P(T;\alpha,\beta)$ of E such that $P(T;\alpha,\beta)^c$ is the spectral projection corresponding to T and the spectral set $\{\alpha + i\beta, \alpha - i\beta\}$. We call $P(T;\alpha,\beta)$ the *spectral projection* corresponding to $a \pm i\beta$. Note that

$$\big(\operatorname{Im} P(T;\alpha,\beta)\big)^c = \operatorname{Ker}(\alpha + i\beta - T^c)^n \dotplus \operatorname{Ker}(\alpha - i\beta - T^c)^n.$$

The image $\operatorname{Im} P(T;\alpha,\beta)$ of $P(T;\alpha,\beta)$ is called the *generalized eigenspace* corresponding to T and $\alpha \pm i\beta$; its dimension is two times the algebraic multiplicity of $\alpha \pm i\beta$ as an eigenvalue of T^c. Write

$$\sigma(T) = \{\lambda_j\}_{j=1}^r \cup \{\alpha_k \pm i\beta_k\}_{k=1}^s,$$

where $\lambda_1, \dots, \lambda_r$ are the distinct real eigenvalues of T and $\alpha_1 + i\beta_1, \dots, \alpha_s + i\beta_s$ are the different eigenvalues of T lying in the upper half-plane. Put

$$P(\lambda_j) = P(T;\lambda_j), \qquad P(\alpha_k,\beta_k) = P(T;\alpha_k,\beta_k),$$
$$N(\lambda_j) = \operatorname{Im} P(T;\lambda_j), \qquad N(\alpha_k,\beta_k) = \operatorname{Im} P(T;\alpha_k,\beta_k).$$

So $N(\lambda_1), \dots, N(\lambda_r)$ and $N(\alpha_1,\beta_1), \dots, N(\alpha_s,\beta_s)$ are the different generalized eigenspaces of T. Note that the projections

$$P(\lambda_1), \dots, P(\lambda_r), \qquad P(\alpha_1,\beta_1), \dots, P(\alpha_s,\beta_s)$$

are mutually disjoint and add up to the identity. Hence

$$E = N(\lambda_1) \dotplus \cdots \dotplus N(\lambda_r) \dotplus N(\alpha_1,\beta_1) \dotplus \cdots \dotplus N(\alpha_s,\beta_s).$$

The invariant subspaces for T are the subspaces of E of the form

$$N = N_1 \dotplus \cdots \dotplus N_r \dotplus \tilde{N}_1 \dotplus \cdots \dotplus \tilde{N}_s, \tag{15.3}$$

where, for $j = 1, \dots, r$, the space N_j is a T-invariant subspace of $N(\lambda_j)$ and, for $k = 1, \dots, s$, the space $\tilde{N}_k$ is a T-invariant subspace of $N(\alpha_k,\beta_k)$.

Theorem 15.9. *A subspace N of E is T-invariant and stable if and only if N is of the form* (15.3)*, where for $j = 1, \dots, r$ and $k = 1, \dots, s$ we have*

(i) *N_j is an arbitrary even-dimensional T-invariant subspace of $N(\lambda_j)$ whenever the algebraic multiplicity of λ_j is even and the geometric multiplicity of λ_j is one;*

(ii) *N_j is an arbitrary T-invariant subspace of $N(\lambda_j)$ whenever the algebraic multiplicity of λ_j is odd and the geometric multiplicity of λ_j is one;*

(iii) *$N_j = \{0\}$ or $N_j = N(\lambda_j)$ whenever λ_j has geometric multiplicity at least two;*

(iv) *$\tilde{N}_k$ is an arbitrary T-invariant subspace of $N(\alpha_k, \beta_k)$ whenever $\alpha_k + i\beta$ and $\alpha_k - i\beta$ have geometric multiplicity one;*

(v) *$\tilde{N}_k = \{0\}$ or $\tilde{N}_k = N(\alpha_k, \beta_k)$ whenever $\alpha_k + i\beta_k$ and $\alpha_k - i\beta_k$ have geometric multiplicity at least two.*

Also, N is an isolated invariant subspace of T if and only if N is of the form (15.3)*, where for $j = 1, \dots, r$ and $k = 1, \dots, s$ the conditions* (iii), (iv), (v) *and, moreover,*

(vi) *N_j is an arbitrary T-invariant subspace of $N(\lambda_j)$ whenever the geometric multiplicity of λ_j is one*

are satisfied.

Proof. Let N be an invariant subspace of T, and write N in the form (15.3), where $N_j = P(\lambda_j)N$, $j = 1, \dots, r$, and $\tilde{N}_k = P(\alpha_k, \beta_k)N$, $k = 1, \dots, s$. For $j = 1, \dots, r$ let T_j be the restriction of T to $N(\lambda_j)$. Also, for $k = 1, \dots, s$, let $\tilde{T}_k$ be the restriction of T to $N(\alpha_k, \beta_k)$. It is easy to see that N is isolated if and only if for $j = 1, \dots, r$, the space N_j is an isolated invariant subspace of T_j and for $k = 1, \dots, s$ the space $\tilde{N}_k$ is an isolated invariant subspace of $\tilde{T}_k$. This statement remains true if isolated is replaced by stable. The proof of this involves the analogues for the real case of Lemmas 14.5 and 14.6. Observe now that $\sigma(T) = \{\lambda_j\}$ and $\sigma(\tilde{T}_k) = \{\alpha_k + i\beta_k, \alpha_k - i\beta_k\}$. Here $j = 1, \dots, r$ and $k = 1, \dots, s$. The desired result is now immediate from Lemmas 15.5–15.8. □

Theorem 15.9 implies that every stable invariant subspace is also isolated. As we already observed, the converse of this is not true. The next theorem is a reformulation of Theorem 15.9.

Theorem 15.10. *Let N be an invariant subspace for T. Then N is stable if and only if N meets the following requirements:*

(i) *If $\{0\} \neq P(\lambda_j)N \neq N(\lambda_j)$, then the geometric multiplicity of λ_j is one, $j = 1, \dots, r$;*

(ii) *If $P(\lambda_j)N$ has odd dimension, then the algebraic multiplicity of λ_j is odd too, $j = 1, \dots, r$;*

(iii) *If* $\{0\} \neq P(\alpha_k, \beta_k)N \neq N(\alpha_k, \beta_k)$, *then the geometric multiplicity of* $\alpha_k \pm i\beta_k$ *is one,* $k = 1, \ldots, s$.

Also N *is isolated if and only if* (i) *and* (iii) *are satisfied.*

In the next section we shall deal with stable and isolated real factorizations of rational matrix functions and of monic matrix polynomials, and also with stable and isolated solutions of the real operator Riccati equation. The version of Theorem 15.9 most fitted for studying these notions reads as follows.

Theorem 15.11. *Let* E_1 *and* E_2 *be finite-dimensional real spaces, and let*

$$T = \begin{bmatrix} T_1 & T_0 \\ 0 & T_2 \end{bmatrix}$$

be a linear operator acting on $E = E_1 \dot{+} E_2$. *Then* E_1 *is a stable invariant subspace for* T *if and only if the following conditions are satisfied:*

(i) *each common eigenvalue of* T_1 *and* T_2 *is an eigenvalue of* T *of geometric multiplicity one;*

(ii) *each common real eigenvalue of* T_1 *and* T_2 *of which the algebraic multiplicity with respect to* T_1 *is odd, has odd algebraic multiplicity with respect to* T.

Also, E_1 *is an isolated invariant subspace for* T *if and only if* (i) *is satisfied.*

Proof. The proof is similar to that of Theorem 14.7. Use Theorem 15.10 instead of Theorem 14.1. □

15.4 Stable and isolated real factorizations

In this section we discuss stable and isolated real factorizations of rational matrix functions and monic matrix polynomials. Also we deal with stable and isolated solutions of the real operator Riccati equation. We begin by considering real rational matrix functions.

Suppose

$$W_0 = W_{01}W_{02} \tag{15.4}$$

is a minimal real factorization. Here W_0, W_{01} and W_{02} are real proper rational $m \times m$ matrix functions which have the value I_m at infinity. For $j = 1, 2$, let

$$W_{0j}(\lambda) = I_m + C_{0j}(\lambda I_{n_j} - A_{0j})^{-1}B_{0j} \tag{15.5}$$

be a minimal real realization for W_{0j}. We say that the factorization (15.4) is *isolated* (*with respect to real perturbations*) if there exists $\varepsilon > 0$ with the following property. If $W_0 = W_1W_2$, where W_1 and W_2 admit minimal real realizations

$$W_j(\lambda) = I_m + C_j(\lambda I_{n_j} - A_j)^{-1}B_j, \qquad j = 1, 2,$$

such that

$$\|A_j - A_{0j}\| + \|B_j - B_{oj}\| + \|C_j - C_{0j}\| < \varepsilon, \qquad j = 1, 2,$$

then $W_1 = W_{01}$ and $W_2 = W_{02}$. By Theorem 7.7, this definition does not depend on the choice of the minimal realizations (15.5). The definition of the notion of a minimal factorization that is *stable* (*with respect to real perturbations*) is analogous to that of a stable minimal factorization given in Section 14.3. The only difference is that now all (minimal) realizations are required to be real. We omit the details.

Closely related to the concepts introduced in the preceding paragraph are those of an isolated and of a stable real supporting projections. Let A_0, B_0 and C_0 be real matrices of appropriate sizes, and consider the corresponding system $\Theta_0 = (A_0, B_0, C_0; \mathbb{C}^n, \mathbb{C}^m)$. Here, as usual, the matrices A_0, B_0 and C_0 are identified with their canonical actions between the corresponding Euclidean spaces. Suppose Π_0 is a real $n \times n$ matrix whose canonical action on $\mathbb{C}^n$ is a supporting projection for Θ_0. In other words, Π_0 is a real supporting projection for Θ_0. We say that Π_0 is *isolated* (*with respect to real perturbations*) if there exists $\varepsilon > 0$ such that each real supporting projection Π for Θ_0 different from Π_0 satisfies $\|\Pi - \Pi_0\| \geq \varepsilon$. Similarly, we call Π_0 *stable* (*with respect to real perturbations*) if, given $\varepsilon > 0$, there exists $\omega > 0$ such that the following is true. If $\Theta_0 = (A_0, B_0, C_0; \mathbb{C}^n, \mathbb{C}^m)$ is a system with real matrices A, B, C and $\|\Theta - \Theta_0\| < \omega$, then Θ has a real supporting projection Π such that $\|\Pi - \Pi_0\| < \varepsilon$.

In the next theorem W_0, W_{01} and W_{02} are real proper rational $m \times m$ matrix functions with value I_m at infinity.

Theorem 15.12. *Suppose $W_0 = W_{01}W_{02}$ is a minimal real factorization. This factorization is stable with respect to real perturbations if and only if the following conditions are satisfied:*

(i) *each common pole (zero) of W_0 and W_{02} is a pole (zero) of W_0 of geometric multiplicity one;*

(ii) *each common real pole of W_{01} and W_{02} of which the order with respect to W_{01} is odd has odd order as a pole of W_0;*

(iii) *each common real pole of W_{01}^{-1} and W_{02}^{-1} of which the order with respect to W_{02}^{-1} is odd has odd order as a pole of W_0^{-1}.*

Also, $W_0 = W_{01}W_{02}$ is isolated with respect to real perturbation if and only if (i) *is satisfied.*

Proof. We only present an outline of the proof. Let

$$W(\lambda) = I_m + C_0(\lambda I_n - A_0)^{-1}B_0$$

be a minimal real realization of W. Denote the supporting projection for $\Theta_0 = (A_0, B_0, C_0; \mathbb{C}^n, \mathbb{C}^m)$ corresponding to the factorization (15.4) by Π_0. From Theorem 15.2 we know that Π_0 is real. Using the techniques of Section 14.3, one can

show that (15.4) is stable (isolated) with respect to real perturbations if and only if the same is true for Π_0. Consider the matrices A_0, B_0, C_0 and Π_0 now as operators from $\mathbb{R}^n$ into $\mathbb{R}^n$, from $\mathbb{R}^m$ into $\mathbb{R}^n$, from $\mathbb{R}^n$ into $\mathbb{R}^m$ and from $\mathbb{R}^n$ into $\mathbb{R}^n$, respectively. Then Π_0 is stable (isolated) with respect to real perturbations if and only if $\operatorname{Ker}\Pi_0$ and $\operatorname{Im}\Pi_0$ are stable (isolated) invariant subspaces for A_0 and $A_0 - B_0C_0$, respectively. From Theorem 15.11 we conclude that Theorem 15.12 is correct if (ii) and (iii) are replaced by

(ii)′ *each common real pole (zero) of W_{01} and W_{02} of which the pole- (zero-) multiplicity with respect to $W_{01}(W_{02})$ is odd, is a pole (zero) of W_0 of odd pole- (zero-) multiplicity,*

Note that a pole of W_0 has geometric multiplicity one if and only if its order and pole-multiplicity are the same (cf., Corollary 8.10). The desired result is now immediate from the fact that if λ_0 is a pole of W_0 of geometric multiplicity one, then the geometric multiplicity of λ_0 with respect to W_{01} does not exceed one. □

In Theorems 14.9 and 15.12 poles of geometric multiplicity one play an important role. If λ_0 is a pole of a rational matrix function W_0 of (positive) order p, then always $p \leq \delta(W_0; \lambda_0)$, where $\delta(W_0; \lambda_0)$ is the local degree (pole-multiplicity) of W_0 at λ_0; equality occurs if and only if the geometric multiplicity of λ_0 is one. This fact was used in the proof of Theorem 15.12. It is also useful in dealing with specific examples.

Example. Consider the case where

$$W_0(\lambda) = \begin{bmatrix} 1 & \dfrac{1}{\lambda} + \dfrac{1}{\lambda^2} \\ 0 & 1 + \dfrac{1}{\lambda} \end{bmatrix}$$

$$W_{01}(\lambda) = \begin{bmatrix} 1 & \dfrac{1}{\lambda} \\ 0 & 1 \end{bmatrix}, \qquad W_{02}(\lambda) = \begin{bmatrix} 1 & 0 \\ 0 & 1 + \dfrac{1}{\lambda} \end{bmatrix}.$$

Then $W_0 = W_{01}W_{02}$ and this factorization is minimal. Indeed, $\delta(W_0) = \delta(W_0; 0) = 2$ and $\delta(W_{01}) = \delta(W_{01}; 0) = \delta(W_{02}) = \delta(W_{02}; 0) = 1$. The (only) common pole of W_{01} and W_{02} is 0 and the order of 0 as a pole of W_0 is equal to $\delta(W_0; 0)$, namely 2. Moreover, W_{01} and W_{02} have no common zeros. So the factorization $W_0 = W_{01}W_{02}$ is isolated with respect to real perturbations. It is clear that (ii) is not satisfied, so this factorization is not stable with respect to real perturbations. Note that it is a stable factorization in the sense of Section 14.3.

Next we consider factorizations of real monic matrix polynomials that are *stable* or *isolated* (*with respect to real perturbations*). The definition of these notions is straightforward, see Section 14.5 and the closing paragraph of Section 14.6.

Therefore we omit the details. The crucial point is that now all factorizations are required to be real. The proof of the next theorem may be based on the material contained in Section 15.3 and involves the techniques of Section 14.5.

Theorem 15.13. *Let L_0, L_{01} and L_{02} be real monic $k \times k$ matrix polynomials, and suppose that $L_0 = L_{02}L_{01}$. This factorization is stable with respect to real perturbations if and only if the following conditions are satisfied:*

(i) *each common eigenvalue of L_{01} and L_{02} is an eigenvalue of L_0 of geometric multiplicity one;*

(ii) *each common real eigenvalue of L_{01} and L_{02} with odd zero-multiplicity relative to L_{01}, is an eigenvalue of L_0 with odd zero-multiplicity.*

Also, the factorization $L_0 = L_{02}L_{01}$ is isolated with respect to real perturbations if and only if condition (i) *is satisfied.*

Note that the zero-multiplicity of an eigenvalue λ_0 of a monic matrix polynomial L is equal to the order of λ_0 as zero of the scalar polynomial $\det L(\lambda)$. Using this one can easily construct examples showing that an isolated factorization of a real monic matrix polynomial need not be stable (with respect to real perturbations). Finally, we consider the real operator Riccati equation

$$XT_{21}X + XT_{22} - T_{11}X - T_{12} = 0. \tag{15.6}$$

Here T_{jk} is a linear operator from the finite-dimensional real space E_k into the finite-dimensional real space E_j $(j, k = 1, 2)$. The definition of a *stable solution* of (15.6) is formally the same as that given in Section 14.6. The only difference is that here the underlying spaces are real instead of complex.

Theorem 15.14. *Let R be a solution of the Riccati equation* (15.6)*, and put*

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} : E_1 \dot{+} E_2 \to E_1 \dot{+} E_2.$$

Then R is stable if and only if the following conditions are satisfied:

(i) *each common eigenvalue of $T_{11} - RT_{21}$ and $T_{22} + T_{21}R$ is an eigenvalue of T of geometric multiplicity one;*

(ii) *each common real eigenvalue of $T_{11} - RT_{21}$ and $T_{22} + T_{21}R$ for which the algebraic multiplicity with respect to $T_{22} + T_{21}R$ is odd, is an eigenvalue of odd algebraic multiplicity of T.*

It is also possible to introduce the notion of an *isolated* solution of of (15.6). The definition goes along the lines indicated at the end of Section 14.6, with the understanding that in the present situation the underlying spaces are real instead of complex. We refrain from further pursuing this point here.

15.5 Stability of stable real factorizations

In order to simplify the following discussion we introduce some terminology. Let T be a linear operator on a finite-dimensional real space E. An invariant subspace N of T is called *perfectly stable* if for each $\varepsilon > 0$ there exists $\delta > 0$ such that $\|S-T\| < \delta$ implies that S has a stable invariant subspace M with gap $(M, N) < \varepsilon$. This terminology will be only of temporary use because we shall show that the notion of a stable and that of a perfectly stable invariant subspace coincide. It is clear that perfect stability implies stability.

Lemma 15.15. *Suppose that all eigenvalues of T have geometric multiplicity one. Put $n = \dim E$. If n is odd, then for each integer k between 0 and n, the operator T has a stable invariant subspace of dimension k. If n is even, then for each even integer m between 0 and n, the operator T has a stable invariant subspace of dimension m.*

Proof. Consider the real Jordan normal form of S (cf., [88], 36.2) and apply Theorem 15.9. □

From the material contained in Section 15.3, we recall the following facts. If $\sigma(T)$ consists of one real eigenvalue of geometric multiplicity one and $\dim E$ is even, then an invariant subspace for T is stable if and only if it has even dimension. If either $\sigma(T)$ consists of one real eigenvalue of geometric multiplicity one and $\dim E$ is odd, or $\sigma(T)$ consists of two non-real eigenvalues of geometric multiplicity one, then each invariant subspace for T is stable. Also note that if T has no real eigenvalue, then each T-invariant subspace (so in particular E itself) has even dimension.

Lemma 15.16. *Suppose that either $\sigma(T)$ consists of exactly one real eigenvalue of geometric multiplicity one, or $\sigma(T)$ consists of two non-real eigenvalues of geometric multiplicity one. Then each stable invariant subspace of T is perfectly stable.*

Proof. Let N be a stable invariant subspace for T, and put $k = \dim N$. Since T is unicellular, N is the only T-invariant subspace of dimension k. Let $T_1, T_2, \ldots$ be a sequence of operators on E converging to T. A simple rank argument (cf., the proof of Theorem 14.24) shows that for n sufficiently large all eigenvalues of T_n have geometric multiplicity one. But then Lemma 15.15 guarantees that for n sufficiently large the operator T_n has a stable invariant subspace M_n of dimension k. The method used to prove Theorem 14.2 can now be employed to show that there exists a subsequence of $M_1, M_2, \ldots$ converging in the gap topology to a k-dimensional invariant subspace M for T. Since M must be equal to N, the proof is complete. □

Theorem 15.17. *Let N be a stable invariant subspace for a linear operator T acting on a finite-dimensional real space E. Then, given $\varepsilon > 0$, there exists $\delta > 0$ such that $\|S - T\| < \delta$ implies that S has a stable invariant subspace M satisfying* gap $(M, N) < \varepsilon$.

Proof. The conclusion of the theorem which we have to establish is that N is perfectly stable. We only present an outline of the proof.

Let us adopt the notation of Section 15.3. Write N in the form (15.3), where $N_j = P(\lambda_j)N,\ j = 1, \dots, r$, and $\tilde{N}_k = P(\alpha_k, \beta_k)N,\ k = 1, \dots, s$. For $j = 1, \dots, r$, let T_k be the restriction of T to $N(\lambda_j)$. Also, for $k = 1, \dots, s$, let $\tilde{T}_k$ be the restriction of T to $N(\alpha_k, \beta_k)$. From the proof of Theorem 15.9 and Lemma 15.16 it follows that N_j is a perfectly stable invariant subspace for T_j, and $\tilde{N}_k$ is a perfectly stable invariant subspace for $\tilde{T}_k$.

Fix j between 1 and r, and let $S_1, S_2, \dots$ be a sequence of operators on E converging to T. Further, let Γ be a circle centered at λ_j such that all eigenvalues of T different from λ_j are outside Γ. For m sufficiently large, the circle Γ will split the spectrum of S_m. Moreover,

$$\lim_{m\to\infty} P(S_m^c; \Gamma) = P(T^c; \Gamma) = P(\lambda_j)^c,$$

where $P(\lambda_j)$ is as defined in Section 15.3. Note that $P(S_m^c; \Gamma)$ is a real operator, so $P(S_m^c; \Gamma) = P_m^c$ for some projection P_m of E. We obviously have

$$\lim_{m\to\infty} P_m = P(\lambda_j).$$

Put $F_m = P(\lambda_j)P_m + \big(I - P(\lambda_j)\big)\big(I - P_m\big)$. Then $F_m \to I$. By passing to a subsequence (if necessary) we may assume that F_m is invertible for all m. It is clear that $F_m P_m = P(\lambda_j)F_m$, so

$$P(\lambda_j) = F_m P_m F_m^{-1}, \qquad m = 1, 2, \dots .$$

Set $T_m = F_m S_m F_m^{-1}$. Then $T_m \to T$. Moreover,

$$P(\lambda_j)^c = P(T_m^c; \Gamma), \qquad m = 1, 2, \dots .$$

Let T_{mj} be the restriction of T_m to $N(\lambda_j) = \operatorname{Im} P(\lambda_j)$. Then $T_{mj} \to T_j$ as $m \to \infty$. Since N_j is a perfectly stable invariant subspace for T_j, there exists a sequence $L_1, L_2, \dots$ of subspaces of $N(\lambda_j)$ such that L_m is a stable T_{mj}-invariant subspace and $\operatorname{gap}(L_m, N_j) \to 0$ as $m \to \infty$. Observe now that L_m is also a stable invariant subspace for T_m. This we know from the real analogue of Lemma 14.5. Put $M_m = F_m^{-1}L_m$. Then M_m is a stable invariant subspace for S_m. From $F_m \to I$ one gets that $\operatorname{gap}(M_m, L_m) \to 0$. Together with $\operatorname{gap}(L_m, N_j) \to 0$, this gives that $\operatorname{gap}(M_m, N_j) \to 0$ as $m \to \infty$.

We have now proved that the spaces N_j are perfectly stable invariant subspaces for T. In the same way one can show that the spaces $\tilde{N}_k$ are of this type. Since N has the form (15.3), it follows that N is a perfectly stable invariant subspace for T, and the proof is complete. □

One might think that Theorem 15.17 could be proved in the same way as Theorem 14.24, using Theorem 15.11 instead of Theorem 14.7. This method of proof however does not work.

With the help of Theorem 15.17 one can sharpen Theorems 15.12, 15.13 and 15.14. By way of example, we present the details concerning the extension of Theorem 15.14 (see also Theorem 14.25 and [13], Theorems 4.2 and 4.3).

Theorem 15.18. *Let R be a stable solution of the Riccati equation* (15.6)*. Then, given $\varepsilon > 0$, there exists $\delta > 0$ such that $\|S_{jk} - T_{jk}\| < \delta$ for $j, k = 0, 1$ implies that the Riccati equation*

$$XS_{21}X + XS_{22} - S_{11}X - S_{12} = 0$$

admits a stable solution Q for which $\|Q - R\| < \varepsilon$.

We emphasize that the solution Q is stable. This can also be expressed as follows. Each common eigenvalue of $S_{11} - QS_{21}$ and $S_{22} + S_{21}Q$ is an eigenvalue of

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

of geometric multiplicity one, and each common real eigenvalue of $S_{11} - QS_{21}$ and $S_{22} + S_{21}Q$ for which the algebraic multiplicity with respect to $S_{22} + S_{21}Q$ is odd, is an eigenvalue of odd algebraic multiplicity of S.

Notes

This chapter is practically identical to Chapter XI in [14], except for the proof of Lemma 15.6 which has been shortened considerably via a reference to the proof of Theorem 14.28. The real invariant subspaces of a real matrix are also discussed in [70]; Sections 14.6 and 15.9 in the latter book study the topology of the set of real invariant subspaces, and the stability issue, while applications are given in Section 17.10. Further developments, including α-stability and stability of special classes of subspaces for matrices that have symmetries with respect to an indefinite inner product, can be found in [99] and the references cited therein.

Bibliography

[1] D. Alpay, I. Gohberg, Unitary rational matrix functions. In: *Topics in Interpolation Theory of Rational Matrix-valued Functions*, OT **33**, Birkhäuser Verlag, Basel, 1988, pp. 175–222.

[2] D.Z. Arov, Scattering theory with dissipation of energy. *Dokl. Akad. Nauk SSSR* **216** (4) (1974), 713–716 (Russian); English translation with addenda: *Sov. Math. Dokl.* **15** (1974), 848–854.

[3] D.Z. Arov, M.A. Kaashoek, D.R. Pik, Optimal time-variant systems and factorization of operators, I, minimal and optimal systems. *Integral Equations and Operator Theory* **31** (1998), 389–420.

[4] D.Z. Arov, M.A. Kaashoek, D.R. Pik, Minimal representation of a contractive operator as a product of two bounded operators. *Acta Sci. Math* (*Szeged*) 71 (2005), 313–336.

[5] D.Z. Arov, M.A. Kaashoek, D.R. Pik, The Kalman-Yakubovich-Popov inequality for discrete time systems of infinite dimension. *J. Operator Theory* **55** (2006), 393–438.

[6] K.R. Baker, *Introduction to sequencing and scheduling.* John Wiley, New York, 1975.

[7] J.A. Ball, N. Cohen, De Branges-Rovnyak operator models and systems theory, a survey. In: *Topics in Matrix and Operator Theory*, OT **50**, Birkhäuser Verlag, Basel, 1991, pp. 93–136.

[8] J.A. Ball, I. Gohberg, L. Rodman, *Interpolation of rational matrix functions.* OT **45**, Birkhäuser Verlag, Basel, 1986.

[9] S. Barnett, *Introduction to mathematical control theory.* Oxford, Clarendon Press, 1975.

[10] H. Bart, Transfer functions and operator theory. *Lin. Alg. Appl.* **84** (1986), 33–61.

[11] H. Bart, I. Gohberg, M.A. Kaashoek, A new characteristic operator function connected with operator polynomials. Wiskundig Seminarium der Vrije Universiteit, Amsterdam, Rapport nr 62, 1977.

[12] H. Bart, I. Gohberg, M.A. Kaashoek, Operator polynomials as inverses of characteristic functions. *Integral Equations and Operator Theory* **1** (1978), 1–12.

[13] H. Bart, I. Gohberg, M.A. Kaashoek, Stable factorizations of monic matrix polynomials and stable invariant subspaces. *Integral Equations and Operator Theory* **1** (1978), 496–517.

[14] H. Bart, I. Gohberg, M.A. Kaashoek, *Minimal factorization of matrix and operator functions*. OT **1**, Birkhäuser Verlag, Basel, 1979.

[15] H. Bart, I. Gohberg, M.A. Kaashoek, Wiener-Hopf integral equations. In: *Toeplitz Centennial*, OT **4**, Birkhäuser Verlag, Basel, 1982, pp. 85–135.

[16] H. Bart, I. Gohberg, M.A. Kaashoek, Explicit Wiener-Hopf factorization and realization. In: *Constructive methods of Wiener-Hopf factorization*, OT **21**, Birkhäuser Verlag, Basel, 1986, pp. 317–355.

[17] H. Bart, I. Gohberg, M.A. Kaashoek, P. Van Dooren, Factorizations of transfer functions. *SIAM J. Control Opt.* **18** (1980), 675–696.

[18] H. Bart, I. Gohberg, M.A. Kaashoek, A.C.M. Ran, Schur complements and realizations of transfer functions. *Lin. Alg. Appl.* **399** (2005), 203–224.

[19] H. Bart, H. Hoogland, Complementary triangular forms of pairs of matrices, realizations with prescribed main matrices, and complete factorization of rational matrix functions. *Lin. Alg. Appl.* **103** (1988), 193–228.

[20] H. Bart, L.G. Kroon, Companion based matrix functions: description and minimal factorization. *Lin. Alg. Appl.* **248** (1996), 1–46.

[21] H. Bart, L.G. Kroon, Factorization and job scheduling: a connection via companion based matrix functions. *Lin. Alg. Appl.* **248** (1996), 111–136.

[22] H. Bart, L.G. Kroon, Variants of the Two Machine Flop Shop Problem connected with factorization of matrix functions. *European J. Operation Research* **91** (1996), 144–159.

[23] H. Bart, L.G. Kroon, R.A. Zuidwijk, Quasicomplete factorization and the two machine flow shop problem. *Lin. Alg. Appl.* **278** (1998), 195–219.

[24] H. Bart, G.Ph.A. Thijsse, Simultaneous reduction to companion and triangular forms of sets of matrices. *Lin. Multilin. Alg.* **26** (1990), 231–241.

[25] H. Bart, H.K. Wimmer, Simultaneous reduction to triangular and companion forms of pairs of matrices: the case rank $(I - AZ) = 1$. *Lin. Alg. Appl.* **150** (1991), 443–461.

[26] H. Bart, R.A. Zuidwijk, Simultaneous reduction to triangular forms after extension with zeroes. *Lin. Alg. Appl.* **281** (1998), 105–135.

[27] V. Belevitch, *Classical Network Theory*. Holden Day, San Francisco, 1968.

[28] H. den Boer, Linearization of operator function on arbitrary open sets. *Integral Equations and Operator Theory* **1** (1978), 19–27.

[29] A. Boettcher, B. Silbermann. *Analysis of Toeplitz operators*, Springer-Verlag, Berlin, 1990.

[30] M.S. Brodskii, *Triangular and Jordan representations of linear operators*. Transl. Math. Monographs, Vol. **32**, Amer. Math. Soc., Providence, R.I., 1970.

[31] M.S. Brodskii, Unitary operator nodes and their characteristic functions. *Uspehi Mat. Nauk* **33** (4) (1978), 141–168 [Russian].

[32] V.M. Brodskii, Some theorems on knots and their characteristic functions. *Funct. Anal. Appl.* **4** (3) (1970), 250–251.

[33] V.M. Brodskii, I.C. Gohberg, M.G. Kreĭn, On characteristic functions of an invertible operator. *Acta. Sci. Math.* (*Szeged*) **32** (1971), 141–164.

[34] S. Campbell, J. Daughtry, The stable solutions of quadratic matrix equations. *Proc. Amer. Math. Soc.* **74** (1979), 19–23.

[35] N. Cohen, On minimal factorizations of rational matrix functions. *Integral Equations and Operator Theory* **6** (1983), 647–671.

[36] M.J. Corless, A.E. Frazho, *Linear systems and control. An operator perspective*. Marcel Dekker, Inc., New York, NY, 2003.

[37] J. Daughtry, Isolated solutions of quadratic matrix equations. *Lin. Alg. Appl.* 21 (1978), 89–94.

[38] P. Dewilde, Cascade scattering matrix synthesis. Tech. Rep. 6560–21, Information Systems Lab., Stanford University, Stanford, 1970.

[39] P. Dewilde, J. Vandewalle, On the factorization of a nonsingular rational matrix. *IEEE Trans. Circuits and Systems*, vol CAS-22 (8) (1975), 387–401.

[40] R.G. Douglas, C. Pearcy, On a topology for invariant subspaces. *J. Funct. Anal.* **2** (1968), 323–341.

[41] R.G. Douglas, H.S. Shapiro, A.L. Shields, Cyclic vectors and invariant subspaces for the backward shift. *Ann. Inst. Fourier* (*Grenoble*), **20**, (1970), 37–76.

[42] I.A. Feldman, Wiener-Hopf operator equation and its application to the transport equation. *Mat. Issled.* **6** (3) (1971), 115–132 [Russian]; English translation in: *Integral Equations and Operator Theory* **3** (1980), 43–61.

[43] C. Foias, A.E. Frazho, I. Gohberg, M.A. Kaashoek, *Metric constrained interpolation, commutant lifting and systems*. OT **100**, Birkhäuser Verlag, Basel, 1998.

[44] M.R. Garey, D.S. Johnson, *Computers and intractability: A guide to the theory of NP-Completeness*. Freeman, San Fransisco, 1979.

[45] I.C. Gohberg, I.A. Feldman, *Convolution equations and projection methods for their solution*. Transl. Math. Monographs, Vol. **41**, Amer. Math. Soc., Providence, R.I., 1974.

[46] I. Gohberg, S. Goldberg, M.A. Kaashoek, *Classes of Linear Operators*, Volume I. OT **49**, Birkhäuser Verlag, Basel, 1990.

[47] I. Gohberg, S. Goldberg, M.A. Kaashoek, *Classes of Linear Operators*, Volume II. OT **63**, Birkhäuser Verlag, Basel, 1993.

[48] I. Gohberg, S. Goldberg, M.A. Kaashoek, *Basic Classes of Linear Operators*. Birkhäuser Verlag, Basel, 2003.

[49] I. Gohberg, M.A. Kaashoek, Block Toeplitz operators with rational symbols. In: *Contributions to Operator Theory and its Applications*, OT **35**, Birkhäuser Verlag, Basel, 1988, pp. 385–440.

[50] I. Gohberg, M.A. Kaashoek, The state space method for solving singular integral equations. In: *Mathematical system theory*, Springer Verlag, Berlin, 1991, pp. 509–523.

[51] I. Gohberg, M.A. Kaashoek, P. Lancaster, General theory of regular matrix polynomials and band Toeplitz operators. *Integral Equations and Operator Theory* **11** (1988), 776–882.

[52] I. Gohberg, M.A. Kaashoek, D.C. Lay, Equivalence, linearization and decompositions of holomorphic operator functions. *J. Funct. Anal.* **28** (1978), 102–144.

[53] I. Gohberg, M.A. Kaashoek, L. Lerer, Minimality and realization of discrete time-varying systems. In: *Time-Variant Systems and Interpolation*, OT **56**, Birkhäuser Verlag, Basel, 1992, pp. 261–296.

[54] I. Gohberg, M.A. Kaashoek, L. Lerer, L. Rodman, Common multiples and common divisors of matrix polynomials, I. Spectral method. *Indiana Univ. Math. J.* **30** (1981), 321–356.

[55] I. Gohberg, M.A. Kaashoek, L. Rodman, Spectral analysis of families of operator polynomials and a generalized Vandermonde matrix, I. The finite dimensional case. In: *Topics in Functional Analysis*, Advances in Mathematics Supplementary Studies, Vol. **3**, Academic Press, New York, 1978, pp. 91–128.

[56] I. Gohberg, M.A. Kaashoek, F. van Schagen, *Partially specified matrices and operators, classification, completion, applications*. OT **79**, Birkhäuser Verlag, Basel, 1995.

[57] I. Gohberg, M.A. Kaashoek, I.M. Spitkovsky, An overview of matrix factorization theory and operator applications. In: *Factorization and integrable systems*. OT **141**, Birkhäuser verlag, Basel, 2003, pp. 1–102.

[58] I.C. Gohberg, M.G. Kreĭn, The basic propositions on defect numbers, root numbers, and indices of linear operators. *Uspehi Mat. Nauk* **12**, no. 2 (74) (1957), 43–118 [Russian]; English transl: *Amer. Math. Soc. Transl.* (Series 2) **13** (1960), 185–265

[59] I.C. Gohberg, M.G. Kreĭn, Systems of integral equations on a half line with kernels depending on the difference of arguments. *Uspehi Mat. Nauk* **13**, no.

2 (80) (1958), 3–72 [Russian]; English transl: *Amer. Math. Soc. Transl.* (2) **14** (1960), 217–287.

[60] I.C. Gohberg, M.G. Kreĭn, *Introduction to the theory of nonselfadjoint operators*. Transl. Math. Monographs, Vol. **18**, Amer. Math. Soc., Providence, R.I., 1969.

[61] I.C. Gohberg, M.G. Kreĭn, *Theory and applications of Volterra operators in Hilbert space*. Transl. Math. Monographs, Vol. **24**, Amer. Math. Soc., Providence, R.I., 1970.

[62] I. Gohberg, N. Krupnik, *Einführung in die Theorie der eindimensionalen singulären Integraloperatoren*. Birkhäuser Verlag, Basel, 1979.

[63] I. Gohberg, N. Krupnik, *One-Dimensional Linear Singular Integral Equations. I. Introduction*. OT **53**, Birkhäuser Verlag, Basel, 1992.

[64] I. Gohberg, N. Krupnik, *One-Dimensional Linear Singular Integral Equations. II. General Theory and Applications*. OT **54**, Birkhäuser Verlag, Basel, 1992.

[65] I. Gohberg, P. Lancaster, L. Rodman, Spectral analysis of matrix polynomials. I. Canonical forms and divisors. *Lin. Alg. Appl.* **20** (1978), 1–44.

[66] I. Gohberg, P. Lancaster, L. Rodman, Spectral analysis of matrix polynomials. II. The resolvent form and spectral divisors. *Lin. Alg. Appl.* **21** (1978), 65–88.

[67] I. Gohberg, P. Lancaster, L. Rodman, Representations and divisibility of operator polynomials. *Can. J. Math.* **30** (1978), 1045–l069.

[68] I. Gohberg, P. Lancaster, L. Rodman, Perturbation theory for divisors of operator polynomials. *Siam J. Math. Anal.* **10** (1979), 1161–1183.

[69] I. Gohberg, P. Lancaster, L. Rodman, *Matrix polynomials*. Acad. Press, Inc., New York, NY, 1882.

[70] I. Gohberg, P. Lancaster, L. Rodman, *Invariant subspaces of matrices with applications*. John Wiley & Sons, Inc., New York, NY, 1986 = Classics in Appl. Math. **51**, SIAM, Philadelphia, PA, 2006.

[71] I.C. Gohberg, A.S. Markus, Two theorems on the gap between subspaces of a Banach space. *Uspehi Mat. Nauk* 14 (1959), 135–140 [Russian].

[72] I. Gohberg, A.C.M. Ran, On pseudo-canonical factorization of rational matrix functions. *Indagationes Mathematica* N.S. **4** (1993), 51–63.

[73] I. Gohberg, L. Rodman, On spectral analysis of non-monic matrix and operator polynomials, I. Reduction to monic polynomials. *Israel J. Math.* **30** (1978), 133–151.

[74] I.C. Gohberg, E.I. Sigal, An operator generalization of the logarithmic residue theorem and the theorem of Rouché. *Mat. Sbornik* **84** (126) (1971), 607-629 [Russian]; English transl.: *Math. USSR, Sbornik* **13** (1971), 603–625.

[75] S. Goldberg, *Unbounded linear operators*. New York, McGraw-Hill, 1966.

[76] G.J. Groenewald, *Wiener-Hopf factorization of rational matrix functions in terms of realizations, an alternative version*. PhD thesis, Vrije Universiteit, Amsterdam, 1993.

[77] P. Halmos, *A Hilbert space problem book*. Graduate Texts in Mathematics, Vol. **19** (2nd edition), Springer Verlag, Berlin, 1082.

[78] A. Heck, *Introduction to Maple*, 3rd edition, Springer-Verlag, New York, 2003.

[79] J.W. Helton, Discrete time systems, operator models, and scattering theory. *J. Funct. Anal.* **16** (1974), 15–38.

[80] J.W. Helton, Systems with infinite-dimensional state space, the Hilbert space approach. *Proc. IEEE* 64(i) (1976), 145–160.

[81] J.W. Helton, J.A. Ball, The cascade decompositions of a given system VS the linear fractional decompositions of its transfer function. *Integral Equations and Operator Theory* **5** (1982), 341–385.

[82] S.M. Johnson, Optimal two- and three stage production schedules with setup times included. *Naval Research Logistics Quarterly* **1** (1) (1954), 61–68.

[83] M.A. Kaashoek, C.V.M. van der Mee, L. Rodman, Analytic operator functions with compact spectrum, II. Spectral pairs and factorization. *Integral Equations and Operator Theory* **5** (1982), 791–827.

[84] T. Kailath, *Linear systems*. Prentice-Hall, Inc., Englewoord Cliffs, N.J., 1980.

[85] T. Kato, Perturbation theory for nullity, deficiency and other quantities of linear operators. *J. Analyse Math.* **6** (1958), 261–322.

[86] T. Kato, *Perturbation theory for linear operators*. Springer, Berlin-Heidelberg-New York, 1966.

[87] N.M. Kostenko, Necessary and sufficient condition for factorization of rational matrix functions. *Funct. Anal. Appl.* 12 (1978), 87–88 [Russian].

[88] H.-J. Kowalsky, *Lineare Algebra*. Walter de Gruyter, Berlin, 1967.

[89] M.G. Kreĭn, M.A. Krasnosel'skii, D.C. Mil'man, On the defect numbers of linear operators in Banach space and on some geometric problems. *Sbornik Trud. Inst. Mat. Akad. Nauk Ukr. SSR* **11** (1948), 97–112 [Russian].

[90] M.G. Kreĭn, Introduction to the geometry of indefinite J-spaces and to the theory of operators in these spaces. *Amer. Math. Soc. Transl.* (2) **93** (1970), 103–176.

[91] P. Lancaster, L. Rodman, *Algebraic Riccati equations*. Clarendon Press, Oxford, 1995.

[92] P. Lancaster, M. Tismenetsky, *The Theory of Matrices*. Academic Press, San Diego, 1985.

[93] Maple, a computer algebra system, version 9, Maplesoft, Waterloo Maple Inc., Ontario, Canada, 2003

[94] A.S. Markus, V.I. Macaev, Two remarks on factorization of matrix functions. *Mat. Issled.* **42** (1976), 216–223.

[95] N.H. McCoy, On the Characteristic Roots of Matric Polynomials. *Bull. Amer. Math. Soc.* **42** (1936), 592–600.

[96] C. van der Mee, Realization and linearization. Wiskundig Seminarium der Vrije Universiteit, Amsterdam, Rapport nr. 109, 1979.

[97] B. Mitiagin, Linearization of holomorphic operator functions. I, II. *Integral Equations and Operator Theory* **1** (1978), 114–131 and 226–249.

[98] A.C.M. Ran, L. Rodman, A.L. Rubin, Stability index of invariant subspaces of matrices. *Linear and Multilinear Algebra* **36** (1993), 27–39.

[99] A.C.M. Ran, L. Rodman, A Class of Robustness Problems in Matrix Analysis. In: *The Harry Dym Anniversary Volume*, OT **134**, Birkhäuser Verlag, Basel 2002, pp. 337–383.

[100] A.C.M. Ran, L. Roozemond, On strong α-stability of invariant subspaces of matrices. In: *The Gohberg Anniversary Volume*, OT **40**, Birkhäuser Verlag, Basel 1989, pp. 427–435.

[101] L. Rodman, *An introduction to operator polynomials*. OT **38**, Birkhäuser Verlag, Basel, 1989.

[102] L. Roozemond, Canonical pseudo-spectral factorization and Wiener-Hopf integral equations. In: *Constructive methods of Wiener-Hopf factorization*, OT **21**, Birkhäuser Verlag, Basel, 1986, pp. 127–156.

[103] L. Roozemond, *Systems of non-normal and first kind Wiener-Hopf equations*, Ph.D. thesis, Vrije Universiteit, Amsterdam, March 5, 1987.

[104] L.A. Sakhnovich, On the factorization of an operator-valued transfer function. *Soviet Math. Dokl.* **17** (1976), 203–207.

[105] M.A. Shubin, On holomorphic families of subspaces of a Banach space. *Mat. Issled.* **5** (1970), 153–165; Letter to the editors, *Mat. Issled.* **6** (1971), 180 [Russian].

[106] O.J. Staffans, *Well-posed linear systems*. Cambridge University Press, Cambridge, 2005.

[107] R.J. Stroeker, J.F. Kaashoek, *Discovering Mathematics with Maple*, Birkhäuser Verlag, Basel etc., 1991.

[108] B. Sz-Nagy, C. Foias, *Analyse harmonique des opérateur de l'espace de Hilbert*. Paris, Masson and Akad. Kiado, Budapest, 1967.

[109] S.H. Tan, J. Vandewalle, On factorization of rational matrices. *IEEE Trans. Circ. Systems* **35** (1988), 1179–1181.

[110] A.E. Taylor, D.C. Lay, *Introduction to Functional Analysis*. Second Edition, John Wiley and Sons, New York, 1980.

[111] G.Ph.A. Thijsse, Spectral criteria for complementary triangular forms. *Integral Equations Operator Theory* **27** (1997), 228–251.

[112] J. Vandewalle, P. Dewilde, On the irreducible cascade synthesis of a system with a real rational transfer matrix. *IEEE Trans. Circuits and Systems*, vol CAS-24 (9)(1977), 481–494.

[113] J. Vandewalle, P. Dewilde, A local i/o structure theory for multivariable systems and its application to minimal cascade realization. *IEEE Trans. Circuits and Systems*, vol. CAS-25 (5)(1978), 279–289.

[114] W.M. Wonham, *Linear multivariable control, a geometric approach*. Springer-Verlag, Berlin, 1979.

[115] D.C. Youla, P. Tissi, An explicit formula for the degree of a rational matrix. Electrophysies Memo, PIBM RI-1273-65, Polytechnic Institute of Brooklyn, Electrophysics Department, 1965.

[116] D.C. Youla, P. Tissi, n-Port synthesis via reactance extraction. Part 1. *IEEE Int. Con. Rec.* Pt 7, vol. 14 (1966), 183–208.

[117] K. Zhou, J.C, Doyle, K. Glover, *Robust and optimal control*. Prentice Hall, Engelwood Cliffs, NJ, 1996.

[118] Y. Zucker, *Constructive factorization and partial indices of rational matrix functions*. Ph. D. thesis, Tel-Aviv University, Tel-Aviv, 1998.

[119] R.A. Zuidwijk, *Complementary Triangular Forms of Pairs of Matrices and Operators*. Ph.D. Thesis, Erasmus University Rotterdam, 1994.

[120] R.A. Zuidwijk, Complementary triangular forms for infinite matrices. In: *Operator theory and boundary eigenvalue problems*, OT **80**, Birkhäuser Verlag, Basel, 1995, pp. 289–299.

[121] R.A. Zuidwijk, Quasicomplete factorizations for rational matrix functions. *Integral Equations and Operator Theory* **27** (1997), 111–124.

[122] R.A. Zuidwijk, Simultaneous similarity of pairs of companions to their transposes. In: *Operator theory and Analysis*, OT **122**, Birkhäuser Verlag, Basel, 2001, pp. 417–425.

List of Symbols

Symbol	Description
$\mathbb{Z}$	set of integers
$\mathbb{Z}_-$	set of (strictly) negative integers
$\mathbb{Z}_+$	set of non-negative integers (including zero)
$\mathbb{R}$	real line
$\mathbb{C}$	complex plane
$\mathbb{D}$	closed unit disc in complex plane
$\mathbb{C}^n$	Euclidean space of complex $n-$vectors
$\Re\lambda$	real part of complex number λ
$\Im\lambda$	imaginary part of complex number λ
$\bar{\lambda}$	complex conjugate of complex number λ
$\sharp V$	number of elements in (finite) set V
$\operatorname{span} V$	span or linear hull of set V
$\overline{V}$	closure of subset V of topological space
$\dim M$	dimension of linear manifold M
$\operatorname{codim} M$	codimension of linear manifold M
$M^\perp$	orthogonal complement of subspace M in Hilbert space
$z \perp M$	element z perpendicular to set M (Hilbert space)
M/N	quotient space of M over N (also denoted by $\frac{M}{N}$)
$\frac{M}{N}$	quotient space of M over N (also denoted by M/N)
$\oplus$	orthogonal direct sum (of subspaces) of Hilbert spaces
$\dot{+}$	algebraic (possibly non-orthogonal) direct sum of linear manifolds or (sub)spaces
X'	conjugate of Banach space X
A'	conjugate of Banach space operator A
I	identity matrix or identity operator on a Hilbert or Banach space
I_n	$n \times n$ identity matrix or identity operator on $\mathbb{C}^n$
$\operatorname{Ker} A$	kernel or null space of operator or matrix A

$\operatorname{Im} A$	range or image of operator or matrix A
$\operatorname{rank} A$	rank of operator or matrix A
$\det A$	determinant of matrix A
$\operatorname{trace} A$	trace of matrix A
$A^\top$	transpose of matrix A
A^*	adjoint of (complex) Hilbert space operator or (complex) matrix
A^{-1}	inverse of invertible operator or matrix
A^{-*}	stands for $(A^*)^{-1}$
D_T	defect operator $(I - T^*T)^{1/2}$
$\lambda - A$	shorthand for $\lambda I - A$ (standard practice)
$\rho(A)$	resolvent set of operator or matrix
$\sigma(A)$	spectrum of operator or matrix A
$m_A(\alpha)$	algebraic multiplicity of α as an eigenvalue of square matrix A; is zero when λ_0 is not an eigenvalue
p_A	spectral polynomial of square matrix A
$P(A;\Gamma)$	stands for $\frac{1}{2\pi i}\int_\Gamma (\lambda - A)^{-1}\, d\lambda$, the spectral or Riesz projection associated with A and Γ
AM	image of M under operator A (also denoted by $A[M]$)
$A[M]$	image of M under operator A (also denoted by AM)
$A^{-1}[M]$	inverse image of M under operator A
$A\vert_M$	restriction of operator A to subspace M
$A(X_1 \to X_2)$	(possibly) unbounded operator A with domain in X_1 and range in X_2
$\mathcal{D}(A)$	domain of (possibly) unbounded operator A
$N(\lambda)$	generalized eigenspace for eigenvalue λ of a matrix
N_R	angular subspace associated with (angular) operator R
$R \dot{+} Q$	diagonal operator built from R and Q
$\mathcal{L}(Y)$	Banach algebra of all bounded linear operators on Banach space Y
$\mathcal{L}(X, Y)$	Banach space of all bounded operators from Banach space X into Banach space Y
$C(\Gamma, Y)$	Banach space of all Y-valued continuous functions on Γ endowed with the supremum norm
$\mathcal{MP}_\ell$	set of all monic $m \times m$ matrix polynomials of degree ℓ
$(A, B, C, D; X, U, Y)$	system
$(A, B, C, D; X, Y)$	system with coinciding input/output space
$(A, B, C; X, Y)$	unital system with coinciding input/output space and the identity operator as external operator
(A, B, C, D)	system (no underlying spaces specified)
Θ	system

W_Θ	transfer function of system Θ
$\Theta^\times$	associate or inverse system
$A^\times$	associate main or state (space) matrix (operator)
$\Theta_1\Theta_2$	product of two systems
$\Theta_1 \cdots \Theta_k$	product of k systems
$\mathrm{pr}_\Pi(\Theta)$	projection of system Θ associated with supporting projection Π
$\mathrm{Ker}\,(C\|A)$	stands for $\mathrm{Ker}\, C \cap \mathrm{Ker}\, CA \cap \mathrm{Ker}\, CA^2 \cap \cdots$
$\mathrm{Im}\,(A\|B)$	stands for $\mathrm{Im}\, B + \mathrm{Im}\, AB + \mathrm{Im}\, A^2B + \cdots$
Θ_φ	Möbius transformation of system Θ
$\|\Theta - \Theta_0\|$	distance between systems Θ and Θ_0
$\simeq$	similarity between systems
W	rational matrix or operator-valued function
W^{-1}	pointwise inverse of W (defined by $W^{-1}(\lambda) = W(\lambda)^{-1}$)
W^*	pointwise adjoint of W (defined by $W^*(\lambda) = W(\overline{\lambda})^*$)
$W^\top$	pointwise transpose of W (defined by $W^\top(\lambda) = W(\lambda)^\top$)
$W^\#$	matrix function defined by $W^\#(\lambda) = W(\lambda^{-1})$
$\delta(W;\lambda_0)$	local degree of W at λ_0
$\delta(W;\infty)$	local degree of W at ∞
$\delta(W)$	McMillan degree of W
$\delta_{\mathrm{q}}(W)$	quasidegree of W
$\nu(W)$	maximal number of non-trivial factors that can occur in a minimal factorization of W
$\mathrm{Ker}\,(W;\lambda_0)$	space of eigenvectors or root vectors of W at λ_0
$\mathrm{Pol}\,(W;\lambda_0)$	space of pole-vectors of W at λ_0
$\kappa_-(W)$	stands for $\delta(W) + \mathrm{codim}\,\big(\mathrm{Ker}\, p_A(A) + \mathrm{Ker}\, p_{A^\times}(A^\times)\big)$; see (10.28)
$\kappa_+(W)$	stands for $\delta(W) + \dim\,\big(\mathrm{Im}\, p_A(A) + \mathrm{Im}\, p_{A^\times}(A^\times)\big)$; see (10.29)
$\kappa(W)$	stands for $\min\{\kappa_-(W), \kappa_+(W)\}$
$\mathbf{V}$	stands for the matrix given by (11.52)
$V(\alpha_1, \ldots, \alpha_n)$	generalized Vandermonde matrix
2MFSP	two machine flow shop problem
$2\mathrm{MFSP}_{\mathrm{red}}$	reduced two machine flow shop problem
O_j^k	operation involving job j in (instance) of 2MFSP to be processed on the kth machine
$s(J)$	sum of processing times of first machine in instance J of 2MFSP
$t(J)$	sum of processing times of second machine in instance J 2MFSP
$\mu(J)$	minimal makespan (i.e., smallest possible makespan) of instance J of 2MFSP
J_{ext}	augmented instance of 2MFSP
$F(O_j^1)$	finish time of operation O_j^1 in (instance of) 2MFSP
$S(O_j^1)$	start time of operation O_j^1 in (instance of) 2MFSP

$\mu_{\mathrm{red}}(J)$	reduced minimal makespan (i.e., smallest possible makespan) of instance J of $2\mathrm{MFSP}_{\mathrm{red}}$
$\nu(J)$	coinciding sums of processing times of the two machines in instance J 2MFSP
$\eta(M_1, M_2)$	minimal opening between subspaces M_1 and M_2
$\varphi_{\min}(M_1, M_2)$	minimal angle between subspaces M_1 and M_2
$\rho(M_1, M_2)$	stands for $\sup_{0 \neq x \in M_1} \inf_{y \in M_2} \frac{\|x-y\|}{\|x\|}$
$\mathrm{gap}\,(M_1, M_2)$	gap or maximal opening between subspaces M_1 and M_2 i.e., $\max\{\rho(M_1, M_2),\, \rho(M_2, M_1)\}$
$\mathrm{GAP}\,(\mathcal{M}, \mathcal{N})$	gap between complete chains $\mathcal{M}$ and $\mathcal{N}$ of subspaces
$\overline{x}$	conjugate $(\overline{x}_1, \ldots, \overline{x}_n)^\top$ of vector $x \in \mathbb{C}^n$
$\overline{M}$	stands for $\{\overline{x} \mid x \in M\}$ of a subspace of $\mathbb{C}^n$ (not to be confused with the closure operation)
$\overline{A}$	conjugate of matrix A
$\overline{W}$	pointwise conjugate of W (defined by $\overline{W}(\lambda) = \overline{W(\overline{\lambda})}$)
E^c	complexification of real Banach space E
$\overline{x - iy}$	conjugate of vector $x + iy$ in complexification of real Banach space
T^c	complexification of (bounded) linear operator between real Banach spaces (defined by $T^c(x + iy) = Tx + iTy$)
$\overline{T}$	conjugate of (bounded) linear operator T between complexifications of real Banach spaces (defined by $\overline{T}(\eta) = \overline{T(\overline{\eta})}$)
$P(T; \lambda_0)$	spectral projection of operator T corresponding to real eigenvalue λ_0 (real spaces) of matrix T
$P(T; \alpha, \beta)$	the range of $P(T; \lambda_0)$

Index