TSHILIDZI MARWALA
ILYES BOULKAIBET
SONDIPON ADHIKARI

PROBABILISTIC FINITE ELEMENT MODEL UPDATING USING BAYESIAN STATISTICS

APPLICATIONS TO AERONAUTICAL AND MECHANICAL ENGINEERING

WILEY

PROBABILISTIC FINITE ELEMENT MODEL UPDATING USING BAYESIAN STATISTICS

PROBABILISTIC FINITE ELEMENT MODEL UPDATING USING BAYESIAN STATISTICS

APPLICATIONS TO AERONAUTICAL AND MECHANICAL ENGINEERING

Tshilidzi Marwala and Ilyes Boulkaibet

University of Johannesburg, South Africa

Sondipon Adhikari

Swansea University, UK

WILEY

This edition first published 2017

Registered Office
John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Library of Congress Cataloging-in-Publication Data

Names: Marwala, Tshilidzi, 1971– author. | Boulkaibet, Ilyes, 1981– author. | Adhikari, Sondipon, author.
Title: Probabilistic finite element model updating using Bayesian statistics: applications to aeronautical and mechanical engineering / Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.
Description: Chichester, UK ; Hoboken, NJ : John Wiley & Sons, 2017. | Includes bibliographical references and index.
Identifiers: LCCN 2016019278| ISBN 9781119153030 (cloth) | ISBN 9781119153016 (epub)
Subjects: LCSH: Finite element method. | Bayesian statistical decision theory. | Engineering–Mathematical models.
Classification: LCC TA347.F5 M3823 2016 | DDC 620.001/51825–dc23
LC record available at https://lccn.loc.gov/2016019278

A catalogue record for this book is available from the British Library.

Cover image: Godruma/Gettyimages

Set in 10/12pt Times by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

Contents

Acknowledgements

We would like to thank the University of Johannesburg and the University of Swansea for contributing towards the writing of this book. We also would like to thank Michael Friswell, Linda Mthembu, Niel Joubert and Ishmael Msiza for contributing towards the writing of this book.

We dedicate this book to the schools that gave us the foundation to always seek excellence in everything we do: the University of Cambridge and the University of Johannesburg.

Tshilidzi Marwala, PhD
Johannesburg
1 February 2016

Ilyes Boulkaibet, PhD
Johannesburg
1 February 2016

Sondipon Adhikari, PhD
Swansea
1 February 2016

Nomenclature

AI	Artificial intelligence
AIC	Akaike information criterion
APEPCS	Adaptive pruned-enriched population control scheme
AR	Acceptance rate
BFGS	Quasi-Newton Broyden–Fletcher–Goldfarb–Shanno
BIC	Bayesian information criterion
CG	Conjugate gradient
c.o.v.	Coefficient of variation
DIC	Deviance information criterion
DOF	Degree of freedom
DWIS	Dynamically weighted importance sampling
FEM	Finite element model
FRF	Frequency response function
GA	Genetic algorithm
GS	Gibbs sampling
HMC	Hybrid Monte Carlo
MC	Markov chain
MCDWIS	Monte Carlo dynamically weighted importance sampling
MD	Molecular dynamics
MCMC	Markov chain Monte Carlo
M-H	Metropolis–Hastings
ML	Maximum likelihood
MAP	Maximum a posteriori
NS	Nested sampling
PDF	Probability distribution function
PSO	Particle swarm optimisation
SA	Simulated annealing
SHMC	Shadow hybrid Monte Carlo
S2HMC	Separable shadow hybrid Monte Carlo

SS	Slice sampling
VV	Velocity verlet
N	Number of degrees of freedom
$\boldsymbol{Z_X}$	Experimental data vector
$\boldsymbol{Z_i}$	Analytical data vector
$\boldsymbol{\theta}$	Uncertain parameter vector
$Dev(\boldsymbol{\theta})$	Deviance of $\boldsymbol{\theta}$
P_D	Posterior mean deviance parameter
$\boldsymbol{S}$	Structure's sensitivity matrix
J	Objective function
$\boldsymbol{Z}$	Evidence
$\ddot{x}$	Acceleration
$\boldsymbol{W}$	Weighting matrix
$\boldsymbol{H}$	Hessian matrix
$\boldsymbol{I}$	Unit matrix
η	Step size used by the conjugate gradient technique
$\boldsymbol{V}$	Variance matrix
$\boldsymbol{\Omega}$	Diagonal matrix with diagonal elements of the natural frequencies
x_i	Chromosome vector or position vector
p_i	Best position
v_i	Velocity
d	Dimension of the updated vector
$\mathbb{R}$	One-dimensional real domain
$\mathbb{R}^n$	n-dimensional real domain
$\mathbb{R}^{m\times n}$	$m\times n$-dimensional real domain
$\boldsymbol{T}$	Transformation matrix
$\boldsymbol{V_{\theta_j}}$	Covariance matrix of the updated vector $\boldsymbol{\theta}$ at the jth iteration
$\boldsymbol{V_{Z_X}}$	Covariance of the measured data
P	Probability function
$\mathcal{D}$	Experimental model data
$P(\boldsymbol{\theta}\|\mathcal{D})$	The posterior probability distribution function
$q(\cdot\|\cdot)$	Proposed probability distribution function
$T_n(\cdot\|\cdot)$	Transition matrix
$\mathcal{N}(\mu, \sigma)$	Normal distribution with mean μ and variance σ
$P(\cdot\|\cdot)$	Joint density
μ_f	Expectation value of the function f
f_i^m	ith measured natural frequency
ω_i^m	ith measured circular natural frequency
N_m	Number of measured modes
f_i	ith analytical frequency obtained from the finite element model
j	Imaginary unit of a complex number
$\|\boldsymbol{\Lambda}\|$	Euclidean norm of $\boldsymbol{\Lambda}$
λ	Lagrange multiplier
K	Bayes factor
$\boldsymbol{E}_i$	Error vector

$E(\cdot)$	Mean value
$E(\mathbf{z}\mathbf{z}^T)$	Variance matrix of $\mathbf{z}$
R_t	Normalisation constant ratio
$\boldsymbol{X}^m$	The Fourier-transformed displacement
$\boldsymbol{F}^m$	Force matrix
W	Kinetic energy
V	Potential energy
∇V	Gradient of V
H	Hamiltonian function
$H_{[2k]}$	Shadow Hamiltonian function of order $2k$
$\mathbf{p}$	Momentum vector
$\nabla(\cdot)$	Gradient
K_B	Boltzmann constant
T	Temperature
δt	Time step
$\{\cdot,\cdot\}$	Poisson bracket of two functions

1

Introduction to Finite Element Model Updating

1.1 Introduction

Finite element model updating methods are intended to correct and improve a numerical model to match the dynamic behaviour of real structures (Marwala, 2010). Modern computers, with their ability to process large matrices at high speed, have facilitated the formulation of many large and complicated numerical models, including the boundary element method, the finite difference method and the finite element models. This book deals with the finite element model that was first applied in solving complex elasticity and structural analysis problems in aeronautical, mechanical and civil engineering. Finite element modelling was proposed by Hrennikoff (1941) and Courant and Robbins (1941). Courant applied the Ritz technique and variational calculus to solve vibration problems in structures (Hastings *et al*., 1985). Despite the fact that the approaches used by these researchers were different from conventional formulations, some important lessons are still relevant. These differences include mesh discretisation into elements (Babuska *et al*., 2004).

The Cooley–Turkey algorithms, which are used to speedily obtain Fourier transformations, have facilitated the development of complex techniques in vibration and experimental modal analysis. Conversely, the finite element model ordinarily predicts results that are different from the results obtained from experimental investigation. Among reasons for the discrepancy between finite element model prediction and experimentally measured data are as the following (Friswell and Mottershead, 1995; Marwala, 2010; Dhandole and Modak, 2011):

- model structure errors resulting from the difficulty in modelling damping and complex shapes such as joints, welds and edges;
- model order errors resulting from the difficulty in modelling non-linearity and often assuming linearity;

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

- model parameter errors resulting in difficulty in identifying the correct material properties;
- errors in measurements and signal processing.

In finite element model updating, it is assumed that the measurements are correct within certain limits of uncertainty and, for that reason, a finite element model under consideration will need to be updated to better reflect the measured data. Additionally, finite element model updating assumes that the difficulty in modelling joints and other complicated boundary conditions can be compensated for by adjusting the material properties of the relevant elements. In this book, it is also assumed that a finite element model is linear and that damping is sufficiently low not to warrant complex modelling (Mottershead and Friswell, 1993; Friswell and Mottershead, 1995). Using data from experimental measurements, the initial finite element model is updated by correcting uncertain parameters so that the model is close to the measured data. Alternatively, finite element model updating is an inverse problem and the goal is to identify the system that generated the measured data (Brincker *et al.*, 2001; Dhandole and Modak, 2010; Zhang *et al.*, 2011; Boulkaibet, 2014; Fuellekrug *et al.*, 2008; Cheung and Beck, 2009; Mottershead *et al.*, 2000).

There are two main approaches to finite element model updating, namely, maximum likelihood and Bayesian methods (Marwala, 2010; Mottershead *et al.*, 2011). In this book, we apply a Bayesian approach to finite element model updating.

1.2 Finite Element Modelling

Finite element models have been applied to aerospace, electrical, civil and mechanical engineering in designing and developing products such as aircraft wings and turbo-machinery. Some of the applications of finite element modelling are (Marwala, 2010): thermal problems, electromagnetic problems, fluid problems and structural modelling. Finite element modelling typically entails choosing elements and basis functions (Chandrupatla and Belegudu, 2002; Marwala, 2010). Generally, there are two types of finite element analysis that are used: two-dimensional and three-dimensional modelling (Solin *et al.*, 2004; Marwala, 2010).

Two-dimensional modelling is simple and computationally efficient. Three-dimensional modelling, on the other hand, is more accurate, though computationally expensive. Finite element analysis can be formulated in a linear or non-linear fashion. Linear formulation is simple and usually does not consider plastic deformation, which non-linear formulation does consider. This book only deals with linear finite element modelling, in the form of a second-order ordinary differential equation of relations between mass, damping and stiffness matrices. A finite element model has *nodes*, with a grid called a *mesh*, as shown in Figure 1.1 (Marwala, 2001). The mesh has material and structural properties with particular loading and boundary conditions. Figure 1.1 shows the dynamics of a cylinder, and the mode shape of the first natural frequency occurring at 433 Hz.

These loaded nodes are assigned a specific density all over the material, in accordance with the expected stress levels of that area (Baran, 1988). Sections which undergo more stress will then have a higher node density than those which experience less or no stress. Points of stress concentration may have fracture points of previously tested materials, joints, welds and high-stress areas. The mesh may be imagined as a spider's web so that, from each node, a mesh

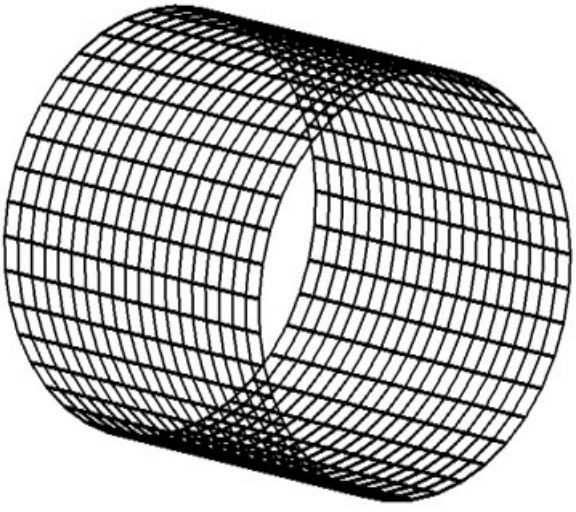

Figure 1.1 A finite element model of a cylindrical shell

element extends to each of the neighbouring nodes. This web of vectors has the material properties of the object, resulting in a study of many elements.

On implementing finite element modelling, a choice of elements needs to be made and these include beam, plate, shell elements or solid elements. A question that needs to be answered when applying finite element analysis is whether the material is isotropic (identical throughout the material), orthotropic (only identical at 90°) or anisotropic (different throughout the material) (Irons and Shrive, 1983; Zienkiewicz, 1986; Marwala, 2010).

Finite element analysis has been applied to model the following problems (Zienkiewicz, 1986; Marwala, 2010):

- vibration analysis for testing a structure for random vibrations, impact and shock;
- fatigue analysis to approximate the life cycle of a material or a structure due to cyclical loading;
- heat transfer analysis to model conductivity or thermal fluid dynamics of the material or structure.

Hlilou *et al.* (2009) successfully applied finite element analysis in softening material behaviour, while Zhang and Teo (2008) successfully applied it in the treatment of a lumbar degenerative disc disease. White *et al*. (2008) successfully applied finite element analysis for shallow-water modelling, while Pepper and Wang (2007) successfully applied it in wind energy assessment of renewable energy in Nevada. Miao *et al.* (2009) successfully applied a three-dimensional finite element analysis model in the simulation of shot peening. Bürg and Nazarov (2015) successfully applied goal-oriented adaptive finite element methods in elliptic problems, while Amini *et al*. (2015) successfully applied finite element modelling in functionally graded piezoelectric harvesters. Haldar *et al.* (2015) successfully applied finite element modelling in the study of the flexural behaviour of singly curved sandwich composite

structures, while Millar and Mora (2015) successfully applied finite element methods to study the buckling in simply supported Kirchhoff plates. Jung *et al.* (2015) successfully used finite element models and computed tomography to estimate cross-sectional constants of composite blades, while Evans and Miller (2015) successfully applied a finite element model to predict the failure of pressure vessels. Other successful applications of finite element analysis are in the areas of metal powder compaction processing (Rahman *et al*., 2009), ferroelectric materials (Schrade *et al.*, 2007), rock mechanics (Chen *et al*., 2009), orthopaedics (Easley *et al.*, 2007), carbon nanotubes (Zuberi and Esat, 2015), nuclear reactors (Wadsworth *et al*., 2015) and elastic wave propagation (Gao *et al.*, 2015; Gravenkamp *et al*., 2015).

1.3 Vibration Analysis

An important aspect to consider when implementing finite element analysis is the kind of data that the model is supposed to predict. It can predict data in many domains, such as the time, modal, frequency and time–frequency domains (Marwala, 2001, 2010). This book is concerned with constructing finite element models to predict measured data more accurately. Ideally, a finite element model is supposed to predict measured data irrespective of the domain in which the data are presented. However, this is not necessarily the case because models updated in the time domain will not necessarily predict data in the modal domain as accurately as they will for data in the time domain. To deal with this issue, Marwala and Heyns (1998) used data in the modal and frequency domains simultaneously to update the finite element model in a multi-criteria optimisation fashion. Again, whichever domain is used, the updated model performs less well on data in a different domain than those used in the updating process. In this book, we use data in the modal domain. Raw data are measured in the time domain and Fourier analysis techniques transform the data into the frequency domain. Modal analysis is applied to transform the data from the frequency domain to the modal domain. All of these domains include similar information, but each domain reveals different data representations.

1.3.1 Modal Domain Data

The modal domain expresses data as natural frequencies, damping ratios and mode shapes. The technique used for extracting the modal properties is a process called *modal analysis* (Ewins, 1995). Natural frequencies are basic characteristics of a system and can be extracted by exciting the structure and analysing the vibration response. Cawley and Adams (1979) used changes in the natural frequencies to identify damage in composite materials. Farrar *et al.* (1994) successfully used the shifts in natural frequencies to identify damage on an I-40 bridge. Other successful applications of natural frequencies include damage detection in tabular steel offshore platforms (Messina *et al*., 1996, 1998), spot welding (Wang *et al.*, 2008) and beam-like structures (Zhong and Oyadiji, 2008; Zhong *et al*., 2008).

A mode shape represents the curvature of a system vibrating at a given mode and a particular natural frequency. West (1982) successfully applied the modal assurance criterion for damage on a Space Shuttle orbiter body flap, while Kim *et al.* (1992) successfully used the coordinate modal assurance criterion of Lieven and Ewins (1988) for damage detection in structures. Further applications of mode shapes include composite laminated plates (Araújo dos Santos *et al.*, 2006; Qiao *et al.*, 2007), linear structures (Fang and Perera, 2009), beam-type structures (Qiao

and Cao, 2008; Sazonov and Klinkhachorn, 2005), optical sensor configuration (Chang and Pakzad, 2015), multishell quantum dots (Vanmaekelbergh *et al.*, 2015) and creep characterisation (Hao *et al*., 2015).

1.3.2 Frequency Domain Data

The measured excitation and response of the structure are converted into the frequency domain using Fourier transforms (Ewins, 1995; Maia and Silva, 1997), and from these the *frequency response function* is extracted. Frequency response functions have, in general, been used to identify faults (Sestieri and D'Ambrogio, 1989; Faverjon and Sinou, 2009). D'Ambrogio and Zobel (1994) used frequency response functions to identify the presence of faults in a truss structure, while Imregun *et al*. (1995) used frequency response functions for damage detection. Lyon (1995) and Schultz *et al*. (1996) used measured frequency response functions for structural diagnostics. Other direct applications of the frequency response functions include the work of Shone *et al*. (2009), Ni *et al*. (2006), X. Liu *et al*. (2009), White *et al.* (2009) and Todorovska and Trifunac (2008). Additional applications include missing-data estimation (Ugryumova *et al*., 2015), identification of a non-commensurable fractional transfer (Valério and Tejado, 2015), as well as damage detection (Link and Zimmerman, 2015).

1.4 Finite Element Model Updating

In real life, it turns out that the predictions of the finite element model are quite different from the measurements. As an example, for a finite element model of a simply suspended beam, the differences between the model-predicticted natural frequencies and the measured frequencies are shown in Table 1.1 (Marwala and Sibisi, 2005; Marwala, 2010). These results are for a fairly easy structure to model, and they demonstrate that the finite element model's data are different from the measured data. Finite element model updating has been studied quite extensively (Friswell and Mottershead, 1995; Mottershead and Friswell, 1993; Maia and Silva, 1997; Marwala, 2010). There are three approaches used in finite element model updating: direct methods, iterative deterministic and uncertainty quantification methods. Direct approaches are computationally inexpensive, but reproduce modal properties that are physically unrealistic.

Although the finite element model can predict measured quantities, the updated model is limited in that it loses the connectivity of nodes, results in populated matrices and in loss of

Table 1.1 Comparison of finite element model and real measurements

Mode number	Finite element frequencies (Hz)	Measured frequencies (Hz)
1	42.30	41.50
2	117.0	114.5
3	227.3	224.5
4	376.9	371.6

matrix symmetry. All these factors are physically unrealistic. Iterative techniques use changes in physical parameters to update the finite element models and, thereby, generate models that are physically realistic (Marwala, 2010). However, since they are based on optimisation techniques, the problems of global versus local optimum solution and over-fitting the measured data, these methods still produce unrealistic results. Esfandiari *et al.* (2009) used the sensitivity approach, frequency response functions and natural frequencies for model updating in structures, while Wang *et al*. (2009) used the Zernike moment descriptor for finite element model updating. Yuan and Dai (2009) updated a finite element model of damped gyroscopic systems, while Kozak *et al*. (2009) used a miscorrelation index for model updating. Arora *et al*. (2009) proposed a finite element model updating approach that used damping matrices, while Schlune *et al*. (2009) implemented finite element model updating to improve bridge evaluation. Yang *et al.* (2009) investigated several objective functions for finite element model updating in structures, while Bayraktar *et al.* (2009) applied modal properties for updating a finite element model of a bridge. Li and Du (2009) used the most sensitive design variable for finite element model updating of stay-cables, while Steenackers *et al*. (2007) successfully applied transmissibility for finite element model updating. Xu Yuan and Ping Yu (2015) proposed finite element model updating of damped structures, while Shabbir and Omenzetter (2015) applied particle swarm optimisation for finite element model updating. The uncertainty quantification techniques, however, include the uncertainties related to the modelled structure (or systems) during the updating procedure. The uncertainty quantification approaches that treat uncertain parameters as random parameters with joint distribution functions are called the probabilistic techniques and these comprise Bayesian and perturbation methods, whereas the non-probabilistic (possibilistic) approaches use the interval method or membership functions (fuzzy technique) to define uncertain parameters. In this book, only the Bayesian approach is used to update structures.

Other successful implementations of finite element model updating methods include applications in bridges (Huang and Zhu, 2008; Jaishi *et al.*, 2007; Niu *et al*., 2015), composite structures (Pavic *et al*., 2007), helicopters (Shahverdi *et al.*, 2006), atomic force microscopes (Chen, 2006), footbridges (Živanović *et al*., 2007), estimating constituent-level elastic parameters of plates (Mishra and Chakraborty, 2015) and identifying temperature-dependent thermal-structural properties (Sun *et al*., 2015). The process of finite element updating is illustrated in Figure 1.2 (Boulkaibet, 2014).

1.5 Finite Element Model Updating and Bounded Rationality

As illustrated in Figure 1.2, optimisation involves minimising the distance between measurements and the model output, whichever way the model is defined, whether deterministically or probabilistically. The minimisation process gives either a local optimum solution or a global one, and one is never sure whether the solution is global or local, particularly for complex problems. Furthermore, the data to be used should be universally represented, meaning that all the domain representations must be used, and this is not possible. A definition of 'rational solution' implies that the solution is optimised, all information is used and an optimal objective function for optimisation is used. In finite element model updating, this is not possible.

In rational theory, the limited availability of information required in making a rational decision, and the limitations of devices for making sense of incomplete decisions, are covered by

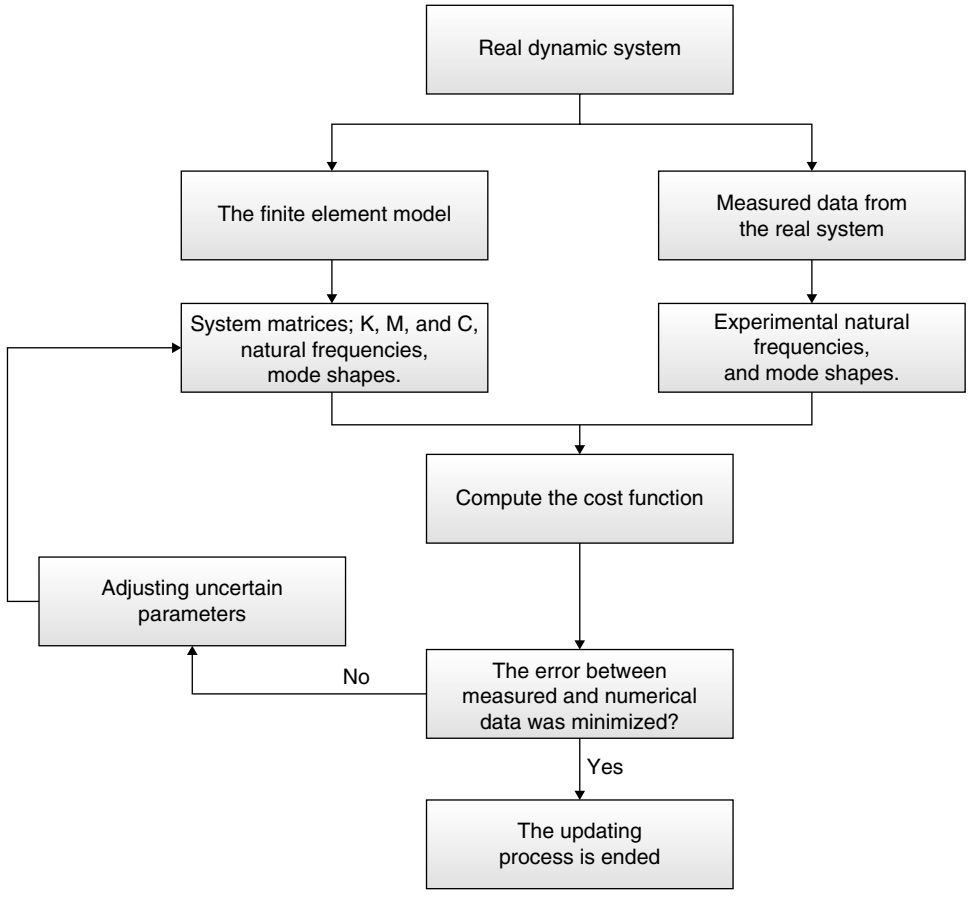

Figure 1.2 Finite element model updating procedure

the theory of bounded rationality, and it was proposed by Herbert Simon (Simon, 1957, 1990, 1991; Tisdell, 1996). The theory of bounded rationality has been used in modelling by researchers such as Lee (2013), Gama (2013), Jiang *et al*. (2013), Stanciu-Viziteu (2012), Aviad and Roy (2012) and Murata *et al*. (2012). Herbert Simon coined the term *satisficing*, by combining the terms 'satisfying' and 'sufficing', which is the concept of making optimised decisions under the limitations that the data used in making such decisions are imperfect and incomplete, while the model used to make such decisions is inconsistent and imperfect. In the same vein, the finite element model updating problem is a satisficing problem, not a process of seeking the correct model.

1.6 Finite Element Model Updating Methods

This section reviews methods that have been used for finite element model updating. They are grouped into classes, and more details on these may be found in Marwala (2010). There are three

categories of finite element model updating techniques: direct methods; iterative methods; and uncertainty quantification methods.

1.6.1 Direct Methods

Direct methods (Friswell and Mottershead, 1995; Marwala, 2010) are one of the earliest strategies used for finite element model updating. They possess the ability to reproduce the exact experimental data and without using iterations, which makes these algorithms computationally efficient. These methods are still used for finite element model updating, and modern instruments and sensors that have lately been used in experiments allow these methods to overcome some of their disadvantages, such as lack of node connectivity and the need for a large amount of data to reproduce the exact experimental matrices. In this subsection, several direct methods – the matrix update methods, the Lagrange multiplier method, the optimal matrix methods and the eigenstructure assignment method – are briefly described.

1.6.1.1 Matrix Update Methods

Matrix update methods operate by modifying structural model matrices, that is, the mass, stiffness and damping matrices (Baruch, 1978). These are obtained by minimising the distance between analytical and measured matrices as follows (Friswell and Mottershead, 1995; Marwala, 2010):

$$\mathbf{E}_i = \left(-\omega_i^2\mathbf{M} + j\omega_i\mathbf{C} + \mathbf{K}\right)\phi_i, \tag{1.1}$$

where $\mathbf{M}$ is the mass matrix, $\mathbf{C}$ is the damping matrix, $\mathbf{K}$ is the stiffness matrix of the structure, $\mathbf{E}_i$ is the error vector (also known as the residual force), $j = \sqrt{-1}$, ω_i is the ith natural frequency and ϕ_i is the ith mode shape. The residual force is the harmonic force with which the unupdated model will have to be excited at a frequency of ω_i so that the structure will respond with the mode shape ϕ_i. The Euclidean norm of $\mathbf{E}_i$ is minimised by updating physical parameters in the model (Ewins, 1995; Marwala, 2010) and choosing an optimisation routine. These techniques are classified as *iterative* since they are employed by iteratively changing the relevant parameters until the error is minimised. Ojalvo and Pilon (1988) minimised the Euclidean norm of the residual force for the ith mode of the structure by using the modal properties. The residual force in the equation of motion in the frequency domain may be minimised as (Friswell and Mottershead, 1995):

$$\mathbf{E} = \left(-\omega^2\mathbf{M} + j\omega\mathbf{C} + \mathbf{K}\right)\mathbf{X}^m - \mathbf{F}^m, \tag{1.2}$$

where $\mathbf{X}^m$ and $\mathbf{F}^m$ are the Fourier-transformed displacement and force matrices, respectively. Each column of the matrix corresponds to a measured frequency point. The Euclidean norm of the error matrix $\mathbf{E}$ is minimised by updating physical parameters in the model. The methods described in this subsection are computationally expensive. In addition, it is challenging to identify a global minimum because of multiple stationary points, which are caused by the

non-unique nature of inverse problems (Janter and Sas, 1990; Mares and Surace, 1996; Friswell *et al*., 1994; Dunn, 1998).

1.6.1.2 Lagrange Multiplier Method

The Lagrange multiplier method is an optimisation technique that deals with the objective function and constraints of an optimisation equation (Rad, 1997). It is implemented by minimising a constrained objective function, where the constraints are imposed by Lagrange multipliers (Marwala, 2010; Minas and Inman, 1988; Heylen and Sas, 1987).

1.6.1.3 Optimal Matrix Methods

Optimal matrix methods employ analytical rather than numerical solutions to obtain matrices from the damaged systems. They are formulated using Lagrange multipliers and perturbation matrices, and the optimisation problem is posed to minimise (Friswell and Mottershead, 1995)

$$E(\Delta\mathbf{M}, \Delta\mathbf{C}, \Delta\mathbf{K}) + \lambda R(\Delta\mathbf{M}, \Delta\mathbf{C}, \Delta\mathbf{K}), \tag{1.3}$$

where E is the objective function, λ is the Lagrange multiplier, R is the constraint of the equation and Δ denotes the perturbation of the system matrices. Different permutations of perturbations are tried until the difference between the finite element model results and the measured results is minimised. Baruch and Bar Itzhack (1978), Berman and Nagy (1983) and Kabe (1985) formulated Equation 1.3by minimising the Frobenius norm of the error, while maintaining the symmetry of the matrices. McGowan *et al*. (1990) introduced an extra constraint that maintained the connectivity of the structure and used measured mode shapes to update the stiffness matrix to locate structural damage. Zimmerman *et al*. (1995) used a partitioning method for matrix perturbations as sums of element or substructural perturbation matrices to reduce the rank of unknown perturbation matrices. Carvalho *et al.* (2007) successfully applied a direct method for model updating with incomplete measured modal data. A limitation of these approaches is that the updated model is physically unrealistic.

1.6.1.4 Eigenstructure Assignment Methods

Eigenstructure assignment approaches are based on control system theory, and the system under consideration is made to respond in a predetermined configuration. An updated finite element model is that with eigenstructure which is obtained from measured data. Zimmerman and Kaouk (1992) applied this approach successfully to update a finite element model of a cantilevered beam based on modal properties, while Schultz *et al*. (1996) updated a finite element model using the measured frequency response functions. The limitation of this technique that the number of sensor locations is less than the degrees of freedom in the finite element model. To deal with this limitation, either the mode shapes and frequency response functions are expanded to the size of the finite element model or the mass and stiffness matrices of the finite element model are reduced to the size of the measured data. The reduction/expansion approaches that are applied are static reduction (Guyan, 1965; Gysin, 1990; Imregun and

Ewins, 1993), dynamic reduction (Paz, 1984), improved reduced systems (O'Callahan, 1989) and the system-equivalent reduction process (O'Callahan *et al*., 1989).

1.6.2 Iterative Methods

Iterative methods (Friswell and Mottershead, 1995; Marwala, 2010) were developed to overcome the weakness of the direct methods and to update finite element models of complex systems. These methods use non-linear equations to deal with the non-convex optimisation problem which arises when a complex system is updated. In these methods, a set of parameters are iteratively adjusted to minimise an objective function (also called a penalty function), where most of the objective functions used in model updating contain only modal and/or response functions data. In this subsection, two popular iterative methods are briefly discussed.

1.6.2.1 Sensitivity Methods

Sensitivity approaches work on the premise that experimentally measured data are perturbations of design data around a finite element model. Therefore, experimentally measured data ought to be approximately equal to data predicted by the finite element model for this approach to work. These approaches uses the derivatives of either the modal properties or the frequency response functions as a basis for finite element model updating. Many procedures have been developed to calculate the derivative of the modal properties and frequency response functions, including Fox and Kapoor (1968), Norris and Meirovitch (1989), Haug and Choi (1984), Chen and Garba (1980) and Adhikari and Friswell (2001). Ben-Haim and Prells (1993) used frequency response function sensitivity to update a finite element model, while Lin *et al*. (1995) used modal sensitivity for finite element model updating and Hemez (1993) used element sensitivity for finite element updating. Alvin (1997) improved the convergence rate by using statistical confidence measurements in finite element model updating.

1.6.2.2 Optimisation Methods

Huang and Zhu (2008) applied optimisation methods for the finite element model updating of bridge structures. The optimisation method was augmented by a sensitivity analysis. Schwarz *et al*. (2007) updated a finite element model which minimised the difference between the modes of a finite element model and those from the experiment. Bakir *et al*. (2007) applied sensitivity approaches for finite element model updating. They used a constrained optimisation method to minimise the differences between the natural frequencies and mode shape.

Jaishi and Ren (2007) applied a multi-objective optimisation approach for finite element model updating. Their multi-objective cost function was based on the differences between eigenvalues and strain energy. Liu *et al*. (2006) updated a finite element model of a 14-bay beam with semi-rigid joints and a boundary using a hybrid optimisation method. Zhang and Huang (2008) applied a gradient descent optimisation method for the finite element model updating of bridge structures. The objective function was formulated as the summation of the frequency difference and modal shapes. Parameter alteration was guided by engineering judgement.

1.6.3 Artificial Intelligence Methods

Finite element modelling updating can be achieved through the use of artificial intelligence techniques. Artificial intelligence techniques are computational tools that are inspired by the way nature and biological systems work. Within the context of finite element model updating, some of the techniques that have been applied are genetic algorithms, particle swarm optimisation, fuzzy logic, neural networks, and support vector machines. A genetic algorithm simulates natural evolution, where the law of the survival of the fittest is applied to a population of individuals. This natural optimisation method is used for optimising a function (Mitchell, 1998). Particle swarm optimisation is an evolutionary optimisation method that was developed by Kennedy and Eberhart (1995), inspired by algorithms that model the flocking behaviour seen in birds. The response surface method is a procedure that functions by generating a response for a given input and then constructs an approximation to a complicated model such as a finite element model (Kamrani *et al.*, 2009).

Finite element models are computationally expensive methods. To manage the computational load, some form of emulator to approximate the finite element model can be implemented. Y. Liu *et al*. (2009) used fuzzy theory, while Jung and Kim (2009) employed the hybrid genetic algorithm for finite element model updating. Tan *et al*. (2009) used support vector machines and wavelet data for finite element model updating in structures, while Zapico *et al.* (2008) applied neural networks. Further successful applications of artificial intelligence methods to finite element model updating include Tu and Lu (2008) and Yan *et al*. (2007), as well as Fei *et al*. (2006) who applied genetic algorithms. Feng *et al*. (2006) applied a hybrid of a genetic algorithm and simulated annealing, and He *et al*. (2008) applied a hybrid of a genetic algorithm and neural networks.

Marwala (2010) used the particle swarm optimisation technique for finite element model updating, and the results were compared to those obtained from the genetic algorithm. Furthermore, simulated annealing was also introduced and applied to finite element model updating, and the results were compared to those from particle swarm optimisation. To deal with the issue of computational efficiency, a response surface method that combines the multi-layer perceptron and particle swarm optimisation was introduced and applied to finite element model updating. The results were compared to those from the genetic algorithm, particle swarm optimisation and simulated annealing.

1.6.4 Uncertainty Quantification Methods

Due to the numerical and experimental uncertainties associated with the updated models, formulating the updating problems as iterative optimisation with constraints may not produce stable and accurate results. Modelling uncertainties are caused by predictions used to model the systems, especially when the physical components used to model the systems are complex and not sufficiently well understood. On the other hand, experimental uncertainties are caused by noise resulting from the measurements or by the variability of the system parameters (Der Kiureghian and Ditlevsen, 2009; Soize, 2010; Walker *et al.*, 2003). In this subsection, the perturbation method, minimum variance method and Bayesian approach are briefly described.

1.6.4.1 Perturbation Method

The perturbation technique uses a Taylor series to extend the terms in model updating equations around a predefined point and then to estimate the mean and variance of the updated parameters (Khodaparast, 2010; Hua *et al.*, 2008; Khodaparast *et al.*, 2008). One type of perturbation technique uses the least-squares method for stochastic finite element model updating, by assuming that the measured data and updating parameters are statistically independent. Another perturbation technique was developed by Hua *et al.* (2008) and assumes that the measured vector $\mathbf{Z}_X$ can be obtained by adding a random component ($\Delta\mathbf{Z}_X$) to a deterministic component (mean value) as follows (Khodaparast, 2010; Hua *et al.*, 2008; Boulkaibet, 2014):

$$\mathbf{Z}_X = \hat{\mathbf{Z}}_X + \Delta\mathbf{Z}_X, \tag{1.4}$$

where the perturbation vector ($\Delta\mathbf{Z}_X$) has zero mean and represents the uncertainty in the measured data. The structural parameters $\boldsymbol{\theta}$, the sensitivity matrix $\mathbf{S}$ and the predictions $\mathbf{Z}$ are defined around the mean value of these vectors and/or matrices as follows (Friswell and Mottershead, 1995; Khodaparast, 2010; Boulkaibet, 2014):

$$\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} + \sum_{i=1}^{n} \frac{\partial \boldsymbol{\theta}}{\partial \Delta \mathbf{Z}_X^i} \Delta \mathbf{Z}_X^i, \tag{1.5}$$

$$\mathbf{S} = \hat{\mathbf{S}} + \sum_{i=1}^{n} \frac{\partial \mathbf{S}}{\partial \Delta \mathbf{Z}_X^i} \Delta \mathbf{Z}_X^i, \tag{1.6}$$

$$\mathbf{Z} = \hat{\mathbf{Z}} + \sum_{i=1}^{n} \frac{\partial \mathbf{Z}}{\partial \Delta \mathbf{Z}_X^i} \Delta \mathbf{Z}_X^i. \tag{1.7}$$

With subscript j denoting iteration number, we obtain (Friswell and Mottershead, 1995; Boulkaibet, 2014):

$$\hat{\mathbf{Z}}_X = \hat{\mathbf{Z}}_j + \hat{\mathbf{S}}_j \left(\hat{\boldsymbol{\theta}}_{j+1} - \hat{\boldsymbol{\theta}}_j \right) \tag{1.8}$$

$$\hat{\mathbf{S}}_j \frac{\partial \boldsymbol{\theta}_{j+1}}{\partial \Delta \mathbf{Z}_X^i} = \hat{\mathbf{S}}_j \frac{\partial \boldsymbol{\theta}_j}{\partial \Delta \mathbf{Z}_X^i} + \left(\boldsymbol{e} - \frac{\partial \mathbf{Z}_j}{\partial \Delta \mathbf{Z}_X^i} - \frac{\partial \mathbf{S}_j}{\partial \Delta \mathbf{Z}_X^i} \right) \left(\hat{\boldsymbol{\theta}}_{j+1} - \hat{\boldsymbol{\theta}}_j \right). \tag{1.9}$$

Here, the vector $\mathbf{e} = [0 \cdots 0\, 1\, 0 \cdots 0]$ has all components equal to zero except for a 1 in ith position. Subtracted from it are (Friswell and Mottershead, 1995):

$$\frac{\partial \mathbf{Z}_j}{\partial \Delta \mathbf{Z}_X^i} = \mathbf{S}_j \frac{\partial \boldsymbol{\theta}_j}{\partial \Delta \mathbf{Z}_X^i}, \tag{1.10}$$

$$\frac{\partial \mathbf{S}_j}{\partial \Delta \mathbf{Z}_X^i} = \sum_{k=1}^{p} \frac{\partial \mathbf{S}_j}{\partial \boldsymbol{\theta}_k} \cdot \frac{\partial \boldsymbol{\theta}_k}{\partial \Delta \mathbf{Z}_X^i}. \tag{1.11}$$

Equation 1.10 defines the approximated mean of the uncertain parameters, while Equation 1.11 defines the covariance matrix that is obtained through (Boulkaibet, 2014):

$$\mathbf{V}_{\boldsymbol{\theta}_j} = \boldsymbol{\Theta}_{j,\Delta\mathbf{Z}_X} \mathbf{V}_{\mathbf{Z}_X} \boldsymbol{\Theta}^T_{j,\Delta\mathbf{Z}_X}, \tag{1.12}$$

where $\mathbf{V}_{\mathbf{Z}_X}$ denotes the covariance of the measured data and

$$\boldsymbol{\Theta}_{j,\Delta\mathbf{Z}_X} = \begin{bmatrix} \dfrac{\partial \boldsymbol{\theta}_j^1}{\partial \Delta \mathbf{Z}_X^1} & \dfrac{\partial \boldsymbol{\theta}_j^1}{\partial \Delta \mathbf{Z}_X^2} & \cdots & \dfrac{\partial \boldsymbol{\theta}_j^1}{\partial \Delta \mathbf{Z}_X^n} \\ \dfrac{\partial \boldsymbol{\theta}_j^2}{\partial \Delta \mathbf{Z}_X^1} & \dfrac{\partial \boldsymbol{\theta}_j^2}{\partial \Delta \mathbf{Z}_X^2} & & \dfrac{\partial \boldsymbol{\theta}_j^2}{\partial \Delta \mathbf{Z}_X^n} \\ \vdots & & \ddots & \vdots \\ \dfrac{\partial \boldsymbol{\theta}_j^p}{\partial \Delta \mathbf{Z}_X^1} & \dfrac{\partial \boldsymbol{\theta}_j^p}{\partial \Delta \mathbf{Z}_X^2} & \cdots & \dfrac{\partial \boldsymbol{\theta}_j^p}{\partial \Delta \mathbf{Z}_X^n} \end{bmatrix}. \tag{1.13}$$

1.6.4.2 Minimum Variance Method

The minimum variance approach is an iterative procedure that takes into account the parameter variability and the uncertainties related to constructing the finite element model (Friswell and Mottershead, 1995; Boulkaibet, 2014). This technique minimises the variance of the uncertain parameters, at each iteration, during the updating process. Suppose $\boldsymbol{\theta}_i$ is the vector of uncertain parameters at the ith iteration of the updating procedure. Then the variance matrix of the parameters at the ith iteration is $E\left(\boldsymbol{\theta}_i \boldsymbol{\theta}_i^T\right) = \mathbf{V}_i$. Subtracting the finite element predicted output $\mathbf{Z}_i$ at the ith iteration from the measurement data $\mathbf{Z}_X$ yields (Friswell and Mottershead, 1995; Boulkaibet, 2014)

$$\delta\mathbf{Z} = \mathbf{Z}_X - \mathbf{Z}_i = \mathbf{S}(\boldsymbol{\theta} - \boldsymbol{\theta}_i). \tag{1.14}$$

Then the approximated uncertain parameter vector at the $(i+1)$th iteration, $\boldsymbol{\theta}_{i+1}$, is written as follows (Friswell and Mottershead, 1995):

$$\boldsymbol{\theta}_{i+1} - \boldsymbol{\theta}_i = \mathbf{T}(\mathbf{Z}_X - \mathbf{Z}_i), \tag{1.15}$$

where $\mathbf{T}$ represents an unknown transformation matrix. The new variance of the estimated parameters, $\boldsymbol{\theta}_{i+1}$, for the $(i+1)$th iteration is given by (Chen, 2001)

$$\mathbf{V}_{i+1} = E\left(\boldsymbol{\theta}_{i+1}\boldsymbol{\theta}_{i+1}^T\right) = \mathbf{V}_i + \left(\mathbf{D}_i - \mathbf{V}_i \mathbf{S}_i^T\right)^T \mathbf{T}^T + \mathbf{T}\left(\mathbf{D}_i^T - V_i.S_i\right) + \mathbf{T}\mathbf{V}_{zi}\mathbf{T}^T, \tag{1.16}$$

where $\mathbf{D}_i = E\left(\boldsymbol{\theta}_i \mathbf{Z}_X^T\right)$ is the correlation between the parameter approximation and the measurement noise. The output error variance is (Chen, 2001)

$$\mathbf{V}_{zi} = \mathbf{S}_i \mathbf{V}_i S_i^T - \mathbf{S}_i \mathbf{D}_i - \mathbf{D}_i^T \mathbf{S}_i^T + \mathbf{V}_e, \tag{1.17}$$

where $\mathbf{V}_e = E\left(\mathbf{Z}_X \mathbf{Z}_X^T\right)$ and the transformation matrix is achieved by minimising the variance at the $(i + 1)$th iteration as follows (Chen, 2001):

$$\mathbf{T} = \left(\mathbf{V}_i \mathbf{S}_i^T - \mathbf{D}_i\right) \mathbf{V}_{zi}^{-1}. \tag{1.18}$$

The updated parameters $\boldsymbol{\theta}_{i+1}$, V_{i+1} and D_{i+1} are obtained as (Chen, 2001):

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i + \left(\mathbf{V}_i \mathbf{S}_i^T - \mathbf{D}_i\right) \mathbf{V}_{zi}^{-1} (\mathbf{Z}_X - \mathbf{Z}_i), \tag{1.19}$$

$$\mathbf{V}_{i+1} = \mathbf{V}_i - \left(\mathbf{V}_i \mathbf{S}_i^T - \mathbf{D}_i\right) \mathbf{V}_{zi}^{-1} \left(\mathbf{V}_i \mathbf{S}_i^T - \mathbf{D}_i\right)^T, \tag{1.20}$$

$$D_{i+1} = \mathbf{D}_i - \left(\mathbf{V}_i \mathbf{S}_i^T - \mathbf{D}_i\right) \mathbf{V}_{zi}^{-1} (\mathbf{S}_i \mathbf{D}_i - \mathbf{V}_e). \tag{1.21}$$

1.6.4.3 Bayesian Approaches

The Bayesian method is a technique based on Bayes' theorem for making statistical inference by using the evidence (observations) to update the probability that a hypothesis is true (Marwala, 2009, 2010). Wong *et al.* (2006) used Bayesian methods to update a bridge model using sensor data, while Marwala and Sibisi (2005) conducted finite element updating in beam structures. Mares *et al.* (2006) used the Monte Carlo method for stochastic model updating, while Lindholm and West (1995) applied a Bayesian parameter approximation for finite element model updating and used this to model experimental dynamic response data. Hemez and Doebling (1999) successfully used a Bayesian approach for finite element model updating and applied this to linear dynamics, while Zheng *et al.* (2009) used a Bayesian approach for finite element model updating of a sky-bridge.

1.7 Bayesian Approach *versus* Maximum Likelihood Method

Finite element model updating is essentially an optimisation problem, where the design variables are the parameters of a finite element model that needs updating. There are two ways to approach this problem: the maximum likelihood technique (also known as the frequentist approach) and the Bayesian approach. The maximum likelihood approach defines an objective function, which is usually some distance between the model and the measured data. Then an optimisation method is applied to identify the optimal design variables. The problem with this approach is that it often overfits the data and it does not offer a probabilistic view of the finite element model updating problem.

Another technique which offers a probabilistic view of the finite element model updating problem is the Bayesian approach. The Bayesian framework is represented mathematically as follows (Bishop, 1995):

$$P(\theta|\mathcal{D}) = \frac{P(\mathcal{D}|\theta)P(\theta)}{P(\mathcal{D})}, \tag{1.22}$$

where $P(\theta)$ is the probability distribution function of the design space in the absence of any data, also called the *prior distribution function* and $\mathcal{D} = (y_1,\ldots,y_N)$ is a matrix containing the data. The expression $P(\theta|\mathcal{D})$ is the *posterior probability distribution function* after the data have been observed, $P(\mathcal{D}|\theta)$ is the likelihood function, and $P(\mathcal{D})$ is the normalisation function (also known as the evidence). This book takes the probability view on finite element model updating.

1.8 Outline of the Book

As stated earlier, finite element models are widely used to model the dynamic behaviour of many systems, including electrical, aerospace and mechanical engineering systems. This book is about probabilistic finite element model updating, which is achieved using Bayesian statistics. The aim of finite element model updating is to ensure that the finite element model better reflects the measured data. The finite element model updating process is limited by the theory of bounded rationality, as the data that could possibly be used for the updating problem are infinite and the number of resulting models that could possibly be identified is infinite because of the infinite starting points in the optimisation of the updating process, and the infinite ways of formulating the updating problem. In this book, the Bayesian framework is employed to estimate the probabilistic finite element models which take into account the uncertainties in the measurements and the modelling procedure. The Bayesian formulation achieves this by setting up the finite element model as a posterior distribution of the model, given the measured data. The data are estimated from the likelihood distribution function, the prior distribution function and the evidence. The finite element model updating posterior distribution function is complex and therefore, even for a fairly simple problem, cannot be estimated analytically. This book describes various sampling techniques based on the Markov chain Monte Carlo (MCMC) method that estimate the posterior probability distribution function of the finite element model updating problem. MCMC is a computational procedure based on the random walk and Markov process. The sampling methods described in this book are slice sampling, nested sampling, the Metropolis–Hastings algorithm, hybrid Monte Carlo (HMC) and shadow hybrid Monte Carlo (SHMC). These sampling methods are applied to estimate the posterior probability distribution function of the finite element model updating problem and are applied to mechanical and aeronautical structures.

This book explains the use of computational statistic techniques in aeronautical and mechanical engineering, a subject that will be of interest and useful to researchers, graduates and postgraduate students.

Chapter 2 discusses model selection in finite element model updating problems. It introduces various methods that can be used to select the best finite element model. A good model satisfies the principle of Occam's razor, which states that the simplest model that describes the observed data is the best one. Furthermore, the chapter studies criteria for model selection: the Akaike information criterion, optimal design, statistical hypothesis testing, Occam's razor, the Bayes factor, structural risk minimisation, cross-validation and the Bayesian information criterion. These techniques are described within the context of finite element model updating. Nested sampling, cross-validation and regularisation techniques are applied for model selection in structures.

Chapter 3 describes Bayesian statistics in structural mechanics. It introduces the concept of Bayesian statistics within the context of structural mechanics. Bayesian statistics basically

states that the probability of an event A happening, given that event B has happened (also called the posterior probability), is equal to the product of the probability of event B happening given that event A has happened (also called the likelihood function), and the probability of event A happening (also called the prior), divided by the probability of event B happening (also called the evidence). A mass and spring system with a single degree of freedom is used to estimate the distribution of the stiffness given the distribution of the measured natural frequency and the mass.

In Chapter 4, the MCMC, which is a statistical procedure for computationally sampling a probability distribution function based on the Markov process, random walk and Monte Carlo simulation, is used for finite element model updating. Two approaches are used to update a finite element model of a mechanical structure: the Metropolis–Hastings approach and slice sampling. Slice sampling is a simple method that offers an adaptive step size, which is automatically adjusted to match the characteristics of the posterior distribution function.

In Chapter 5, Monte Carlo dynamically weighted importance sampling (MCDWIS) is applied for finite element model updating. An aeronautical structure application is presented. The motivation for applying MCDWIS is in the complexity of computing normalising constants in higher-dimensional or multimodal systems. MCDWIS accounts for this intractability by analytically computing importance sampling estimates at each time step of the algorithm, thus removing the need for perfect sampling. In addition, a dynamic weighting step with an adaptive pruned-enriched population control scheme allows for further control over weighted samples and population size. The performance of the MCDWIS simulation is graphically illustrated for all algorithm dependent parameters and shows unbiased, stable sample estimates. MCDWIS is then compared to the Metropolis–Hastings technique.

In Chapter 6, the adaptive Metropolis–Hastings (AMH) algorithm and Bayesian statistics are used for finite element model updating. In the AMH method the Gaussian proposal distribution is adapted using the full information gathered hitherto; because of the adaptive characteristics of the method, this technique is non-Markovian but also possesses full ergodic properties. The AMH method is implemented to update a finite element model of a cantilevered beam, an H-shaped structure, as well as an aircraft structure and the results are compared to the results from the MCDWIS method.

In Chapter 7, HMC and Bayesian finite element model updating is discussed. MCMC basically operates by moving from one state to another, through the random walk procedure, where the transition between one state and another is determined using the Markov chain and the acceptance or rejection of a state is decided using the Metropolis–Hastings method. HMC improves the search by using the gradient information to move from one state to another. In this way, the acceptance rate is greatly improved. Formally, HMC is implemented by calculating the Hamiltonian, which is the sum of the potential energy (position) and the kinetic energy (velocity).

In Chapter 8, a shadow HMC is applied for finite element model updating. To deal with this constraint, that the HMC acceptance rate is influenced by the system size and the time step used to estimate the molecular dynamics trajectory, the SHMC algorithm is used. The SHMC algorithm improves sampling for large system sizes and time steps by sampling from a modified Hamiltonian function instead of the normal Hamiltonian function. The SHMC is implemented to update a finite element model of an aircraft structure.

In Chapter 9, the separable shadow hybrid Monte Carlo (S2HMC) method is implemented for finite element model updating. The HMC method is a powerful sampling method for

solving higher-dimensional complex problems. It uses the molecular dynamics (MD) as a global Monte Carlo move to reach areas of high probability. However, the HMC acceptance rate is sensitive to the system size, as well as the time step used to evaluate the MD trajectory. To overcome this, we propose the use of the S2HMC method. This method generates samples from a separable shadow Hamiltonian. The accuracy and the efficiency of this sampling method are tested on the updating of an aeronautical structure.

In Chapter 10, an evolutionary method for sampling a posterior probability density function for updating finite element models is discussed. The evolutionary sampling algorithm hybridises the concepts of genetic algorithms, simulated annealing and MCMC methods and, accordingly, these techniques are described in this chapter. The evolutionary sampling method uses concepts such as reproduction, mutation and crossover to construct the Markov chain to obtain samples. This method is then tested on the updating of a truss structure.

In Chapter 11 the adaptive hybrid Monte Carlo method is used for finite element model updating. The convergence rate of the HMC algorithm is high compared to the Metropolis–Hastings method because its trajectory is augmented by the derivative of the posterior probability distribution function. Nevertheless, the performance of the HMC method deteriorates when sampling from the posterior probability functions of high dimension and exhibits strong correlations between the uncertain parameters. The adaptive hybrid Monte Carlo approach facilitates efficient sampling from complex posterior distribution functions in high dimensions. The performance of the adaptive hybrid Monte Carlo method is tested for finite element model updating.

In Chapter 12, various issues associated with Bayesian sampling are discussed, including the formulation of the posterior probability distribution function. Sampling methods, nested sampling, Metropolis–Hastings, HMC, SHMC and adaptive hybrid Monte Carlo are discussed and compared and conclusions are drawn. Outstanding issues with regard to the application of Bayesian statistics for finite element model updating are extensively discussed. In particular, reversible jump Monte Carlo, the Dirichlet distribution, the expectation–maximisation algorithm and the distribution of optimal posterior probability models are described and proposed for future studies of finite element model updating.

References

Adhikari S, Friswell MI (2001) Eigenderivative analysis of asymmetric non-conservative systems. *International Journal for Numerical Methods in Engineering* **51**: 709–733.

Alvin KF (1997) Finite element model updating via Bayesian estimation and minimisation of dynamic residuals. *AIAA Journal*, **35**(5): 879–886.

Amini Y, Emdad H, Farid M (2015) Finite element modeling of functionally graded piezoelectric harvesters. *Composite Structures* **129**: 165–176.

Araújo dos Santos JV, Lopes HMR, Vaz M, Mota Soares CM, Mota Soares CA, de Freitas MJM (2006) Damage localization in laminated composite plates using mode shapes measured by pulsed TV holography. *Composite Structures* **76**: 272–281.

Arora V, Singh SP, Kundra TK (2009) Finite element model updating with damping identification. *Journal of Sound and Vibration* **324**: 1111–1123.

Aviad B, Roy G (2012) A decision support method, based on bounded rationality concepts, to reveal feature saliency in clustering problems. *Decision Support Systems* **54**: 292–303.

Babuska I, Banerjee U, Osborn JE (2004) Generalized finite element methods: main ideas, results, and perspective. *International Journal of Computing Methods* **1**: 67–103.

Bakir PG, Reynders E, De Roeck G (2007) Sensitivity-based finite element model updating using constrained optimization with a trust region algorithm. *Journal of Sound and Vibration* **305**: 211–225.

Baran NM (1988) *Finite Element Analysis on Microcomputers*. New York: McGraw-Hill.

Baruch M (1978) Optimisation procedure to correct stiffness and flexibility matrices using vibration data. *American Institute of Aeronautics and Astronautics Journal* **16**: 1208–1210.

Baruch M, Bar Itzhack IY (1978) Optimum weighted orthogonalization of measured modes. *American Institute of Aeronautics and Astronautics Journal* **20**: 1623–1626.

Bayraktar A, Altunişik AC, Sevim B, Türker T (2009) Finite element model updating of Kömürhan highway bridge. *Technical Journal of Turkish Chamber of Civil Engineers* **20**: 4675–4700.

Ben-Haim Y, Prells U (1993) Selective sensitivity in the frequency domain, part I: Theory. *Mechanical Systems and Signal Processing* **7**: 461–475.

Berman A, Nagy EJ (1983) Improvement of large analytical model using test data. *American Institute of Aeronautics and Astronautics Journal* **21**: 1168–1173.

Bishop CM (1995) *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.

Boulkaibet, I. (2014) Finite element model updating using Markov chain Monte Carlo techniques. PhD thesis, University of Johannesburg.

Brincker R, Zhang L, Andersen, P (2001) Modal identification of output-only systems using frequency domain decomposition. *Smart Materials and Structures* **10**: 441–445.

Bürg M, Nazarov M (2015) Goal-oriented adaptive finite element methods for elliptic problems revisited. *Journal of Computational and Applied Mathematics* **287**: 125–147.

Carvalho J, Datta BN, Gupta A, Lagadapati M (2007) A direct method for model updating with incomplete measured data and without spurious modes. *Mechanical Systems and Signal Processing* **21**: 2715–2731.

Cawley P, Adams RD (1979) The location of defects from measurements of natural frequencies. *Journal of Strain Analysis* **14**: 49–57.

Chandrupatla TR, Belegudu AD (2002) *Introduction to Finite Elements in Engineering*. Upper Saddle River, NJ: Prentice Hall.

Chang M, Pakzad SN (2015) Optimal sensor configuration for flexible structures with multi-dimensional mode shapes. *Smart Materials and Structures* **24**, article no. 055012.

Chen, G. (2001) Finite element model validation for structural dynamics. PhD thesis, University of London.

Chen JC, Garba JA (1980) Analytical model improvement using modal test results. *American Institute of Aeronautics and Astronautics Journal* **18**: 684–690.

Chen KN (2006) Model updating and optimum designs for V-shaped atomic force microscope probes. *Engineering Optimization* **38**: 755–770.

Chen S, Fu C, Isam S (2009) Finite element analysis of jointed rock masses reinforced by fully-grouted bolts and shotcrete lining. *International Journal of Rock Mechanics and Mining Sciences* **46**: 19–30.

Cheung SH, Beck JL (2009) Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters. *Journal of Engineering Mechanics ASCE*, **135**(4): 243–255.

Courant R, Robbins H (1941) *What is Mathematics?* New York: Oxford University Press.

D'Ambrogio W, Zobel PB (1994). Damage detection in truss structures using a direct updating technique. Proceedings of the 19th International Seminar for Modal Analysis on Tools for Noise and Vibration Analysis, Leuven: Katholieke Universiteit, Leuven, **2**: 657–667.

Der Kiureghian A, Ditlevsen O (2009), Aleatory or epistemic? Does it matter? *Structural Safety* **31**(2): 105–112.

Dhandole SD, Modak SV (2010) A comparative study of methodologies for vibro-acoustic FE model updating of cavities using simulated data. *International Journal of Mechanics and Materials in Design* **6**(1): 27–43.

Dhandole SD, Modak SV (2011) A constrained optimization based method for acoustic finite element model updating of cavities using pressure response. *Applied Mathematical Modelling* **36**(1): 399–413.

Dunn SA (1998) The use of genetic algorithms and stochastic hill-climbing in dynamic finite-element model identification. *Computers & Structures* **66**: 489–497.

Easley SK, Pal S, Tomaszewski PR, Petrella AJ, Rullkoetter PJ, Laz PJ (2007) Finite element-based probabilistic analysis tool for orthopaedic applications. *Computer Methods and Programs in Biomedicine* **85**: 32–40.

Esfandiari A, Bakhtiari-Nejad F, Rahai A, Sanayei M (2009) Structural model updating using frequency response function and quasi-linear sensitivity equation. *Journal of Sound and Vibration* **326**: 557–573.

Evans CJ, Miller TF (2015) Failure prediction of pressure vessels using finite element analysis. *Journal of Pressure Vessel Technology* **137**, article no. 051206.

Ewins DJ (1995) *Modal Testing: Theory and Practice*. Letchworth: Research Studies Press.
Fang S, Perera R (2009) Power mode shapes for early damage detection in linear structures. *Journal of Sound and Vibration* **324**: 40–56.
Farrar CR, Baker WE, Bell TM *et al.* (1994) Dynamic characteristics and damage detection in the I-40 bridge over the Rio Grande. Report LA-12767-MS, Los Alamos National Laboratory.
Faverjon B, Sinou JJ (2009) Identification of an open crack in a beam using a posteriori error estimator of the frequency response functions with noisy measurements. *European Journal of Mechanics A/Solids* **28**: 75–85.
Fei Q, Li A, Miao C (2006) Dynamic finite element model updating using meta-model and genetic algorithm. *Journal of Southeast University (English Edition)* **22**: 213–217.
Feng FZ, Kim YH, Yang BS (2006) Applications of hybrid optimization techniques for model updating of rotor shafts. *Structural and Multidisciplinary Optimisation* **32**: 65–75.
Fox RL, Kapoor MP (1968) Rates of change of eigenvalues and eigenvectors. *American Institute of Aeronautics and Astronautics Journal* **6**: 2426–2429.
Friswell MI, Mottershead JE (1995) *Finite Element Model Updating in Structural Dynamics*. Norwell, MA: Kluwer Academic.
Friswell MI, Penny JET, Wilson DAL (1994) Using vibration data and statistical measures to locate damage in structures. *Modal Analysis: The International Journal of Analytical and Experimental Modal Analysis* **9**: 239–254.
Fuellekrug U, Boeswald M, Goege D, Govers Y (2008) Measurement of FRFs and modal identification in case of correlated multi-point excitation. *Shock and Vibration*, **15**(3–4): 435–445.
Gama J (2013) Data stream mining: the bounded rationality. *Informatica* **37**: 21–25.
Gao K, Fu S, Gibson RL, Chung ET, Efendiev Y (2015) Generalized multiscale finite-element method (GMsFEM) for elastic wave propagation in heterogeneous, anisotropic media. *Journal of Computational Physics* **295**: 161–188.
Gravenkamp H, Birk C, Song C (2015) Simulation of elastic guided waves interacting with defects in arbitrarily long structures using the scaled boundary finite element method. *Journal of Computational Physics* **295**: 438–455.
Guyan RJ (1965) Reduction of stiffness and mass matrices. *American Institute of Aeronautics and Astronautics Journal* **3**: 380.
Gysin H (1990) Comparison of expansion methods for FE model localization. Proceedings of the 8th International Modal Analysis Conference, Schenectady, NY: Union College, pp. 195–204.
Haldar S, Caputo D, Buesking K, Bruck HA (2015) Flexural behavior of singly curved X-Cor® sandwich composite structures: experiment and finite element modeling. *Composite Structures* **129**: 70–79.
Hao L, Chen Y, Sun Z (2015) The sliding mode control for different shapes and dimensions of IPMC on resisting its creep characteristics. *Smart Materials and Structures* **24**, article no. 045040.
Hastings JK, Juds MA, Brauer JR (1985) Accuracy and economy of finite element magnetic analysis. Proceedings of the 33rd Annual National Relay Conference, Stillwater, OK, pp. 45–50.
Haug EF, Choi KK (1984) Structural design sensitivity with generalized global stiffness and mass matrices. *American Institute of Aeronautics and Astronautics Journal* **22**: 1299–1303.
He HX, Yan WM, Wang Z (2008) Stepwise model updating method based on substructures and GA-ANN. *Engineering Mechanics* **25**: 99–105.
Hemez FM (1993) Theoretical and experimental correlation between finite element models and modal tests in the context of large flexible structures. PhD thesis, University of Colorado, Boulder.
Hemez FM, Doebling SW (1999) Validation of Bayesian finite element model updating for linear dynamics. Proceedings of the International Modal Analysis Conference, Vol. **2**, pp. 1545–1555.
Heylen W, Sas P (1987) Review of model optimisation techniques. Proceedings of the 5th International Modal Analysis Conference, Schenectady, NY: Union College, pp. 1177–1182.
Hlilou A, Ben Naceur I, Saï K, Gérard C, Forest S, Cailletaud G (2009) Generalization of the polycrystalline β-model: finite element assessment and application to softening material behavior. *Computational Materials Science* **45**: 1104–1112.
Hrennikoff A (1941) Solution of problems of elasticity by the frame-work method. *ASME Journal of Applied Mechanics* **8**: A619–A715.
Hua XG, Ni YQ, Chen ZQ, Ko JM (2008) An improved perturbation method for stochastic finite element model updating. *International Journal for Numerical Methods in Engineering* **73**: 1845–1864.
Huang M, Zhu H (2008) Finite element model updating of bridge structures based on sensitivity analysis and optimization algorithm. *Wuhan University Journal of Natural Sciences* **13**: 87–92.
Imregun M, Ewins DJ (1993) An investigation into mode shape expansion techniques. Proceedings of the 11th International Modal Analysis Conference, pp. 168–175.

Imregun M, Visser WJ, Ewins DJ (1995) Finite element model updating using frequency response function data I. Theory and initial investigation. *Mechanical Systems and Signal Processing* **9**: 187–202.

Irons B, Shrive N (1983) *Finite Element Primer*. Chichester: Ellis Horwood.

Jaishi B, Kim HJ, Kim MK, Ren WX, Lee SH (2007) Finite element model updating of concrete-filled steel tubular arch bridge under operational condition using modal flexibility. *Mechanical Systems and Signal Processing* **21**: 2406–2426.

Jaishi B, Ren WX (2007) Finite element model updating based on eigenvalue and strain energy residuals using multiobjective optimization technique. *Mechanical Systems and Signal Processing* **21**: 2295–2317.

Janter T, Sas P (1990) Uniqueness aspects of model-updating procedure. *American Institute of Aeronautics and Astronautics Journal* **28**: 538–543.

Jiang Z, Jiao W, Meng S (2013) Fault diagnosis method of time domain and time-frequency domain based on information fusion. *Applied Mechanics and Materials* **300–301**: 635–639.

Jung DS, Kim CY (2009) FE model updating based on hybrid genetic algorithm and its verification on numerical bridge model. *Structural Engineering and Mechanics* **32**: 667–683.

Jung SN, Dhadwal MK, Kim YW, Kim JH, Riemenschneider J (2015) Cross-sectional constants of composite blades using computed tomography technique and finite element analysis. *Composite Structures* **129**: 132–142.

Kabe AM (1985) Stiffness matrix adjustment using mode data. *American Institute of Aeronautics and Astronautics Journal* **23**: 1431–1436.

Kamrani B, Berbyuk V, Wäppling D, Stickelmann U, Feng X (2009) Optimal robot placement using response surface method. *International Journal of Advanced Manufacturing Technology* **44**: 201–210.

Kennedy JE, Eberhart RC (1995) Particle swarm optimization. Proceedings of the IEEE International Conference on Neural Networks, Piscataway, NJ: IEEE, pp. 1942–1948.

Khodaparast HH (2010) Stochastic finite element model updating and its applications in aeroelasticity. PhD thesis, University of Liverpool.

Khodaparast HH, Mottershead JE, Friswell MI (2008) Perturbation methods for the estimation of parameter variability in stochastic model updating. *Mechanical Systems and Signal Processing* **22**: 1751–1773.

Kim JH, Jeon HS, Lee SW (1992) Application of modal assurance criteria for detecting and locating structural faults. Proceedings of the 10th International Modal Analysis Conference, Schenectady, NY: Union College, pp. 536–540.

Kozak MT, Öztürk M, Özgüven HN (2009) A method in model updating using miscorrelation index sensitivity. *Mechanical Systems and Signal Processing* **23**: 1747–1758.

Lee IH (2013) Speculation under bounded rationality. *Journal of Economic Theory and Econometrics* **24**: 37–53.

Li H, Liu F, Hu SLJ (2008) Employing incomplete complex modes for model updating and damage detection of damped structures. *Science in China, Series E Technological Sciences* **51**: 2254–2268.

Li YQ, Du YL (2009) Dynamic finite element model updating of stay-cable based on the most sensitive design variable. *Journal of Vibration and Shock* **28**: 141–143.

Lieven NAJ, Ewins DJ (1988) Spatial correlation of mode shapes, the co-ordinate modal assurance criterion. Proceedings of the 6th International Modal Analysis Conference, Schenectady, NY: Union College, pp. 690–695.

Lin RM, Lim MK, Du H (1995) Improved inverse eigensensitivity method for structural analytical model updating. *Journal of Vibration and Acoustics* **117**: 192–198.

Lindholm BE, West RL (1995) Updating finite element models with experimental dynamic response data using bayesian parameter estimation. Collection of Technical Papers – AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Vol. **2**, pp. 794–802.

Link RJ, Zimmerman DC (2015) Structural damage diagnosis using frequency response functions and orthogonal matching pursuit: theoretical development. *Structural Control and Health Monitoring* **22**: 889–902.

Liu X, Lieven NAJ, Escamilla-Ambrosio PJ (2009) Frequency response function shape-based methods for structural damage localization. *Mechanical Systems and Signal Processing* **23**: 1243–1259.

Liu Y, Duan Z, Liu H (2006) Updating finite element model of structures with semi-rigid joints and boundary. *Proceedings of SPIE – The International Society for Optical Engineering* **6174 II**, Article No. 61743L.

Liu Y, Duan Z, Liu H (2009) Updating of finite element model in considering mode errors with fuzzy theory. *Key Engineering Materials*, **413–414**: 785–792.

Lyon R (1995) Structural diagnostics using vibration transfer functions. *Sound and Vibration Magazine* **29**: 28–31.

Maia NMM, Silva JMM (1997) *Theoretical and Experimental Modal Analysis*. Letchworth: Research Studies Press.

Mares C, Mottershead JE, Friswell MI (2006) Stochastic model updating, part 1: Theory and simulated example. *Mechanical Systems and Signal Processing* **20**: 1674–1695.

Mares C, Surace C (1996) An application of genetic algorithms to identify damage in elastic structures. *Journal of Sound and Vibration* **195**: 195–215.

Marwala T (2001) Fault identification using neural networks and vibration data. Unpublished PhD thesis, University of Cambridge.

Marwala T (2009) *Computational Intelligence for Missing Data Imputation, Estimation and Management: Knowledge Optimization Techniques*. New York: IGI Global Publications.

Marwala T (2010) *Finite Element Model Updating Using Computational Intelligence Techniques*. London: Springer-Verlag.

Marwala T, Heyns PS (1998) Multiple criterion method for determining structural damage. *American Institute of Aeronautics and Astronautics Journal* **36**: 1494–1501.

Marwala T, Sibisi S (2005) Finite element updating using Bayesian framework and modal properties. *Journal of Aircraft* **42**: 275–278.

McGowan PE, Smith SW, Javeed M (1990) Experiments for locating damage members in a truss structure. Proceedings of the 2nd USAF/NASA Workshop on System Identification and Health Monitoring of Precision Space Structures, pp. 571–615.

Messina A, Jones IA, Williams EJ (1996) Damage detection and localization using natural frequency changes. Proceedings of the 1st International Conference on Identification in Engineering Systems, Swansea, pp. 67–76.

Messina A, Williams EJ, Contursi T (1998) Structural damage detection by a sensitivity and statistical-based method. *Journal of Sound and Vibration* **216**: 791–808.

Miao HY, Larose S, Perron C, Lévesque M (2009) On the potential applications of a 3D random finite element model for the simulation of shot peening. *Advances in Engineering Software* **40**: 1023–1038.

Millar F, Mora D (2015) A finite element method for the buckling problem of simply supported Kirchhoff plates. *Journal of Computational and Applied Mathematics* **286**: 68–78.

Minas C, Inman D (1988) Correcting finite element models with measured modal results using eigenstructure assignment methods. Proceedings of the International Modal Analysis Conference, Schenectady, NY: Union College, pp. 583–587.

Mishra AK, Chakraborty S (2015) Development of a finite element model updating technique for estimation of constituent level elastic parameters of FRP plates. *Applied Mathematics and Computation* **258**: 84–94.

Mitchell M (1998) *An Introduction to Genetic Algorithms (Complex Adaptive Systems)*. Cambridge, MA: MIT Press.

Mottershead JE, Friswell MI (1993) Model updating in structural dynamics: a survey. *Journal of Sound and Vibration* **167**: 347–375.

Mottershead JE, Link M, Friswell MI (2011). The sensitivity method in finite element model updating: a tutorial. *Mechanical Systems and Signal Processing*, **25**(7): 2275–2296.

Mottershead JE, Mares C, Friswell MI, James S (2000) Selection and updating of parameters for an aluminium spaceframe model. *Mechanical Systems and Signal Processing*, **14**(6): 923–944.

Murata A, Kubo S, Hata N (2012) Study on promotion of cooperative behavior in social dilemma situation by introduction of bounded rationality – effects of group heuristics on cooperative behavior. Proceedings of the SICE Annual Conference, Akita, Japan, pp. 261–266.

Ni YQ, Zhou XT, Ko JM (2006) Experimental investigation of seismic damage identification using PCA-compressed frequency response functions and neural networks. *Journal of Sound and Vibration* **290**: 242–263.

Niu J, Zong Z, Chu F (2015) Damage identification method of girder bridges based on finite element model updating and modal strain energy. *Science China Technological Sciences* **58**: 701–711.

Norris MA, Meirovitch L (1989) On the problem of modelling for parameter identification in distributed structures. *International Journal for Numerical Methods in Engineering* **28**: 2451–2463.

O'Callahan JC (1989) A procedure for improved reduced system (IRS) model. Proceedings of the 7th International Modal Analysis Conference, Schenectady, NY: Union College, pp. 17–21.

O'Callahan JC, Avitabile P, Riemer R (1989) System equivalent reduction expansion process. Proceedings of the 7th International Modal Analysis Conference, Schenectady, NY: Union College, pp. 17–21.

Ojalvo IU, Pilon D (1988) Diagnosis for geometrically locating structural mathematical model errors from modal test data. Proceedings of the 29th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, pp. 1174–1186.

Pavic A, Miskovic Z, Reynolds P (2007) Modal testing and finite-element model updating of a lively open-plan composite building floor. *Journal of Structural Engineering* **133**: 550–558.

Paz M (1984) Dynamic condensation. *American Institute of Aeronautics and Astronautics Journal* **22**: 724–727.

Pepper W, Wang X (2007) Application of an H-adaptive finite element model for wind energy assessment in Nevada. *Renewable Energy* **32**: 1705–1722.

Qiao P, Cao M (2008) Waveform fractal dimension for mode shape-based damage identification of beam-type structures. *International Journal of Solids and Structures* **45**: 5946–5961.

Qiao P, Lu K, Lestari W, Wang J (2007) Curvature mode shape-based damage detection in composite laminated plates. *Composite Structures* **80**: 409–428.

Rad SZ (1997) Methods for updating numerical models in structural dynamics. PhD thesis, University of London.

Rahman MM, Ariffin AK, Nor SSM (2009) Development of a finite element model of metal powder compaction process at elevated temperature. *Applied Mathematical Modelling* **33**: 4031–4048.

Sazonov E, Klinkhachorn P (2005) Optimal spatial sampling interval for damage detection by curvature or strain energy mode shapes. *Journal of Sound and Vibration* **285**: 783–801.

Schlune H, Plos M, Gylltoft K (2009) Improved bridge evaluation through finite element model updating using static and dynamic measurements. *Engineering Structures* **31**: 1477–1485.

Schrade D, Mueller R, Xu BX, Gross D (2007) Domain evolution in ferroelectric materials: a continuum phase field model and finite element implementation. *Computer Methods in Applied Mechanics and Engineering* **196**: 4365–4374.

Schultz M, Pai PF, Abdelnaser AS (1996) Frequency response function assignment technique for structural damage identification. Proceedings of the 14th International Modal Analysis Conference, Schenectady, NY: Union College, pp. 105–111.

Schwarz B, Richardson M, Formenti DL (2007) FEA model updating using SDM. *Sound and Vibration Magazine* **41**: 18–23.

Sestieri A, D'Ambrogio W (1989) Why be modal: how to avoid the use of modes in the modification of vibrating systems. Proceedings of the 7th International Modal Analysis Conference, Schenectady, NY: Union College, pp. 25–30.

Shabbir F, Omenzetter P (2015) Particle swarm optimization with sequential niche technique for dynamic finite element model updating. *Computer-Aided Civil and Infrastructure Engineering* **30**: 359–375.

Shahverdi H, Mares C, Wang W, Greaves CH, Mottershead JE (2006) Finite element model updating of large structures by the clustering of parameter sensitivities. *Applied Mechanics and Materials* **5–6**: 85–92.

Shone SP, Mace BR, Waters TP (2009) Locating damage in waveguides from the phase of point frequency response measurements. *Mechanical Systems and Signal Processing* **23**: 405–414.

Simon H (1957) A behavioral model of rational choice. In: *Models of Man, Social, and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: John Wiley & Sons, Inc.

Simon H (1990) A mechanism for social selection and successful altruism. *Science* **250**: 1665–1668.

Simon H (1991) Bounded rationality and organizational learning. *Organization Science* **2**: 125–134.

Soize C (2010) Generalized probabilistic approach of uncertainties in computational dynamics using random matrices and polynomial chaos decompositions. *International Journal for Numerical Methods in Engineering* **81**(8): 939–970.

Solin P, Segeth K, Dolezel I (2004) *Higher-Order Finite Element Methods*. Boca Raton, FL: Chapman & Hall/CRC Press.

Stanciu-Viziteu LD (2012) The shark game: equilibrium with bounded rationality. In *Managing Market Complexity* (ed. Teglio A, Alfarano S, Camacho-Cuena E, Ginés-Vilar M), Lecture Notes in Economics and Mathematical Systems **662**, pp. 103–111. Heidelberg: Springer-Verlag.

Steenackers G, Devriendt C, Guillaume P (2007) On the use of transmissibility measurements for finite element model updating. *Journal of Sound and Vibration* **303**: 707–722.

Sun K, Zhao Y, Hu H (2015) Identification of temperature-dependent thermal-structural properties via finite element model updating and selection. *Mechanical Systems and Signal Processing* **52–53**: 147–161.

Tan D, Qu W, Wang J (2009) The finite element model updating of structure based on wavelet packet analysis and support vector machines. *Journal of Huazhong University of Science and Technology (Natural Sciences Edition)* **37**: 104–107.

Tisdell C (1996) *Bounded Rationality and Economic Evolution: A Contribution to Decision Making, Economics, and Management*. Cheltenham: Brookfield.

Todorovska MI, Trifunac MD (2008) Earthquake damage detection in the Imperial County Services Building III: Analysis of wave travel times via impulse response functions. *Soil Dynamics and Earthquake Engineering* **28**: 387–404.

Tu Z, Lu Y (2008) FE model updating using artificial boundary conditions with genetic algorithms. *Computers and Structures* **86**: 714–727.

Ugryumova D, Pintelon R, Vandersteen G (2015) Frequency response function estimation in the presence of missing output data. *IEEE Transactions on Instrumentation and Measurement* **64**: 541–553.

Valério D, Tejado I (2015) Identifying a non-commensurable fractional transfer function from a frequency response. *Signal Processing* **107**: 254–264.

Vanmaekelbergh D, Van Vugt LK, Bakker HE, Rabouw FT, Nijs BD, Van Dijk-Moes RJA, Van Huis MA, Baesjou PJ, Van Blaaderen A (2015) Shape-dependent multiexciton emission and whispering gallery modes in supraparticles of CdSe/multishell quantum dots. *ACS Nano*: **9**(4),3942–3950.

Wadsworth M, Kyaw ST, Sun W (2015) Finite element modelling of the effect of temperature and neutron dose on the fracture behaviour of nuclear reactor graphite bricks. *Nuclear Engineering and Design* **280**: 1–7.

Walker W, Harremoes P, Rotmans J, van der Sluijs J, van Asselt M, Janssen P and von Krauss MK (2003) Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment* **4**(1): 5–17.

Wang W, Mottershead JE, Mares C (2009) Mode-shape recognition and finite element model updating using the Zernike moment descriptor. *Mechanical Systems and Signal Processing* **23**: 2088–2112.

Wang R, Shang D, Li L, Li C (2008) Fatigue damage model based on the natural frequency changes for spot-welded joints. *International Journal of Fatigue* **30**: 1047–1055.

West WM (1982) Single point random modal test technology application to failure detection. *Shock and Vibration Bulletin* **52(4)**: 25–31.

White L, Deleersnijder E, Legat V (2008) A three-dimensional unstructured mesh finite element shallow-water model, with application to the flows around an island and in a wind-driven, elongated basin. *Ocean Model* **22**: 26–47.

White C, Li HCH, Whittingham B, Herszberg I, Mouritz AP (2009) Damage detection in repairs using frequency response techniques. *Composite Structures* **87**: 175–181.

Wong JM, Mackie K, Stojadinovic B (2006) Bayesian updating of bridge fragility curves using sensor data. Proceedings of the 3rd International Conference on Bridge Maintenance, Safety and Management, London: Taylor & Francis, pp. 613–614.

Xu Yuan Z, Ping Yu K (2015) Finite element model updating of damped structures using vibration test data under base excitation. *Journal of Sound and Vibration* **340**: 303–316.

Yan GR, Duan ZD, Ou JP (2007) Application of genetic algorithm on structural finite element model updating. *Journal of Harbin Institute of Technology* **39**: 181–186.

Yang Z, Wang L, Li B, Liu J (2009) Objective functions and algorithms in structural dynamic finite element model updating. *Chinese Journal of Applied Mechanics* **26**: 288–296.

Yuan Y, Dai H (2009) The direct updating of damping and gyroscopic matrices. *Journal of Computational and Applied Mathematics* **231**: 255–261.

Zapico JL, Gonzlez-Buelga A, Gonzlez MP, Alonso R (2008) Finite element model updating of a small steel frame using neural networks. *Smart Materials and Structures* **17**, article no. 045016.

Zhang L, Tamura Y, Wang T, Sun X (2011) A new broadband modal identification technique with applications. In *Modal Analysis Topics, Volume 3* (ed. Proulx T), pp. 261–271. New York: Springer-Verlag.

Zhang LZ, Huang Q (2008) Updating of bridge finite element model based on optimization design theory. *Journal of Harbin Institute of Technology* **40**: 246–249.

Zhang QH, Teo EC (2008) Finite element application in implant research for treatment of lumbar degenerative disc disease. *Medical Engineering and Physics* **30**: 1246–1256.

Zheng YM, Sun HH, Zhao X, Chen W, Zhang RH, Shen XD (2009) Finite element model updating of a long-span steel skybridge. *Journal of Vibration Engineering* **22**: 105–110.

Zhong S, Oyadiji SO (2008) Analytical predictions of natural frequencies of cracked simply supported beams with a stationary roving mass. *Journal of Sound and Vibration* **311**: 328–352.

Zhong S, Oyadiji SO, Ding K (2008) Response-only method for damage detection of beam-like structures using high accuracy frequencies with auxiliary mass spatial probing. *Journal of Sound and Vibration* **311**: 1075–1099.

Zienkiewicz OC (1986) *The Finite Element Method*. New York: McGraw-Hill.

Zimmerman DC, Kaouk M (1992) Eigenstructure assignment approach for structural damage detection. *American Institute of Aeronautics and Astronautics Journal* **30**: 1848–1855.

Zimmerman DC, Kaouk M, Simmermacher T (1995) Structural damage using frequency response functions. Proceedings of the 13th International Modal Analysis Conference, Bethel, CT: SPIE, pp. 179–184.

Živanović S, Pavic A, Reynolds P (2007) Finite element modelling and updating of a lively footbridge: the complete process. *Journal of Sound and Vibration* **301**: 126–145.

Zuberi MJS, Esat V (2015) Investigating the mechanical properties of single walled carbon nanotube reinforced epoxy composite through finite element modelling. *Composites Part B: Engineering* **71**: 1–9.

2

Model Selection in Finite Element Model Updating

2.1 Introduction

Finite element model updating is a computational technique intended to correct and improve the numerical model to match the dynamic behaviour of real structures. As explained in the Chapter 1, this process does not seek to identify the correct model – such a goal is often difficult to achieve and may even by fruitless – but to seek a satisfying solution. The term satisficing, which is a hybrid of the words 'satisfactory' and 'sufficient', was coined by Herbert Simon to illustrate decision-making with limited data and limited data processing infrastructure (Simon, 1956, 1957, 1991). In this chapter, we seek to identify a satisficing finite element model which better predicts the measured data.

What is this satisficing finite element model? Firstly, the model should accurately predict the measured data. Secondly, it must be the simplest model, in line with the principle of Occam's razor which states that the preferred model is the simplest model which accurately reflects the measured data (Hoffmann *et al*., 1997; Gernert, 2007). Occam's razor was popularised by Sir William Hamilton and originated from Wiliam of Ockham (Soklakov, 2002; Gauch, 2003; Carey and Lewis, 2010). Of importance in Occam's razor is how the complexity of the model is managed in reproducing the observed data. For example, in regression problems where the aim is to estimate free parameters it is generally assumed that a simple model is the one where the free parameters are smooth, that is, have the same order of magnitude.

Many models have been proposed for dealing with model selection and these include using criteria such as the Akaike information criterion (AIC), Bayes factor, Bayesian information criterion (BIC), cross-validation, regularisation, nested sampling and the evidence information criterion. These techniques are reviewed in this chapter. Furthermore, cross-validation, regularisation and nested sampling are applied for finite element model updating model selection

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

(Aho *et al*., 2014; Ando, 2010; Burnham and Anderson, 2002; Chamberlin, 1890; Claeskens and Hjort, 2008; Konishi and Kitagawa, 2008; Lahiri, 2001).

2.2 Model Selection in Finite Element Modelling

In this section, a number of model selection approaches that have been used in finite element model updating are briefly described. These methods are: the AIC, BIC, Bayes factor, deviance information criterion (DIC), cross-validation and regularisation. Moreover, the use of the particle swarm optimisation (PSO) and nested sampling methods in model selection are described.

2.2.1 Akaike Information Criterion

The AIC is a statistical procedure for assessing the quality of a model (Akaike, 1973). The AIC was proposed by a Japanese statistician by the name of Hirotugu Akaike. Hashiyama *et al*. (2014) successfully studied the jackknife bias correction of the AIC for choosing variables in canonical correlation analysis under model misspecification, while Jordanger and Tjøstheim (2014) compared the AIC and cross-validation for model selection of copulas and observed that the two approaches gave similar results. Kawakubo and Kubokawa (2014) proposed the modified AIC to cover both the underspecified and overspecified cases, while Fujikoshi *et al*. (2014) studied the consistency of high-dimensional AIC-type and C_p-type criteria in multivariate linear regression. Nishimura *et al*. (2015) proposed a multivariate control chart based on variable selection and the AIC. The AIC is estimated as follows (Akaike, 1974):

$$AIC = 2k - 2\ln(L), \tag{2.1}$$

where k is the number of free parameters to be estimated, ln(.) is the logarithm and L is the maximised value of the likelihood function, which is a measure of the goodness of fit. The preferred model is the one with the smallest AIC.

2.2.2 Bayesian Information Criterion

The BIC is a model selection concept which is based on Bayesian statistics (Findley, 1991; McQuarrie and Tsai, 1998). It penalises models which overfit the data. It is mathematically expressed as follows (Kass and Wasserman, 1995; Bhat and Kumar, 2010):

$$BIC = k\ln(n) - 2\ln(L), \tag{2.2}$$

where n is the number of observations, k is the number of free parameters to be estimated, $L = P(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M})$ is the maximised value of the likelihood function ($\mathcal{D}$ is the data, $\mathcal{M}$ represents the model and $\boldsymbol{\theta}$ is vector of parameters to be estimated). Zhao *et al*. (2015) successfully applied the BIC and hierarchical Bayesian information criterion (HBIC) for mixture model selection; they found that the HBIC performs considerably better than the BIC and that BIC is prone to underestimation. Luo and Chen (2013) successfully applied the BIC in linear regression models with a number of features. Zou and Chen (2012) successfully applied the BIC for sparse model selection, while Penny (2012) applied the AIC, BIC and free energy to compare

dynamic causal models and observed that free energy selects models the best. Shen and Ghosh (2011) successfully applied the BIC to detect change-points, while Chakrabarti and Ghosh (2011) applied the BIC and AIC and observed that the BIC identifies a correct model better while the AIC better identifies the best model. Wei and Zhou (2010) applied the AIC and BIC for model selection in the joint modelling of paired functional data and observed that the AIC and BIC are computationally fast while giving similar results.

2.2.3 Bayes Factor

The Bayes factor is a Bayesian statistical hypothesis test that is used to compare models $\boldsymbol{\mathcal{M}} = \{\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_k\}$. Suppose that $\mathcal{M}_1$ and $\mathcal{M}_2$ are two different models with two different vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively. The Bayes factor compares models $\mathcal{M}_1$ and $\mathcal{M}_2$ by using the likelihood function as a basis and is written as follows (Goodman, 1999; Kass and Raftery, 1995):

$$\begin{aligned} K &= \frac{P(\mathcal{D}|\mathcal{M}_1)}{P(\mathcal{D}|\mathcal{M}_2)} \\ &= \frac{\int P(\boldsymbol{\theta}_1|\mathcal{M}_1)P(\mathcal{D}|\boldsymbol{\theta}_1,\mathcal{M}_1)d\boldsymbol{\theta}_1}{\int P(\boldsymbol{\theta}_2|\mathcal{M}_2)P(\mathcal{D}|\boldsymbol{\theta}_2,\mathcal{M}_2)d\boldsymbol{\theta}_2} \\ &= \frac{\int P(\boldsymbol{\theta}_1|\mathcal{M}_1)P(\mathcal{D}|\boldsymbol{\theta}_1,\mathcal{M}_1)d\boldsymbol{\theta}_1}{\int P(\boldsymbol{\theta}_2|\mathcal{M}_2)P(\mathcal{D}|\boldsymbol{\theta}_2,\mathcal{M}_2)d\boldsymbol{\theta}_2}. \end{aligned} \tag{2.3}$$

Here K is the Bayes factor, $\mathcal{D}$ is the observed data and $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are model parameters. Mulder (2014) successfully used Bayes factors to test order-constrained hypotheses on correlations, while Cabras *et al.* (2015) successfully applied a minimal training sample structure for Bayes factors in censored data. The Bayes factor is a well-known likelihood-ratio test especially if the maximum likelihood is used to calculate the ratio. Wang *et al.* (2011) applied Bayes factors successfully to deal with hidden population stratification, while Morey *et al*. (2011) successfully applied the Markov chain Monte Carlo (MCMC) technique to estimate Bayes factors and successfully applied this to one-sample and one-way designs. Other applications of the Bayes factor are in assessing handwriting features (Taroni *et al.*, 2014), to study genome-wide association (Zheng *et al*., 2011) and for model selection based on objective probabilities (Pericchi, 2005).

2.2.4 Deviance Information Criterion

The DIC is a generalisation of the AIC and BIC (Ando, 2010). The deviance can be estimated as follows (van der Linde, 2005):

$$Dev(\boldsymbol{\theta}) = -2\ln(P(\mathcal{D}|\boldsymbol{\theta})) + C, \tag{2.4}$$

where $P(\mathcal{D}|\boldsymbol{\theta})$ is the likelihood function, $\mathcal{D}$ is the observed data, $\boldsymbol{\theta}$ is the vector of parameters to be estimated and C is a constant. The expectation of the error, which is a measure of how well a model fits the data, can be written as (van der Linde, 2005)

$$\overline{Dev} = E(Dev(\boldsymbol{\theta})). \tag{2.5}$$

The posterior mean deviance parameter P_D can thus be written as follows (van der Linde, 2005):

$$P_D = \overline{Dev} - Dev(\bar{\boldsymbol{\theta}}), \tag{2.6}$$

where $\bar{\boldsymbol{\theta}}$ is the expectation value of $\boldsymbol{\theta}$. Thus the DIC can thus be written as follows (van der Linde, 2005):

$$DIC = Dev(\bar{\boldsymbol{\theta}}) + 2P_D. \tag{2.7}$$

Shriner and Yi (2009) applied the DIC in Bayesian multiple quantitative trait locus mapping and found that the DIC offers a computationally efficient approach to conducting sensitivity analysis and is applicable in environmental effects, gene–gene interactions and gene–environment interactions. Chang *et al.* (2015) applied the DIC and Bayesian model averaging (BMA) for fish stock assessment and observed that the BMA gave a more accurate estimation of uncertainty in the model than the DIC. Wheeler *et al*. (2010) applied the DIC to assess local model adequacy in Bayesian hierarchical models. Other applications of the DIC include Nasir and Pan (2015) for model discrimination in reliability studies, McGrory and Titterington (2007) for variational approximations in Bayesian model selection for finite mixture distributions and Worden and Hensman (2012) in a class of hysteretic systems.

2.2.5 Particle Swarm Optimisation for Model Selection

Another method which has been applied for model selection is the PSO technique (Mthembu *et al.*, 2011a). PSO is an optimisation practice that is based on the interchange between individual and group intelligence to identify an optimal solution (Kennedy and Eberhart, 1995). It is fundamentally a population-based stochastic search algorithm stimulated by the behaviour of biological objects in nature when they are foraging for resources. Mthembu *et al.* (2011a) proposed the application of PSO to the problem of finite element model selection. This problem arises when a choice of the best model for a system has to be made from a set of competing models, each developed in advance from engineering judgement. Each candidate model is characterised as a particle that shows both individualistic and group behaviour. Each particle moves within the model search space seeking the best solution by updating the parameter values that define it. The most important step in the particle swarm algorithm is the method of representing models which should take into account the number, location and variables of parameters to be updated.

2.2.6 Regularisation

Finite element model updating is essentially an optimisation method where minimising the distance between the model and the measurements is the objective, with the model being variable and the measurements fixed. One method of identifying a suitable model is to use a procedure called regularisation where the optimisation problem is framed such that the simplest model is identified in line with the Occam's razor principle which states that the simplest model should be the one preferred (Tsuruoka *et al*., 2009; Bishop, 2007; Alpaydin, 2004). Regularisation was introduced by Andrey Tikhonov and essentially uses the norm which makes the optimisation process smooth by finding a balance between fitting the model and reducing the norm (Tikhonov and Asenin, 1977). Regularisation was first introduced to finite element model updating by Natke (1992).

There are many regularisation methods, and in this chapter we apply the sum of squares of errors. In this chapter, as in Marwala (2010), the finite element model updating problem is formulated by minimising the distance between *measured* natural frequencies and the *calculated* finite element model natural frequencies and the differences between the measured and calculated mode shapes:

$$E = \|\mathbf{\Omega M \Phi} + \mathbf{K \Phi}\|, \tag{2.8}$$

where $\mathbf{M}$ is the mass matrix, $\mathbf{\Omega}$ is a diagonal matrix of squares of the natural frequencies of each mode, $\mathbf{\Phi}$ is the mode shape matrix, $\mathbf{K}$ is the stiffness matrix and $\| \, . \, \|$ denotes the Euclidean norm. In this chapter, the modal properties, that is, natural frequencies and mode shapes, are used as a basis for finite element model updating using vibration data. In this chapter, it is assumed that the reason why the finite element model is unable to predict the measured data is that the moduli of elasticity are not correct. Therefore, the aim is to correctly identify the moduli of elasticity that would give the updated finite element model. To regularise Equation 2.8, we can rewrite it as

$$E = \| -\mathbf{\Omega M \Phi} + \mathbf{K \Phi}\| + \gamma \|\mathbf{\Lambda}\| \tag{2.9}$$

where γ is a hyperparameter which is chosen arbitrarily and the vector $\mathbf{\Lambda}$ contains all the parameters in the finite element model that are deemed to be inaccurately modelled. Pereira *et al.* (2015) applied Bayesian regularisation to an acoustic problem, while Yang (2015) applied regularisation in stereo correspondence. Ding and Selesnick (2015) applied regularisation to artefact-free wavelet denoising, while Shi and Qi (2015) applied regularisation in kernel-based face hallucination. Afonso and Sanches (2015) applied regularisation to blind inpainting, while Simoes *et al.* (2015) applied regularisation in hyperspectral image super-resolution. Filipovic (2015) applied regularisation in recursive identification of multivariable autoregressive models with exogenous input, while Wingen *et al.* (2015) applied regularisation on soft X-ray imaging.

2.2.7 Cross-Validation

Another method for model identification is cross-validation, which guards the process from overfitting and thus produces a model that is less complex, an ideal that is in line with Occam's razor. There are many types of cross-validation techniques, and these include cross-validation,

leave-*k*-out cross-validation and repeated random subsampling cross-validation (Stone, 1977; Geisser, 1993; Yu and Clarke, 2015). In this subsection we study a multifold cross-validation technique (Devijver and Kittler, 1982; Kohavi, 1995; Wang *et al.*, 2015). The multifold method implemented is shown in Figure 2.1 (Marwala, 2001). Each column in Figure 2.1 shows a partition of the data to be used for finite element model updating and each row represents an updating case. The shaded box for a given updating case is the partition that is left out of the training process for that particular updating stage, while the rest of the boxes in one row are used to update a finite element model.

When the multifold cross-validation method is applied, the data set with N modes is segmented into K partitions. Here it is assumed that N is divisible by K and that $K > 1$. In Figure 2.1, N is equal to 5. For each model updating process, the model is identified with the data from all partitions excluding one and the validation set is the subset that is left out. The partition that is left out is shaded. For example, in updating stage 1 the model is identified using partitions 2–5 and partition 1 is used as a validation set. The procedure is repeated for five stages, by leaving the shaded partition for validation and using the remaining partitions for model identification. It should be noted that the type of multifold cross-validation technique applied in this chapter resets the model once in training case 1. The finite element model attained after training case 1 provides the initial parameters for training case 2, and so on.

Rushing *et al.* (2015) applied a leave-one-out cross-validation for the identification of markers associated with survival, while Gavett *et al.* (2015) applied cross-validation for modelling latent dementia phenotype. Other applications of cross-validation include unsupervised adaptive sign language recognition (Zhou *et al.*, 2015), modelling automobile crash response (Acar, 2015), heart stroke classification (Niranjana Murthy and Meenakshi, 2015) and in metabolomics (Triba *et al.*, 2015).

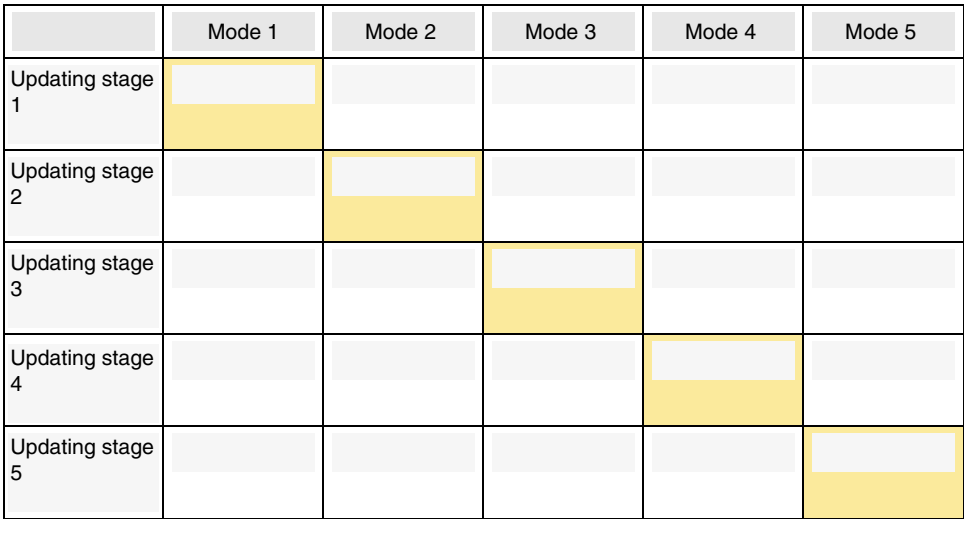

Figure 2.1 The multifold cross-validation technique applied where the model is updating K times, each time leaving out the data indicated by the shaded area and not initialising the finite element model parameter when moving to the next updating stage

2.2.8 Nested Sampling for Model Selection

In this section we describe a finite element model selection procedure based on nested sampling (Skilling, 2006; Feroz and Hobson, 2008). Nested sampling is a sampling technique used for estimating probability distribution functions which was introduced by Skilling (2004) and first introduced into finite element model updating by Mthembu *et al.* (2011b). One way of estimating a finite element model instead of using the maximum likelihood approach which is achieved by minimising Equation 2.8 or 2.9 is to apply a Bayesian approach. Bayes' theorem can be written as follows (Bishop, 2007):

$$P(\mathcal{M}_i|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M}_i)P(\mathcal{M}_i)}{P(\mathcal{D})}, \tag{2.10}$$

where $P(\mathcal{M}_i|\mathcal{D})$ is the posterior probability distribution function of the model $\mathcal{M}_i$ given data $\mathcal{D}$, $P(\mathcal{D}|\mathcal{M}_i)$ is the likelihood function (or evidence), $P(\mathcal{M}_i)$ is the prior distribution function of the hypothesis $\mathcal{M}_i$ and $P(\mathcal{D})$ is the probability of the data. This can be rewritten as follows (Bishop, 2007):

$$P(\mathcal{M}_i|\mathcal{D}) \propto P(\mathcal{D}|\mathcal{M}_i)P(\mathcal{M}_i). \tag{2.11}$$

The first term in Equation 2.11 is the evidence from Equation 2.10. Given that each model is equally likely to fit the data, the evidence term is the determining factor as to which model is the most believable for a particular observed data set. One typical technique for comparing models in Bayesian analysis is the Bayes factor, which can thus be written as follows using the expanded version of the evidence (Mthembu et al., 2011b):

$$K = \frac{P(\mathcal{D}|\mathcal{M}_i)}{P(\mathcal{D}|\mathcal{M}_j)} = \frac{\int P(\mathcal{D}|\boldsymbol{\theta}_i, \mathcal{M}_i)P(\boldsymbol{\theta}_i|\mathcal{M}_i)d\boldsymbol{\theta}_i}{\int P(\mathcal{D}|\boldsymbol{\theta}_j, \mathcal{M}_j)P(\boldsymbol{\theta}_j|\mathcal{M}_i)d\boldsymbol{\theta}_j}. \tag{2.12}$$

By estimating each model's evidence we can compute their ratios and thus execute a model selection process (Beck and Yuen, 2004; Ching and Chen, 2007; Muto and Beck, 2008). Beck and Yuen (2004) assumed that the posterior distribution of the model was Gaussian, and this is only valid for certain types of models (Bishop, 2007). An alternative method proposed by Ching and Chen (2007) approximates the model evidence by sampling the posterior probability distribution of the model by a sequence of non-normalised intermediate probability functions. The shortcoming of this algorithm is that it uses many free parameters. The nested sampling method is simpler and has fewer free parameters. The evidence may be written as follows (Mthembu *et al.*, 2011b):

$$P(\mathcal{D}|\mathcal{M}_i) = \int P(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M}_i)P(\boldsymbol{\theta}|\mathcal{M}_i)d\boldsymbol{\theta}_i. \tag{2.13}$$

$P(\boldsymbol{\theta}|\mathcal{M}_i)$, which is also known as the marginal likelihood, is the predictive probability of the data given the model $\mathcal{M}_i$. This means that $P(\boldsymbol{\theta}|\mathcal{M}_i)$ measures how well the model predicts the

data. The likelihood function $P(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M}_i)$ is a normalised exponential error between measured and real data (the error given in Equation 2.8). The posterior probability will be highest at the most probable parameter set which occupies a small region of the original parameter space. In simple and smaller models, where only few parameters are unknown (with small variations of these parameters), the prior parameter space would almost be fully utilised to fit the data as the posterior probability would occupy a large portion of the prior parameter space, resulting in a correspondingly large evidence value (Mthembu *et al*., 2011a). Complex models have larger parameter spaces for the reason that they have many free parameters that permit them to fit virtually any data, and this frequently results in overfitting. This bigger parameter space is underutilised in explaining the narrow posterior density. The Bayesian approach automatically penalises complex models without the need for a model regularisation term as used in maximum likelihood approaches. The evidence can be rewritten in the following form (Mthembu *et al.*, 2011a):

$$Z = \int L(\boldsymbol{\theta})P(\boldsymbol{\theta})d\boldsymbol{\theta} = \int_0^1 L d\mathbf{X}. \tag{2.14}$$

Calculating this integral analytically is close to impossible because the product of the prior and likelihood is often complicated, particularly when the parameter space is high-dimensional, requiring computing multidimensional likelihoods. The most common method for estimating such integrals is to apply numerical methods such as importance sampling and thermodynamic integration, but these require that the prior parameters are distributed in regions where the likelihood function is not highly concentrated (Murray, 2007). Skilling (2006) proposed a nested sampling method to compute integrals efficiently, and this technique transforms the multidimensional parameter space integral into a one-dimensional problem and then a numerical estimation method approximates the area under the function. In the context of finite element updating, the algorithm achieves the above approximation in the following manner as described by Mthembu *et al.* (2011b):

1. Sample N finite element model updating parameters from the prior probability distribution and calculate their likelihoods.
2. Using the N samples, choose the sample with the lowest likelihood L_i.
3. Increase the evidence by $Z_i = \frac{L_i}{2}(X_{i-1} - X_{i+1})$.
4. Remove the sample with the lowest likelihood and substitute it with a new point from within the residual prior volume $[0, X_i]$, ensuring that the new sample satisfies the constraint new $L_{\text{new}} > L_i$.
5. Repeat steps 2–4 until some stopping criterion such as the desired precision of the evidence is satisfed.

In this chapter three model selection methods are considered: cross-validation, regularisation and nested sampling. To implement cross-validation and regularisation the maximum likelihood technique is used, and the optimisation technique that is used is simulated annealing which is the subject of the next section.

2.3 Simulated Annealing

Simulated annealing is a Monte Carlo method that is used to identify an optimal solution given an objective function. It was inspired by the process of annealing where objects such as metals recrystallise or liquids freeze (Kirkpatrick *et al*., 1983; Marwala, 2010). For example, a metal is heated until it melts and then its temperature is gradually decreased such that at any given time it is nearly in thermodynamic equilibrium. As the temperature of the object is reduced, the system becomes more ordered and approaches a frozen state at $T = 0$. If the cooling process is done ineffectively or the initial temperature of the object is not satisfactorily high, the system may form defects thus freezing out in meta-stable states, and this shows that the system is stuck in a local minimum energy state. Simulated annealing was first applied to optimisation problems by Kirkpatrick *et al*. (1983). Marwala (2010) successfully used simulated annealing for finite element model updating to ensure that the finite element model calculated data do predict the measured data better. Xiao *et al.* (2015) successfully applied a hybrid Lagrangian simulated annealing-based heuristic for scheduling with sequence-dependent set-up times, while García-Villoria *et al*. (2015) applied simulated annealing procedures to an assembly line problem. Palubeckis (2015) applied simulated annealing to a single-row equidistant facility layout, while Matai (2015) applied it in solving a multi-objective facility layout problem. Trivedi *et al.* (2015) applied simulated annealing to improve the static and dynamic travel range of electrostatically actuated microbeams, while Li *et al.* (2015) applied it in atmospheric compensation for free space optical communication. Shokouhifar and Jalali (2015) applied simulated annealing for symbolic simplification of analogue circuits, while Moschakis and Karatza (2015) applied it in scheduling for Internet of Things applications on clouds.

To simulate the annealing process, the procedure proposed by Metropolis *et al.* (1953) needs to be followed. This involves selecting an initial state and temperature and, keeping the temperature constant, disturbing the initial arrangement and calculating the error at the new state. If the new error is lower than the old error then the new state is accepted, otherwise the new state is accepted with a low probability. This is essentially a Monte Carlo technique.

In simulated annealing algorithm the current state is replaced with a random solution that depends on the difference between the matching objective function values and the temperature. The temperature is decreased during the course of the process, and when the temperature approaches zero there are fewer random alterations to the solution. Similar to greedy algorithms, simulated annealing generally moves in the direction of the best solution, except that it has reversal in fitness (the algorithm allows moving to a solution with worse fitness than the current solution). The main advantage of using reversal in fitness is to ensure that the solution is not found at a local optimum, but rather a global optimum. Theoretically, simulated annealing is able to find the global optimum if an infinite amount of time is allowed. The probability of accepting the solution is given by Boltzmann's equation (Bryan *et al*., 2006; Marwala, 2010),

$$P(\Delta E) = \frac{1}{Z}\exp\left(-\frac{\Delta E}{T}\right), \tag{2.15}$$

where ΔE is the energy difference – in this chapter the error between measurements and the results from the finite element model. The state indicates the possible updated finite element model, T is the temperature of the system, and Z is a normalisation factor to ensure that the

probability function integrates to 1 when the bounds approach infinity. The schedule chosen determines the rate at which the temperature decays, and there are many different temperature schedules. Simulated annealing is implemented by choosing certain parameters and within the context of finite element model updating (Marwala, 2010):

- The state space, which is a set of variables, such as moduli of elasticity, that constitutes a candidate updated finite element model.
- The objective function defining the difference between the measured data and the finite element model results. There are many ways in which such an objective function can be chosen, and in this chapter we use Equations 2.8 and 2.9.
- The candidate generator mechanism, which is a random number generator that generates a set of design variables, such as moduli of elasticity.
- The acceptance probability scheme, which is a process through which a set of design variables, such as moduli of elasticity, may be accepted.
- The annealing temperature schedule.

The selection of these parameters is of the utmost importance for the efficacy of the simulated annealing technique as far as identifying an optimal solution is concerned. However, there is no way to select these parameters that is perfect for all problems and there is no methodical practice for optimally selecting these parameters for a given problem. Accordingly, the selection of these parameters is subjective, and trial and error is generally used.

Simulated annealing is applied by a random walk process for a given temperature by moving from one temperature to another. What is known as the transition probability is the probability of transiting from one state to another where a state represents a given finite element model with all the parameters specified. This probability depends on the present temperature, the order of producing the candidate moves and the acceptance probability function. This chapter uses an MCMC technique as a transition mechanism from one state to another. The technique produces a chain of candidate finite element models and accepts or rejects them using the Metropolis algorithm (Metropolis *et al.*, 1953; Meer, 2007).

The MCMC method is a version of the Monte Carlo technique, a computational routine that applies recurrent random sampling to compute results (Mathe and Novak, 2007; Akhmatskaya *et al.*, 2009; Ratick and Schwarz, 2009; Marwala, 2010). Monte Carlo approaches are useful in, for example, solving matrix and integral problems (Lai, 2009), modelling time-dependent radiative transfer with adaptive material coupling (McClarren and Urbatsch, 2009), particle coagulation (Zhao and Zheng, 2009), diffusion problems (Liu *et al.*, 2009), the design of radiation detectors (Dunn and Shultis, 2009), modelling bacterial activities (Oliveira *et al.*, 2009), vehicle detection (Jia and Zhang, 2009), modelling the bystander effect (Xia *et al.*, 2009) and modelling nitrogen absorption (Rahmati and Modarress, 2009).

MCMC is a procedure for generating a chain of states through a random walk process using two major techniques: the Markov process and a Monte Carlo simulation (Liesenfeld and Richard, 2008). It has been widely used in problems such as the tracking of manoeuvering objects (Jing and Vadakkepat, 2010), identifying optimal models (Gallagher *et al.*, 2009), DNA profiling (Curran, 2008), environmental modelling (Gaucherel *et al.*, 2008), medical imaging (Jun *et al.*, 2008), lake water quality modelling (Malve *et al.*, 2007), economics (Jacquier *et al.*, 2007) and statistics (Lombardi, 2007).

To explain the MCMC procedure, a system is considered whose evolution is represented by a stochastic process consisting of random variables $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_i\}$ where a variable $\boldsymbol{x}_i$ inhabits a state $\boldsymbol{x}$ at discrete time *i.* The set of all possible states orientation that all random variables can occupy is called the state space. If the probability that the system occupies state x_{i+1} at time $i+1$ depends entirely on the fact that system was in state $\boldsymbol{x_i}$ at time i (the new state of the system depends only on its previous state), then the random variables $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_\mathrm{i}\}$ form a Markov chain. The transition between states in the MCMC is attained by adding a random component ($\boldsymbol{\varepsilon}$) to the present state as follows (Marwala, 2010):

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \boldsymbol{\varepsilon}. \tag{2.16}$$

When the present state has been realised, it is either accepted or rejected, and one mechanism for achieving this is to use the Metropolis algorithm (Metropolis *et al.*, 1953; Moskovkin and Hou, 2007; Tiana *et al.*, 2007; Bedard, 2008; Sacco *et al.*, 2008; Meyer *et al.*, 2008; Bazavov *et al.*, 2010).

On using the Metropolis algorithm to sample a stochastic process with random variables $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_i\}$, random changes to $\boldsymbol{x}$ are evaluated and either accepted or rejected according to the following criterion (Metropolis *et al.*, 1953):

$$\begin{cases} \text{If } F_{\text{new}} < F_{\text{old}} \text{ accept state } (\boldsymbol{x}_{i+1}) \\ \text{Else} \\ \text{Accept state } (\boldsymbol{x}_{i+1}) \text{ with probability } \exp\{-(F_{\text{new}} - F_{\text{old}})\} \end{cases} \tag{2.17}$$

Here F_{new} and F_{old} are the fitness function (posterior distribution function) values corresponding to $\boldsymbol{x}_{i+1}$ and $\boldsymbol{x}_i$, respectively. A cooling schedule is the procedure by which the temperature T is reduced during simulated annealing (De Vicente *et al.*, 2003). The cooling rate should be satisfactorily low for the probability distribution of the current state to be always approximately equal to the thermodynamic equilibrium (Das and Chakrabarti, 2005). As described by Marwala (2010), the time taken for the equilibrium to be restored, known as the relaxation time, after an alteration in temperature is determined by the shape of the objective function, the current temperature and by the random number generator. The perfect cooling rate is experimentally attained for each problem. In this chapter we use the cooling model given by (Salazar and Toral, 2006),

$$T(i) = \frac{T(i-1)}{1+\sigma}, \tag{2.18}$$

where $T(i)$ is the present temperature, $T(i-1)$ is the previous temperature and σ the cooling rate. The temperature is decreased and the process is repeated until a frozen state is achieved where $T = 0$. In this book, the present state is the current finite element model with all assumed variables, the energy equation is the objective function indicating the difference between the measured data and the finite element model predicted data, and the ground state is the global optimum solution.

2.4 Asymmetrical H-Shaped Structure

This chapter explores model selection using cross-validation, regularisation and the Bayes factor estimated using nested sampling. To test these three methods, we use the asymmetrical H-shaped aluminium structure shown in Appendix A, which was constructed and measured by Marwala (1997). The structure was used so that the initial finite element model gave data that were far from the measured data and thus tested the proposed procedure with a difficult finite element model updating problem. The structure was suspended using elastic bands. It was excited using an electromagnetic shaker, and the response was measured using an accelerometer. The structure was divided into 12 elements. It was excited at a chosen position and the acceleration was measured at 15 positions. The structure was tested freely suspended, and a set of 15 frequency response functions were calculated. A roving accelerometer was used for measuring the response.

The mass of the accelerometer was found to be negligible compared to the mass of the structure. The finite element model was constructed using the Structural Dynamics Toolbox SDT® (Balmes, 1997), with Euler–Bernoulli beam elements (Zienkiewicz, 1971). The finite element model of the structure contained 12 elements. The moduli of elasticity of these elements were used as updating parameters, which were restricted to fall in the interval from 6.00×10^{10} to 8.00×10^{10} N m^{-2}. The cross-validation, regularisation and Bayes factor estimated using nested sampling were implemented. The measured natural frequencies of this structure occurred at 53.9, 117.3, 208.4, 254 and 445 Hz, which correspond to modes 7, 8, 10, 11 and 13, respectively.

2.4.1 Regularisation

We begin by applying regularisation and simulated annealing for finite element model updating. In particular, Equation 2.9, which is a regularised objective function, is applied. When simulated annealing was applied, the scale of the cooling schedule was set to 4, and the number of individual annealing runs was set to 3. The results obtained are shown in Table 2.1. These results show that the error between the measured and the finite element model for the first natural frequency improved from 4.3% to 0.4%, for the second natural frequency from 8.4% to 1.7%, for the third natural frequency from 9.6% to 0.9%, for the fourth natural frequency from 3.4% to 0.7%, and for the fifth natural frequency from 1.6% to 1.7%. The overall average error was 1.1%.

Table 2.1 Finite element model updating of asymmetrical H-shaped structure using simulated annealing and regularised objective function

Measured frequency (Hz)	Initial finite element frequency (Hz)	Final finite element frequency (Hz)
53.9	56.2	54.1
117.3	127.1	119.3
208.4	228.4	210.3
254.8	263.4	252.9
445.1	452.4	437.3

Table 2.2 Finite element model updating of asymmetrical H-shaped structure using simulated annealing and cross-validation

Measured frequency (Hz)	Initial finite element frequency (Hz)	Final finite element frequency (Hz)
53.9	56.2	52.4
117.3	127.1	120.2
208.4	228.4	209.3
254.8	263.4	256.0
445.1	452.4	446.9

2.4.2 Cross-Validation

We now apply cross-validation and simulated annealing for finite element model updating. In particular, Equation 2.8, which is not regularised, is applied. Again, when simulated annealing is applied, the scale of the cooling schedule was set to 4 and the number of individual annealing runs was set to 3. In this example the first updating process naturally excluded the first frequency, then without initialising the finite element model the optimisation process continues by excluding only the second natural frequency, then without initialising the finite element model excluding only the third natural frequency, then without initialising the finite element model excluding only the fourth natural frequency, and then without initialising the finite element model excluding only the fifth natural frequency. The results obtained are shown in Table 2.2. These results show that the error between the measured and the finite element model for first natural frequency improved from 4.3% to 2.8%, and for the second natural frequency from 8.4% to 2.5%, for the third natural frequency from 9.6% to 0.4%, for the fourth natural frequency from 3.4% to 0.5% and for the fifth natural frequency from 1.6% to 0.4%. The overall average error was 1.3%.

2.4.3 Bayes Factor and Nested Sampling

In order to authenticate that model evidence estimation can expose the most probable finite element model(s), four randomly designed models of one beam structure are established and the evidence of each is estimated (Mthembu *et al.*, 2011a). This process accepts that no prior knowledge of which updating parameters should be selected is obtainable. The goal is then to establish from evidence ratios the most probable model from this random set. The finite element model updating parameters were randomly sampled from a Gaussian prior probability distribution with a mean of 7.2×10^{10} Nm^{-2} and an exponential inverse variance of 4.0×10^{-20} for Young's modulus values between 6.8×10^{10} and 8.0×10^{10} $\mathrm{N\,m}^{-2}$ for aluminium. The number of samples, N, was set to 100 (to sample the distribution well) and the sampling algorithm stopping criterion is experimentally set to a maximum of 250 iterations. Several finite element models were constructed, models 1A, 1B, 2A and 2B, the details of which are in Mthembu *et al.* (2011a). It is observed that the models with the fewest updating parameters produced the best evidence, and this is because of their relative simplicity (recall Occam's razor). As shown in Table 2.3, models 1A and 2B are shown to be relatively similar, while there is strong evidence against model 2A from model 1B. So evidence calculation can provide a

Table 2.3 Bayes factors for asymmetrical H-shaped structure

Models	Bayes factor	Evidence
1A/1B	10 122	Very strong
1A/2A	349 760	Very strong
1A/2B	1	Weak
1B/2A	35	Strong

mechanism for eliminating poor models from the outset. It also provides a platform to determine salient parameters to consider in the updating process.

2.5 Conclusion

In this chapter we have explored three model selection procedures: regularisation, cross-validation and the Bayes factor. All these were sampled through the use of nested sampling concepts for the problem of finite element model updating. The regularisation method performed better than the cross-validation technique, while the Bayes factor is used for ensuring that plausible model evidences be calculated before models can be updated.

References

Acar E (2015) Increasing automobile crash response metamodel accuracy through adjusted cross validation error based on outlier analysis. *International Journal of Crashworthiness* **20**: 107–122.

Afonso MV, Sanches JMR (2015) Blind inpainting using ℓ_0 and total variation regularization. *IEEE Transactions on Image Processing* **24**: 2239–2253.

Aho K, Derryberry D, Peterson T (2014) Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* **95**: 631–636.

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. Proceedings of the 2nd International Symposium on Information Theory, Budapest: Akadémiai Kiadó, pp. 267–281.

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**: 716–723.

Akhmatskaya E, Bou-Rabee N, Reich S (2009) A comparison of generalized hybrid Monte Carlo methods with and without momentum flip. *Journal of Computational Physics* **228**: 2256–2265.

Alpaydin E (2004) *Introduction to Machine Learning*. Cambridge, MA: MIT Press.

Ando T (2010) *Bayesian Model Selection and Statistical Modeling*. Boca Raton, FL: CRC Press.

Balmes E (1997) *Structural Dynamics Toolbox User's Manual Version 2.1*. Paris: SDTools.

Bazavov A, Berg BA, Zhou H (2010) Application of biased Metropolis algorithms: from protons to proteins. *Mathematics and Computers in Simulation* **80**: 1056–1067.

Beck JL, Yuen K-V (2004) Model selection using response measurements: Bayesian probabilistic approach. *Journal of Engineering Mechanics* **130**: 192–203.

Bedard M (2008) Optimal acceptance rates for Metropolis algorithms: moving beyond 0.234. *Stochastic Processes and their Applications* **118**: 2198–2222.

Bhat HS, Kumar N (2010) *On the Derivation of the Bayesian Information Criterion*. School of Natural Sciences, University of California.

Bishop CM (2007) *Pattern Recognition and Machine Learning*. New York: Springer.

Bryan K, Cunningham P, Bolshkova N (2006) Application of simulated annealing to the biclustering of gene expression data. *IEEE Transactions on Information Technology in Biomedicine* **10**: 519–525.

Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn. London: Springer-Verlag.

Cabras S, Castellanos ME, Perra S (2015) A new minimal training sample scheme for intrinsic Bayes factors in censored data. *Computational Statistics & Data Analysis* **81**: 52–63.

Carey, TV and Lewis, R. (eds) (2010) Parsimony (in as few words as possible). *Philosophy Now* (UK), issue 81. http://bit.ly/288pmPi, Retrieved 27 October 2012.

Chakrabarti A, Ghosh JK (2011) AIC, BIC and recent advances in model selection. *Handbook of the Philosophy of Science* **7**: 583–605.

Chamberlin TC (1890) The method of multiple working hypotheses. *Science* **15**: 93 (reprinted 1965, *Science* **148**: 754–759).

Chang Y-J, Brodziak J, O'Malley J, Lee H-H, DiNardo G, Sun C-L (2015) Model selection and multi-model inference for Bayesian surplus production models: a case study for pacific blue and striped marlin. *Fisheries Research* **166**: 129–139.

Ching JY, Chen YC (2007) Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. *Journal of Engineering Mechanics – ASCE* **133**: 816–832.

Claeskens G, Hjort NL (2008) *Model Selection and Model Averaging*. Cambridge: Cambridge University Press.

Curran JM (2008) A MCMC method for resolving two person mixtures. *Science and Justice* **48**: 168–177.

Das A, Chakrabarti BK (2005) *Quantum Annealing and Related Optimization Methods*. Lecture Notes in Physics Vol. 679. Heidelberg: Springer.

De Vicente J, Lanchares J, Hermida R (2003) Placement by thermodynamic simulated annealing. *Physics Letters A* **317**: 415–423.

Devijver PA, Kittler J (1982) *Pattern Recognition: A Statistical Approach*. London: Prentice Hall.

Ding Y, Selesnick IW (2015) Artifact-free wavelet denoising: Non-convex sparse regularization, convex optimization. *IEEE Signal Processing Letters* **22**: 1364–1368.

Dunn WL, Shultis JK (2009) Monte Carlo methods for design and analysis of radiation detectors. *Radiation Physics and Chemistry* **78**: 852–858.

Feroz F, Hobson MP (2008) Multimodal nested sampling: an efficient and robust alternative to Markov chain Monte Carlo methods for astronomical data analyses. *Monthly Notices of the Royal Astronomical Society* **384**: 449–463.

Filipovic VZ (2015) Recursive identification of multivariable ARX models in the presence of a priori information: robustness and regularization. *Signal Processing* **116**: 68–77.

Findley DF (1991) Counterexamples to parsimony and BIC. *Annals of the Institute of Statistical Mathematics* **43**: 505–514.

Fujikoshi Y, Sakurai T, Yanagihara H (2014) Consistency of high-dimensional AIC-type and C_p-type criteria in multivariate linear regression. *Journal of Multivariate Analysis* **123**: 184–200.

Gallagher K, Charvin K, Nielsen S, Sambridge M, Stephenson J (2009) Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for earth science problems. *Marine and Petroleum Geology* **26**: 525–535.

García-Villoria A, Corominas A, Pastor R (2015) Heuristics and simulated annealing procedures for the accessibility windows assembly line problem level 1 (AWALBP-L1). *Computers and Operations Research* **62**: 1–11.

Gauch HG (2003) *Scientific Method in Practice*. Cambridge: Cambridge University Press.

Gaucherel C, Campillo F, Misson L, Guiot J, Boreux JJ (2008) Parameterization of a process-based tree-growth model: comparison of optimization. MCMC and particle filtering algorithms. *Environmental Modelling and Software* **23**: 1280–1288.

Gavett BE, Vudy V, Jeffrey M, John SE, Gurnani AS, Adams JW (2015) The δ latent dementia phenotype in the uniform data set: cross-validation and extension. *Neuropsychology* **29**: 344–352.

Geisser S (1993) *Predictive Inference*. New York: Chapman & Hall.

Gernert D (2007) Ockham's razor and its improper use. *Journal of Scientific Exploration* **21**: 135–140.

Goodman S (1999) Toward evidence-based medical statistics 2: The Bayes factor. *Annals of Internal Medicine* **130**: 1005–1013.

Hashiyama Y, Yanagihara H, Fujikoshi Y (2014) Jackknife bias correction of the AIC for selecting variables in canonical correlation analysis under model misspecification. *Linear Algebra and Its Applications* **455**: 82–106.

Hoffmann R, Minkin VI, Carpenter BK (1997) Ockham's razor and chemistry. *Journal for Philosophy of Chemistry* **3**: 3–28.

Jacquier E, Johannes M, Polson N (2007) MCMC maximum likelihood for latent state models. *Journal of Economics* **137**: 615–640.

Jia Y, Zhang C (2009) Front-view vehicle detection by Markov chain Monte Carlo method. *Pattern Recognition* **42**: 313–321.

Jing L, Vadakkepat P (2010) Interacting MCMC particle filter for tracking maneuvering target. *Digital Signal Processing* **20**: 561–574.

Jordanger LA, Tjøstheim D (2014) Model selection of copulas: AIC versus a cross validation copula information criterion. *Statistics & Probability Letters* **92**: 249–255.

Jun SC, George JS, Kim W, Pare-Blagoev J, Plis S, Ranken DM, Schmidt DM (2008) Bayesian brain source imaging based on combined MEG/EEG and fMRI using MCMC. *NeuroImage* **40**: 1581–1594.

Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association* **90**: 791.

Kass RE, Wasserman L (1995) A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association* **90**: 928–934.

Kawakubo Y, Kubokawa T (2014) Modified conditional AIC in linear mixed models. *Journal of Multivariate Analysis* **129**: 44–56.

Kennedy J, Eberhart R (1995) Particle swarm optimization. Proceedings of the IEEE International Conference on Neural Networks, Piscataway, NJ: IEEE, pp. 1942–1948.

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. *Science, New Series* **220**: 671–680.

Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence, San Francisco: Morgan Kaufmann, pp. 1137–1143.

Konishi S, Kitagawa G (2008) *Information Criteria and Statistical Modeling*. London: Springer.

Lahiri P (2001) *Model Selection*. Beachwood, OH: Institute of Mathematical Statistics.

Lai Y (2009) Adaptive Monte Carlo methods for matrix equations with applications. *Journal of Computational and Applied Mathematics* **231**: 705–714.

Li Z, Cao J, Zhao X, Liu W (2015) Atmospheric compensation in free space optical communication with simulated annealing algorithm. *Optics Communications* **338**: 11–21.

Liesenfeld R, Richard J (2008) Improving MCMC, using efficient importance sampling. *Computational Statistics and Data Analysis* **53**: 272–288.

Liu X, Newsome D, Coppens M (2009) Dynamic Monte Carlo simulations of binary self-diffusion in ZSM-5. *Microporous and Mesoporous Materials* **125**: 149–159.

Lombardi MJ (2007) Bayesian inference for α-stable distributions: a random walk MCMC approach. *Computational Statistics and Data Analysis* **51**: 2688–2700.

Luo S, Chen Z (2013) Extended BIC for linear regression models with diverging number of relevant features and high or ultra-high feature spaces. *Journal of Statistical Planning and Inference* **143**: 494–504.

Malve O, Laine M, Haario H, Kirkkala T, Sarvala J (2007) Bayesian modelling of algal mass occurrences – using adaptive MCMC methods with a lake water quality model. *Environmental Modelling and Software* **22**: 966–977.

Marwala T (1997) A multiple criterion updating method for damage detection on structures. MEng thesis, University of Pretoria.

Marwala T (2001) Fault identification using neural networks and vibration data. PhD thesis, University of Cambridge.

Marwala T (2010) *Finite Element Model Updating Using Computational Intelligence Techniques*. Heidelberg: Springer-Verlag.

Matai R (2015) Solving multi objective facility layout problem by modified simulated annealing. *Applied Mathematics and Computation* **261**: 302–311.

Mathe P, Novak E (2007) Simple Monte Carlo and the Metropolis algorithm. *Journal of Complexity* **23**: 673–696.

McClarren RG, Urbatsch TJ (2009) A modified implicit Monte Carlo method for time-dependent radiative transfer with adaptive material coupling. *Journal of Computational Physics* **228**: 5669–5686.

McGrory CA, Titterington DM (2007) Variational approximations in Bayesian model selection for finite mixture distributions. *Computational Statistics & Data Analysis* **51**: 5352–5367.

McQuarrie ADR, Tsai C-L (1998) *Regression and Time Series Model Selection*. River Edge, NJ: World Scientific.

Meer K (2007) Simulated annealing versus Metropolis for a TSP instance. *Information Processing Letters* **104**: 216–219.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**: 1087–1092.

Meyer R, Cai B, Perron F (2008) Adaptive rejection Metropolis sampling using Lagrange interpolation polynomials of degree 2. *Computational Statistics and Data Analysis* **52**: 3408–3423.

Morey RD, Rouder JN, Pratte MS, Speckman PL (2011) Using MCMC chain outputs to efficiently estimate Bayes factors. *Journal of Mathematical Psychology* **55**: 368–378.

Moschakis IA, Karatza HD (2015) Towards scheduling for internet-of-things applications on clouds: a simulated annealing approach. *Concurrency Computation* **27**: 1886–1899.

Moskovkin P, Hou M (2007) Metropolis Monte Carlo predictions of free Co-Pt nanoclusters. *Journal of Alloy and Compounds* **434–435**: 550–554.

Mthembu L, Marwala T, Friswell MI, Adhikari S (2011a) Finite element model selection using particle swarm optimization. *Dynamics of Civil Structures* **4**: 41–52.

Mthembu L, Marwala T, Friswell MI, Adhikari S (2011b) Model selection in finite element model updating using the Bayesian evidence statistic. *Mechanical Systems and Signal Processing* **25**: 2399–2412.

Mulder J (2014) Bayes factors for testing order-constrained hypotheses on correlations. *Journal of Mathematical Psychology* (doi:10.1016/j.jmp.2014.09.004).

Murray I (2007) Advances in Markov Chain Monte Carlo methods. PhD thesis, University College London.

Muto M, Beck LJ (2008) Bayesian updating and model class selection for hysteretic structural models using stochastic simulation. *Journal of Vibration and Control* **14**: 7–34.

Nasir EA, Pan R (2015) Simulation-based Bayesian optimal ALT designs for model discrimination. *Reliability Engineering & System Safety* **134**: 1–9.

Natke HG (1992) On regularisation methods applied to the error localisation of mathematical models. Proceedings of the 9th International Modal Analysis Conference, Schenectady, NY: Union College, pp. 70–73.

Niranjana Murthy HS, Meenakshi M (2015) Comparison between ANN-based heart stroke classifiers using varied folds data set cross-validation. *Advances in Intelligent Systems and Computing* **308**: 693–699.

Nishimura K, Matsuura S, Suzuki H (2015) Multivariate EWMA control chart based on a variable selection using AIC for multivariate statistical process monitoring. *Statistics & Probability Letters* **104**: 7–13.

Oliveira RG, Schneck E, Quinn BE, Konovalov OV, Brandenburg K, Seydel U, Gill T, Hanna CB, Pink DA, Tanaka M (2009) Physical mechanisms of bacterial survival revealed by combined grazing-incidence X-ray scattering and Monte Carlo simulation. *Comptes Rendus Chimie* **12**: 209–217.

Palubeckis G (2015) Fast simulated annealing for single-row equidistant facility layout. *Applied Mathematics and Computation* **263**: 287–301.

Penny WD (2012) Comparing dynamic causal models using AIC, BIC and free energy. *NeuroImage* **59**: 319–330.

Pereira A, Antoni J, Leclère Q (2015) Empirical Bayesian regularization of the inverse acoustic problem. *Applied Acoustics* **97**: 11–29.

Pericchi LR (2005) Model selection and hypothesis testing based on objective probabilities and Bayes factors. In Dey DK, Thrift N (eds) *Handbook of Statistics, Vol. 25: Bayesian Thinking: Modeling and Computation*, Amsterdam: Elsevier, pp. 115–149.

Rahmati M, Modarress H (2009) Nitrogen adsorption on nanoporous zeolites studied by grand canonical Monte Carlo simulation. *Journal of Molecular Structures: THEOCHEM* **901**: 110–116.

Ratick S, Schwarz G (2009) Monte Carlo simulation. In Kitchin R, Thrift N (ed) *International Encyclopedia of Human Geography*, Oxford: Elsevier.

Rushing C, Bulusu A, Hurwitz HI, Nixon AB, Pang H (2015) A leave-one-out cross-validation SAS macro for the identification of markers associated with survival. *Computers in Biology and Medicine* **57**: 123–129.

Sacco WF, Lapa CMF, Pereira CMNA, Filho HA (2008) A Metropolis algorithm applied to a nuclear power plant auxiliary feedwater system surveillance tests policy optimization. *Programming in Nuclear Energy* **50**: 15–21.

Salazar R and Toral R (2006) Simulated annealing using hybrid Monte Carlo. arXiv:cond-mat/9706051.

Shen G, Ghosh JK (2011) Developing a new BIC for detecting change-points. *Journal of Statistical Planning and Inference* **141**: 1436–1447.

Shi J, Qi C (2015) Kernel-based face hallucination via dual regularization priors. *IEEE Signal Processing Letters* **22**: 1189–1193.

Shokouhifar M, Jalali A (2015) An evolutionary-based methodology for symbolic simplification of analog circuits using genetic algorithm and simulated annealing. *Expert Systems with Applications* **42**: 1189–1201.

Shriner D, Yi N (2009) Deviance information criterion (DIC) in Bayesian multiple QTL mapping. *Computational Statistics & Data Analysis* **53**: 1850–1860.

Simoes M, Bioucas-Dias J, Almeida LB, Chanussot J (2015) A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Transactions on Geoscience and Remote Sensing* **53**: 3373–3388.

Simon H (1956). Rational choice and the structure of the environment. *Psychological Review* **63**(2): 129–138.

Simon H (1957) *Models of Man: Social and Rational-Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: John Wiley & Sons, Inc.

Simon H (1991) Bounded rationality and organizational learning. *Organization Science* **2**: 125–134.
Skilling, J. (2004) Nested sampling. *AIP Conference Proceedings*, **735**: 395–405.
Skilling J (2006) Nested sampling for general Bayesian computation. *Bayesian Analysis* **1**: 833–860.
Soklakov AN (2002) Occam's Razor as a formal basis for a physical theory. *Foundations of Physics Letters* **15**: 107–135.
Stone M (1977) Asymptotics for and against cross-validation. *Biometrika* **64**: 29–35.
Taroni F, Marquis R, Schmittbuhl M, Biedermann A, Thiéry A, Bozza S (2014) Bayes factor for investigative assessment of selected handwriting features. *Forensic Science International* **242**: 266–273.
Tiana G, Sutto L, Broglia RA (2007) Use of the Metropolis algorithm to simulate the dynamics of protein chains. *Physica A: Statistical Mechanics and Its Applications* **380**: 241–249.
Tikhonov AN, Asenin VY (1977) *Solutions of Ill-Posed Problems*. New York: John Wiley & Sons, Inc.
Triba MN, Le Moyec L, Amathieu R, Goossens C, Bouchemal N, Nahon P, Rutledge DN, Savarin P (2015) PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the *K*-fold cross-validation quality parameters. *Molecular BioSystems* **11**: 13–19.
Trivedi RR, Bhushan A, Joglekar MM, Pawaskar DN, Shimpi RP (2015) Enhancement of static and dynamic travel range of electrostatically actuated microbeams using hybrid simulated annealing. *International Journal of Mechanical Sciences* **98**: 93–110.
Tsuruoka Y, Tsujii J, Ananiadou S (2009) Stochastic gradient descent training for L1-regularized log-linear models with cumulative penalty. Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, pp. 477–485.
van der Linde A (2005) DIC in variable selection. *Statistica Neerlandica* **59**: 45–56.
Wang L, Li Q, Li Z, Zheng G (2011) Bayes factors in the presence of population stratification. *Statistics & Probability Letters* **81**: 836–841.
Wang Y, Li J, Li Y, Wang R, Yang X (2015) Confidence interval for F1 measure of algorithm performance based on blocked 3×2 cross-validation. *IEEE Transactions on Knowledge and Data Engineering* **27**: 651–659.
Wei J, Zhou L (2010) Model selection using modified AIC and BIC in joint modeling of paired functional data. *Statistics & Probability Letters* **80**: 1918–1924.
Wheeler DC, Hickson DA, Waller LA (2010) Assessing local model adequacy in Bayesian hierarchical models using the partitioned deviance information criterion. *Computational Statistics & Data Analysis* **54**: 1657–1671.
Wingen A, Shafer MW, Unterberg EA, Hill JC, Hillis DL (2015) Regularization of soft-X-ray imaging in the DIII-D tokamak. *Journal of Computational Physics* **289**: 83–95.
Worden K, Hensman JJ (2012) Parameter estimation and model selection for a class of hysteretic systems using Bayesian inference. *Mechanical Systems and Signal Processing* **32**: 153–169.
Xia J, Liu L, Xue J, Wang Y, Wu L (2009) Modeling of radiation-induced bystander effect using Monte Carlo methods. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interaction with Materials and Atoms* **267**: 1015–1018.
Xiao J, Yang H, Zhang C, Zheng L, Gupta JND (2015) A hybrid Lagrangian-simulated annealing-based heuristic for the parallel-machine capacitated lot-sizing and scheduling problem with sequence-dependent setup times. *Computers and Operations Research* **63**: 72–82.
Yang Q (2015) Local smoothness enforced cost volume regularization for fast stereo correspondence. *IEEE Signal Processing Letters* **22**: 1429–1433.
Yu C-W, Clarke B (2015) Regular, median and Huber cross-validation: a computational comparison. *Statistical Analysis and Data Mining* **8**: 14–33.
Zhao J, Jin L, Shi L (2015) Mixture model selection via hierarchical BIC. *Computational Statistics & Data Analysis* **88**: 139–153.
Zhao H, Zheng C (2009) Correcting the multi-Monte Carlo method for particle coagulation. *Powder Technology* **193**: 120–123.
Zheng G, Yuan A, Jeffries N (2011) Hybrid Bayes factors for genome-wide association studies when a robust test is used. *Computational Statistics & Data Analysis* **55**: 2698–2711.
Zhou Y, Yang X, Zhang Y, Xu X, Wang Y, Chai X, Lin W (2015) Unsupervised adaptive sign language recognition based on hypothesis comparison guided cross validation and linguistic prior filtering. *Neurocomputing* **149**: 1604–1612.
Zienkiewicz OC (1971) *The Finite Element Method in Engineering Science*. London: McGraw-Hill.
Zou C, Chen X (2012) On the consistency of coordinate-independent sparse estimation with BIC. *Journal of Multivariate Analysis* **112**: 248–255.

3

Bayesian Statistics in Structural Dynamics

3.1 Introduction

Uncertainty in structural mechanics is an important subject for constructing models that are essential for the design and control of complex structures. Uncertainty has been studied widely by numerous researchers. DiazDelaO *et al.* (2013) applied Gaussian processes to study the stochastic characteristics of structural dynamics. They formulated the frequency response function and applied Bayesian statistics to quantify uncertainty of the frequency response of a non-proportionally damped carbon fibre/epoxy composite plate. Pascual and Adhikari (2012a) used random matrix theory and polynomial chaos expansion to develop a parametric–non-parametric uncertainty quantification method. The parametric technique was implemented through the Karhunen–Loève expansion (Karhunen, 1947; Loève, 1955), while the non-parametric technique was implemented using the Wishart random matrix (Wishart, 1928). The authors were able to derive analytical equations for the first two moments. Chowdhury and Adhikari (2012) applied fuzzy logic for parametric quantification of uncertainty in linear systems. Fuzzy logic is a technique for modelling uncertainty, particularly for situations where the availability of data is limited (Zadeh, 1965; Pelletier, 2000). This technique was then applied for modal analysis of an aircraft wing, and the results compared favourably to the results from Monte Carlo simulation using random variables to simulate the dynamics of the system (Del Moral, 2013). Pascual and Adhikari (2012b) applied the hybrid perturbation–polynomial chaos method, which used the Rayleigh quotient, the power method, the inverse power method and the eigenvalue equation in random algebraic eigenvalue problems of an Euler–Bernoulli beam and a plate (Horn and Johnson, 1985). Murugan *et al.* (2012) studied the effects of uncertain material properties on the aeroelastic response predictions of a helicopter rotor, while DiazDelaO and

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

Table 3.1 Types of uncertainty problems (Adhikari, 2015)

Input	System	Output	Problem name
Deterministically known	Incorrect deterministic model	Deterministically known	Model updating
Deterministically known	Unknown	Deterministically known	System identification
Known	Partially known	Known	Structural health monitoring (SHM)
Random and known	Incompletely known	Random and known	Probabilistic updating

Adhikari (2012) applied Bayesian statistics which was implemented using a Gaussian distribution for finite element modelling. Other works studying uncertainty modelling in structural dynamics include a random matrix model (Adhikari *et al.*, 2012), a reduced spectral approach (Adhikari, 2011), correlation functions in reliability analysis (Chowdhury and Adhikari, 2011), a joint diagonalisation method (Li *et al.*, 2010), sensitivity analysis (Adhikari, 2010), as well as matrix variate distributions (Adhikari, 2007).

There are several types of uncertainty problems, and some of these are shown in Table 3.1 and Figure 3.1 (Adhikari, 2015).

There are several sources of these uncertainties. These include parameter uncertainty, modelling inadequacy, experimental error, computational uncertainty and model uncertainty (Kennedy and O'Hagan, 2001; Matthies, 2007; Lee and Chen, 2009; Adhikari, 2015):

1. *Parameter uncertainty*: when finite element models are created, there is a need to specify certain parameters such as density, modulus of elasticity, and Poisson's ratios, and these parameters are almost uncertain.
2. *Parametric uncertainty*: variability of input variables of a model such as the shape and size of the model.
3. *Modelling uncertainty*: this is due to the fact that there are many approaches that can be used. For example, the choice of elements in finite element modelling is not an exact science, but an approximation.
4. *Experimental error*: for example, if one were to measure the natural frequency of a beam, then one would realise that each measurement is not an exact replica of the previous measurements because of variation in measurement conditions.
5. *Algorithmic error*: this is due to the fact that, on constructing the model, there are many numerical considerations to be taken into account, such as matrix manipulations and optimisation computations.
6. *Interpolation uncertainty*: this is due to the fact that some of the data needed for modelling may be missing.

There are two kinds of uncertainty quantification problem: the forward problem and the inverse problem. The forward problem is where, given the input, the model and the output, and given that the input is uncertain, one studies how this uncertainty is propagated to the output. Theinverse uncertainty is where, given the input and the output, which are both uncertain, one needs to identify an appropriate model and its uncertainty characteristics that map the input

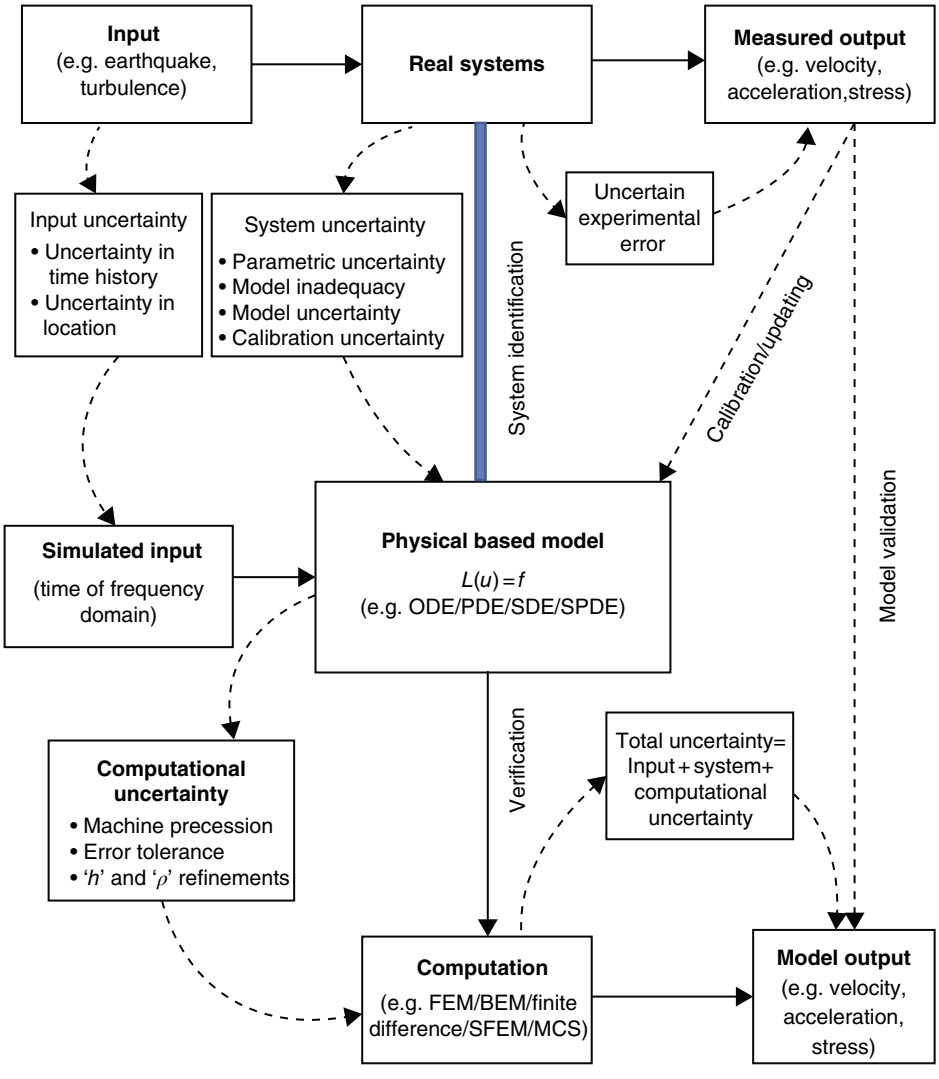

Figure 3.1 How to handle uncertainty (Adhikari, 2015)

to the output. There are a number of models that have been proposed to quantify uncertainty in the forward and inverse problems:

1. *Forward problem*: simulation methods such as Monte Carlo simulation (Rosenbluth and Rosenbluth, 1955); expansion techniques such as the perturbation method and Taylor series (Hazewinkel, 2001; Martínez-Carranza *et al*., 2012); the methods based on the most probable point (or design point) such as reliability techniques; as well as numerical methods such as model reduction techniques (Guyan, 1965; Roweis and Saul, 2000).
2. *Inverse problems*: maximum likelihood or frequentist method (Charnes *et al*., 1976) and the Bayesian approach (Kennedy and O'Hagan, 2000).

This chapter is concerned with the application of Bayesian statistics in applied mechanics (Giagopoulos *et al.*, 2009). In particular, it begins by introducing Bayes' rule and the practical methods that can be used to approximate Bayesian distributions. In particular, a single-degree-of-freedom system is used to illustrate these approaches. Papadimitriou and Papadioti (2013) proposed a fast computing method based Bayesian statistics for uncertainty quantification in structural dynamics, while Giagopoulos *et al.* (2013) used Bayesian statistics to quantify uncertainty and propagation in non-linear structural dynamics. Soize (2013) applied Bayesian posteriors to quantify uncertainty in structural dynamics for low- and medium-frequency ranges, while Yuen (2010) applied Bayesian statistics for structural dynamics and Calanni *et al.* (2007) applied Bayesian neural networks for quantifying uncertainty in structural dynamics. The next chapter studies the dynamics of a simple one-degree-of-freedom structure. At the end of this chapter, a simple automobile suspension system is used to compare some of the methods that have been discussed in this chapter.

3.2 Bayes' Rule

Bayes' theorem offers the ability to identify uncertain parameters (Bishop, 1995; Marwala, 2009) in the finite element model updating problem. Bayesian approaches are governed by the following rule (Boulkaibet *et al.*, 2012, 2015):

$$P(\boldsymbol{\theta}|\mathcal{D},\boldsymbol{\mathcal{M}}) = \frac{P(\mathcal{D}|\boldsymbol{\theta},\boldsymbol{\mathcal{M}})P(\boldsymbol{\theta}|\boldsymbol{\mathcal{M}})}{P(\mathcal{D}|\boldsymbol{\mathcal{M}})}, \tag{3.1}$$

where $\boldsymbol{\mathcal{M}}$ is the model class for the target system, and each model class is defined by the updated parameters of the model, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^D$. Note that different vectors $\boldsymbol{\theta}$ represent a different model classes $\boldsymbol{\mathcal{M}}$. $\mathcal{D}$ represents the modal properties obtained from experiments (natural frequencies f_i^m and mode shapes ϕ_i^m). $P(\boldsymbol{\theta}|\boldsymbol{\mathcal{M}})$ is the prior probability density function (PDF) of the updating parameters when the model class $\boldsymbol{\mathcal{M}}$ is known and the data $\mathcal{D}$ are absent. $P(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\mathcal{M}})$ is the posterior PDF of the updated parameters in the presence of the data $\mathcal{D}$ and the model class $\boldsymbol{\mathcal{M}}$. $P(\mathcal{D}|\boldsymbol{\theta}, \boldsymbol{\mathcal{M}})$ is the likelihood function of the data $\mathcal{D}$ in the presence of the uncertain parameters $\boldsymbol{\theta}$ and the model class $\boldsymbol{\mathcal{M}}$. $P(\mathcal{D}|\boldsymbol{\mathcal{M}})$ is a normalising factor in the presence of the model class. Since only one model class is updated, the dependence on the model class $\boldsymbol{\mathcal{M}}$ is omitted to simplify the notation.

Equation 3.1 can be simplified further to be represented by the prior density and the likelihood functions. This is always true when the measured data $\mathcal{D}$ are considered to be a constant, where the marginal distribution of the data $\mathcal{D}$ does not depend on the model parameters $\boldsymbol{\theta}$. Equation 3.1 is then written as

$$P(\boldsymbol{\theta}|\mathcal{D}) \propto P(\mathcal{D}|\boldsymbol{\theta})P(\boldsymbol{\theta}). \tag{3.2}$$

The posterior expectation of a function $f(\boldsymbol{\theta})$ is given by

$$E[f(\boldsymbol{\theta})|\mathcal{D}] = \frac{\int f(\boldsymbol{\theta})P(\mathcal{D}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathcal{D})}. \tag{3.3}$$

Equation 3.3 can be used to obtain several properties of the posterior inference such as the mean value of the updated parameters ($f(\boldsymbol{\theta}) = \boldsymbol{\theta}$). However, this integral depends on the posterior PDF and the analytical solution of this integral might be unavailable for complex systems with large sizes.

To solve the integral in Equation 3.3, several alternative approaches may be used:

- *Numerical evaluations*, which may not be accurate when the size of posterior distribution function is large.
- *Analytic approximations* such as maximum likelihood (ML), maximum a posteriori (MAP) and Laplace approximation. These approaches are sometimes appropriate; however, they converge to a local minimum.
- *Sampling techniques* such as the Markov chain Monte Carlo (MCMC) methods are the most popular methods that have been used to solve complex posterior distributions.

In the next sections, several analytic approximations that have been used to approximate the posterior function are presented.

3.3 Maximum Likelihood Method

The ML method is the most popular method used to approximate the distribution functions. The approximation obtained by solving the maximum likelihood method represents a local solution, while the results obtained by approximating the posterior distribution function are a global solution. However, when the likelihood function is well defined, the results obtained by ML are very close to the results obtained by solving the posterior function. The ML method can be summarised as

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \left(\log\left(P(\mathcal{D}|\boldsymbol{\theta}, \boldsymbol{\mathcal{M}})\right)\right). \tag{3.4}$$

This ML method can be seen as an optimisation problem where the objective is to minimise the difference between the measured frequency and the analytical natural frequency. The likelihood function is then defined as the error between measurements and analytical frequencies, while the estimated solution is obtained by maximising $\log(P(\mathcal{D}|\boldsymbol{\theta}, \boldsymbol{\mathcal{M}})$.

3.4 Maximum a Posteriori Parameter Estimates

The MAP parameter estimate is the simplest approximation to the posterior density distribution function. This method is similar to the maximum likelihood method, but contains information from the prior density function. The idea is then to estimate the updated parameters as follows:

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \left(\log\left(P(\mathcal{D}|\boldsymbol{\theta}, \boldsymbol{\mathcal{M}})P(\boldsymbol{\theta}|\boldsymbol{\mathcal{M}})\right)\right). \tag{3.5}$$

Solving both ML and MAP methods might be hard when the updated system is complex, and iterative optimisation methods will be needed to obtain $\boldsymbol{\theta}^*$.

3.5 Laplace's Method

The Laplace approximation method (Kass and Raftery, 1995) creates a local Gaussian approximation around the estimated $\boldsymbol{\theta}^*$ (obtained by the MAP estimation method). To create a local Gaussian approximation, first we define the logarithm of the posterior distribution function:

$$T(\boldsymbol{\theta}) = \log(P(\mathcal{D}|\boldsymbol{\theta}, \boldsymbol{\mathcal{M}})P(\boldsymbol{\theta}|\boldsymbol{\mathcal{M}})). \tag{3.6}$$

Then Taylor series are used to expend the function $T(\boldsymbol{\theta})$ to second order:

$$T(\boldsymbol{\theta}) \approx T(\boldsymbol{\theta}^*) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\boldsymbol{H}(\boldsymbol{\theta}^*)(\boldsymbol{\theta}-\boldsymbol{\theta}^*), \tag{3.7}$$

where $\boldsymbol{H}(\boldsymbol{\theta}^*)$ is the Hessian of the log posterior function,

$$\boldsymbol{H}(\boldsymbol{\theta}^*) = \left.\frac{\partial^2 \log(P(\boldsymbol{\theta}|\mathcal{D}.\boldsymbol{\mathcal{M}}))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = \left.\frac{\partial^2 T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}. \tag{3.8}$$

Then the approximation of the posterior density function is defined by the following Gaussian distribution:

$$\begin{aligned} P(\boldsymbol{\theta}|\mathcal{D},\boldsymbol{\mathcal{M}}) &\approx \exp\left(T(\boldsymbol{\theta}^*) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\boldsymbol{H}(\boldsymbol{\theta}^*)(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\right) \\ &\approx M(\boldsymbol{\theta}^*)\exp\left(\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\boldsymbol{H}(\boldsymbol{\theta}^*)(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\right) \end{aligned}, \tag{3.9}$$

where $M(\boldsymbol{\theta}^*) - \exp(T(\boldsymbol{\theta}^*))$ and the Hessian matrix can be seen as the inverse of the variance matrix of the approximate distribution.

It is very important to choose prior, likelihood and posterior functions of the correct form, where this can simplify the updating process. In the next section, a simple dynamic example will be studied, and the prior, likelihood and posterior functions will be provided.

3.6 Prior, Likelihood and Posterior Function of a Simple Dynamic Example

As indicated above, finite element models have been applied to aerospace, electrical, civil and mechanical engineering in designing and developing products such as aircraft wings and turbo-machinery. Finite element models represent a system using mass and stiffness matrices, with some assumption about the damping characteristics of the system being modelled. This section studies the system illustrated in Figure 3.2. A brick with mass M is attached to a wall by a spring with modulus of elasticity K. The brick is excited with a force f and moves with a distance x.

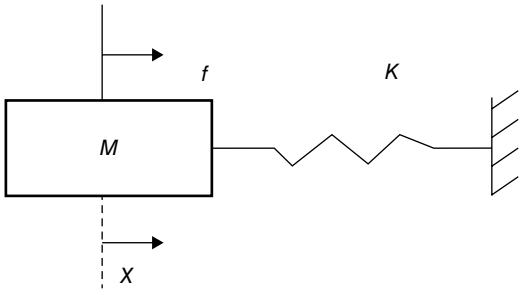

Figure 3.2 A single-degree-of-freedom system

Using Newton's second law of motion, we can express the relationship between the forces acting on the brick and relate these to the acceleration as

$$M\ddot{x} = \sum f, \tag{3.10}$$

where $\ddot{x}$ is the acceleration and F is the force. Equation 3.10 can then be rewritten as

$$M\ddot{x} = -Kx + f \Rightarrow M\ddot{x} + Kx = f. \tag{3.11}$$

Equation 3.11 is in the time domain because the acceleration $\ddot{x}$ and the displacement x are a function of time. This equation can be transformed into the modal domain, that is, mode shape and natural frequency, using a technique called modal analysis, and then Equation 3.11 is rewritten as follows (Ewins, 2000; He and Fu, 2001):

$$-\omega^2 M\phi + K\phi = 0, \tag{3.12}$$

where ω is the natural frequency and ϕ is the mode shape of the system. The natural frequency can thus be rewritten as follows:

$$\omega = \sqrt{\frac{K}{M}}. \tag{3.13}$$

The aim of this chapter is study the inverse quantification of uncertainty modelling, and we begin this by studying a technique called the frequentist or maximum likelihood approach to system identification.

The problem of identifying K in the example in Figure 3.2 may be posed as follows by setting $\boldsymbol{\theta} = K$ and $\mathcal{D} = \omega$ (the measured frequencies in this case are are just scalars). The posterior distribution function of the generalization of the n-degrees-of-freedom undamped dynamic system is given by

$$p(\mathbf{K}|\boldsymbol{\omega}) = \frac{p(\boldsymbol{\omega}|\mathbf{K})p(\mathbf{K})}{p(\boldsymbol{\omega})}. \tag{3.14}$$

for $n \times n$ mass and stiffness matrices $\mathbf{M}$ and $\mathbf{K}$. Here the eigenvectors have the following orthogonality properties: $\boldsymbol{\phi}_i^T \mathbf{M} \boldsymbol{\phi}_j = 0 \quad (i \neq j)$, $\boldsymbol{\phi}_i^T \mathbf{M} \boldsymbol{\phi}_i = M_i$, $\boldsymbol{\phi}_i^T \mathbf{K} \boldsymbol{\phi}_j = 0 \quad (i \neq j)$ and $\boldsymbol{\phi}_i^T \mathbf{K} \boldsymbol{\phi}_i = K_i = \omega_i^2 M_i$.

In the next subsections, the likelihood, prior and posterior functions are proposed.

3.6.1 Likelihood Function

The likelihood function expresses the unknown vector $\mathbf{K}$ in terms of known and fixed data $\boldsymbol{\omega}$. The likelihood can be written mathematically as follows (Bishop, 1995; Marwala, 2009):

$$p(\mathbf{K}|\boldsymbol{\omega}) = \frac{1}{Z_D} \exp(-\beta E_D) = \frac{1}{Z_D} \exp\left(-\beta \sum_{i=1}^{n} \left(\omega_i - \sqrt{\frac{K_i}{M_i}} \right)^2 \right), \tag{3.15}$$

where the parameters $(K_1,\ldots,K_m)$, which are obtained from the stifness matrix $\mathbf{K}$ as mentioned above, are the unknown parameters $(n \geq m)$. The values $(M_1,\ldots,M_n)$ were obtained from the mass matrix. In the next sections, we will consider $\mathbf{K} = (K_1,\ldots,K_m)$ as a vector of the unknown parameters. $\boldsymbol{\omega}$ is the vector of the natural frequencies andE_D is the error function

$$E_D = \sum_{i=1}^{n} \left(\omega_i - \sqrt{\frac{K_i}{M_i}} \right)^2. \tag{3.16}$$

In Equation 3.15, β represents the hyperparameters and $Z_D = Z_D(\beta)$ is a normalisation constant which can be estimated as follows (Bishop, 1995; Marwala, 2009):

$$Z_D(\beta) = \int_{-\infty}^{+\infty} \exp(-\beta E_D) dD. \tag{3.17}$$

3.6.2 Prior Function

The prior probability distribution is the presumed probability of the design variables. A prior is normally a subjective approximation by a knowledgeable expert. There are many types of priors, including informative and non-informative priors. An informative prior indicates certain information about a variable, while a non-informative prior shows general information about a variable. A prior distribution that assumes that the model parameters are of the same order of magnitude can be expressed as follows (Berger, 1985; Bishop, 1995):

$$p(\mathbf{K}) = \frac{1}{Z_K} \exp(-\alpha E_K) = \frac{1}{Z_K} \exp\left(-\alpha \sum_{j=1}^{m} K_j^2 \right), \tag{3.18}$$

where $m \leq n$. Here E_K represents the sum of squares of the design variables K_j,

$$E_K = \alpha \sum_{j=1}^{m} K_j^2, \tag{3.19}$$

α represents the hyperparameters. and $Z_K = Z_K(\alpha)$ is the normalisation constant, which can be approximated as

$$Z_K(\alpha) = \int_{-\infty}^{\infty} \exp(-\beta E_K) dK. \tag{3.20}$$

The prior term E_K helps separate the noise from the data, which is similar to the concept of regularisation that has been implemented in optimisation problems (Tikhonov and Arsenin, 1977; Tibshirani, 1996).

The prior distribution of a Bayesian method is then a regularisation parameter that brings extra information to the objective function (in our case the posterior function), using a penalty function, to solve an ill-posed problem or to prevent overfitting to smooth the objective function to balance complexity with accuracy.

3.6.3 Posterior Function

In this chapter, the posterior probability is the probability of the model design variable K given the observed data ω. Essentially, it is a conditional probability assigned after the relevant evidence, that is, the probability distribution of the observed data is taken into account (Rubin *et al.*, 2003). It is estimated by multiplying the likelihood function with the prior function, and dividing it by a normalisation function which is also called the evidence. By combining both Equations 3.15 and 3.18, the posterior distribution can be expressed as (Berger, 1985; Bishop, 1995)

$$p(\mathbf{K}|\boldsymbol{\omega}) = \frac{1}{Z_S} \exp\left(-\beta \sum_{i=1}^{n} \left(\omega_i - \sqrt{\frac{K_i}{M_i}} \right)^2 - \alpha \sum_{j}^{m} K_j^2 \right), \tag{3.21}$$

where

$$Z_S = Z_S(\alpha, \beta) = Z_K(\alpha) \times Z_D(\beta) = \left(\frac{2\pi}{\beta} \right)^{n/2} \left(\frac{2\pi}{\alpha} \right)^{m/2}. \tag{3.22}$$

The distribution in Equation 3.19 is a canonical distribution (Haykin, 1999). Using the Bayesian method produces an identification of the probability distribution of the design variables. The Bayesian method, by design, penalises highly complex models and consequently can identify an optimal model (Bishop, 1995).

3.6.4 Gaussian Approximation

Using the Taylor expansion (the same as described in Section **3.5**), the $-\log(p(\mathbf{K}|\boldsymbol{\omega}))$ function, which can be seen as an objective function, can be expressed as follows (Bishop, 1995):

$$\begin{aligned} E(\mathbf{K}) &= -\beta \sum_{i=1}^{n} \left(\omega_i - \sqrt{\frac{K_i}{M_i}} \right)^2 - \alpha \sum_{j}^{m} K_j^2 \\ &\approx E(\mathbf{K}_{\mathrm{MP}}) + \frac{1}{2} (\mathbf{K} - \mathbf{K}_{\mathrm{MP}})^T \mathbf{A} (\mathbf{K} - \mathbf{K}_{\mathrm{MP}}) \end{aligned}, \tag{3.23}$$

where

$$\mathbf{A} = \beta \nabla \nabla E(\mathbf{K}) + \alpha \mathbf{I}. \tag{3.24}$$

Here, the subscript MP denotes the most probable design variable, the superscript T indicates transposition and $\mathbf{A}$ stands for the Hessian matrix. The evidence can be expressed as follows (Bishop, 1995):

$$\begin{aligned} p(\mathbf{K}|\alpha,\beta) &= \frac{1}{Z_D Z_K} \int \exp(-E(\mathbf{K}))d\mathbf{K} \\ &= \frac{Z_E}{Z_D Z_K} = \frac{(2\pi/\beta)^{n/2} + (2\pi/\alpha)^{m/2}}{(2\pi/\beta)^{n/2}(2\pi/\alpha)^{m/2}}. \end{aligned} \tag{3.25}$$

Maximisation of the log evidence gives the following estimates for the hyperparameters (Bishop, 1995):

$$\beta_{\mathrm{MP}} = \frac{n-\gamma}{2E_D(\mathbf{K}_{\mathrm{MP}})} \tag{3.26}$$

$$\alpha_{\mathrm{MP}} = \frac{\gamma}{2E_K(\mathbf{K}_{\mathrm{MP}})}, \tag{3.27}$$

where $2E_K = \mathbf{K}^T \mathbf{I} \mathbf{K}$ and

$$\gamma = \sum_j \left(\frac{\pi_j - \alpha}{\eta_j} (\mathbf{V}^T \mathbf{I} \mathbf{V})_{jj} \right), \tag{3.28}$$

in which η_j are the eigenvalues of the Hessian $\mathbf{A}$, and $\mathbf{V}$ are the eigenvectors such that $\mathbf{V}^T\mathbf{V} = \mathbf{I}$. This procedure can be summarised as follows (MacKay, 1991, 1992):

1. Randomly select the initial values for the hyperparameters.
2. Estimate the design variables using an optimisation method such as the Broyden–Fletcher–Goldfarb–Shanno (BFGS) or the scaled conjugate gradient algorithm to minimise the objective function, to obtain the parameter $\mathbf{K}_{\mathrm{MP}}$.
3. Apply the evidence framework to approximate the hyperparameters using Equations 3.28 and 3.29.
4. Repeat steps 2 and 3 until convergence.

The posterior distribution can also be approximated by the MAP parameter estimate, which is related to the ML method, but includes the prior knowledge in its optimisation objective. In this simple dynamic example, the numerical approximation methods are not needed since the example is very simple and only one variable is updated. However, later a few numerical optimisation approaches are discussed where these methods can be implemented in the investigation of a complex example.

3.7 The Posterior Approximation

Suppose, for the example given in the previous section, that we take a measurement of the natural frequency and we know the mass, and we want to estimate the modulus of elasticity from these measurements. One way is to use the ML approach to estimate the modulus of elasticity. This can lead to an optimisation problem, where the objective is to minimise the difference between the measured frequency and the model predicted natural frequency, with the modulus of elasticity as the design variable.

3.7.1 Objective Function

The objective function obtained from the log-likelihood is given by Equation 3.16 where E_D is the error between the measured natural frequency ω_i, the model predicted natural frequency $\sqrt{K_i/M_i}$ and n represents the number of measurements or data points collected (in our example $n=1$). Unfortunately, merely minimising Equation 3.16 identifies not only the correct model, but also the errors in measurements as part of the model. To separate the noise from the data, the prior distribution terms will be needed. This also can be called as regularisation (Tikhonov and Arsenin, 1977; Tibshirani, 1996). The regularisation terms obtained from the log-prior are given in Equation 3.19, and the optimisation equation may thus be written as follows:

$$f(\mathbf{K}) = E_D + E_K = \sum_{i=1}^{n}\left(\omega_i - \sqrt{\frac{K_i}{M_i}}\right)^2 + \alpha\sum_{j=1}^{m} K_j^2. \tag{3.29}$$

3.7.2 Optimisation Approach

Equation 3.29 needs to be optimised in order to identify the design variable K, and this section describes two techniques that could be used to achieve this goal. Optimisation is a mathematical technique for identifying a minimum or a maximum point (Fletcher, 2000). There are a number of issues that require careful attention in optimisation, and chief among these is the issue of the global versus the local optimum point. The general idea is that, on identifying an optimal point, classical methods identify a stationary point where the gradient of the objective function is zero. Many such stationary points are merely local optima. What complicates matters is that often a global optimum is that point which overfits the data, including noise. The identification of an optimal point normally involves calculating the derivative of the objective function with respect to the design variable and, for Equation 3.29, this can be written as follows (Fletcher, 2000; Marwala, 2010):

$$\begin{aligned}
\frac{df(\mathbf{K})}{d\mathbf{K}} &= 0,\\
\frac{d\left(\sum_{i=1}^{n}\left(\omega_i - \sqrt{K_i/M_i}\right)^2 + \alpha\sum_{j=1}^{m} K_j^2\right)}{d\mathbf{K}} &= 0,\\
\sum_{i=1}^{n} 2\left(\omega_i - \sqrt{\frac{K_i}{M_i}}\right)\left(-\frac{1}{2\sqrt{K_i M_i}}\right) + \alpha\sum_{j=1}^{m} 2K_j &= 0,\\
\sum_{i=1}^{n}\left(\sqrt{\frac{K_i}{M_i}} - \omega_i\right)\left(\frac{1}{\sqrt{K_i M_i}}\right) + \alpha\sum_{j=1}^{m} 2K_j &= 0.
\end{aligned} \tag{3.30}$$

Then Equation 3.30 is solved to identify **K** as other parameters are known. If the example is large and complex, iterative optimisation will be required.

3.7.2.1 Quasi-Newton Broyden-Fletcher-Goldfarb-Shanno Method

One popular numerical optimisation method is the quasi-Newton BFGS method. The quasi-Newton optimisation technique is a robust and quadratically convergent optimisation routine that uses a gradient of the objective function to identify the optimal solution. The method is a variant of the Newton–Raphson technique, where the inverse of the second derivative is updated using a one-dimensional or multi-dimensional Hessian estimation technique. Ghosal and Chaki (2010) used the quasi-Newton method to optimise the laser welding process. For a one-dimensional function, the Newton–Raphson algorithm can be expressed as (Ransome, 2006; Marwala, 2010):

$$x_{n+1} = x_n - \eta \frac{\dot{f}(x_n)}{\ddot{f}(x_n)} \tag{3.31}$$

where $f(x_n)$ is the objective function (the log-likelihood), $\dot{f}(x_n)$ is the Jacobian (first derivative) and $\ddot{f}(x_n)$ is the Hessian (second derivative). For a one-dimensional function, $\ddot{f}(x_n)$ can be updated by using the following equation (Ransome, 2006; Marwala, 2010):

$$\ddot{f}(x_{n+1}) = \frac{s_n}{y_n}, \tag{3.32}$$

where $s_n = x_{n+1} - x_n$ and $y_n = \dot{f}(x_{n+1}) - \dot{f}(x_n)$. The method widely used for identifying the Hessian is the BFGS method (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970). It has been applied widely in machining processes (Sun *et al.*, 2009), fluid dynamics (Tan *et al.*, 2009), distribution process optimisation (Du *et al.*, 2009), aerodynamics (Papadimitriou and Giannakoglou, 2009) and non-convex problems (Xiao *et al.*, 2009). For multi-dimensional problems, the BFGS method approximates the inverse Hessian, $\boldsymbol{H}_{n+1}$, which can be expressed as follows:

$$\mathbf{H}_{n+1} = \mathbf{H}_n + \frac{\mathbf{q}_n \mathbf{q}_n^T}{\mathbf{q}_n^T \mathbf{s}_n} - \frac{\mathbf{H}_n^T \mathbf{s}_n^T \mathbf{s}_n \mathbf{H}_n}{\mathbf{s}_n^T \mathbf{H}_n \mathbf{s}_n}. \tag{3.33}$$

Starting with an initial estimation, K_0 and inverse Hessian matrix, $\boldsymbol{H}_0$, repeat the following steps (Nocedal and Wright, 2006):

1. Calculate $\mathbf{s}_n = -\mathbf{H}_n^{-1} \nabla f(\mathbf{K}_n)$.
2. Perform a line search to identify an optimal step size λ_n in the direction of the initial step and use this to estimate $\mathbf{K}_{n+1} = \mathbf{K}_n + \lambda_n \mathbf{s}_n$.
3. $\mathbf{q}_n = \nabla f(\mathbf{K}_{n+1}) - \nabla f(\mathbf{K}_n)$.
4. Compute Equation 3.33.

3.7.2.2 Conjugate Gradient Method

The conjugate gradient technique (Hestenes and Stiefel, 1952) is a gradient-based optimisation method. It is an improvement of the gradient descent method. In the gradient descent method, the step size is expressed as $-\eta\partial f(\mathbf{K})/\partial\mathbf{K}$, where the parameter η is the step size and $\partial f(\mathbf{K})/\partial\mathbf{K}$ is the gradient of the objective function. Thus, the gradient descent method can be written as follows (Haykin, 1999):

$$\mathbf{K}_{n+1} = \mathbf{K}_{n+1} - \eta\partial f(\mathbf{K})/\partial\mathbf{K}. \tag{3.34}$$

If the step size is sufficiently small, the value of the error decreases at each step until a minimum value is attained for the objective function. The disadvantage of this technique is that it is computationally more expensive than other procedures. In the conjugate gradient method, a quadratic function of the error is minimised at each iteration over a gradually expanding linear vector space that includes the global minimum of the error by using this procedure (Luenberger, 1984; Fletcher, 2000; Bertsekas, 1995):

1. Select the initial weight vector $\mathbf{K}_0$.
2. Compute the gradient vector $\left.\frac{\partial f(\mathbf{K})}{\partial\mathbf{K}}\right|_{\mathbf{K}_0}$.
3. At each step n apply a line search to identify the $\eta(n)$ that minimises $f(\eta)$, which represents the objective function, which is a function of η for fixed values of $\mathbf{K}$ and $-\left.\frac{\partial f(\mathbf{K})}{\partial\mathbf{K}}\right|_{\mathbf{K}_n}$.
4. Establish whether the Euclidean norm of the vector $-\left.\frac{\partial f(\mathbf{K})}{\partial\mathbf{K}}\right|_{\mathbf{K}_n}$ is adequately less than that of $-\left.\frac{\partial f(\mathbf{K})}{\partial\mathbf{K}}\right|_{\mathbf{K}_0}$.
5. Update $\mathbf{K}$.
6. For $\mathbf{K}_{n+1}$, compute the updated gradient $\left.\frac{\partial f(\mathbf{K})}{\partial\mathbf{K}}\right|_{\mathbf{K}_{n+1}}$.
7. Apply the Polak–Ribière technique to estimate

$$\beta(n+1) = \frac{\nabla f\left(\mathbf{K}_{n+1}^T\right)\left(\nabla f(\mathbf{K}_{n+1}) - \nabla f(\mathbf{K}_n)\right)}{\nabla f\left(\mathbf{K}_n^T\right)\nabla f(\mathbf{K}_n)}.$$

8. Update the direction vector:

$$\left.\frac{\partial f(\mathbf{K})}{\partial\mathbf{K}}\right|_{\mathbf{K}_{n+2}} = \left.\frac{\partial f(\mathbf{K})}{\partial\mathbf{K}}\right|_{\mathbf{K}_{n+1}} - \beta(n+1)\left.\frac{\partial f(\mathbf{K})}{\partial\mathbf{K}}\right|_{\mathbf{K}_n}.$$

9. Set $n = n+1$ and go to step 3.
10. Stop when the condition $\left.\frac{\partial f(\mathbf{K})}{\partial\mathbf{K}}\right|_{\mathbf{K}_{n+2}} = \varepsilon\left.\frac{\partial f(\mathbf{K})}{\partial\mathbf{K}}\right|_{\mathbf{K}_{n+1}}$ is satisfied, where ε is a small number.

3.7.3 Case Example

In the single-degree-of-freedom example presented in Section 3.6, suppose the mass of the brick is 2 kg and the natural frequencies were measured several times, giving 5.3, 5.2, 5.5 and 5.4 Hz. So one can identify the K that will give the natural frequency of 5.3 Hz using Equation 3.29 and assuming that $\alpha = 0$ by minimising:

$$E_D = \left(5.3 - \sqrt{\frac{K}{2}}\right)^2.$$

This gives $K = 56.2$. If one uses two measurements, 5.3 and 5.2 Hz, then the objective function becomes

$$E_D = \left(5.3 - \sqrt{\frac{K}{2}}\right)^2 + \left(5.2 - \sqrt{\frac{K}{2}}\right)^2.$$

If three measurements are used, then the objective function becomes

$$E_D = \left(5.3 - \sqrt{\frac{K}{2}}\right)^2 + \left(5.2 - \sqrt{\frac{K}{2}}\right)^2 + \left(5.5 - \sqrt{\frac{K}{2}}\right)^2.$$

If all measurements are used, then the objective function becomes

$$E_D = \left(5.3 - \sqrt{\frac{K}{2}}\right)^2 + \left(5.2 - \sqrt{\frac{K}{2}}\right)^2 + \left(5.5 - \sqrt{\frac{K}{2}}\right)^2 + \left(5.4 - \sqrt{\frac{K}{2}}\right)^2.$$

Whether one measurement or two or three or four are used to estimate K, the optimisation process through the use of the conjugate gradient method or the BFGS returns a single estimate of K. If it is to be assumed that the reason for these four different measurements is uncertainty in measurements, the ML method returns K without uncertainty. To correct this issue; a Bayesian method, which is the subject of the next section, can be used.

3.8 Sampling Approaches for Estimating Posterior Distribution

If the updated system/structure is too large, the posterior probability may be estimated using the Monte Carlo method, the MCMC method, genetic MCMC sampling, simulated annealing or Gibbs sampling.

3.8.1 Monte Carlo Method

Monte Carlo approaches are numerical methods that are applied by repetitive random sampling to estimate the probability distribution of any system. Because of their dependence on repeated

computation of random or simulated random numbers, these methods are appropriate for approximation using computers and are used when it is impractical to estimate a solution with a deterministic approach (Kandela *et al.*, 2010; Mathe and Novak, 2007; Akhmatskaya *et al.*, 2009; Ratick and Schwarz, 2009). Monte Carlo methods have been applied in solving integral problems (Lai, 2009), radiative transfer problems (McClarren and Urbatsch, 2009), particle coagulation (Zhao and Zheng, 2009), diffusion problems (Liu *et al.*, 2009), the design of radiation detectors (Dunn and Shultis, 2009), modelling bacterial activities (Oliveira *et al.*, 2009), vehicle detection (Jia and Zhang, 2009), modelling the bystander effect (Xia *et al.*, 2009) and modelling nitrogen absorption (Rahmati and Modarress, 2009). Monte Carlo techniques are normally implemented as follows (Robert and Casella, 2004):

1. Describe the input space.
2. Randomly create input from the input space by applying a selected probability distribution. In this chapter, this is the posterior distribution function.
3. The new impute will be accepted or rejected randomly.
4. Apply the produced input for the deterministic computation.
5. After obtaining the desired number of samples (or inputs), the mean value will be computed to approximate the results.

3.8.2 Markov Chain Monte Carlo Method

The MCMC technique is a random walk Monte Carlo procedure that is implemented by producing a Markov chain to identify a probability distribution function. The MCMC uses the Markov process and a Monte Carlo simulation by simulating many random walk steps, retaining some states and converging on a posterior distribution function (Liesenfeld and Richard, 2008). The MCMC has been applied in many complex problems such as renal diseases (Rodina *et al.*, 2010), spectrum sensing (Wang *et al.*, 2010), water distribution systems (Wang and Harrison, 2010), tracking of manoeuvring objects (Jing and Vadakkepat, 2010), DNA profiling (Curran, 2008), environmental modelling (Gauchere *et al.*, 2008), medical imaging (Jun *et al.*, 2008), lake water quality modelling (Malve *et al.*, 2007), economics (Jacquier *et al.*, 2007) and statistics (Lombardi, 2007).

In using the MCMC method, a system is handled whose evolution is characterised by a stochastic process consisting of random variables $\{\boldsymbol{K}_1, \boldsymbol{K}_2, \boldsymbol{K}_3, \ldots, \boldsymbol{K}_i\}$, where a random variable $\boldsymbol{K}_i = \left\{K_1^i, K_2^i, \ldots, K_n^i\right\}$, a vector of n variables that need to be updated, occupies state vector $\boldsymbol{K}$ at discrete time i. All the possible states that all random variables occupy are referred to as the *state space*. In the case of MCMC, the probability that the system is in state $\boldsymbol{K}_{i+1}$ at time $i+1$ depends completely on the information that it occupied state $\boldsymbol{K}_i$ at time i, because the random variables $\{\boldsymbol{K}_1, \boldsymbol{K}_2, \boldsymbol{K}_3, \ldots, \boldsymbol{K}_i\}$ form a Markov chain. The transition between states in the MCMC is attained by adding a random noise ($\boldsymbol{\varepsilon}$) to the current state as follows (Marwala, 2010):

$$\boldsymbol{K}_{i+1} = \boldsymbol{K}_i + \boldsymbol{\varepsilon}. \tag{3.35}$$

The current state is either accepted or rejected by applying the Metropolis algorithm (Bedard, 2008; Meyer *et al.*, 2008; Metropolis *et al.*, 1953), a popular means of solving problems of statistical mechanics. The Metropolis algorithm has been applied in protein simulation

(Bazavov *et al*., 2010; Tiana *et al*., 2007) and to predict free Co-Pt nanoclusters (Moskovkin and Hou, 2007). The Metropolis algorithm is mathematically represented as follows (Marwala, 2009, 2010):

$$\begin{aligned}
&\text{If } \sum_{i=1}^{n}\left(\omega_i - \sqrt{\frac{K_i^{\text{new}}}{M}}\right)^2 < \sum_{i=1}^{n}\left(\omega_i - \sqrt{\frac{K_i^{\text{old}}}{M}}\right)^2 \text{ accept } \mathbf{K}^{\text{new}} \\
&\text{else} \\
&\text{accept } \mathbf{K}^{\text{new}} \text{with probability} \\
&\exp\left\{\left(\sum_{i=1}^{n}\left(\omega_i - \sqrt{\frac{K_i^{\text{old}}}{M}}\right)^2 - \sum_{i=1}^{n}\left(\omega_i - \sqrt{\frac{K_i^{\text{new}}}{M}}\right)^2\right)\right\}.
\end{aligned} \tag{3.36}$$

3.8.3 Simulated Annealing

Simulated annealing is a Monte Carlo method inspired by the procedure of annealing where metals recrystallise or liquids freeze; it can be used to sample a probability distribution. It has been applied in scheduling problems (Naderi *et al*., 2009; Torabzadeh and Zandieh, 2010), modelling asphalt (Ozgan and Saruhan, 2010), cellular manufacturing (Paydar *et al*., 2010), protein structures (Kannan and Zacharias, 2009) and the design of acoustic structures (Cretu and Pop, 2009).

In the annealing technique, a material is heated until it is molten, and then its temperature is slowly decreased such that the metal is almost in thermodynamic equilibrium. As the temperature of the object falls, the system becomes more ordered and approaches a *frozen* state at $T = 0$. If the cooling procedure is regulated inefficiently or the initial temperature of the object is insufficiently high, the system forms defects or freezes out in meta-stable states, representing that the system is stuck in a local minimum energy state. The probability of accepting the reversal is given by Boltzmann's equation (Bryan *et al*., 2006; Marwala, 2010):

$$\begin{aligned}
P(\Delta E) &= \frac{1}{Z}\exp\left(-\frac{\Delta E}{T}\right) \\
&= \frac{1}{Z}\exp\left(\frac{\sum_{i=1}^{n}\left(\omega_i - \sqrt{K_i^{\text{old}}/M}\right)^2 - \sum_{i=1}^{n}\left(\omega_i - \sqrt{K_i^{\text{new}}/M}\right)^2}{T}\right).
\end{aligned} \tag{3.37}$$

Here, ΔE is the difference in the objective function between the old and the new states. The state designates the possible updated finite element models. T is the temperature of the system, and Z is a normalisation constant that guarantees that when the probability function is integrated to infinity it equals to 1.

3.8.4 Gibbs Sampling

Gibbs sampling is a technique used to generate a sequence of samples from the joint probability distribution of at least two random variables in order to estimate a joint distribution, approximate the marginal distribution of one of the variables, and compute an integral that approximates the expected value of a variable (Casella and George, 1992; Natesan *et al.*, 2010). Gibbs sampling, which can be regarded as a kind of MCMC method, is an instrument for statistical inference and it is a random technique. Gibbs sampling has been applied in genetics of dairy cows and sheep (Hossein-Zadeh and Ardalan, 2010), mapping ambiguous short-sequence tags (Wang *et al.*, 2010), and the decentralised coordination of autonomous swarms (Tan *et al.*, 2010).

Gibbs sampling is applied when the joint distribution is not completely known or is difficult to sample, yet the conditional distribution of each variable is known or can be easily sampled. The Gibbs sampling technique generates a sample from the distribution of each variable, conditional on the current values of the other variables and thus produces a Markov chain, and the stationary distribution of the Markov chain (Gelman *et al.*, 1995). Gibbs sampling is applied as follows, by obtaining k samples of $\boldsymbol{K} = \{K_1,\ldots,K_n\}$ from a joint distribution $p(K_1, \ldots, K_n)$ and denoting the ith sample by $\boldsymbol{K}^{(i)} = \left\{K_1^{(i)},\ldots,K_n^{(i)}\right\}$ (Gelman *et al.*, 1995):

- Start with an initial value $\boldsymbol{K}^{(0)}$ for each variable.
- For each sample $i = \{1,\ldots,n\}$, sample each variable $K_j^{(i)}$ from the conditional distribution $p\left(K_j^{(i)} \middle| K_1^{(i)},\ldots,K_{j-1}^{(i)},K_{j+1}^{(i-1)},\ldots,K_n^{(i-1)}\right)$ by sampling each variable from the distribution of that variable conditional on all other variables and applying the current values, and update the variable with a new value once it has been sampled.

These samples then approximate the joint distribution of all the variables, and the marginal distribution of any variable can be calculated from the samples of those variables and discarding the others.

3.9 Comparison between Approaches

This chapter has studied three techniques that can be used to update a finite element model. These three techniques are the ML method, Gaussian approximation of the posterior probability, and by directly sampling the posterior probability function. These techniques are applied in the context of system identification of a one-dimensional mass and spring system. From this study, the remarks in Table 3.2 can be presented, and it is concluded that directly sampling the posterior probability is the best approach to performing finite element model updating. The remaining chapters will therefore pursue this approach.

3.9.1 Numerical Example

In this example, a simplified model of an automobile suspension system is updated (see Figure 3.3). This system has two degrees of freedom, where the mass M is attached to the

Table 3.2 Final remarks regarding the overall study

Method	Techniques	Advantages	Disadvantages
Maximum-likelihood method	BFGS, conjugate gradient method, steepest descent	• Simple to implement • Easy to understand • Computationally efficient	• Does not quantify uncertainty simply • Does not offer a probabilistic insight
Approximation of the posterior distribution	Gaussian approximation, Laplace approximation, conjugate gradient method	• Simple to implement • Estimates the error bars of the solution • Quantifies uncertainty	• Is an approximation of the posterior distribution • Too many approximations
Directly sampling the posterior distribution	Monte Carlo method, Markov chain Monte Carlo, Gibbs sampling, simulated annealing, hybrid Monte Carlo	• Offers probability interpretation • Quantifies uncertainty	• Computationally expensive

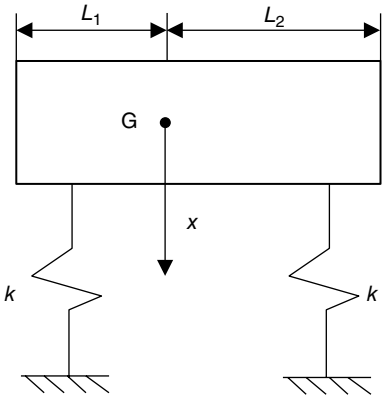

Figure 3.3 A two-degrees-of-freedom system

ground through two springs with modulus of elasticity k, x is the displacement and θ is the angular rotation of the body from the horizontal position. The automobile weight is $M = 600\,\text{kg}$, moment of inertia $I = 1100\ \text{kg m}^2$, the spring stiffness $k = 20\,000\ \text{N/m}$, and L_1 and L_2 are 1.4 and 1.47 m, respectively. The system in Figure 3.3 can be represented by the following differential equation:

$$\begin{bmatrix} m & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \ddot{x} \\ \ddot{\theta} \end{bmatrix} + \begin{bmatrix} 2k & (l_2 - l_1)k \\ (l_2 - l_1)k & (l_2^2 - l_1^2)k \end{bmatrix} \begin{bmatrix} x \\ \theta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{3.38}$$

Three different methods are used to update the system described in Equation 3.38: the ML method, Gaussian approximation and simulated annealing, where the spring stiffness is

Table 3.3 The updated value of the spring stiffness using maximum likelihood, Gaussian approximation and simulated annealing techniques

	Spring stiffness (N/m)			
	Initial	Maximum likelihood	Gaussian approximation	Simulated annealing
k	2.7000×10^4	2.0034×10^4	2.0515×10^4	1.9934×10^4

Table 3.4 Frequencies and errors when maximum likelihood, Gaussian approximation and simulated annealing techniques used to update spring stiffness

Measured frequency (Hz)	Initial frequency (Hz)	Frequencies maximum likelihood (Hz)	Frequencies Gaussian approximation (Hz)	Frequencies simulated annealing (Hz)
8.65	10.05	8.66	8.76	8.64
25.57	29.72	25.60	25.90	25.53
Total average error	16.21	0.12	1.28	0.14

adjusted so that the initial model matches the experimental results. The updated values of the spring stiffness are given in Table 3.3, while Table 3.4 presents the natural frequencies obtained by all three algorithms.

As shown in Table 3.3, all algorithms updated the spring stiffness and gave different results. However, the ML method gave better results than both simulated annealing and Gaussian approximation (see the total average error in Table 3.4). These results are obvious since only one variable was updated,the representation of the system matrices is simple and the frequencies can be analytically obtained. However, sampling methods give better results than others when complex systems are updated. In the next chapters, more complex system will be updated.

3.10 Conclusions

This chapter has discussed the Bayesian theory and approaches used to approximate the posterior probability density function such as the maximum-likelihood approach, maximum a posteriori, Laplace approximation and sampling the posterior distribution function directly. Bayesian statistics states that the probability of an event A happening given that event B has happened, also called the posterior probability, is equal to the product of the probability of B happening given that A has happened, also called the likelihood function, and the probability of A happening, also called the prior, divided by the probability of B happening, also called the evidence. These techniques were studied using a single-degree-of-freedom mass and spring system, where the distribution of the measured natural frequency and the mass are assumed to be known and Bayesian statistics is used to estimate the distribution of the stiffness. At the end of this chapter, a simplified model of an automobile suspension system is updated using maximum likelihood, Gaussian approximation of frequencies and simulated annealing. It is concluded that the best approach is to sample the posterior probability function directly when the updated system has a large size.

References

Adhikari S (2007) Matrix variate distributions for probabilistic structural mechanics. *AIAA Journal* **45**: 1748–1762.

Adhikari S (2010) Sensitivity based reduced approaches for structural reliability analysis. *Proceedings of the Indian Academy of Engineering Sciences* **35**: 319–339.

Adhikari S (2011) A reduced spectral function approach for the stochastic finite element analysis. *Computer Methods in Applied Mechanics & Engineering* **200**: 1804–1821.

Adhikari S (2015) Uncetainty Quantification in Structural Dynamics. http://engweb.swan.ac.uk/~adhikaris/stochastic.html (accessed July 6 2015).

Adhikari S, Pastur L, Lytova A, Du Bois JL (2012) Eigenvalue density of linear stochastic dynamical systems: a random matrix approach. *Journal of Sound & Vibration* **331**: 1042–1058.

Akhmatskaya E, Bou-Rabee N, Reich S (2009) A comparison of generalized hybrid Monte Carlo methods with and without momentum flip. *Journal of Computational Physics* **228**: 2256–2265.

Bazavov A, Berg BA, Zhou H (2010) Application of biased Metropolis algorithms: from protons to proteins. *Mathematics and Computers in Simulation* **80**: 1056–1067.

Bedard M (2008) Optimal acceptance rates for Metropolis algorithms: moving beyond 0.234. *Stochastic Processes and their Applications* **118**: 2198–2222.

Berger JO (1985) *Statistical Decision Theory and Bayesian Analysis*. Berlin:Springer-Verlag.

Bertsekas DP (1995) *Nonlinear Programming*. Belmont, MA:Athenas Scientific.

Bishop CM (1995) *Neural Networks for Pattern Recognition*. Oxford:Oxford University Press.

Boulkaibet I, Marwala T, Mthembu L, Friswell MI, Adhikari S (2012) Sampling techniques in Bayesian finite element model updating. *Proceedings of the Society for Experimental Mechanics*, **29**: 75–83.

Boulkaibet I, Marwala T, Mthembu L, Friswell MI, Adhikari S (2015) Finite element model updating using the shadow hybrid Monte Carlo technique. *Mechanical System and Signal Processing* **52**: 115–132.

Broyden CG (1970) The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and its Applications* **6**: 76–90.

Bryan K, Cunningham P, Bolshkova N (2006) Application of simulated annealing to the biclustering of gene expression data. *IEEE Transactions on Information Technology in Biomedicine* **10**:519–525.

Calanni G, Volovoi V, Ruzzene M, Vining C, Cento P (2007) Application of Bayesian belief nets for modeling uncertainty in structural dynamics. In *Proceedings of the ASME Turbo Expo: Power for Land, Sea, and Air*, Vol. 5, pp. 1135–1149.

Casella G, George EI (1992) Explaining the Gibbs sampler. *American Statistics* **46**: 167–174.

Charnes A, Frome EL, Yu PL (1976) The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. *Journal of the American Statistical Association* **71**(353): 169–171.

Chowdhury R, Adhikari S (2011) Reliability analysis of uncertain dynamical systems using correlated function expansion. *International Journal of Mechanical Sciences* **53**: 281–285.

Chowdhury R, Adhikari S (2012) Fuzzy parametric uncertainty analysis of linear dynamical systems: a surrogate modeling approach. *Mechanical Systems & Signal Processing* **32**: 5–17.

Cretu N, Pop M (2009) Acoustic behavior design with simulated annealing. *Computational Material Sciences* **44**: 1312–1318.

Curran JM (2008) A MCMC method for resolving two person mixtures. *Science and Justice* **48**: 168–177.

Del Moral P (2013) *Mean Field Simulation for Monte Carlo Integration*. Boca Raton, FL: Chapman & Hall/CRC Press.

DiazDelaO FA, Adhikari S (2012) Bayesian assimilation of multi-fidelity finite element models. *Computers & Structures* **92–93**: 206–215.

DiazDelaO FA, Adhikari S, Flores EIS, Friswell MI (2013) Stochastic structural dynamic analysis using Gaussian process emulators. *Computers & Structures* **120**: 24–32.

Du N, Fan J, Wu H, Sun W (2009) Optimal porosity distribution of fibrous insulation. *International Journal of Heat and Mass Transfer* **52**: 4350–4357.

Dunn WL, Shultis JK (2009) Monte Carlo methods for design and analysis of radiation detectors. *Radiative Physics and Chemistry* **78**: 852–858.

Ewins DJ (2000) *Modal Testing: Theory, Practice and Application*. Baldock: Research Studies Press.

Fletcher R (2000) *Practical Methods of Optimization*. Chichester: John Wiley & Sons, Ltd.

Fletcher RA (1970) New approach to variable metric algorithms. *Computing Journal* **13**: 317–322.

Gauchere C, Campillo F, Misson L, Guiot J, Boreux JJ (2008) Parameterization of a process-based tree-growth model: comparison of optimization, MCMC and particle filtering algorithms. *Environmental Model and Software* **23**: 1280–1288.

Gelman A, Carlin JB, Stern HS, Rubin DB (1995) *Bayesian Data Analysis*. London: Chapman & Hall.

Ghosal S, Chaki S (2010) Estimation and optimization of depth of penetration in hybrid CO2 LASER-MIG welding using ANN-optimization hybrid model. *International Journal of Advanced Manufacturing Technology* **47**: 1149–1157.

Giagopoulos, D., Papadioti, D.-C., Papadimitriou, C. and Natsiavas, S. (2009) *Bayesian uncertainty quantification and propagation in nonlinear structural dynamics*. Proceedings of Civil-Comp, p. 92.

Giagopoulos D, Papadioti D-C, Papadimitriou C, Natsiavas S (2013) Bayesian uncertainty quantification and propagation in nonlinear structural dynamics. In Simmermacher T, Cogan S, Moaveni B, Papadimitriou C, eds, *Topics in Model Validation and Uncertainty Quantification, Volume* **5***: Proceedings of the 31st IMAC, A Conference on Structural Dynamics*, pp. 33–41. New York: Springer.

Goldfarb DA (1970) Family of variable metric updates derived by variational means. *Mathematics of Computing* **24**: 23–26.

Guyan RJ (1965) Reduction of stiffness and mass matrices. *AIAA Journal* **3**: 380.

Haykin S (1999) *Neural Networks*. Upper Saddle River, NJ: Prentice Hall.

Hazewinkel M ed. (2001), Taylor series. In *Encyclopedia of Mathematics*. Heidelberg: Springer.

He J, Fu Z-F (2001) *Modal Analysis*. Butterworth:Heinemann.

Hestenes MR, Stiefel E (1952) Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards* **6**: 409–436.

Horn RA, Johnson CA (1985) *Matrix Analysis*. Cambridge:Cambridge University Press.

Hossein-Zadeh NG, Ardalan M (2010) Estimation of genetic parameters for body weight traits and litter size of Moghani sheep, using a Bayesian approach via Gibbs sampling. *Journal of Agricultural Science* **148**: 363–370.

Jacquier E, Johannes M, Polson N (2007) MCMC maximum likelihood for latent state models. *Journal of Economics* **137**: 615–640.

Jia Y, Zhang C (2009) Front-view vehicle detection by Markov chain Monte Carlo method. *Pattern Recognition* **42**: 313–321.

Jing L, Vadakkepat P (2010) Interacting MCMC particle filter for tracking maneuvering target. *Digital Signal Processing* **20**: 561–574.

Jun SC, George JS, Kim W, Pare-Blagoev J, Plis S, Ranken DM, Schmidt DM (2008) Bayesian brain source imaging based on combined MEG/EEG and fMRI using MCMC. *NeuroImage* **40**: 1581–1594.

Kandela B, Sheorey U, Banerjee A, Bellare J (2010) Study of tablet-coating parameters for a pan coater through video imaging and Monte Carlo simulation. *Powder Technology* **204**: 103–112.

Kannan S, Zacharias M (2009) Simulated annealing coupled replica exchange molecular dynamics – an efficient conformational sampling method. *Journal of Structural Biology* **166**: 288–294.

Karhunen K (1947) Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Annales Academiæ Scientiarum Fennicæ, Series A.I Mathematics & Physics*. **37**: 1–79.

Kass RE, Raftery AE (1995). Bayes factors. *Journal of the American Statistical Association*, **90**: 773–795.

Kennedy MC, O'Hagan, A. (2000) *Supplementary details on Bayesian calibration of computer models*. Technical Report, University of Sheffield.

Kennedy MC, O'Hagan A (2001) Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B* **63**: 425–464.

Lai Y (2009) Adaptive Monte Carlo methods for matrix equations with applications. *Journal of Computational and Applied Mathematics* **231**: 705–714.

Lee SH, Chen W (2009) A comparative study of uncertainty propagation methods for black-box-type problems. *Structural & Multidisciplinary Optimization* **37**: 239–253.

Li CF, Adhikari S, Cen S, Feng YT, Owen DRJ (2010) A joint diagonalisation approach for linear stochastic systems. *Computers & Structures* **88**: 1137–1148.

Liesenfeld R, Richard J (2008) Improving MCMC, using efficient importance sampling. *Computational Statistics and Data Analysis* **53**: 272–288.

Liu X, Newsome D, Coppens M (2009) Dynamic Monte Carlo simulations of binary self-diffusion in ZSM-5. *Microporous and Mesoporous Materials* **125**: 149–159.

Loève M (1955) *Probability Theory*. Princeton, NJ: Van Nostrand.

Lombardi MJ (2007) Bayesian inference for α-stable distributions: a random walk MCMC approach. *Computational Statistics and Data Analysis* **51**: 2688–2700.

Luenberger DG (1984) *Linear and Non-linear Programming*, 2nd edn. Reading, MA:Addison-Wesley.

MacKay DJC (1991) Bayesian methods for adaptive models. PhD thesis, California Institute of Technology.

MacKay DJC (1992) A practical Bayesian framework for backpropagation networks. *Neural Computing* **4**: 448–472.

Malve O, Laine M, Haario H, Kirkkala T, Sarvala J (2007) Bayesian modelling of algal mass occurrences – using adaptive MCMC methods with a lake water quality model. *Environmental Modelling and Software* **22**: 966–977.

Martínez-Carranza J, Soto-Eguibar F, Moya-Cessa H (2012) Alternative analysis to perturbation theory in quantum mechanics. *European Physical Journal D* **66**: 22.

Marwala T (2009) *Computational Intelligence for Missing Data Imputation, Estimation and Management: Knowledge Optimization Techniques*. New York: IGI Global Publications.

Marwala T (2010) *Finite Element Model Updating Using Computational Intelligence Techniques: Applications to Structural Dynamics*. Heidelberg: Springer.

Mathe P, Novak E (2007) Simple Monte Carlo and the Metropolis algorithm. *Journal of Complexity* **23**: 673–696.

Matthies HG (2007) Quantifying uncertainty: modern computational representation of probability and applications. In *Extreme Man-Made and Natural Hazards in Dynamics of Structures*. NATO Security through Science Series, pp. 105–135. Dordrecht: Springer.

McClarren RG, Urbatsch TJ (2009) A modified implicit Monte Carlo method for time-dependent radiative transfer with adaptive material coupling. *Journal of Computational Physics* **228**: 5669–5686.

Metropolis N, Rosenbluth A, Rosenbluth M, Teller, A, Teller, E (1953) Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21**: 1087–1092.

Meyer R, Cai B, Perron F (2008) Adaptive rejection Metropolis sampling using Lagrange interpolation polynomials of degree 2. *Computational Statistics and Data Analysis* **52**: 3408–3423.

Moskovkin P, Hou M (2007) Metropolis Monte Carlo predictions of free Co-Pt nanoclusters. *Journal of Alloys and Compounds* **434–435**: 550–554.

Murugan S, Chowdhury R, Adhikari S, Friswell MI (2012) Helicopter aeroelastic analysis with spatially uncertain rotor blade properties. *Aerospace Science and Technology* **16**: 29–39.

Naderi B, Zandieh M, Khaleghi A, Balagh G, Roshanaei V (2009) An improved simulated annealing for hybrid flowshops with sequence-dependent setup and transportation times to minimize total completion time and total tardiness. *Expert Systems with Applications* **36**: 9625–9633.

Natesan P, Limbers C, Varni JW (2010) Bayesian estimation of graded response multilevel models using Gibbs sampling: formulation and illustration. *Educational and Psychological Measurement* **70**: 420–439.

Nocedal J, Wright SJ (2006) *Numerical Optimization*. Berlin:Springer-Verlag.

Oliveira RG, Schneck E, Quinn BE, Konovalov OV, Brandenburg K, Seydel U, Gill T, Hanna CB, Pink DA, Tanaka M (2009) Physical mechanisms of bacterial survival revealed by combined grazing-incidence x-ray scattering and Monte Carlo simulation. *Comptes Rendus Chimie* **12**: 209–217.

Ozgan E, Saruhan H (2010) Modeling of asphalt concrete via simulated annealing. *Advances in Engineering Software* **41**: 680–683.

Papadimitriou DI, Giannakoglou KC (2009) The continuous direct-adjoint approach for second order sensitivities in viscous aerodynamic inverse design problems. *Computers and Fluids* **38**: 1539–1548.

Papadimitriou C, Papadioti D-C (2013) Fast computing techniques for Bayesian uncertainty quantification in structural dynamics. In Simmermacher T, Cogan S, Moaveni B, Papadimitriou C, eds, Topics in Model Validation and Uncertainty Quantification, Volume 5: Proceedings of the 31st IMAC, A Conference on Structural Dynamics, pp. 25–31.

Pascual B, Adhikari S (2012a) Combined parametric-nonparametric uncertainty quantification using random matrix theory and polynomial chaos expansion. *Computers & Structures* **112–113**: 364–379.

Pascual B, Adhikari S (2012b) Hybrid perturbation-polynomial chaos approaches to the random algebraic eigenvalue problem. *Computer Methods in Applied Mechanics & Engineering* **217–220**: 153–167.

Paydar MM, Mahdavi I, Sharafuddin I, Solimanpur M (2010) Applying simulated annealing for designing cellular manufacturing systems using MDmTSP. *Computers & Industrial Engineering* **59**: 929–936.

Pelletier FJ (2000) Review of metamathematics of fuzzy logics. *Bulletin of Symbolic Logic* **6**: 342–346.

Rahmati M, Modarress H (2009) Nitrogen adsorption on nanoporous zeolites studied by grand canonical Monte Carlo simulation. *Journal of Molecular Structures: THEOCHEM* **901**: 110–116.

Ransome, T.M. (2006) Automatic minimisation of patient setup errors in proton beam therapy. Master's thesis, University of the Witwatersrand.

Ratick S, Schwarz G (2009) Monte Carlo simulation. In Kitchin R, Thrift N, eds. *International Encyclopedia of Human Geography*. Oxford: Elsevier.

Robert CP, Casella G (2004) *Monte Carlo Statistical Methods*. London: Springer.

Rodina A, Bliznakova K, Pallikarakis N (2010) End stage renal disease patients' projections using Markov chain Monte Carlo simulation. *MEDICON 2010: XII Mediterranean Conference on Medical and Biological Engineering and Computing, Satellite Event: 7th European Symposium on Biomedical Engineering –ESBME 2010, Chalkidiki, Greece, IFMBE Proceedings* Vol. 29, pp. 796–779.

Rosenbluth MN, Rosenbluth AW (1955) Monte-Carlo calculations of the average extension of macromolecular chains. *Journal of Chemical Physics* **23**: 356–359.

Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**: 2323–2326.

Rubin DB, Gelman A, Carlin JB, Stern H (2003) *Bayesian Data Analysis* (2nd edn). Boca Raton, FL: Chapman & Hall/CRC Press.

Shanno DF (1970) Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computing* **24**: 647–656.

Soize C (2013) Bayesian posteriors of uncertainty quantification in computational structural dynamics for low- and medium-frequency ranges. *Computers and Structures* **126**: 41–55.

Sun YW, Xu JT, Guo DM, Jia ZY (2009) A unified localization approach for machining allowance optimization of complex curved surfaces. *Precision Engineering* **33**: 516–523.

Tan X, Xi W, Baras JS (2010) Decentralized coordination of autonomous swarms using parallel Gibbs sampling. *Automatica* **46**: 2068–2076.

Tan Z, Lim KM, Khoo BC (2009) An immersed interface method for stokes flows with fixed/moving interfaces and rigid boundaries. *Journal of Computational Physics* **228**: 6855–6881.

Tiana G, Sutto L, Broglia RA (2007) Use of the Metropolis algorithm to simulate the dynamics of protein chains. *Physica A: Statistical Mechanics and its Applications* **380**: 241–249.

Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* **58**:267–288.

Tikhonov AN, Arsenin VY (1977) *Solutions of Ill-Posed Problems*. New York: Winston.

Torabzadeh E, Zandieh M (2010) Cloud theory-based simulated annealing approach for scheduling in the two-stage assembly flowshop. *Advances in Engineering Software* **41**: 1238–1243.

Wang, H. and Harrison, K.W. (2010) Adaptive Bayesian contaminant source characterization in water distribution systems via a parallel implementation of Markov chain Monte Carlo (MCMC). Proceedings of the World Environment and Water Research Congress, American Society of Civil Engineers, pp. 4323–4329.

Wang J, Huda A, Lunyak VV, Jordan IK (2010) A Gibbs sampling strategy applied to the mapping of ambiguous short-sequence tags. *Bioinformatics* **26**: 2501–2508.

Wishart J (1928) The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* **20A** (1–2): 32–52.

Xia J, Liu L, Xue J, Wang Y, Wu L (2009) Modeling of radiation-induced bystander effect using Monte Carlo methods. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interaction with Materials and Atoms* **267**: 1015–1018.

Xiao Y, Sun H, Wang Z (2009) A globally convergent BFGS method with non-monotone line search for non-convex minimization. *Journal of Computers and Applied Mathematics* **230**: 095–106.

Yuen K-V (2010) *Bayesian Methods for Structural Dynamics and Civil Engineering*. Singapore: John Wiley & Sons (Asia).

Zadeh LA (1965) Fuzzy sets. *Information & Control* **8**: 338–353.

Zhao H, Zheng C (2009) Correcting the multi-Monte Carlo method for particle coagulation. *Powder Technology* **193**: 120–123.

4

Metropolis–Hastings and Slice Sampling for Finite Element Updating

4.1 Introduction

This chapter uses Bayesian statistics, the Metropolis–Hastings (M-H) approach and slice sampling (SS) for probabilistic finite element model updating. Bayesian statistics is used to formulate the probability of the uncertain finite element model parameters given the observed data, which is called the posterior probability density function (PDF), and relate this to the likelihood function, which measures the distance between the finite element prediction and the measured data, the prior probability function and the evidence. Chapter 3 has described how the posterior probability function is obtained, and several methods such as the Gaussian approximation, maximum a posteriori parameter method and Monte Carlo sampling. The Gaussian approximation, which is based on the Laplace approximation method, is an estimation of the posterior probability function, and is often difficult to solve analytically. Monte Carlo sampling is more accurate than the Gaussian approximation, but is computationally more expensive.

In this chapter, we pursue two methods to approximate the posterior PDF: the M-H and the SS methods. The M-H method is a technique for sampling a PDF (Renshaw, 2004; Strid, 2010; Johnson and Flegal, 2014) which has been applied to sample the Bose–Einstein condensates (Grišins and Mazets, 2014), to efficiently sample a proposal distribution (Shao *et al.*, 2013), in malaria diagnosis (Bauer *et al.*, 2014) and in a jet-milling model (Kastner *et al.*, 2013). SS is also a method which has been used to sample a PDF, and one example of this method is the parallel SS (Pietrabissa and Rusconi, 2014) and elliptical SS (Nishihara *et al.*, 2014). It has been applied to assess ventricular functions in magnetic resonance imaging (Mazonakis *et al.*, 2011), for parameter estimation (Hatjispyros *et al.*, 2007), for modelling the gold price (Rostami *et al.*, 2013) and for rail inspection (Nieto *et al.*, 2012).

This chapter is organised as follows. The posterior probability function within the finite element model updating problem is formulated. Thereafter, the M-H and the SS algorithms

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

are described. Then the statistical measures used in this chapter are discussed, and the sampling methods are used to update a cantilevered beam and an asymmetrical H-shaped structure.

4.2 Likelihood, Prior and the Posterior Functions

The likelihood function is the probability of measurements in the presence of uncertain parameters (Canary *et al.*, 2014; Butler *et al.*, 2014). Likelihood functions have been formulated in a number of ways, including the use of fuzzy logic (Coletti *et al.*, 2014), and have been applied in environmental modelling (Joseph and Guillaume, 2013) and for model calibration in hydrology models (Cheng *et al.*, 2014). This function can be defined as the normalised exponent of the error function that represents the differences between the measured and the analytical modal properties (Marwala, 2010; Boulkaibet, 2014):

$$P(\mathcal{D}|\boldsymbol{\theta}) = \frac{1}{(2\pi/\beta)^{N_m/2}\prod_{i=1}^{N_m} f_i^m} \exp\left(-\frac{\beta_c}{2}\sum_{i}^{N_m}\left(\frac{f_i^m - f_i}{f_i^m}\right)^2\right), \tag{4.1}$$

where $\boldsymbol{\theta}$ is the vector representing parameters to be updated, β_c is a constant, f_i^m is the ith measured natural frequency ($\omega_i^m = 2\pi f_i^m$ and ω_i^m is the ith measured circular natural frequency), N_m is the number of measured modes and $f_i = f_i(\boldsymbol{\theta})$ is the ith analytical frequency obtained from the finite element model. The likelihood function can be modified by including the error function that represents the differences between the analytical and measured eigenvectors. However, the eigenvector sensitivities are less than the eigenvalue (natural frequencies) sensitivities in model updating problems. In this book, only natural frequencies are involved in the model updating process.

The prior PDF denotes the prior knowledge about the updating parameters ($\boldsymbol{\theta}$). For structural systems, the prior knowledge could be the observation that parameters near joints should be updated more intensely than for those corresponding to smooth surface areas far from joints. The prior PDF for parameters $\boldsymbol{\theta} = (\theta_i, \ldots, \theta_Q)$ in Equation 4.1 is assumed to be Gaussian (for more information on this, see Appendix C) and is given by (Marwala, 2010; Boulkaibet, 2014)

$$P(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{Q/2}\prod_{i=1}^{Q}\left(1/\sqrt{\alpha_i}\right)} \exp\left(-\sum_{i}^{d}\frac{\alpha_i}{2}\left\|\theta_i - \theta_i^0\right\|^2\right), \tag{4.2}$$

where d is the number of parameters to be updated, $\boldsymbol{\theta}_0 = \left(\theta_1^0, \ldots, \theta_d^0\right)$ represents the initial values of the updated vector and α_i is the coefficient of the prior PDF for the ith updated parameter. The notation $\| * \|$ denotes the Euclidean norm. If α_i is constant for all the updating parameters in Equation 4.1 then the updated parameters will be of the same order of magnitude. Equation 4.1 may be regarded as a regularisation parameter that controls the amount of change of the updated parameters at each iteration (Marwala, 2010).

In Equation 4.1, Gaussian priors are conveniently chosen as various natural processes tend to have a Gaussian distribution. The posterior distribution function of the parameters $\boldsymbol{\theta}$ given the observed data $\mathcal{D}$ is denoted by $P(\boldsymbol{\theta}|\mathcal{D})$, and is obtained by applying Bayes' theorem (Bishop,

2006). The distribution $P(\boldsymbol{\theta}|\mathcal{D})$ is estimated by substituting Equations 4.1 and 4.2 into the Bayesian equation (see Equation 3.2) to give (Marwala, 2010; Boulkaibet, 2014)

$$P(\boldsymbol{\theta}|\mathcal{D}) \propto \frac{1}{Z_s(\alpha,\beta)} \exp\left(-\frac{\beta_c}{2} \sum_{i}^{N_m} \left(\frac{f_i^m - f_i}{f_i^m} \right)^2 - \sum_{i}^{d} \frac{\alpha_i}{2} \left\| \theta_i - \theta_i^0 \right\|^2 \right), \tag{4.3}$$

where

$$Z_s(\alpha,\beta) = \left(\frac{2\pi}{\beta_c} \right)^{N_m/2} \prod_{i=1}^{N_m} f_i^m (2\pi)^{d/2} \prod_{i=1}^{d} \frac{1}{\sqrt{\alpha_i}}. \tag{4.4}$$

Equation 4.3 can be written in the following normal distribution form:

$$P(\boldsymbol{\theta}|\mathcal{D}) = \frac{1}{(2\pi)^{(N_m+d)/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2} (\boldsymbol{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\mu}) \right), \tag{4.5}$$

where $\boldsymbol{X} = [f_1, \ldots, f_{N_m}, \theta_1, \ldots, \theta_d]$, $\boldsymbol{\mu} = \left[f_1^m, \ldots, f_{N_m}^m, \theta_1^0, \ldots, \theta_d^0 \right]$ and

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \frac{\beta_c}{f_1^{m^2}} & 0 & \cdots & & & \\ 0 & \ddots & 0 & & 0 & \\ \vdots & 0 & \frac{\beta_c}{f_{N_m}^{m^2}} & & & \\ & & & \alpha_1 & 0 & \vdots \\ & 0 & & 0 & \ddots & 0 \\ & & & \cdots & 0 & \alpha_d \end{bmatrix}. \tag{4.6}$$

In most cases, the analytical form of the posterior PDF solution is not obtainable because of the complexity of the problem, as well as because of the high dimensionality of the parameter search space. If Y is the observation of certain parameters at different discrete time instances, then the total probability theorem provides the probabilistic information for the prediction of the future responses Y at different time instances, and this can be written as follows (Bishop, 2006):

$$P(Y|\mathcal{D}) = \int P(Y|\boldsymbol{\theta}) P(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta}. \tag{4.7}$$

Equation 4.7 depends on the posterior distribution function, and the dimension of the updating parameters and the complexity of the modelled system make it very difficult to obtain an analytical solution.

Different methods have been proposed to estimate the probability distribution, with the maximum likelihood being the most commonly used method. As described in the previous chapter, the maximum likelihood method represents an asymptotic approximation to the full Bayesian solution and can be applied in the case where the unknown parameters are modelled as Gaussian. The most probable values are obtained by maximising the likelihood function, and the covariance matrix is obtained by using the Hessian of the likelihood function. However, in many cases, this approach may not give an accurate prediction due to the complexity of the problem. Moreover, this approach does not involve the prior knowledge of the updating process. On the other hand, the sampling methods provide a practical solution to estimating this function (posterior). Sampling techniques can simplify the Bayesian inference by providing a set of random samples from the posterior distribution (Marwala, 2010; Boulkaibet *et al.*, 2014). Given a set of N_s random parameter vectors drawn from $P(\boldsymbol{\theta}|D)$, the expectation value of any observed function Y can be easily estimated (Bishop, 2006).

These algorithms are used to generate a sequence of vectors $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{N_s}\}$, where N_s is the number of samples drawn from the posterior function. This generated vector is then used to predict the form of the posterior distribution function $P(\boldsymbol{\theta}|D)$. The integral in Equation 4.7 can be approximated as (Boulkaibet, 2014)

$$\tilde{Y} \cong \frac{1}{N_s} \sum_{i=1}^{N_s} G(\boldsymbol{\theta}_i), \tag{4.8}$$

where G is a function that depends on the updated parameters $\boldsymbol{\theta}_i$. As an example, if $\mathbf{G} = \boldsymbol{\theta}$ then $\tilde{Y}$ becomes the expected value of $\boldsymbol{\theta}$. Generally, $\tilde{Y}$ is the vector that contains the modal properties while N_s is the number of retained states.

Several sampling techniques such as multivariate normal sampling (Khodaparast, 2010), Latin hypercube sampling (LHS; Iman, 2008) and orthogonal array (OA) sampling (Zhang *et al.*, 2009) have been used in finite element model updating. Multivariate normal sampling techniques can only be applied for uncertain parameters which belong to Gaussian posterior distribution function. In the case where the parameters are uncorrelated, the covariance matrix is then diagonal. But when the uncertain parameters are correlated the sampling process becomes more difficult, and the accuracy of this method rapidly decreases when a complex system is tested. The LHS method was introduced by McKay *et al.* (1979). Its objective is to divide the parameter space into subspaces of equal probability. Samples are then taken from each subspace, ensuring that every parameter is covered equally. This method is very effective in the case where only one parameter is sampled. This method becomes more complex, impractical and computationally expensive when dealing with high-dimensional problems since all possible combinations of the parameters need to be considered. The OA sampling method has the same basics as the LHS method (Koehler and Owen, 1996). It produces uniform samples in multi-dimensional search space, while the LHS method produces a uniform sampling in one dimension only (LHS is special case of OA sampling).

However, the results obtained from all three sampling algorithms are greatly affected by the complexity of the system and the size of the search space. The most popular class of sampling methods are the Markov chain Monte Carlo (MCMC) methods, which allow sampling from a large class of distributions and are able to handle large search spaces. More information on the MCMC is given in Appendix B. The MCMC has been applied in structure learning (Larjo and

Lähdesmäki, 2015), in big data analytics (Mahani and Sharabiani, 2015), in genetics data (Tasaki *et al.*, 2015) and in astrophysics (Greig and Mesinger, 2015).

4.3 The Metropolis–Hastings Algorithm

The M-H algorithm is one of the simplest MCMC methods that can draw samples from multivariable densities (Metropolis *et al.*, 1953; Hastings, 1970; Roberts and Smith, 1994; Chib and Greenberg, 1995; Heckman and Leamer, 2001; Gilks, 2005; Calderhead, 2014; Vu *et al.*, 2014; Van Dyk and Jiao, 2015). This algorithm is related to rejection and importance sampling, and the general idea is to propose a PDF and then use it to generate proposed values (Marwala, 2010; Boulkaibet *et al.*, 2012). The proposed distribution is also used to obtain a move probability which is used to determine whether or not the drawn value should be accepted as the next state of the Markov chain. The move probability is defined by the product of the ratio of the target density and the ratio of the proposed density. This means that a normalising constant of the target density distribution function is not required in this algorithm.

In the M-H algorithm, to sample from the posterior distribution function $P(\boldsymbol{\theta}|\mathcal{D})$, which is the target distribution function, where $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_d\}$ is a d-dimensional parameter vector, the proposal density distribution $q(\boldsymbol{\theta}|\boldsymbol{\theta}_{t-1})$ is introduced in order to generate a random vector $\boldsymbol{\theta}$ given the value at the previous iteration of the algorithm. The M-H algorithm consists of two basic steps: the sample drawn from the proposed density step and the accept or reject step of the proposed state. The M-H algorithm can be summarised as follows (Hastings, 1970; Brooks *et al.*, 2011; Joubert, 2015; Boulkaibet, 2014):

1. Initiate the algorithm with a value $\boldsymbol{\theta}_0$.
2. At iteration t, $\boldsymbol{\theta}^*$ is drawn from the proposed probability distribution density $q(\boldsymbol{\theta}|\boldsymbol{\theta}_{t-1})$, where $\boldsymbol{\theta}_{t-1}$ is the parameter value of the previous step.
3. Update the finite element model to obtain the new analytical frequencies, then compute the acceptance probability, given by

$$a(\boldsymbol{\theta}^*, \boldsymbol{\theta}_{t-1}) = \min\left\{1, \frac{P(\boldsymbol{\theta}^*|\mathcal{D})q(\boldsymbol{\theta}_{t-1}|\boldsymbol{\theta}^*)}{P(\boldsymbol{\theta}_{t-1}|\mathcal{D})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_{t-1})}\right\}.$$

4. Draw u from a uniform distribution $U(0, 1)$.
5. If $u \leq a(\boldsymbol{\theta}^*, \boldsymbol{\theta}_{t-1})$ then accept state $\boldsymbol{\theta}^*$. Otherwise, reject state $\boldsymbol{\theta}^*$.
6. Return to Step 2.

Figures 4.1 and 4.2 show the results of running a single variable sampling using the M-H algorithm (Boulkaibet, 2014). The proposal distribution is a normal distribution with mean and variance equal to 1, $q(\theta) = \mathcal{N}(1,1)$, and the single variable target distribution, which is given by $P(\theta) \propto e^{-\theta^2}(2 + \sin(5\theta) + \sin(2\theta))$, is plotted for 100 and 1000 iterations (samples), respectively.

The samples for both cases are also plotted. As expected, the histogram of the samples approximates the target distribution in both cases. However, 100 iterations (or samples) are not enough to give a good representation of the target distribution because the acceptance rate (AR) of the M-H algorithm is only 59%, which means only 59 samples were used to

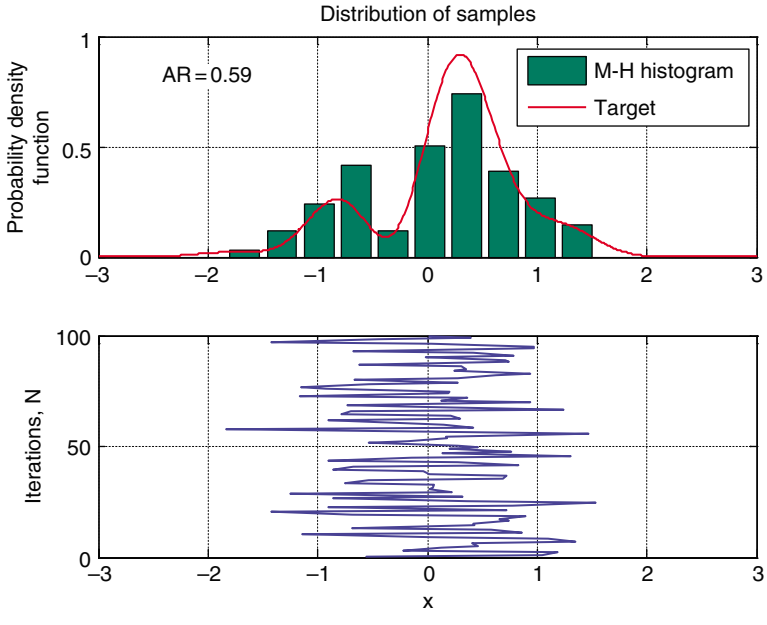

Figure 4.1 Target distribution and histogram of the M-H samples with 100 iterations

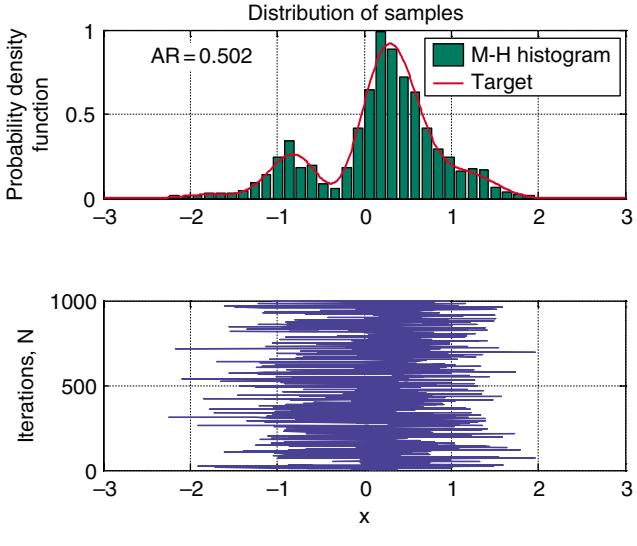

Figure 4.2 Target distribution and histogram of the M-H samples with 1000 iterations

approximate the target distribution function. The AR of the M-H algorithm can be improved by reducing the move step of the algorithm (decreasing the standard deviation of the proposal distribution). Increasing the number of iterations to 1000 leads to an almost perfect prediction of the form of the target distribution.

4.4 The Slice Sampling Algorithm

The general idea of the SS algorithm was first introduced by Swendsen and Wang (1987). In this algorithm auxiliary variables are implemented to improve the efficiency of the MCMC sampling technique. The SS algorithm was advanced by a number of researchers. Besag *et al.* (1993) proposed a similar algorithm and applied it in agriculture. Higdon (1996) introduced an improved auxiliary variable technique for the MCMC method and applied it to Bayesian image analysis. Additional information on the SS algorithm can be found in Edwards and Sokal (1988), Fishman (1999), Roberts and Rosenthal (1999), Mira *et al.* (2001) and Lu (2008). The SS technique is a type of MCMC algorithm that provides adaptive step size which automatically adjusts to match the characteristics of the posterior PDF (Neal, 2003; Bishop, 2006; Tibbits *et al.*, 2011; Thompson, 2011). The SS technique does not involve any extra distribution to help in drawing samples, and the idea of the method is to uniformly sample from a region under the target distribution curve. Supposing that the probability function $f(x)$ of a variable x is the target distribution, a set of samples can be drawn from a region that is under the plot of $f(x)$.

This can be organised by introducing an auxiliary variable y to create an extended uniform joint distribution $P(x, y)$ as follows (Swendsen and Wang, 1987):

$$P(x, y) = \begin{cases} \dfrac{1}{Z}, & \text{if } 0 < y < P(x), \\ 0, & \text{otherwise}, \end{cases} \tag{4.9}$$

where $Z = \int f(x)dx$ and the density of x is given by $P(x) = f(x)/Z$. The new samples for x are attained by drawing candidates from the joint distribution $P(x, y)$ while the y values are ignored.

Figure 4.3 explains the drawing of a new sample in an example with a single variable. Initially, the x_i is drawn from a uniform distribution between 0 and $f(x_i)$ and called y_{i+1}. Then the

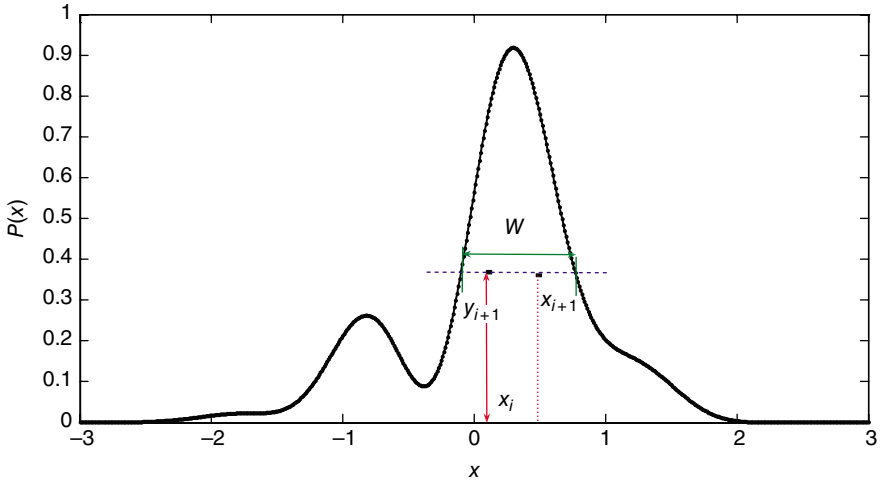

Figure 4.3 Slice sampling approach

interval w is created as presented in Figure 4.3 and a new sample x_{i+1} is uniformly drawn from the interval.

On the whole, the objective of the SS technique is to sample uniformly from the area under the posterior distribution $P(\boldsymbol{\theta}|\mathcal{D})$ where $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_d\}$. The multivariate SS with hyper-rectangles algorithm can be explained as follows (Kyprianou *et al.*, 2001; Cheung and Beck, 2010; Boulkaibet, 2014):

1. Sample Y from the uniform distribution $U(0, P(\boldsymbol{\theta}_0|\mathcal{D}))$.
2. Initiate the interval around $\boldsymbol{\theta}_0$ as by following these steps:
   For i = 1 to d
   $$U_i \sim U(0,1)$$
   $$L_i \leftarrow \theta_{0,i} - w_i U_i$$
   $$R_i \leftarrow L_i + w_i$$
   End
3. Draw sample from the interval $I = (R, L)$ and perform the following steps:
   Repeat:
   For i = 1 to d
   $$U_i \sim U(0,1)$$
   $$\theta_{j,i} \leftarrow L_i + U_i(R_i - L_i)$$
   End
   If $Y \leq P(\boldsymbol{\theta}_j|\boldsymbol{\theta})$ then exit loop
   For $i = 1$ to N
   $$\text{If } \theta_{j,i} < \theta_{o,i} \text{ then } L_i \leftarrow \theta_{j,i} \text{ else } R_i \leftarrow \theta_{j,i}$$
   End
4. Repeat Step 3 to obtain N_s samples.

The SS algorithm has excellent AR; normally a 100% AR is achieved at the end of the algorithm run. Nevertheless, the efficiency of this algorithm depends on the selected move step and is influenced by the interval w_i used in the pseudo-code. Large moves may be computationally expensive because it is difficult for the algorithm to identify the correct sampling interval.

The efficiency of this algorithm for high-dimensional problems can be improved by reducing the number of random moves. In the SS algorithm, the random walk suppression is conducted by reflecting from the edges of the slice (Neal, 2003).

4.5 Statistical Measures

The advantages of using Bayesian methods in finite element model updating are that the uncertainty of the attained parameters is represented and the correlations between the uncertain

parameters are measured. In this section the mathematical tools used to calculate the estimated mean value, variance, covariance and correlation matrices are explained. The approximated mean value of the estimated parameters is expressed as follows (Boulkaibet, 2014):

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}) \cong \frac{1}{N_s}\sum_{i=1}^{N_s} \boldsymbol{\theta}^i, \tag{4.10}$$

where $E(\boldsymbol{\theta})$ is the mean value of $\boldsymbol{\theta}$ and N_s is the number of samples. The uncertainty of the estimated mean parameters can be characterised by the variance, computed as follows (Boulkaibet, 2014):

$$V\left(\hat{\boldsymbol{\theta}}\right) = E\left(\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)^2\right) \cong \frac{1}{N_s}\sum_{i=1}^{N_s}\left(\boldsymbol{\theta}^i - \hat{\boldsymbol{\theta}}\right)^2, \tag{4.11}$$

where the standard deviation is given by $\boldsymbol{\sigma}_{\boldsymbol{\theta}} = \sqrt{V\left(\hat{\boldsymbol{\theta}}\right)}$. The covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is used to identify the correlation between the estimated parameters and is written as follows (Boulkaibet, 2014):

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \begin{bmatrix} c_{11} & c_{21} & \cdots & c_{d1} \\ c_{12} & c_{22} & \cdots & c_{d2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{d1} & c_{d2} & \cdots & c_{dd} \end{bmatrix}, \tag{4.12}$$

where $c_{ij} = \operatorname{cov}\left(\theta_i, \theta_j\right) = E\left(\left(\theta_i - \hat{\theta}_i\right).\left(\theta_j - \hat{\theta}_j\right)\right) \cong \frac{1}{N_s}\sum_{i=1}^{N_s}\left(\theta_i - \hat{\theta}_i\right).\left(\theta_j - \hat{\theta}_j\right)$.

A different way to explain the correlation between the predicted parameters is by computing the statistical correlation matrix. and this is similar to the covariance matrix. However, the computed parameters values are bounded between –1 and 1:

$$\boldsymbol{\Gamma}_{\boldsymbol{\theta}} = \begin{bmatrix} r_{11} & r_{21} & \cdots & r_{d1} \\ r_{12} & r_{22} & \cdots & r_{d2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{d1} & r_{d2} & \cdots & r_{dd} \end{bmatrix}, \tag{4.13}$$

where $r_{ij} = \dfrac{\operatorname{cov}\left(\theta_i, \theta_j\right)}{\sqrt{E\left(\left(\theta_i - \hat{\theta}_i\right)^2\right)E\left(\left(\theta_j - \hat{\theta}_j\right)^2\right)}}$.

In the following sections, we apply the M-H and SS algorithms to update a simple cantilever beam and an asymmetrical H-shaped aluminium structure, the details of which are given in Appendix A.

4.6 Application 1: Cantilevered Beam

An experimental cantilever steel beam which was designed and measured by Kraaij (2006) was modelled using finite element modelling, and this model was updated based on the measurements made on the cantilevered beam. The beam had length 500 mm, width 60 mm and thickness 10 mm, and the material properties were modelled using $E = 2.1 \times 10^{11}\,\mathrm{N/m^2}$, $\nu = 0.3$ and $\rho = 7850\,\mathrm{kg/m^3}$. Three accelerometers were used for measurements and were positioned 490 mm from the clamped end. This position was chosen because the response at this position was large. Each accelerometer had a mass of 40 g; the middle accelerometer was of type 303*A*3, the outer accelerometers were of type 303*A*2. To test the updating techniques, the beam was modelled using Version 6.3 of the Structural Dynamics Toolbox SDT® for MATLAB (Boulkaibet, 2014).

The beam was segmented into 50 Euler–Bernoulli beam elements and excited at a number of positions. The measured natural frequencies of interest of this structure were: 31.9, 197.9, 553, 1082.2 and 1781.5 Hz (Kraaij, 2006; Boulkaibet, 2014). The Young's modulus of the beam elements was used as an updating parameter, where for every 10 elements a different Young's modulus was assigned. Therefore, the parameters to be updated were represented by a vector of dimension $d = 5$, represented as $\boldsymbol{\theta} = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$.

This vector of five parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ was updated using the Bayesian approach as described in this chapter. N_s samples of the vector $\boldsymbol{\theta}$ were generated from the posterior distribution function, $P(\boldsymbol{\theta}|D)$, as given in Equation 4.3. The constant β_c was set equal to 1 and the coefficients α_i were set equal to $1/\sigma_i^2$, where σ_i^2 was the variance of the parameter θ_i, and this means that parameter $\boldsymbol{\theta}$ is uncertain with a standard deviation of $\boldsymbol{\sigma}$. Because the updating parameter vector comprises only the Young's modulus, all σ_i were set equal to 2×10^{11}. A large value of σ_i was used to allow the algorithm more freedom and thus to focus more on the likelihood term during sampling. The updating parameters θ_i were bounded with a maximum value of 2.5×10^{11} and a minimum value of 1.7×10^{11}.

The number of samples N_s was 1000 for all methods, and the initial value of $\boldsymbol{\theta}$ was set at $\boldsymbol{\theta}_0 = \{2.4 \times 10^{11}, 2.4 \times 10^{11}, 2.4 \times 10^{11}, 2.4 \times 10^{11}, 2.4 \times 10^{11}\}$. As an alternative to using the mean steel value of $\boldsymbol{\theta}$ as an initial value, a large value of the initial parameter vector was selected to emphasise the updating process. Each algorithm was implemented over 30 independent runs, and the final results are shown in Tables 4.1 and 4.2 which are the average of these 30 runs.

Table 4.1 The updated vector of Young's modulus using the M-H and SS techniques

	Young's modulus ($\mathrm{N/m^2}$)				
	Initial	M-H method	$\frac{\sigma_i}{\mu_i}$ (c.o.v., %)	Slice sampling method	$\frac{\sigma_i}{\mu_i}$ (c.o.v., %)
E_1	2.4×10^{11}	2.111×10^{11}	7.00	2.125×10^{11}	7.34
E_2	2.4×10^{11}	2.110×10^{11}	7.09	2.094×10^{11}	8.10
E_3	2.4×10^{11}	2.100×10^{11}	7.10	2.095×10^{11}	8.07
E_4	2.4×10^{11}	2.098×10^{11}	7.13	2.093×10^{11}	8.27
E_5	2.4×10^{11}	2.101×10^{11}	7.14	2.077×10^{11}	7.69

Table 4.2 Frequencies and errors when the M-H and SS techniques used to update Young's modulus

Measured frequency (Hz)	Initial frequency (Hz)	Frequencies (errors) M-H method (Hz)	Frequencies (errors) slice sampling method (Hz)
31.90	32.70	30.70 (2.28%)	30.7 (2.45%)
197.90	209.40	196.10 (1.78%)	196.0 (2.08%)
553.00	594.80	556.90 (1.74%)	556.0 (2.43%)
1082.20	1177.80	1102.70 (1.52%)	1100.90 (2.02%)
1781.50	1961.70	1.836.90 (1.51%)	1833.70 (2.09%)

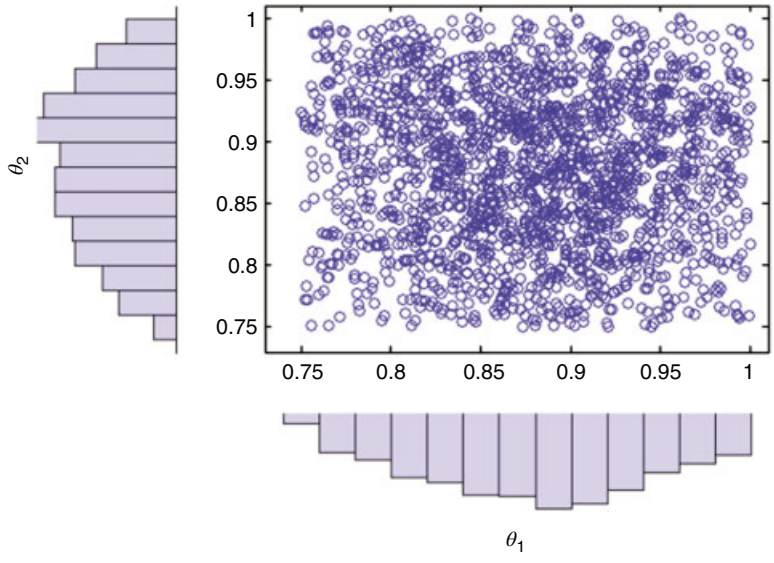

Figure 4.4 Scatter plot of θ_1 versus $\theta_2$with marginal histograms using the M-H method

Figures 4.4–4.7 show the scatter plots with marginal histograms for four uncertain parameters using the M-H and SS algorithms. The plots show that all algorithms found the area of high probability. It was observed that both the M-H and SS algorithms require very large samples ($N_s >> 1000$) to reach their equilibrium.

The updated parameters, also presented in Table 4.1, demonstrate that the M-H method gave results that are close to the mean value for steel of $2.1 \times 10^{11}\,\mathrm{N/m^2}$. The SS technique gave updated parameters that were close to the mean value. The SS method applied in this chapter updated individual vector entries as opposed to updating all entries simultaneously. The SS algorithm took a larger number of samples to converge and required more samples when compared to the M-H algorithm.

The coefficient of variation (c.o.v.) denotes the errors related to the updated parameters, and it is computed by dividing the standard deviation vector by the approximated $\boldsymbol{\theta}$ vector.

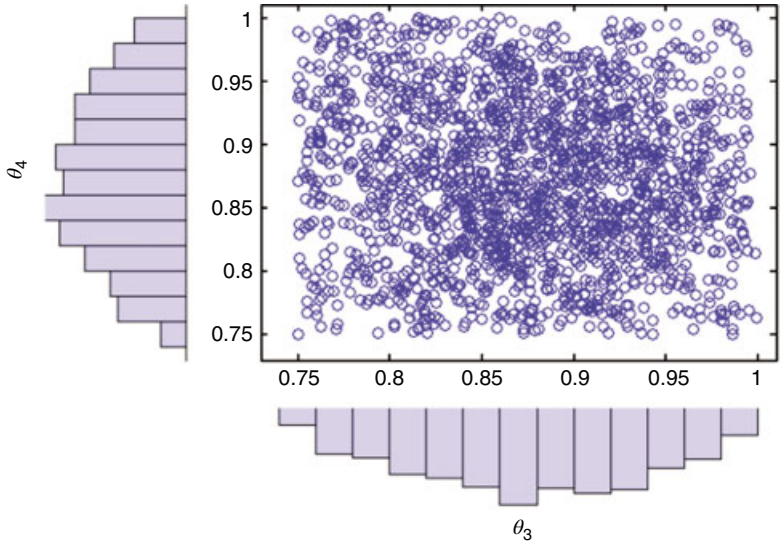

Figure 4.5 Scatter plot of θ_3 versus θ_4 with marginal histograms using the M-H method

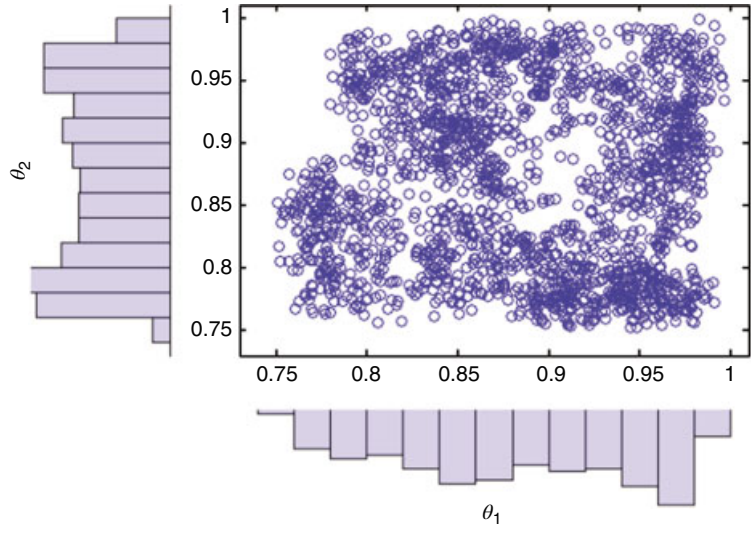

Figure 4.6 Scatter plot of θ_1 versus θ_2 with marginal histograms using the SS method

Table 4.1 indicates that the c.o.v. values were small for all algorithms (less than 8% and 9% for the M-H and SS algorithms, respectively), and that the two algorithms estimated the uncertain parameters efficiently (Boulkaibet, 2014). Figures 4.8 and 4.9 demonstrate the correlation between all updated parameters for the M-H and SS algorithms (Boulkaibet, 2014). Small values indicate that both parameters are weakly correlated (<0.3), while large values

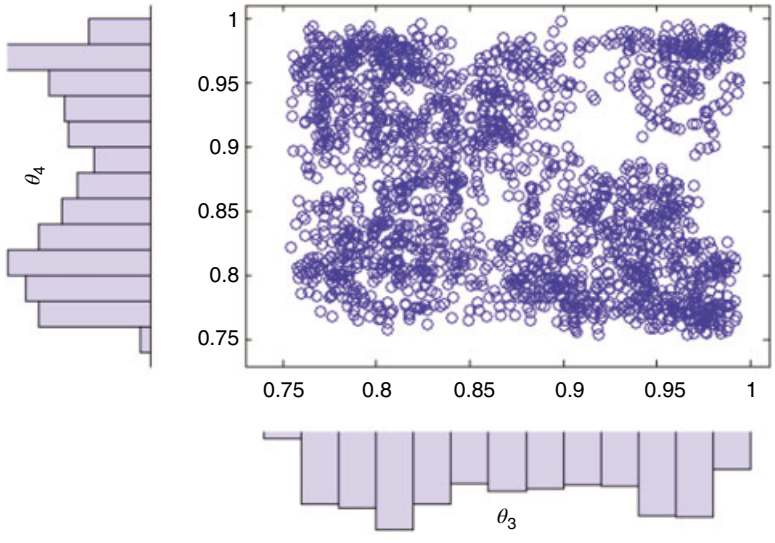

Figure 4.7 Scatter plot of θ_3 versus θ_4 with marginal histograms using the SS method

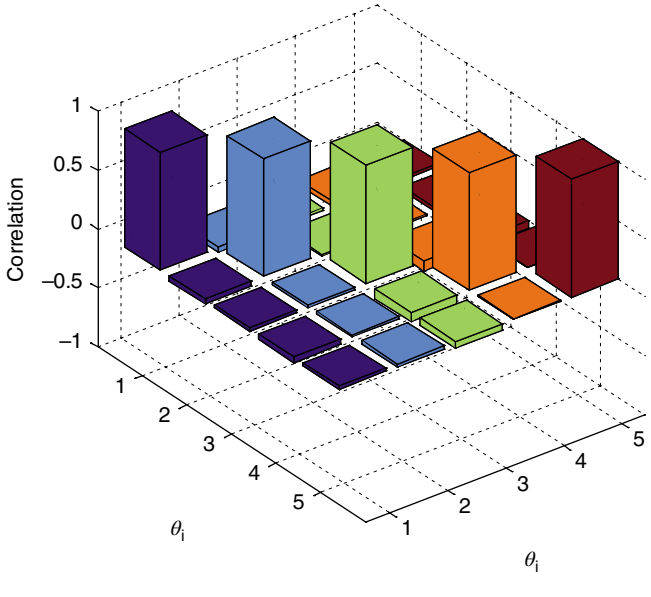

Figure 4.8 The correlation between the updated parameters using the M-H algorithm

(>0.7) indicate that the parameters are highly correlated, and a zero value indicates that the parameters are uncorrelated. A positive correlation indicates that the variables are positively related, while a negative correlation indicates the opposite. The results indicate that for all algorithms, all parameters are weakly correlated.

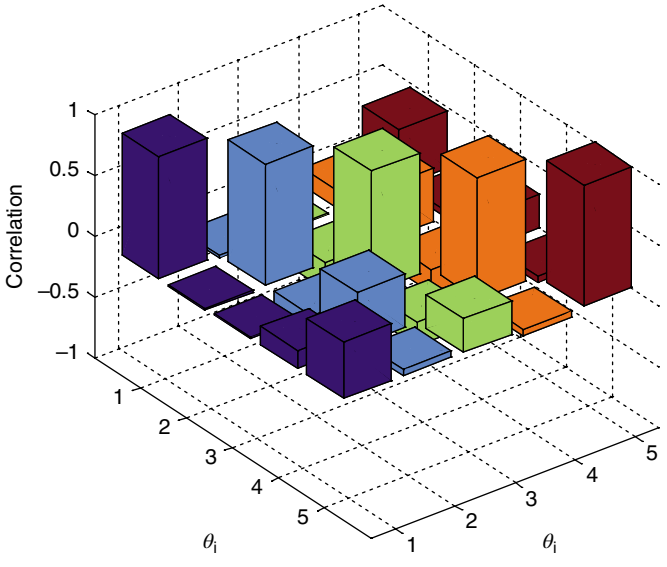

Figure 4.9 The correlation between the updated parameters using the SS algorithm

The results for the updated modes given in Table 4.2 demonstrate that the two sampling methods give results which are, on average, better than the initial finite element model and use the posterior density function in different ways since each algorithm has a different way of generating samples. In the SS technique, each variable is changed one at a time, which is not the case with the M-H algorithm where all the parameters are varied at once.

The absolute mode errors ($\left|f_i^m - f_i\right|/f_i^m$) and the final model percentage error for both algorithms demonstrate that the SS algorithm gives better results than those obtained using the M-H method. Both the M-H and SS algorithms gave relatively high error percentages (3.90% for the M-H algorithm and 3.84% for the SS algorithm). The error between the third measured natural frequency and that of the initial model was 7.55%; when the M-H method was used this error was reduced to 0.71%, whereas the SS method reduced it to 0.54%. The results obtained show that the M-H method was faster than the SS method as the M-H algorithms required a few samples to converge.

4.7 Application 2: Asymmetrical H-Shaped Structure

The asymmetrical H-shaped aluminium structure was updated using the previous two algorithms (Marwala, 1997) and the details are in Appendix A. The structure was divided into 12 elements, each modelled as an Euler–Bernoulli beam. The beams were modelled using Version 6.3 of the Structural Dynamics Toolbox SDT® for MATLAB. The structure was excited at a specified position and the acceleration was measured at 15 different positions. The structure was excited using an electromagnetic shaker, while a roving accelerometer was used to measure the response. A set of 15 frequency-response functions were calculated. The measured natural frequencies were 53.9, 117.3, 208.4, 254.0 and 445.0 Hz. In this example, the moments of

inertia and the cross-sectional areas of the left, middle and right subsections of the beam were updated. The updating parameter vector was $\boldsymbol{\theta} = \{I_{x1}, I_{x2}, I_{x3}, A_{x1}, A_{x2}, A_{x3}\}$. The Young's modulus for the beam was fixed at $7.2 \times 10^{10} \mathrm{N/m^2}$, and the density was set to $2785\ \mathrm{kg/m^3}$. The updating parameters θ_i were bounded by maximum values equal to $[3.73 \times 10^{-8}, 3.73 \times 10^{-8}, 3.73 \times 10^{-8}, 4.16 \times 10^{-4}, 4.16 \times 10^{-4}, 4.16 \times 10^{-4}]$, and the minimum values were set at $[1.73 \times 10^{-8}, 1.73 \times 10^{-8}, 1.73 \times 10^{-8}, 2 \times 10^{-4}, 2.16 \times 10^{-4}, 2.16 \times 10^{-4}]$. The boundaries helped to keep the updated vector physically realistic. The constant β_c of the posterior distribution was fixed equal to 10. The coefficients α_i were set equal to $1/\sigma_i^2$, where σ_i^2 is the variance of the ith parameter, and the variance vector was defined as $\boldsymbol{\sigma} = [5 \times 10^{-8}, 5 \times 10^{-8}, 5 \times 10^{-8}, 5 \times 10^{-4}, 5 \times 10^{-4}, 5 \times 10^{-4}]$. The constant β_c and the variance $\boldsymbol{\sigma}$ were chosen so that the weight of the likelihood terms was greater than the second term in the posterior distribution function. The number of samples N_s was set to 1000.

The Bayesian simulation results are presented in terms of the mean values of the obtained samples for each method. Each algorithm was implemented over 20 independent runs and the results are shown in Tables 4.3 and 4.4.The initial and updated values of the updating vector obtained by each method show that both algorithms successfully updated the $\boldsymbol{\theta}$ vector where the updated values are different than the initial $\boldsymbol{\theta}_0$. The c.o.v. values obtained by M-H and SS algorithms were large, and the reason for this is that large move steps were used for both M-H and SS algorithms to ensure a fast convergence and to improve the total average errors. Figures 4.10 and 4.11 show the correlation between all updated parameters for both algorithms. The figures

Table 4.3 Initial and updated parameters using the SS and M-H algorithms

	Initial $\boldsymbol{\theta}_0$	$\boldsymbol{\theta}$ vector SS method	$\frac{\sigma_i}{\mu_i}$ (c.o.v., %)	$\boldsymbol{\theta}$ vector M-H method	$\frac{\sigma_i}{\mu_i}$ (c.o.v., %)
I_{x1}	2.73×10^{-8}	2.33×10^{-8}	13.37	2.31×10^{-8}	22.59
I_{x2}	2.73×10^{-8}	2.43×10^{-8}	19.71	2.68×10^{-8}	15.25
I_{x3}	2.73×10^{-8}	2.40×10^{-8}	22.37	2.17×10^{-8}	13.96
A_{x1}	3.16×10^{-4}	2.82×10^{-4}	20.40	2.85×10^{-4}	14.36
A_{x2}	3.16×10^{-4}	2.83×10^{-4}	19.93	2.83×10^{-4}	14.36
A_{x3}	3.16×10^{-4}	2.84×10^{-4}	18.09	2.77×10^{-4}	13.08

Table 4.4 Natural frequencies and errors when the SS and M-H algorithms were used to update the parameters

Measured frequency (Hz)	Initial frequency (Hz)	Frequencies (errors) SS method (Hz)	Frequencies (errors) M-H method (Hz)
53.90	51.40	53.87 (7.01%)	53.92 (3.96%)
117.30	116.61	121.71 (7.69%)	122.05 (4.28%)
208.40	201.27	206.80 (7.94%)	210.93 (4.95%)
254.00	247.42	258.56 (8.16%)	258.94 (4.81%)
445.00	390.33	407.19 (8.22%)	410.33 (4.74%)

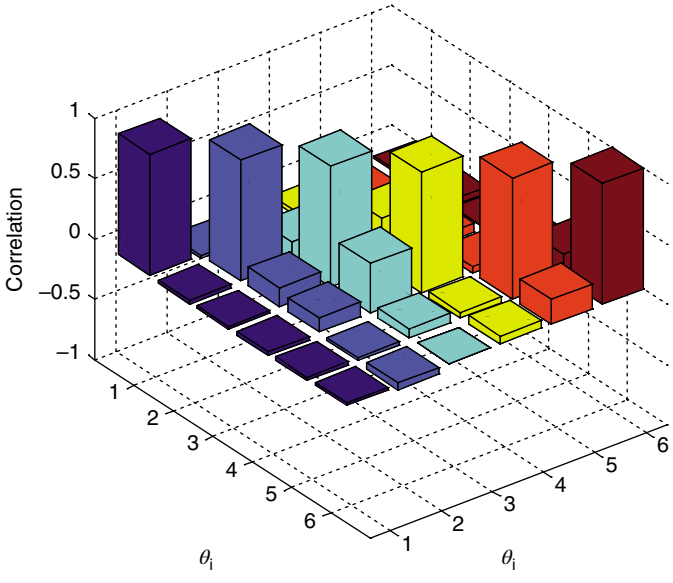

Figure 4.10 The correlation between the updated parameters using the SS algorithm

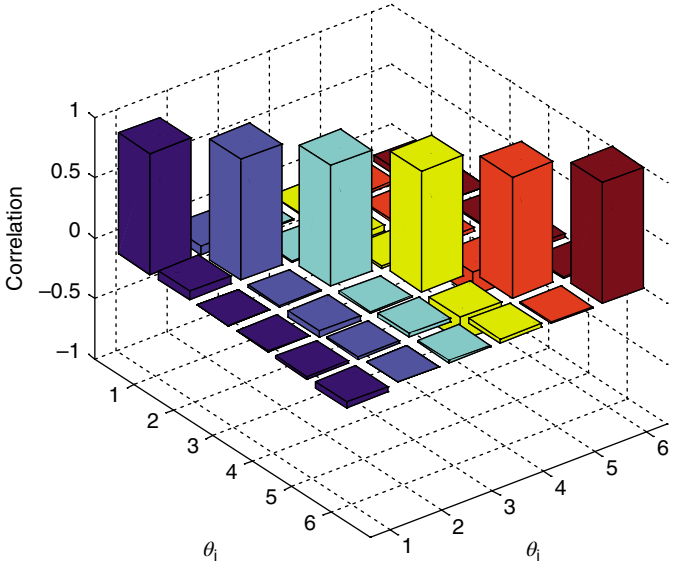

Figure 4.11 The correlation between the updated parameters using the M-H algorithm

indicate that all parameters are correlated for both algorithms (all values are non-zero), where most of the parameters are weakly correlated.

Table 4.4 shows the finite element updated frequencies, and the c.o.v. values in brackets. The error between the first measured natural frequency and that of the initial model was 4.63%, and

when the M-H and SS algorithms were used it was reduced to 0.04 and 0.06%, respectively. The overall updated finite element model natural frequencies for all three algorithms were better than the initial ones. Both the M-H and SS algorithms had high convergence rate because large move steps were used in these algorithms, but the move steps used in these two algorithms affected the acceptance rate of the M-H algorithm, giving an AR of 46.9%, while the running time of the SS algorithm was almost five times that of the M-H algorithm.

4.8 Conclusions

This chapter presented the M-H and SS algorithms for finite element model updating of two structures. The M-H algorithm's performance decreased with the complexity of the system and the size of the uncertain vector. The SS algorithm exhibited poorer performance than the M-H algorithm and was found to be computationally less efficient than the M-H algorithm. Both the M-H and SS algorithms performed poorly with small move steps. A large move step was needed in order to reduce the average errors for these algorithms. The SS algorithm with large move step performed worse and with very long running time compared to the M-H method.

References

Bauer R, Mentré F, Kaddouri H, Le Bras J, Le Nagard H (2014) Benefits of a new Metropolis–Hasting based algorithm, in non-linear regression for estimation of ex vivo antimalarial sensitivity in patients infected with two strains. *Computers in Biology and Medicine* **55**:16–25.

Besag J, Green PJ (1993) Spatial statistics and Bayesian computation (with discussion). *Journal of the Royal Statistics Society, Series B* **55**:25–37.

Bishop CM (2006) *Pattern Recognition and Machine Learning*. New York: Springer-Verlag.

Boulkaibet I (2014) Finite element model updating using the Markov chain Monte Carlo technique. PhD thesis, University of Johannesburg.

Boulkaibet I, Marwala T, Mthembu L, Friswell MI, Adhikari S (2012) Sampling techniques in Bayesian finite element model updating. *Proceedings of the Society for Experimental Mechanics* **29**:75–83.

Boulkaibet I, Mthembu L, Marwala T, Friswell MI, Adhikari S (2014) Finite element model updating using the shadow hybrid Monte Carlo technique. *Mechanical System and Signal Processing* **52**:115–132.

Brooks S, Gelman A, Jones GL, Meng X-L (2011) *Handbook of Markov Chain Monte Carlo*. Boca Raton, FL: Chapman & Hall/CRC.

Butler A, Haynes RD, Humphries TD, Ranjan P (2014) Efficient optimization of the likelihood function in Gaussian process modelling. *Computational Statistics & Data Analysis* **73**:40–52.

Calderhead B (2014) A general construction for parallelizing Metropolis–Hastings algorithms. *Proceedings of the National Academy of Sciences of the United States of America* **111**:17408–17413.

Canary H, Russell G, Taylor II M, Quammen C, Pratt S, Gómez FA, O′Shea B, Healey CG (2014) Visualizing likelihood density functions via optimal region projection. *Computers & Graphics* **41**:62–71.

Cheng Q-B, Chen X, Xu C-Y, Reinhardt-Imjela C, Schulte A (2014) Improvement and comparison of likelihood functions for model calibration and parameter uncertainty analysis within a Markov chain Monte Carlo scheme. *Journal of Hydrology* **519**:2202–2214.

Cheung SH, Beck JL (2010) Calculation of posterior probabilities for Bayesian model class assessment and averaging from posterior samples based on dynamic system data. *Computer-Aided Civil and Infrastructure Engineering* **25**:304–321.

Chib S, Greenberg E (1995) Understanding the Metropolis–Hastings algorithm. *American Statistician* **49**:327–335.

Coletti G, Petturiti D, Vantaggi B (2014) Possibilistic and probabilistic likelihood functions and their extensions: common features and specific characteristics. *Fuzzy Sets and Systems* **250**:25–51.

Edwards RG, Sokal AD (1988) Generalization of the Fortuin–Kasteleyn–Swendsen–Wang representation and Monte Carlo algorithm. *Physical Review Letters* **38**: 2009–2012.

Fishman, G (1999) An analysis of Swendsen Wang and related sampling methods. *Journal of the Royal Statistical Society, Series B* **61**: 623–641.

Gilks WR (2005) Markov chain Monte Carlo. In *Encyclopedia of Biostatistics*. Chichester: John Wiley & Sons, Ltd.

Greig B, Mesinger A (2015) 21CMMC: an MCMC analysis tool enabling astrophysical parameter studies of the cosmic 21 cm signal. *Monthly Notices of the Royal Astronomical Society* **449**:4246–4263.

Grišins P, Mazets IE (2014) Metropolis–Hastings thermal state sampling for numerical simulations of Bose–Einstein condensates. *Computer Physics Communications* **185**:1926–1931.

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**:97–109.

Hatjispyros SJ, Nicoleris T, Walker SG (2007) Parameter estimation for random dynamical systems using slice sampling. *Physica A: Statistical Mechanics and Its Applications* **381**:71–81.

Heckman JJ, Leamer E (2001) *Handbook of Econometrics*, Vol. **5**. Amsterdam: Elsevier.

Higdon, D.M. (1996) Auxiliary variable methods for Markov Chain Monte Carlo with applications. ISDS Discussion Paper, Duke University.

Iman RL (2008) *Latin Hypercube Sampling*. London: John Wiley & Sons, Ltd.

Johnson AA, Flegal JM (2014) A modified conditional Metropolis–Hastings sampler. *Computational Statistics & Data Analysis* **78**:141–152.

Joseph JF, Guillaume JHA (2013) Using a parallelized MCMC algorithm in R to identify appropriate likelihood functions for SWAT. *Environmental Modelling & Software* **46**:292–298.

Joubert, D.J. (2015) *Markov chain Monte Carlo method for finite element model updating*. MEng thesis, University of Johannesburg.

Kastner CA, Braumann A, Man PLW, Mosbach S, Brownbridge GPE, Akroyd J, Kraft M, Himawan C (2013) Bayesian parameter estimation for a jet-milling model using Metropolis–Hastings and Wang–Landau sampling. *Chemical Engineering Science* **89**:244–257.

Khodaparast HH (2010) Stochastic finite element model updating and its application in aeroelasticity. PhD thesis, University of Liverpool.

Koehler JR, Owen AB (1996) Computer experiments. In: *Handbook of Statistics*, Vol. **13** (ed. Ghosh S, Rao CR), pp. 261–308. Amsterdam: Elsevier.

Kraaij, C.S. (2006) Model updating of a 'clamped' free beam system using FEM tools. Technical Report, Technische Universiteit Eindhoven.

Kyprianou A, Worden K, Panet M (2001) Identification of hysteretic systems using the differential evolution algorithm. *Journal of Sound and Vibration* **248**:289–314.

Larjo A, Lähdesmäki H (2015) Using multi-step proposal distribution for improved MCMC convergence in Bayesian network structure learning. *EURASIP Journal on Bioinformatics and Systems Biology* **2015**(1), article no. 6.

Lu, J. (2008) Multivariate slice sampling. PhD thesis, Drexel University.

Mahani AS, Sharabiani MTA (2015) SIMD parallel MCMC sampling with applications for big-data Bayesian analytics. *Computational Statistics and Data Analysis* **88**:75–99.

Marwala, T. (1997) A multiple criterion updating method for damage detection on structures. MEng thesis, University of Pretoria.

Marwala T (2010) *Finite Element Model Updating Using Computational Intelligence Techniques*. London: Springer-Verlag.

Mazonakis M, Sahin B, Pagonidis K, Damilakis J (2011) Assessment of left ventricular function and mass by MR imaging: a stereological study based on the systematic slice sampling procedure. *Academic Radiology* **18**:738–744.

McKay MD, Conover WJ, Beckman RJ (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**:239–245.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**:1087–1092.

Mira A, Moller J, Roberts GO (2001) Perfect slice samplers. *Journal of the Royal Statistical Society, Series B* **63**:593–606.

Neal RM (2003) Slice sampling. *Annals of Statistics* **31**:705–741.

Nieto, M., Cortés, A., Otaegui, O. and Etxabe, I. (2012) MCMC particle filter with over-relaxated slice sampling for accurate rail inspection. *Proceedings of the International Conference on Computer Vision Theory and Applications*, Vol. 2, pp. 164–172.

Nishihara R, Murray I, Adams RP (2014) Parallel MCMC with generalized elliptical slice sampling. *Journal of Machine Learning Research* **15**:2087–2112.

Pietrabissa T, Rusconi S (2014) Parallel slice sampling. In: *The Contribution of Young Researchers to Bayesian Statistics* (ed. Lazarone E, Ieva F), Springer Proceedings in Mathematics & Statistics **63**, pp. 81–84. Cham: Springer.

Renshaw E (2004) Metropolis–Hastings from a stochastic population dynamics perspective. *Computational Statistics & Data Analysis* **45**:765–786.

Roberts GO, Rosenthal JS (1999) Convergence of slice sampler Markov chains. *Journal of the Royal Statistical Society, Series B* **61**:643–660.

Roberts GO, Smith AF (1994) Simple conditions for the convergence of the Gibbs sampler and Metropolis–Hastings algorithms. *Stochastic Processes and Their Applications* **49**:207–216.

Rostami M, Adam MB, Ibrahim NA and Yahya MH (2013) Slice sampling technique in Bayesian extreme of gold price modelling. *AIP Conference Proceedings*, **1557**:473–477.

Shao W, Guo G, Meng F, Jia S (2013) An efficient proposal distribution for Metropolis–Hastings using a B-splines technique. *Computational Statistics & Data Analysis* **57**:465–478.

Strid I (2010) Efficient parallelisation of Metropolis–Hastings algorithms using a prefetching approach. *Computational Statistics & Data Analysis* **54**:2814–2835.

Swendsen RH, Wang JS (1987) Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters* **58**:86–88.

Tasaki S, Sauerwine B, Hoff B, Toyoshiba H, Gaiteri C, Neto EC (2015) Bayesian network reconstruction using systems genetics data: comparison of MCMC methods. *Genetics* **199**:973–989.

Thompson, M.B. (2011) Slice sampling with multivariate steps. PhD dissertation, University of Toronto.

Tibbits MM, Haran M, Liechty JC (2011) Parallel multivariate slice sampling. *Statistics and Computing* **21**:415–430.

Van Dyk DA, Jiao X (2015) Metropolis–Hastings within partially collapsed Gibbs samplers. *Journal of Computational and Graphical Statistics* **24**:301–327.

Vu T, Vo B-N, Evans R (2014) A particle marginal Metropolis–Hastings multi-target tracker. *IEEE Transactions on Signal Processing* **62**:3953–3964.

Zhang Z, Koh CG and Zhang J (2009) System identification via orthogonal arrays sampled genetic algorithms. Proceedings of the 27th International Modal Analysis Conference, Bethel, CT: SPIE.

5

Dynamically Weighted Importance Sampling for Finite Element Updating

5.1 Introduction

In the previous chapter, the Metropolis–Hastings (M-H) and slice sampling techniques were used for finite element updating within the context of Bayesian statistics and Monte Carlo simulation. In this chapter, we apply dynamically weighted importance sampling (DWIS) for finite element model updating as was done by Joubert (2015) and Joubert and Marwala (2015). Finite element model updating is a numerical procedure that is used to bring the finite element model closer to the measured data. The use of Bayesian statistics essentially makes the distance between the measured data and finite element models probabilistic, and this makes sense because both the measured data and the finite element models are uncertain. In order to construct the probability distribution of the distance between the finite element model and measured data, Bayesian statistics is used, because it estimates the probability of unknown parameters, that is, the distance between the model and measurement. This distance is also known as the posterior distribution function, and it is estimated on the basis of the information on the parameters that describe such distance (the prior), the distance between the model and the data (the likelihood) and the observed data (the evidence). The posterior distribution function is very difficult to estimate analytically, and the Monte Carlo simulation method can be used. With unlimited computational capability, the Monte Carlo method is able to solve such problems, but this is not practical. Monte Carlo approaches have been explored in the finite element updating problem (Boulkaibet *et al.*, 2015; Marwala, 2010; Friswell, 2001; Mottershead and Friswell, 1995). Unfortunately, computational power is limited and some approximation of the Monte Carlo method should be explored. The techniques that were explored in the previous chapter, M-H and slice sampling, are versions of the Monte Carlo method. This version is known as the Markov chain Monte Carlo (MCMC) method and it is merely a short-cut version of the Monte Carlo method, as it is intended to manage the reality that computational

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

capability is expensive (Larjo and Lähdesmäki, 2015; Mahani and Sharabiani, 2015; Papaioannou *et al.*, 2015).

Another technique that can be used to estimate the posterior probability distribution function is the importance sampling method. On sampling a probability distribution function one often realises that it is important to know where to sample because there are many local distribution functions which are not necessarily global and therefore correct. In optimisation we often learn about the relationship between global and local optimum points (Wang *et al.*, 2016; Morales-Enciso and Branke, 2015). The same dilemma of the local versus the global optimum posterior distribution is relevant here.

The Monte Carlo importance sampling algorithm is a Monte Carlo procedure which is used in this chapter to update a finite element model and basically allows one to know where to sample, hopefully to increase the likelihood of identifying an optimal posterior probability distribution function; it was first used in finite element model updating by Joubert and Marwala (2015) and Joubert (2015). The justification for implementing importance weights in the Monte Carlo procedure is to ensure smooth transitions between balanced states, which is not the case in the Metropolis transition rules. DWIS substantially performs better than other Monte Carlo method, for instance the M-H process and the dynamic weighting procedure (Liang *et al*., 2010; Liang, 2002). In the DWIS method the Markov chain state is added to a population, and at each step two mechanisms are used: dynamic weighting and population control. These steps guarantee that the DWIS technique moves through energy barriers using controlled weights as well as finite expectation. An additional advantage of the DWIS technique when compared to the M-H technique (Hastings, 1970) is that the DWIS method offers the possibility of varying the parameters of the procedure in line with the problem being solved. The DWIS method has been observed to give more stable approximation of the stationary distribution than the dynamic weighting, as well as converging fast (Wong and Liang, 1997).

5.2 Bayesian Modelling Approach

The Bayesian approach is governed by Bayes' rule, which is written as (Bishop, 2006)

$$P(\boldsymbol{\theta}|\mathcal{D}) \propto P(\mathcal{D}|\boldsymbol{\theta})P(\boldsymbol{\theta}), \tag{5.1}$$

where $\boldsymbol{\theta}$ represents the vector of updating parameters, $P(\mathcal{D}|\boldsymbol{\theta})$ is the likelihood probability distribution function and $P(\boldsymbol{\theta})$ denotes the prior probability distribution. The likelihood distribution can be expressed as the probability of the measured modal frequencies given the observed data corresponding to the variable parameters as follows (Boulkaibet *et al.*, 2015):

$$P(\mathcal{D}|\boldsymbol{\theta}) = \frac{1}{(2\pi/\beta)^{N_m}\prod_{i=1}^{N_m} f_i^m} \exp\left(-\beta \sum_{i=1}^{N_m} \left(\frac{f_i^m - f_i^c}{f_i^m}\right)^2\right), \tag{5.2}$$

where N_m denotes the variables in the updating vector, β is a constant, and f_i^m and f_i^c represent the measured and computed frequencies, respectively. The prior probability distribution can be assumed to be Gaussian and is expressed as follows (Bishop, 2006; Boulkaibet *et al.*, 2015):

$$P(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{Q/2}\prod_{i=1}^{Q}\left(1/\sqrt{\alpha_i}\right)}\exp\left(-\sum_{i}^{Q}\frac{\alpha_i}{2}||\boldsymbol{\theta}-\boldsymbol{\theta}_0||^2\right). \tag{5.3}$$

In Equation 5.3, the Euclidean norm of the updating vector is used to quantify the difference between the results from the finite element model and the measured results. The coefficient on the *i*th component of the updating parameters is denoted by α_i and the number of parameters to be updated is Q. The posterior probability distribution function obtained using Bayes' rule can then be expressed as follows (Joubert and Marwala, 2015):

$$P(\boldsymbol{\theta}|\mathcal{D}) = \frac{1}{Z(\alpha,\beta)}\exp\left(-\beta\sum_{i=1}^{N_m}\left(\frac{f_i^m - f_i^c}{f_i^m}\right)^2 - \sum_{i}^{Q}\frac{\alpha_i}{2}||\boldsymbol{\theta}-\boldsymbol{\theta}_0||^2\right), \tag{5.4}$$

where the normalising constant is

$$Z(\alpha,\beta) = \left(\frac{2\pi}{\beta}\right)^{N_m}\prod_{i=1}^{N_m} f_i^m (2\pi)^{Q/2}\prod_{i=1}^{Q}\frac{1}{\sqrt{\alpha_i}}.$$

Equation 5.4 can be written as a normal distribution as follows (Boulkaibet, 2014; Joubert and Marwala, 2015):

$$P(\boldsymbol{\theta}|\mathcal{D}) = \frac{1}{(2\pi)^{(N_m+d)/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right), \tag{5.5}$$

where

$$\boldsymbol{X} = [f_1, \ldots, f_{N_m}, \theta_1, \ldots, \theta_d], \quad \boldsymbol{\mu} = \left[f_1^m, \ldots, f_{N_m}^m, \theta_1^0, \ldots, \theta_d^0\right],$$

$$\Sigma^{-1} = \begin{bmatrix} \frac{\beta_c}{(f_1^m)^2} & 0 & \cdots & & & \\ 0 & \ddots & 0 & & 0 & \\ \vdots & 0 & \frac{\beta_c}{(f_{N_m}^m)^2} & & & \\ & & & \alpha_1 & 0 & \vdots \\ & 0 & & 0 & \ddots & 0 \\ & & & \cdots & 0 & \alpha_d \end{bmatrix}. \tag{5.6}$$

5.3 Metropolis–Hastings (M-H) Algorithm

The posterior probability distribution function can be sampled using the M-H algorithm (Metropolis *et al.*, 1953; Hastings, 1970). Neal (2001) proposed annealed importance sampling where the transition between states were achieved using a Markov chain. Bassetti and Diaconis (2006) compared importance sampling to the Metropolis algorithm and found that these methods gave two different stationary distributions. Berg (2003) proposed Metropolis importance sampling for rugged dynamical variables by introducing a mechanism for guiding the simulation from higher to lower temperatures for constructing Metropolis weights, and this technique was found to be useful in large problems. Takaishi (2013) compared the MCMC method to importance sampling in an application to Bayesian inference of the generalised autoregressive conditional heteroscedasticity (GARCH) model. Takaishi observed that the statistical errors of the GARCH parameters obtained from the implementation of importance sampling were statistically similar to but smaller than those attained from the MCMC technique.

The MCMC sampling approach uses the theory of Markov chains. A set of randomly created variables, $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$, is deemed to be a Markov chain if conditional independence between the variables is sustained and, for $n \in \{1, \ldots, N-1\}$ (Bishop, 2006; Boulkaibet, 2014),

$$P(\boldsymbol{\theta}_{n+1}|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n) = P(\boldsymbol{\theta}_{n+1}|\boldsymbol{\theta}_n). \tag{5.7}$$

As a result, the transition probability matrix which takes $\boldsymbol{\theta}_n$ as input and gives $\boldsymbol{\theta}_{n+1}$ as output can be mathematically expressed as follows (Bishop, 2006):

$$T_n(\boldsymbol{\theta}_{n+1}|\boldsymbol{\theta}_n) \equiv \mathrm{P}(\boldsymbol{\theta}_{\mathrm{n}+1}|\boldsymbol{\theta}_{\mathrm{n}}). \tag{5.8}$$

The transition matrix $\boldsymbol{T}$ is stochastic and, on condition that the state space is a periodic and irreducible, the simulation will always converge to a desired stationary distribution function and, in this chapter, the posterior distribution function. The transition matrix is a mechanism for prescribing how the simulation moves from the current state to the next state. The Markov process does not necessitate memory of states prior to the current state. As an irreducible chain the states are always connected in that there is a transition probability between the states. The state space is a connected graph of these states and the aperiodicity property is useful for creating a Markov chain so as to avoid oscillation effects in the simulation. Normally, if all the elements of the transition matrix are greater than 0, then the chain is periodic and a procedure to prompt convergence of the Markov chain is the detailed balance condition which is expressed as follows (Bishop, 2006):

$$P(\boldsymbol{\theta}_n)P(\boldsymbol{\theta}_{n+1}|\boldsymbol{\theta}_n) = P(\boldsymbol{\theta}_{n+1})P(\boldsymbol{\theta}_n|\boldsymbol{\theta}_{n+1}). \tag{5.9}$$

By integrating Equation 5.9 with respect to $\boldsymbol{\theta_n}$ we obtain the following property of the Markov chain (Bishop, 2006):

$$\int P(\boldsymbol{\theta}_n)P(\boldsymbol{\theta}_{n+1}|\boldsymbol{\theta}_n) = P(\boldsymbol{\theta}_{n+1}). \tag{5.10}$$

Equation 5.10 exhibits ergodic characteristics and gives sufficient conditions for constructing $P(\theta_{n+1}|\theta_n)$ to create samples from $P(\boldsymbol{\theta}_n)$.

In the M-H algorithm, a proposal distribution function $q(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})$ is used to obtain samples, where $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_t\}$ is the finite element model updating parameter vector, and the M-H algorithm was described earlier in Chapter 4 (Metropolis *et al.*, 1953; Hastings, 1970; Brooks *et al.*, 2011; Joubert, 2015; Boulkaibet *et al.*, 2015).

5.4 Importance Sampling

Before we discuss the DWIS method, it is important to describe the importance sampling (IS) technique (Dempster, 2006; Zhang *et al.*, 2009; Parpas, 2012; Elvira *et al.*, 2015; Martino *et al.*, 2015). IS has been applied in the Monte Carlo technique to reduce the variance of the samples generated. It is a generalised method applied to approximate properties of a specified probability distribution function. It emphasises the intuitive foundation of obtaining samples from important sections of the search space. This is accomplished by sampling from a simple distribution and then using the weights of the samples from important regions to estimate the probability distribution of a complicated distribution function. In so doing, IS reduces the variance of the samples. Therefore, IS is applied instead of acceptance–rejection sampling, for sensitivity analysis and for calculating normalising constants of probability densities, and this is a mechanism used by the Monte Carlo dynamically weighted importance sampling (MCDWIS) technique to compute IS estimation.

In IS some random values in the Monte Carlo sampling process are deemed more important than others. For this reason the sampling process is designed in such way that the important samples are visited more frequently than those deemed less important. Li and Lin (2015) proposed an adaptive importance method and successfully applied it to the problem of Bayesian inversion with multimodal distributions, while Fan and Liang (2015) applied IS to efficiently analyse the reliability of axially loaded piles. Yilmaz and Ozev (2015) successfully applied adaptive IS in analogue circuit estimation, while Xu *et al.* (2015) applied the IS technique in portfolio risk estimation. Blitzstein and Diaconis (2015) applied the sequential IS method to produce random graphs, while Nadjafi *et al.* (2015) applied the IS method together with fuzzy logic to improve uncertainty and applied this to clustering problems (Tokdar and Kass, 2009). If we assume that $P(\boldsymbol{\theta})$ is the probability distribution function of the finite element model parameters $\boldsymbol{\theta}$, then the expectation μ_f of some integrand $f(\boldsymbol{\theta})$ can be calculated as follows (Tokdar and Kass, 2009):

$$\mu_f = \int f(\boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{5.11}$$

If we assume that we have the proposed probability distribution function $q(\boldsymbol{\theta})$ which is such that $q(\boldsymbol{\theta}) > 0$ whenever $q(\boldsymbol{\theta})P(\boldsymbol{\theta}) \neq 0$, then the following expression can be estimated (Tokdar and Kass, 2009):

$$\mu_f = E_q[w(\boldsymbol{\theta}) f(\boldsymbol{\theta})]. \tag{5.12}$$

In this expression, f is the integrand and $w(\boldsymbol{\theta}) = P(\boldsymbol{\theta})/q(\boldsymbol{\theta})$ is the likelihood ratio, in which q is the importance distribution and P is the density distribution of interest. We can therefore estimate the posterior probability in Equation 5.4 (now deemed to be $P(\boldsymbol{\theta})$) with the normal distribution in Equation 5.5 (now deemed to be $q(\boldsymbol{\theta})$). It is therefore now possible to estimate the probability distribution function $P(\boldsymbol{\theta})$ by merely sampling the simple probability distribution function $q(\boldsymbol{\theta})$, and then the expectation is estimated using the following equation (Tokdar and Kass, 2009; Joubert, 2015):

$$\mu_f = \frac{1}{m}\sum_{j=1}^{m} w(\boldsymbol{\theta}_j) f(\boldsymbol{\theta}_j). \tag{5.13}$$

The IS technique for estimating the finite element model parameters $\boldsymbol{\theta}$ is implemented as follows (Tokdar and Kass, 2009; Joubert and Marwala, 2015):

1. Gather a small number of samples from f.
2. Create the importance distribution function q.
3. Sample from q.
4. Approximate the expected values of the finite element model parameters $\boldsymbol{\theta}$.

5.5 Dynamically Weighted Importance Sampling

Liang and Cheon (2009) applied Monte Carlo based dynamically weighted importance sampling in models that contained intractable normalising constants. This method applied the Monte Carlo method within the context of the MCMC technique while retaining the invariance of the target distribution in DWIS. In so doing, this method eliminates the requirement for perfect sampling. The variable finite element updating parameter vector is characterised by $\boldsymbol{\theta}$, which are inputs to the finite element model to calculate natural frequencies using a vibrational modal analysis approach. In this section, we use different symbols than P to represent the distribution functions. The prior distribution is denoted by $P(\boldsymbol{\theta})$ and is the prior knowledge of the problem at hand. The likelihood function is denoted by $P(\mathcal{D}|\boldsymbol{\theta})$, while the posterior density function is denoted by $P(\boldsymbol{\theta}|\mathcal{D})$. The idea of the DWIS algorithm is to generate auxiliary variables (or data) to integrate out as many variables as possible in order to accelerate convergence. In this case, an auxiliary data vector $\boldsymbol{y}$ will be generated, to complete the missing data, and the likelihood function will be defined as $P(\boldsymbol{y}|\boldsymbol{\theta})$.

This method is particularly applicable in problems with incomplete posterior distribution and missing data. In the DWIS approach, given simulated data $\boldsymbol{y}$, the likelihood function can be expressed as follows (Liang *et al*., 2010; Joubert and Marwala, 2015):

$$P(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{P(\boldsymbol{\theta},\boldsymbol{y})}{Z(\boldsymbol{\theta})}, \tag{5.14}$$

where $\boldsymbol{\theta}$ indicates the model parameters under consideration which are inputs to the finite element model to compute the natural frequencies, $P(\boldsymbol{\theta}, \boldsymbol{y})$ is the joint density of both $\boldsymbol{\theta}$ and $\boldsymbol{y}$, and Z

($\boldsymbol{\theta}$) is the normalising constant dependent on the parameter $\boldsymbol{\theta}$. The posterior probability distribution function of the variable $\boldsymbol{\theta}$ is expressed as follows (Liang *et al*., 2010; Joubert and Marwala, 2015):

$$P(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{1}{Z(\boldsymbol{\theta})} P(\boldsymbol{\theta},\boldsymbol{y}) P(\boldsymbol{\theta}), \tag{5.15}$$

where $P(\boldsymbol{\theta})$ represents the prior distribution function, and a state in a Markov chain of population size N is expressed as a joint function $(\boldsymbol{\theta},\boldsymbol{w}) = \{\theta_1, w_1; \ldots; \theta_N, w_N\}$. An iteration of the technique entails the following (Liang *et al*., 2010):

1. *Dynamic weighting*. Each state is updated by a dynamic weighting transition step with the purpose of calculating a new population.
2. *Population control*. A population control system where samples connected to small weights with reference to successful finite element model updating are eliminated, whereas weighted samples with stronger significance to the model objective function are reproduced in the new population. This results in biased samples but is compensated by allocating new weights to sample data in the new population.

5.5.1 Markov Chain

If we simulate the states $\boldsymbol{\theta} = \{\theta_{0:i-1}; i = 1, \ldots, N\}$ via the MCMC technique, then we calculate the importance sampling approximation and the normalisation constant ratio $R_t(\boldsymbol{\theta},\boldsymbol{\theta}^*) = Z(\boldsymbol{\theta})/Z(\boldsymbol{\theta}^*)$ as follows (Liang *et al*., 2010; Joubert and Marwala, 2015):

$$R_t(\boldsymbol{\theta},\boldsymbol{\theta}^*) = \frac{w_t}{M} \sum_{i=1}^{M} \frac{P(y_i,\boldsymbol{\theta})}{P(y_i,\boldsymbol{\theta}^*)}. \tag{5.16}$$

Thus the Monte Carlo dynamic weighting ratio may be estimated as follows (Liang *et al*., 2010; Joubert and Marwala, 2015):

$$r_d = r_d(\boldsymbol{\theta},\boldsymbol{\theta}^*,w) = R_t(\boldsymbol{\theta},\boldsymbol{\theta}^*) \frac{P(\boldsymbol{y},\boldsymbol{\theta}^*)}{P(\boldsymbol{y},\boldsymbol{\theta})} \frac{q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}{q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}, \tag{5.17}$$

where $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$ is the proposal distribution function.

5.5.2 Adaptive Pruned-Enriched Population Control Scheme

The adaptive pruned-enriched population control scheme (APEPCS) controls the steps of the algorithm in the dynamic weighting and population. The weight of the population that is located outside the weight bounds W_{up} and W_{low} is handled by applying the control parameter d_s in the enriching stage where $w_{t,i}$ is too large, and the probability density parameter q during

the pruning stage ($w_{t,i} < W_{\text{low}}$), of the population control scheme. Weighting is controlled by the ratio of the upper to the lower weight control bounds and is written as $k = W_{\text{up}}/W_{\text{low}}$. This ratio controls the movement of the system and is known as the freedom parameter (Liang *et al.*, 2010; Joubert and Marwala, 2015; Joubert, 2015).

To summarise the concept of the APEPCS, several parameters will be defined. First, the samples are drawn first using the M-H algorithm. The current population is $(\boldsymbol{\theta}_t, \boldsymbol{w}_t)$, where the ith state of the population is $(\boldsymbol{\theta}_{t,i}, w_{t,i})$. Let n_t and n_t' represent the current and new population sizes. The minimum and maximum population sizes allowed by the user are $n_{\min}$ and $n_{\max}$, respectively. The APEPCS consists of the following steps (Liang, 2002; Joubert and Marwala, 2015; Joubert, 2015):

1. *Initialise weight control bounds.* The upper and lower weight control bounds are written as (Liang, 2002)

$$W_{\text{low},t} = \sum_{i=1}^{n} w_{t,i}/n_{\text{up}}, \tag{5.18}$$

$$W_{\text{up},t} = \sum_{i=1}^{n} w_{t,i}/n_{\text{low}}, \tag{5.19}$$

   where n_{up} and n_{low} are reference bounds on the population size.
2. *Pruning.* The state is accepted with probability $q = 1 - w_{t,i}/W_{\text{low},t}$ when the $i$thindividual weight of the population is less than the lower weight control bound and therefore $n_{t+1}' = n_t' + d$.
3. *Enrich.* If $w_{t,i} > W_{\text{up},t}$, the ith state weight is substituted by an enriched state $w_{t,i}' = w_{t,i}/d$, where $d = \left[w_{t,i}/W_{\text{up},t} + 1\right]$ and the new population is adapted.
4. *Unchanged.* If $W_{\text{low},t} < w_{t,i} < W_{\text{up},t}$, the state remains the same and the new population is increased by 1 and therefore $n_{t+1}' = n_t' + 1$.
5. *Check.* If $n_t' > n_{\max}$, the upper and lower control bounds are then adapted and therefore $W_{\text{low},t} \leftarrow \lambda W_{\text{low},t}$ and $W_{\text{up},t} \leftarrow \lambda W_{\text{up},t}$. Steps 2–4 are then reiterated for $i = 1, 2, \ldots, n_t'$. If $n_t' < n_{\min}$, $W_{\text{low},t} \leftarrow W_{\text{low},t}/\lambda$ and $W_{\text{up},t} \leftarrow W_{\text{up},t}/\lambda$, then steps 2–4 are repeated.

There are two population control techniques: the W-scheme and the R-scheme. In this chapter the R-scheme is used (Liang, 2002; Joubert and Marwala, 2015; Joubert, 2015).

The dynamic weighting and population control of the R-scheme are as follows:

1. Dynamic weighting (Liang, 2002; Joubert and Marwala, 2015; Joubert, 2015). The procedure is applied to $(\boldsymbol{\theta}_{t-1}, \boldsymbol{w}_{t-1})$ where W_c is the dynamic switching parameter that alters the value of φ_t to either 0 or 1, depending on the value of $W_{\text{up},t-1}$. If $W_{\text{up},t-1} \leq W_c$, then $\varphi_t = 1$, otherwise $\varphi_t = 0$ and the new population is then $(\boldsymbol{\theta}_t', \boldsymbol{w}_t')$. The parameter φ_t is selected as a function of $(\boldsymbol{\theta}_t, \boldsymbol{w}_t)$, and it should be noted that if $\varphi_t = 0$ then the sampler is basically a random walk process, and if $\varphi_t = 1$ then the sampler performs the R-type move. The introduction of the φ_t parameter is to prevent the weighting technique from converging to zero or infinity.

2. Population control (Liang, 2002; Joubert and Marwala, 2015; Joubert, 2015). The APEPCS is then applied to the new population and the new population then becomes (θ_t, w_t).

The W-scheme consists of the following techniques:

1. Initial weight control (Liang, 2002; Joubert and Marwala, 2015; Joubert, 2015). If $n_{t-1} < n_{\text{low}}$, then $W_{\text{low},t} = W_{\text{low},t-1}/\lambda$ and $W_{\text{up},t} = W_{\text{up},t-1}/\lambda$. If $n_{t-1} > n_{\text{up}}$, then $W_{\text{low},t} = \lambda W_{\text{up},t-1}$ and $W_{\text{up},t} = \lambda W_{\text{up},t-1}$. Or else $W_{\text{low},t} = W_{\text{up},t-1}$ and $W_{\text{up},t} = W_{\text{up},t-1}$.
2. Dynamic weighting (Liang, 2002; Joubert and Marwala, 2015; Joubert, 2015). Use the previous population $(\boldsymbol{\theta}_{t-1}, \boldsymbol{w}_{t-1})$ with $\delta_t = 1/\left(\alpha + \beta W_{\text{up},t}^{1+\epsilon}\right)$ for some positive integer $\epsilon > 0$, and then $\left(\boldsymbol{\theta}_t', \boldsymbol{w}_t'\right)$ is regarded as the new population.
3. Population control (Liang, 2002; Joubert and Marwala, 2015; Joubert, 2015). APEPCS is applied to the population $\left(\boldsymbol{\theta}_t', \boldsymbol{w}_t'\right)$, and the resulting new population is then denoted by $(\boldsymbol{\theta}_t, \boldsymbol{w}_t)$.

For finite element model updating Joubert and Marwala (2015) used the R-scheme and set the M-H ratio as $r(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) = \hat{R}(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)\alpha(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)$, where $\hat{R}(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)$ is the approximated normalising constant. Then the expectation for w_t can be estimated as follows (Liang *et al.*, 2010; Joubert and Marwala, 2015; Joubert, 2015):

$$E[w_t|\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t, \boldsymbol{w}_{t-1}] = w_{t-1}[1 + r(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)]. \tag{5.20}$$

From Equation 5.20, it is evident that because the M-H ratio is always greater than zero the weight process is characteristically monotonically increasing, and therefore it is important to ensure that the weight does not drift to zero or infinity. This is achieved by analysing the weight behaviour at every time step and constraining the upper limit of the weight. When $\varphi_t = 0$ the weight is regulated by (Liang *et al.*, 2010; Joubert and Marwala, 2015; Joubert, 2015)

$$w_t = \exp\left(\log w_0 + \sum_{s=1}^{t} \log r(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) - \sum_{s=1}^{t} \log d_s\right), \tag{5.21}$$

where d_s is the integer set at the pruning stage of the APEPCS.

5.5.3 Monte Carlo Dynamically Weighted Importance Sampling

First, a proposed sample $\boldsymbol{\theta}^*$ is drawn from the proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_t)$ and sampled state trajectories are generated from $f(\boldsymbol{y}|\boldsymbol{\theta}^*)$ using the M-H technique $\boldsymbol{y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_M\}$. Then the importance sampling estimate is calculated from the normalisation constant ratio $\hat{R}_t(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = Z(\boldsymbol{\theta})/Z(\boldsymbol{\theta}^*)$ (Liang *et al.*, 2010; Joubert and Marwala, 2015; Joubert, 2015):

$$\hat{R}_t(\boldsymbol{\theta}_t, \boldsymbol{\theta}^*) = \frac{1}{M}\sum_{j=1}^{M} \frac{P(\boldsymbol{y}_j, \boldsymbol{\theta})}{P(\boldsymbol{y}_j, \boldsymbol{\theta}^*)}. \tag{5.22}$$

The Monte Carlo dynamic weighting ratio is then estimated from Equation 5.22 and the conditional probability distribution is written as $P(\boldsymbol{y}, \boldsymbol{\theta}_t) = P(\boldsymbol{\theta}_t|\boldsymbol{y})P(\boldsymbol{y})$ (Liang *et al.*, 2010; Joubert and Marwala, 2015; Joubert, 2015):

$$r_d = r_d(\theta_t, \theta^*, w) = w\hat{R}_t(\boldsymbol{\theta_t}, \boldsymbol{\theta^*})\frac{P(\boldsymbol{y}, \boldsymbol{\theta}^*)}{P(\boldsymbol{y}, \boldsymbol{\theta}_t)}\frac{q(\boldsymbol{\theta}_t|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_t)}. \tag{5.23}$$

The joint probability density function is updated via random numbers from the uniform distribution $U \sim U(0,1)$ as follows (Liang *et al.*, 2010; Joubert and Marwala, 2015; Joubert, 2015):

$$(\boldsymbol{\theta}', \boldsymbol{w}') = \begin{cases} \left(\boldsymbol{\theta}^*, \frac{r_d}{a}\right), & \text{if } U \le a, \\ \left(\boldsymbol{\theta}, \frac{w}{1-a}\right), & \text{otherwise,} \end{cases} \tag{5.24}$$

where $a = r_d/(r_d + \varphi_t)$.

Given the dynamic weights as well as the corresponding sampled parameters, $(\boldsymbol{\theta}_1, w_1)$, $(\boldsymbol{\theta}_2, w_2)$, …, $(\boldsymbol{\theta}_N, w_N)$, the weighted average of the samples is written as follows (Liang *et al.*, 2010; Joubert and Marwala, 2015; Joubert, 2015):

$$\hat{\mu} = \sum_{t=1}^{N}\sum_{i=\text{burn-in}+1}^{n'}\left(\frac{w_{t,i}\rho(\theta_{t,i})}{w_{t,i}}\right), \tag{5.25}$$

where $\rho(\theta_t)$ denotes a state function over all the samples.

5.6 Application 1: Cantilevered Beam

The objective of finite element model updating is to minimise the distance between the final element model results and the measured data. In this section we use the cantilevered beam data studied and reported in Marwala (1997). The five natural frequencies obtained from the measurements of this cantilevered beam are represented as a vector, $\boldsymbol{f}_{\text{modes}} = [f_1, f_2, f_3, f_4, f_5]$. The MCDWIS and M-H techniques are applied for finite element model updating of the cantilevered beam (Marwala, 1997, 2010). On implementing the MCDWIS technique the parameters stipulated in Table 5.1 are used. The auxiliary samples were then drawn using the M-H sampling technique.

The approximation procedure involves computing an importance sampling estimate, as described in Section 5.4, and then applying the dynamic weighting procedure and the APEPCS procedures. The population is controlled in the population range $[n_{\min}, n_{\max}]$. Figure 5.1 shows

Table 5.1 MCDWIS algorithm parameters

Algorithm parameter	N	Burn-in	W_c	n_{min}	n_{max}	n_{low}	n_{up}	λ	k
Values	1000	250	e^5	100	1000	200	500	2	$\log_{10}W_c$

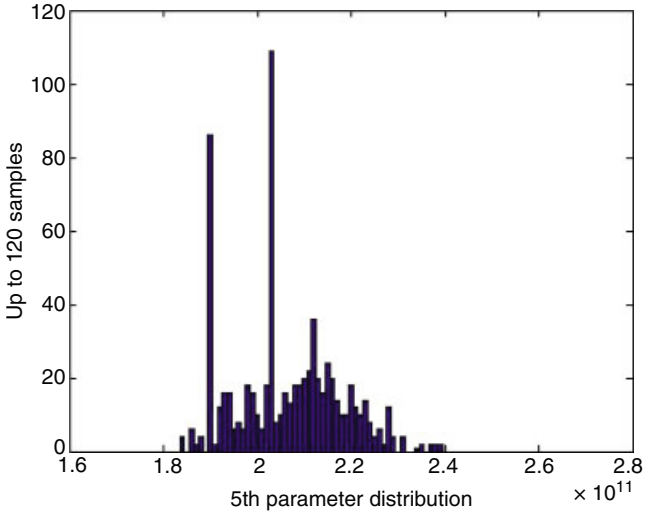

Figure 5.1 Sample distribution

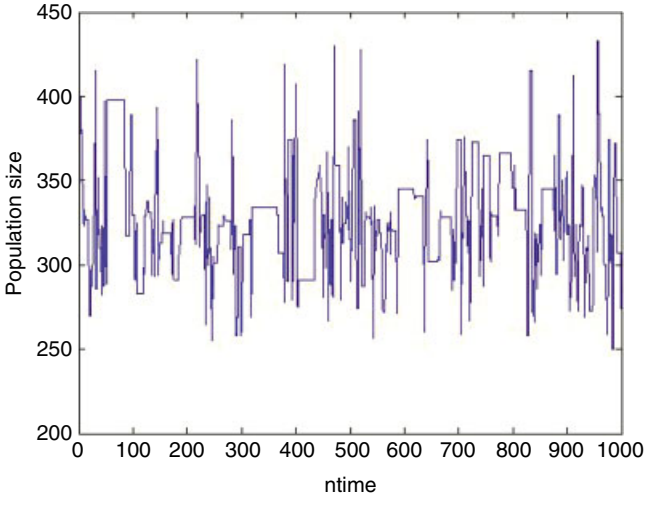

Figure 5.2 Population size adaptation

the distribution of the fifth variable parameter (Joubert and Marwala, 2015; Joubert, 2015). It is shown that the dynamic weighting procedure and population control method give distributions of Gaussian shape. Using the dynamic weighting, preference is assigned to the samples with the highest weight distributions.

Figure 5.2 demonstrates that the adaptation of the population size in the simulation is influenced by the APEPCS procedure. Figures 5.3 and 5.4 demonstrate the log weight distributions at different states *i* in the simulation, and the Gaussian shapes of these distributions demonstrate

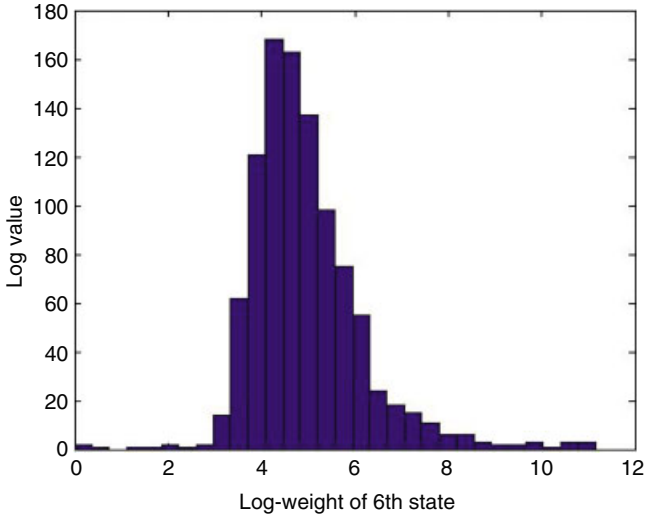

Figure 5.3 Log weight of sixth state

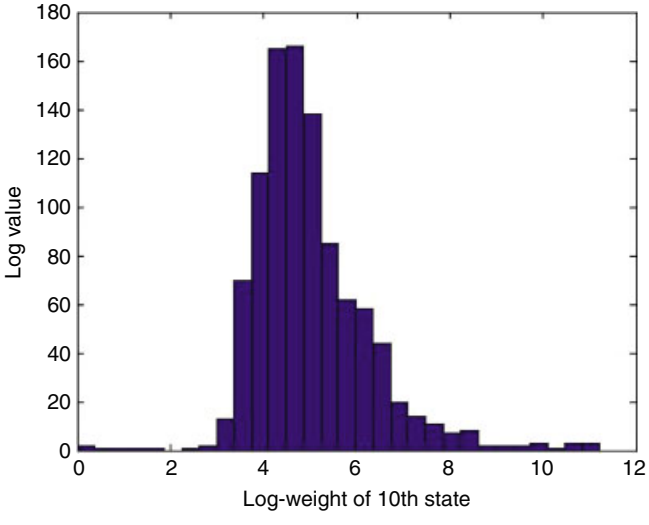

Figure 5.4 Log weight of 10th state

the functionality of the population control and dynamic weighting techniques. Figure 5.5 demonstrates the adaptation of the upper weight control bound, and in the simulation the weight control bounds were initialised as follows: $W_{\mathrm{up},0} = k$ and $W_{\mathrm{low},0} = 1$.

The population control parameters were adjusted to facilitate adequate transitional ability between states and in such a way that the upper and lower weight control bounds avoid drifting exponentially away from the controlled range. These techniques prevent the sampling sequence from merely becoming a random walk, which weakens the ability of the dynamic weighting and population control procedures.

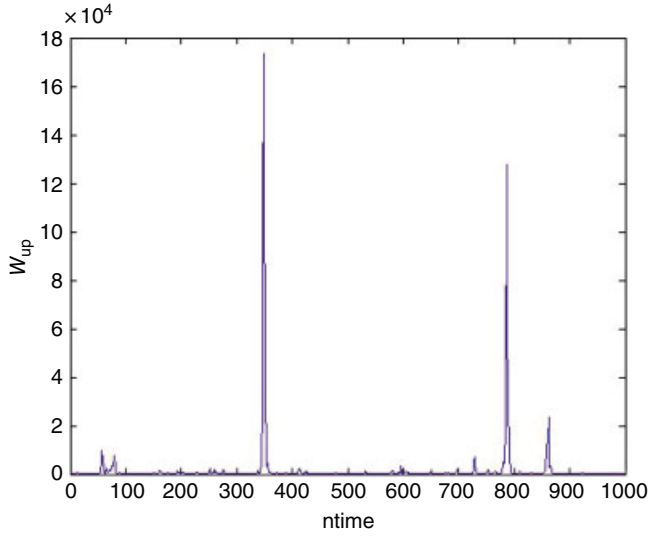

Figure 5.5 W_{up} adaptation

Table 5.2 Variable parameter vector results

D	Initial $\boldsymbol{\theta}_0$	$\boldsymbol{\theta}$ vector M-H method	$\boldsymbol{\theta}$ vector MCDWIS method
E_1	2.4×10^{11}	2.18×10^{11}	2.1524×10^{11}
E_2	2.4×10^{11}	2.19×10^{11}	2.1091×10^{11}
E_3	2.4×10^{11}	2.19×10^{11}	2.0430×10^{11}
E_4	2.4×10^{11}	2.20×10^{11}	2.0342×10^{11}
E_5	2.4×10^{11}	2.19×10^{11}	2.0336×10^{11}

Table 5.3 Cantilever frequency results

Measured frequency (Hz)	Initial frequency (Hz)	M-H frequency (Hz)	MCDWIS frequency (Hz)
31.9	32.7	30.7	30.8
197.9	209.4	196.1	194.8
553.0	559.8	556.9	553.2
1082.2	1177.8	1102.7	1094.4
1781.5	1961.7	1836.9	1824.0

Tables 5.2 and 5.3 show the results for the parameter vector $\boldsymbol{\theta} = \{E_1, E_2, E_3, E_4, E_5\}$ and the resulting vector of computed natural frequencies $\boldsymbol{f}_{\text{modes}} = \{f_1, f_2, f_3, f_4, f_5\}$. It is evident that the MCDWIS method is more accurate than the M-H method and is reliable and computationally efficient. Figure 5.6 demonstrates the correlation between new samples chosen at the 100th state.

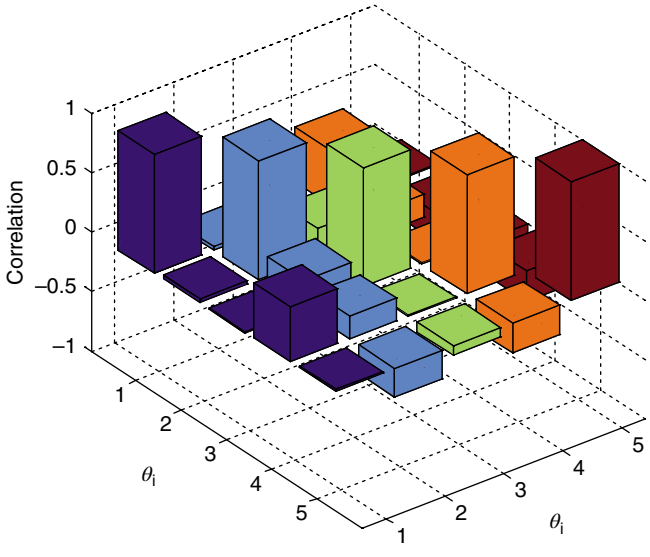

Figure 5.6 Correlation of samples

5.7 Application 2: H-Shaped Structure

The asymmetrical H-shaped beam aluminium structure studied and reported in Marwala (1997) was modelled using the finite element model technique, and modal frequencies were extracted using modal analysis. This structure was excited using an electromagnetic shaker and modal hammer, and the response was measured using strategically placed accelerometers. Data acquisition was conducted and data were recorded. The uncertainty parameters in the finite element model were the corresponding areas and moments of inertia, expressed as the vector $\boldsymbol{\theta} = [A_{x1}, A_{x2}, A_{x3}, I_{x1}, I_{x2}, I_{x3}]$. The statistical input parameters (covariance) and population control parameters were tuned to meet the model needs in the MCDWIS algorithm. The results from the MCDWIS and M-H algorithms were compared. The measured frequencies for the mode shapes were 53.9, 117, 208.4 and 445 Hz. Physical properties not updated in the finite element model included the Young's modulus, the material density and dimensions of the structure. Young's modulus was fixed at 7.2×10^{10} N/m^2 and density at 2785 kg/m^3. The covariance started with a value of $\sigma^2 = 5 \times 10^{-8}$ on the diagonals of a 6×6 matrix. The proposal was truncated to fall between a lower limit of $\theta_{\min} = [1.7265 \times 10^{-8},\ 1.7265 \times 10^{-8}, 1.7265 \times 10^{-8}, 2.1556 \times 10^{-4}, 2.1556 \times 10^{-4}, 2.1556 \times 10^{-4}]$ and an upper limit of $\theta_{\max} = [3.7265 \times 10^{-8},\ 3.7265 \times 10^{-8},\ 3.7265 \times 10^{-8},\ 4.1556 \times 10^{-4},\ 4.1556 \times 10^{-4},\ 4.1556 \times 10^{-4}]$. These methods were initialised at $\boldsymbol{\theta}_0 = [2.7265 \times 10^{-8},\ 2.7265 \times 10^{-8}, 2.7265 \times 10^{-8},\ 3.1556 \times 10^{-4},\ 3.1556 \times 10^{-4}, 3.1556 \times 10^{-4}]$.

The state of the Markov chain in the MCDWIS algorithm was improved to a population of weighted samples $(\theta_{t,i},\ w_{t,i})$, where $t = \{1, \ldots, N\}$. The Monte Carlo dynamic weighting step updates each individual state of the current population by a dynamically weight correlated value. The second step is APEPCS which replicates states with large weights and rejects

Table 5.4 MCDWIS algorithm parameters

Algorithm parameter	N	Burn-in	W_c	n_{min}	n_{max}	n_{low}	n_{up}	λ	k
Values	1200	250	e^5	100	1000	200	500	2	$\log_{10} W_c$

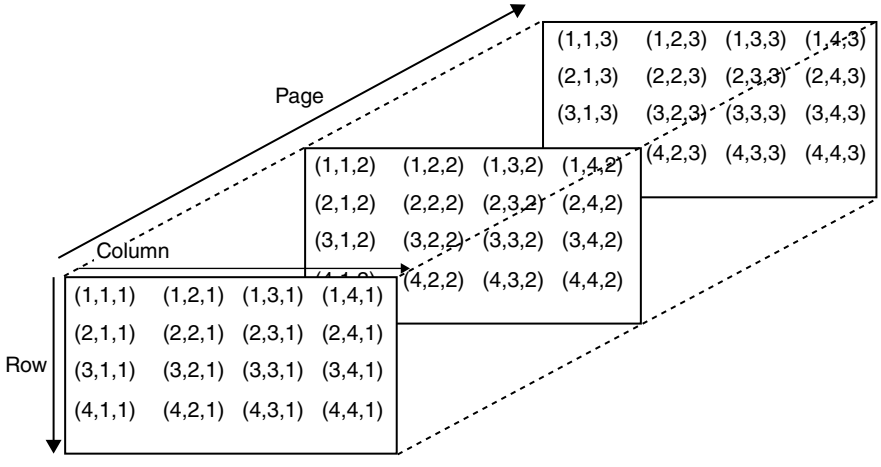

Figure 5.7 Multidimensional array with each page a matrix of samples for each variable parameter

the individual states with small weights. The algorithm parameters were chosen as shown in Table 5.4.

The samples were stored in a multidimensional array as: $samples = zeros(N, n_{\max}, 6)$. Figure 5.7 demonstrates how the tables were construed, with each page a matrix of samples for each variable parameter (Joubert and Marwala, 2015; Joubert, 2015). The index $n_{\max}$ is the maximum population size which differs in the simulation and is expressed by n' as described in the APEPCS, and this n' variation is regulated within the range $[n_{\min}, n_{\max}]$. The upper weight control bound W_{up} was started at k, and the starting value of W_{low} was made equal to 1. The subsequent samples were pseudo-random variables sampled from an abridged Gaussian distribution. Figures 5.8 and 5.9 show the log-weight for the 10th and 100th states respectively. Figure 5.10 shows the overall log weight results.

The population size at time t, given by n', was enhanced by the pruning and enriching steps and using the checking routine, and $W_{\text{low},t+1}$ and $W_{\text{up},t+1}$ were calculated. The dynamic move switching parameter φ_t, was estimated using the R-scheme in the dynamic weighting step. The initial population was created from the M-H algorithm to offer the main run of the MCDWIS a good initial sample distribution. Three thousand M-H method iterations were carried out, of which the first 1000 were eliminated as burn-in. Figures 5.8 and 5.9 show the log-weight for the 6th and 10th states respectively and Figure 5.10 the overall log weight results, while Figure 5.11 shows the population size. Figure 5.10 demonstrated the adaptation characteristics of $W_{\text{up},t}$. The results in Tables 5.5 and 5.6 show that the MCDWIS results are more accurate than the M-H results. The MCDWIS method was found to be faster and more accurate than the

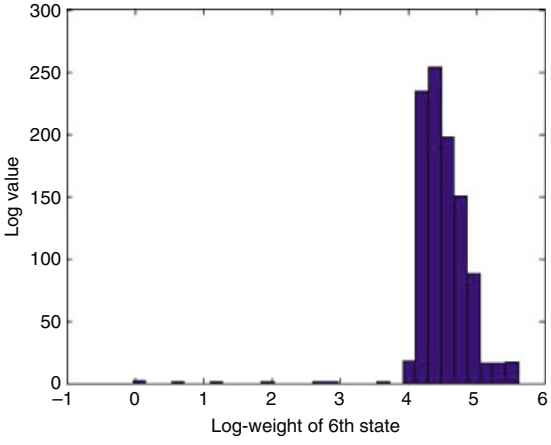

Figure 5.8 Log weight of the sixth state

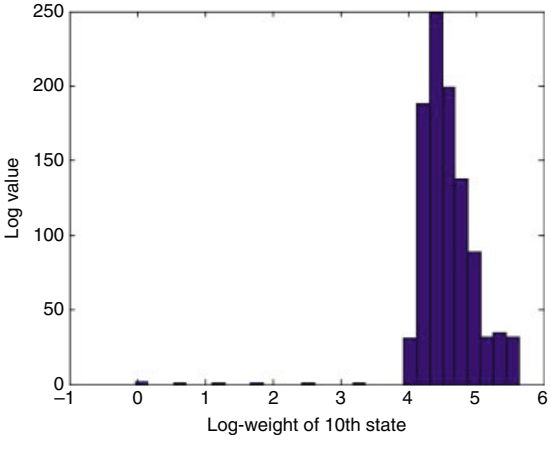

Figure 5.9 Log weight of the 10th state

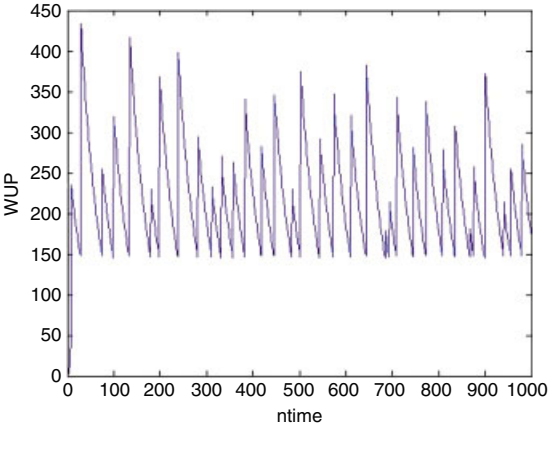

Figure 5.10 W_{up} adaptation

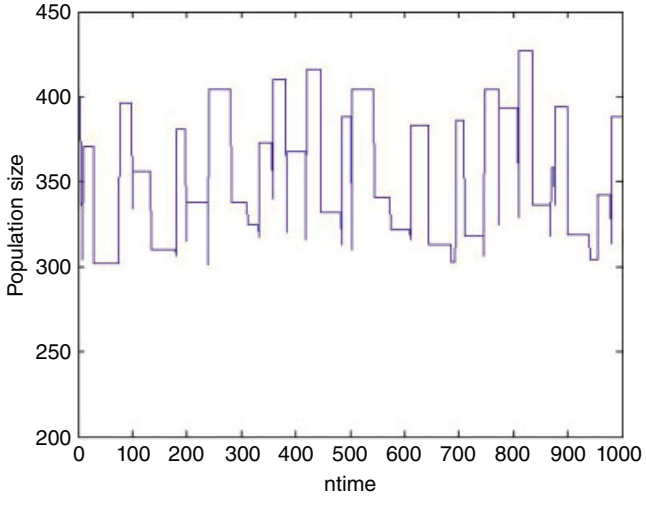

Figure 5.11 Population size

Table 5.5 Updating vector

D	Initial $\boldsymbol{\theta}_0$	$\boldsymbol{\theta}$ vector M-H method	$\boldsymbol{\theta}$ vector MCDWIS method
A_{x1}	2.7×10^{-8}	2.31×10^{-8}	2.2366×10^{-8}
A_{x2}	2.7×10^{-8}	2.68×10^{-8}	2.4680×10^{-8}
A_{x3}	2.7×10^{-8}	2.17×10^{-8}	2.7389×10^{-8}
I_{x1}	3.1×10^{-4}	2.85×10^{-4}	3.9876×10^{-4}
I_{x2}	3.1×10^{-4}	2.83×10^{-4}	2.3254×10^{-4}
I_{x3}	3.1×10^{-4}	2.77×10^{-4}	2.1756×10^{-4}

Table 5.6 Natural frequencies

Mode	Measured frequency	Initial frequency	M-H frequency	MCDWIS frequency
1	53.9	51.0389	53.92	53.05
2	117.3	115.7929	122.05	118.68
3	208.4	199.8772	210.93	208.15
4	253.0	246.0752	258.94	256.08
5	445	389.1767	410.33	446.96

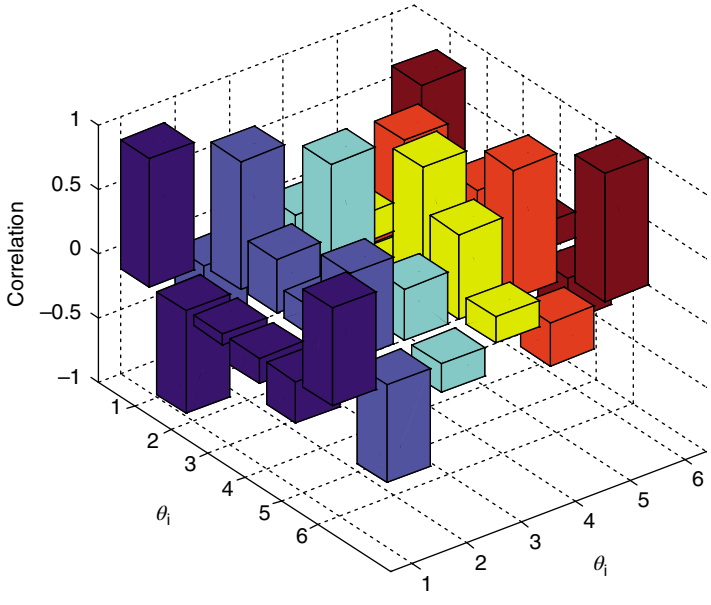

Figure 5.12 Correlation between samples

M-H algorithm. Additionally, the MCDWIS method was observed to be stable and reliable for higher-dimensional problems. The weight characteristics were applied to estimate the unbiased normalisation constants using the IS estimate, $\hat{R}_t(\boldsymbol{\theta}_t,\boldsymbol{\theta}^*) = R_t(\boldsymbol{\theta}_t,\boldsymbol{\theta}^*)$. Figure 5.12 demonstrates the correlation between samples of the 100th state, and the symmetry over the diagonals of the correlation shows the correlation of the samples.

5.8 Conclusions

The MCDWIS algorithm was applied for finite element model updating to approximate unbiased estimates controlled within a targeted reduced variance range from the originally drawn samples. By applying population control procedure the system was made adaptive to the needs of the environment and fixing the posterior distribution function. The MCDWIS was compared to the M-H method and found to be computationally more efficient and accurate than the M-H method. The MCDWIS approach provided a number of advantages compared to the M-H algorithm: the DWIS allows for adjusting the mixing abilities of the system (allowing for high or low correlated samples) by tuning freedom parameters according to the specific problem, and it offered stable estimates and converged faster.

References

Bassetti F, Diaconis P (2006) Examples comparing importance sampling and the Metropolis algorithm. *Illinois Journal of Mathematics* **50**:67–91.

Berg BA (2003) Metropolis importance sampling for rugged dynamical variables. *Physical Review Letters* **90**:180601/1–180601/4.

Bishop CM (2006) *Pattern Recognition and Machine Learning*. New York: Springer-Verlag.
Blitzstein J, Diaconis P (2015) A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Mathematics* **6**:489–522.
Boulkaibet I (2014) Finite element model updating using Markov chain Monte Carlo techniques. PhD thesis. University of Johannesburg.
Boulkaibet I, Mthembu L, Marwala T, Friswell M, Adhikari S (2015) Finite element model updating using the shadow hybrid Monte Carlo technique. *Mechanical Systems and Signal Processing* **52–53**:115–132.
Brooks S, Gelman A, Jones GL, Meng X-L (2011) *Handbook of Markov Chain Monte Carlo*. Boca Raton, FL: Chapman & Hall/CRC.
Dempster M (2006) Sequential importance sampling algorithms for dynamic stochastic programming. *Journal of Mathematical Sciences* **133**:1422–1444.
Elvira V, Martino L, Luengo D, Bugallo MF (2015) Efficient multiple importance sampling estimators. *IEEE Signal Processing Letters* **22**:1757–1761.
Fan H, Liang R (2015) Importance sampling based algorithm for efficient reliability analysis of axially loaded piles. *Computers and Geotechnics* **65**:278–284.
Friswell MI (2001) Finite element model updating using experimental test data. *Transactions of the Royal Society of London, Series A: Mathematical, Physical & Engineering Science* **359**:169–186.
Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**:97–109.
Joubert, D.J. (2015) Markov chain Monte Carlo method for finite element model updating. MEng thesis, University of Johannesburg.
Joubert DJ, Marwala T (2015) Monte Carlo dynamically weighted importance sampling for finite element model updating. arXiv:1510.04632.
Larjo A, Lähdesmäki H (2015) Using multi-step proposal distribution for improved MCMC Convergence in Bayesian network structure learning. *EURASIP Journal on Bioinformatics and Systems Biology*, **2015**(1), article no. 6.
Li W, Lin G (2015) An adaptive importance sampling algorithm for Bayesian inversion with multimodal distributions. *Journal of Computational Physics* **294**:173–190.
Liang F (2002) Dynamically weighted importance sampling in Monte Carlo computation. *Journal of the American Statistical Association*, **97**,(459), 807–821.
Liang F, Cheon S (2009) Monte Carlo dynamically weighted importance sampling for spatial models with intractable normalizing constants. *Journal of Physics: Conference Series*, **197**, article no. 012004.
Liang F, Liu C, Carroll R (2010) *Advanced Markov Chain Monte Carlo Methods*. Chichester: John Wiley & Sons, Ltd.
Mahani AS, Sharabiani MTA (2015) SIMD parallel MCMC sampling with applications for big-data Bayesian analytics. *Computational Statistics and Data Analysis* **88**:75–99.
Martino L, Elvira V, Luengo D, Corander J (2015) MCMC-driven adaptive multiple importance sampling. In: *Interdisciplinary Bayesian Statistics* (ed. Polpo A, Louzada F, Rifo LLR, Stern JM, Lauretto M), *Springer Proceedings in Mathematics and Statistics* **118**, pp. 97–109. Cham: Springer.
Marwala, T. (1997) A multiple criterion updating method for damage detection on structures. MEng thesis, University of Pretoria.
Marwala T (2010) *Finite Element Updating Using Computational Intelligence*. London: Springer.
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**:1087–1092.
Morales-Enciso S, Branke J (2015) Tracking global optima in dynamic environments with efficient global optimization. *European Journal of Operational Research* **242**:744–755.
Mottershead JE, Friswell MI (1995) *Finite Element Model Updating in Structural Dynamics* (1st edition). Dordrecht: Kluwer Academic Publishers.
Nadjafi M, Farsi, MA, Najafi A (2015) Uncertainty improving in importance sampling: an integrated approach with fuzzy-cluster sampling. Methodology and Applications: Proceedings of the European Safety and Reliability Conference, pp. 2107–2112.
Neal RM (2001) Annealed importance sampling. *Statistics and Computing* **11**:125–139.
Papaioannou I, Betz W, Zwirglmaier K, Straub D (2015) MCMC algorithms for subset simulation. *Probabilistic Engineering Mechanics* **41**:89–103.
Parpas P (2012) Importance sampling in stochastic programming: A Markov chain Monte Carlo approach. *INFORMS Journal on Computing* **27**:358–377.
Takaishi T (2013) Markov chain Monte Carlo versus importance sampling in Bayesian inference of the GARCH model. *Procedia Computer Science* **22**:1056–1064.

Tokdar ST, Kass RE (2009) Importance sampling: a review. *WIREs Computational Statistics* **2**:54–60.
Wang X, Shi Y, Ding D, Gu X (2016) Double global optimum genetic algorithm–particle swarm optimization-based welding robot path planning. *Engineering Optimization*, **48**(2):299–316.
Wong WH, Liang F (1997) Dynamic weighting in Monte Carlo and optimization. *Proceedings of the National Academy of Sciences of the United States of America* **94**:14220–14224.
Xu C, Wu Q, Sun L (2015) Importance sampling method for portfolio risk. *Journal of Tongji University* **43**:633–638.
Yilmaz E, Ozev S (2015) Adaptive-learning-based importance sampling for analog circuit DPPM estimation. *IEEE Design and Test* **32**:36–43.
Zhang D, Nie C, Xu B (2009) Importance sampling method of software reliability estimation. *Journal of Software* **20**:2859–2866.

6

Adaptive Metropolis–Hastings for Finite Element Updating

6.1 Introduction

This book is concerned with model selection to identify the most appropriate finite element model updating parameters given the measured data. A good model is one that satisfies the principle of Occam's razor, which states that the simplest model that describes the observed data is the best one (Anderson, 2008; Ando, 2010). This book has also studied criteria for model selection, such as the Akaike information criterion, optimal design, statistical hypothesis testing, Occam's razor, the Bayes factor, structural risk minimisation, cross-validation, and the Bayesian information criterion (Breiman, 2001; Burnham and Anderson, 2002). Nested sampling, cross-validation and regularisation techniques have been applied for finite element model selection in structures.

This book implements Bayesian statistics as a mechanism for model selection and finite element model updating to produce probabilistic finite element model updating. In this regard, given the distribution of the measured natural frequencies and some known physical properties, the Bayesian technique is used to estimate the distribution of the unknown parameters – referred to as the stochastic system identification procedure (Gelman *et al.*, 2013; Albert, 2009). The Bayesian formulation was solved using the Markov chain Monte Carlo method, which is a statistical procedure for computationally sampling a probability distribution function based on the Markov process, random walk and Monte Carlo simulation. Thus far, in this book, the following methods have been used to sample the posterior probability distribution function: the Metropolis–Hastings (M-H) algorithm, slice sampling and the Monte Carlo dynamically weighted importance sampling (MCDWIS) algorithm (Berg, 2004; Asmussen and Glynn, 2007; Joubert, 2015).

In this chapter, the adaptive Metropolis–Hastings (AMH) method is described and applied to update finite element models of a cantilevered beam, as well as an asymmetrical H-shaped

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

structure. The results of using the AMH method are compared to those obtained from using the MCDWIS algorithm.

6.2 Adaptive Metropolis–Hastings Algorithm

Sejdinovic *et al.* (2014) proposed the kernel based AMH method for sampling from a target distribution with significant non-linear characteristics. In this proposed method, the kernel Hilbert space is used in the trajectory of the Markov chain in such a way that the feature space covariance of the samples informs the choice of the proposal (Berlinet and Thomas, 2004). The proposed method was found to be computationally efficient and simple to design and use because the kernel Hilbert space moves were calculated analytically. In this technique, the proposal distribution was a Gaussian distribution with a mean and covariance influenced by the current sample which was in support of the target distribution and updated the local covariance. This method does not necessitate the gradients of the target distribution and performs better than fixed and adaptive methods.

Luengo and Martino (2013) proposed an AMH algorithm to sample from general multimodal and multi-dimensional target distributions. In the proposed technique, the proposal density was a mixture of Gaussian densities with weights, mean vectors and covariance matrices identified using recursive rules. Griffin and Walker (2013) proposed AMH methods which used the product of a proposal density and K duplicates of the target density to describe a joint density. The Gibbs technique was used with the Metropolis algorithm to sample the target distribution (Casella and George, 1992). Cui *et al.* (2011) applied the AMH method for Bayesian approximation of the parameters of a numerical model of a geothermal reservoir. This procedure was proposed for the sake of the sampling efficiency and was observed to offer significant improvement, particularly in problems with large degrees of freedom.

Giordani and Kohn (2010) proposed an AMH method that used a mixture of normal distributions as a proposal distribution and was applied to real situations and simulated data, while Holden *et al.* (2009) proposed an AMH algorithm that learned from all previous states (not just one previous state) in the chain, except the current state. Using previous sates in sampling helps the proposed distribution to be a better approximation of the target distribution. This adaptive algorithm is useful in problems of Bayesian estimation. Cai *et al.* (2008) proposed a general type of AMH method based on M-H and Gibbs sampling, which was more efficient than other adaptive methods such as the normal kernel coupler. Wolpert and Lee (2006) applied the AMH method to sample and optimise a number of spin-glasses and found it to perform better than the classical M-H method. Behmanesh and Moaveni (2013) applied the AMH algorithm in finite element model updating of a footbridge structure on Tufts University campus, and used this to simulate structural damage using the measured modal properties and generate a convergent Markov chain for the proposal probability density function.

In order to understand the AMH method, it is important to locate it within the framework of the M-H algorithm (Calderhead, 2014; Vu *et al.*, 2014; Monroe and Cai, 2014; Eberle, 2014; Van Dyk and Jiao, 2015). The M-H algorithm is a simple Markov chain Monte Carlo method, which is utilised to accept or reject generated random samples. Usually, these samples are generated simply by adding some random noise from some proposal distribution to the previous state in the Markov chain with this noise. The most frequently used proposal distribution is the normal distribution with a prescribed variance, and this can be expressed mathematically as

$\boldsymbol{\theta}^* = \boldsymbol{\theta}_t + \mathcal{N}(0, \sigma^2)$ The M-H algorithm then accepts the current state if the target probability of the current state is higher than the target probability of the previous state. If the target probability of the current state is lower than the target probability of the previous state, then the M-H algorithm conducts a random guess to accept the current state with a lower probability. The Markov chain is generated such that all accepted states converge to the target stationary probability distribution.

In the M-H algorithm, samples are drawn from the proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_t)$ where $\boldsymbol{\theta}^*$ indicates the proposed value. The M-H algorithm is summarised by the following steps (Metropolis *et al.*, 1953; Hastings, 1970; Roberts *et al.*, 1997; Gelman *et al.*, 1996; Teller, 2001; Robert and Casella, 2004; Gubernatis, 2005; Brooks *et al.*, 2011; Joubert, 2015):

1. Initialise $\boldsymbol{\theta}_0$.
   For $j = 0 : M-1$ where $M-1$ is the number of samples
2. Propose a new sample from the proposal distribution $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_j)$.
3. Generate uniform random variables, $u \sim U(0,1)$.
4. Transit to the next stage according to the following transition probability:

$$\text{If } u < \alpha(\boldsymbol{\theta}_{j+1}, \boldsymbol{\theta}^*) = \min\left\{1, \frac{P(\boldsymbol{\theta}^*|\mathcal{D})q(\boldsymbol{\theta}_j|\boldsymbol{\theta}^*)}{P(\boldsymbol{\theta}_{j-1}|\mathcal{D})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_j)}\right\}$$

$$\boldsymbol{\theta}_{j+1} = \boldsymbol{\theta}^*$$

else

$$\boldsymbol{\theta}_{j+1} = \boldsymbol{\theta}_j .$$

5. Return to step 2.

The AMH algorithm works by perturbing the proposal distribution through continuously scaling the system covariance to attain convergence. The adaptation mechanism impacts both the size and spatial orientation of the proposal probability density function, and a scaling factor is augmented to the covariance matrix to identify the optimal covariance matrix that correlates to the target probability distribution (Haario *et al.*, 2001; Joubert, 2015). Therefore, the covariance matrix is a function of the scaling factor.

In this chapter, we review the use of the AMH method in finite element model updating to estimate the finite element model updating posterior probability distribution function (Joubert, 2015):

$$P(\boldsymbol{\theta}|\mathcal{D}) \propto \frac{1}{Z_s(\alpha,\beta)} \exp\left(-\frac{\beta_c}{2}\sum_{i}^{N_m}\left(\frac{f_i^m - f_i}{f_i^m}\right)^2 - \sum_{i}^{d}\frac{\alpha_i}{2}\left\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_i^0\right\|^2\right), \tag{6.1}$$

where the normalisation constant is

$$Z_s(\alpha,\beta) = \left(\frac{2\pi}{\beta_c}\right)^{N_m/2} \prod_{i=1}^{N_m} f_i^m (2\pi)^{d/2} \prod_{i=1}^{d} \frac{1}{\sqrt{\alpha_i}}, \tag{6.2}$$

f_i^m is the ith measured frequency, f_i is the ith calculated frequency, and α_i and β_c are hyperparameters.

In this regard, the AMH method was applied in the same way as by Haario *et al*. (2001) where the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ was approximated from the output of the available Monte Carlo output density functions, and this permits the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ to adapt as the simulation progresses.

For the random walk Gaussian proposal distribution, $q(\boldsymbol{\theta}'|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}', \boldsymbol{\theta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ which is the probability density function of the multivariate Gaussian with mean $\boldsymbol{\theta}$ and a covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$. Gelman *et al*. (1996) observed that the desired optimal covariance matrix can be written as $(2.38^2/d)\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ under particular conditions where d is the dimension of the stochastic variables: the mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ of the target probability distribution $P(.)$.

The proposed distribution can be used to sample the states $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{m-1}$ at time $m-1$, where $\boldsymbol{\theta}_0$ is the initially estimated state, and then the covariance $\boldsymbol{\Sigma}_m = \mathrm{Cov}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{m-1})$ is calculated from the mean of the current state $\boldsymbol{\theta}_{m-1}$ in the Gaussian proposal distribution. Note that the target distribution is bounded in the subset $S \in \mathbb{R}^d$. In the adaptive algorithm, the covariance matrix of the proposed distribution depends on the previous chain of states, and this can be represented, after an initial period, by $s_d \mathrm{Cov}(\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_{h-1}) + s_d \varepsilon \boldsymbol{I}_d$ where s_d is the scaling factor, ε is a constant greater than zero but smaller than the size of the search space and $\boldsymbol{I}_d$ is the d-dimensional identity matrix. To implement the adaptation of the covariance matrix, an initial covariance matrix is chosen such that $\boldsymbol{\Sigma}_0 > 0$, and the length of the initial period is selected such that $t_0 > 0$; then the covariance can be written as (Gelman *et al*., 1996; Newman and Barkema, 1999; Joubert, 2015)

$$\boldsymbol{\Sigma}_m = \begin{cases} \boldsymbol{\Sigma}_0 & m \le t_0, \\ s_d \mathrm{Cov}(\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_{h-1}) + s_d \varepsilon \boldsymbol{I}_d, & m \ge t_0. \end{cases} \tag{6.3}$$

The covariance matrix $\boldsymbol{\Sigma}_m$ is consequently a function of m variables from the search space $\mathbb{R}^d$ and is a uniform positive definitive matrix (Horn and Johnson, 1990; Bhatia, 2007). If we consider the definition of the empirical covariance matrix which is established by the observed states in the search space and suppose that the elements are column vectors, $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m \in \mathbb{R}^d$, then (Gelman *et al*., 1996; Joubert, 2015)

$$\mathrm{Cov}(\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_m) = \frac{1}{m}\left(\sum_{i=1}^{m} \boldsymbol{\theta}_i \boldsymbol{\theta}_i^T - (h-1)\bar{\boldsymbol{\theta}}_t \bar{\boldsymbol{\theta}}_t^T\right), \tag{6.4}$$

where $\bar{\boldsymbol{\theta}} = (m+1)^{-1} \sum_{j=0}^{i} \boldsymbol{\theta}_j$. The recursive formula to solve the consecutive covariance, for $m \ge t_0 + 1$, can then be mathematically expressed as follows (Gelman *et al*., 1996; Joubert, 2015):

$$\boldsymbol{\Sigma}_{m+1} = \frac{m-1}{m}\boldsymbol{\Sigma}_m + \frac{s_d}{m}\left(m\bar{\boldsymbol{\theta}}_{t\,m-1}\bar{\boldsymbol{\theta}}_{t\,m-1}^T - (m+1)\bar{\boldsymbol{\theta}}_m\bar{\boldsymbol{\theta}}_m^T + \boldsymbol{\theta}_m\boldsymbol{\theta}_m^T + \varepsilon \boldsymbol{I}_d\right). \tag{6.5}$$

This can then be additionally abridged by presenting a gain factor sequence and expressing it as $\{y_m\}$. Accordingly the following conditions must then be satisfied as is needed by the standard stochastic estimation (Haario *et al*., 2001; Joubert, 2015):

$$\sum_{m=1}^{\infty}\gamma_m = \infty, \qquad \sum_{m=1}^{\infty}\gamma_m^{1+\delta} < \infty \quad \text{for some } \delta \in (0,1]. \tag{6.6}$$

The AMH algorithm can therefore be described mathematically as follows (Haario *et al*., 2001; Joubert, 2015):

1. Given $\boldsymbol{\theta}_0$, $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$.
2. Begin the Markov chain process and for iteration $h+1$ ensure the availability of $\boldsymbol{\theta}_m$, $\boldsymbol{\mu}_m$ and $\boldsymbol{\Sigma}_m$.
   a. Establish the M-H kernel to estimate the probability density distribution at $m+1$, $P_{\mu_m,\,\Sigma_m}(\boldsymbol{\theta}_m, \cdot)$.
   b. Update the mean vector and covariance matrix

$$\boldsymbol{\mu}_{m+1} = \boldsymbol{\mu}_m + \gamma_{m+1}\,(\boldsymbol{\theta}_{m+1} - \boldsymbol{\mu}_m)$$

$$\boldsymbol{\Sigma}_{m+1} = \boldsymbol{\Sigma}_m + \gamma_{m+1}\left[(\boldsymbol{\theta}_{m+1} - \boldsymbol{\mu}_m)(\boldsymbol{\theta}_{m+1} - \boldsymbol{\mu}_m)^T - \boldsymbol{\Sigma}_m\right].$$

3. Go to step 2.

The AMH method is applied to a cantilevered beam and an asymmetrical H-shaped structure and the results are compared to those obtained from the MCDWIS described in Chapter 5.

6.3 Application 1: Cantilevered Beam

In this chapter finite element model updating based on the AMH method is used to resolve the differences between the results from a finite element model of a cantilevered beam with experimental data measured by Marwala (1997). The results of the AMH method are compared to those obtained from the MCDWIS method as was done in Chapter 5. The AMH algorithm is applied by perturbing the covariance matrix of the system with the expectation that it will converge to the unknown target distribution. The AMH algorithm begins by using cumulated observations through pre-run and burn-in stages to update the covariance and scaling factor, making the sampling is more effective at an early stage of the simulation thus diminishing the number of evaluations needed.

The proposal distribution for the simulated states is a Gaussian distribution with mean at the current state $\boldsymbol{\theta}_t$ ($\boldsymbol{\mu}_t$ as presented in the previous AHM algorithm) and covariance $s_d\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is the covariance matrix determined by the spatial distribution of the states, $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m \in \mathbb{R}^d$. The scaling parameter s_d depends only on the dimension d of the sample space, and this implementation guides the proposal distribution to approach an appropriately scaled Gaussian approximation of the target distribution and thus improves the efficiency of the simulation. The proposal distribution was set to fall between $x_{\min} = [1.8\times10^{11},\ 1.8\times10^{11},\ 1.8\times10^{11},$

$1.8\times10^{11}1.8\times10^{11}]$ and $x_{\max}=[2.4\times10^{11}, 2.4\times10^{11}, 2.4\times10^{11}, 2.4\times10^{11}, 2.4\times10^{11}]$. The initial vector was initialised as $\theta_0=[2.4\times10^{11}, 2.4\times10^{11}, 2.4\times10^{11}, 2.4\times10^{11}, 2.4\times10^{11}]$. The initial covariance matrix, $\boldsymbol{\Sigma}_0$ was set to 1.2×10^{10} on the diagonals of a 5×5 matrix. The adaptation on the scaling parameter s_d and the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ implemented the pre-run and burn-in stages of the algorithm. Figures 6.1 and 6.2 show the adaptation of the scaling parameter s_d and the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$. Figure 6.3 shows the correlation representation between the updating parameters and Figure 6.4 show the adaptation and the probability density estimate of the first updating parameter.

Tables 6.1 and 6.2 show the results for the parameter vector $\boldsymbol{\theta}=\{E_1,E_2,E_3,E_4,E_5\}$ and the resulting vector of computed natural frequencies $\boldsymbol{f}=\{f_1,f_2,f_3,f_4,f_5\}$.

The total mean error is the average between the calculated frequency and the measured frequencies, and the results show that the MCDWIS algorithm is 0.38% more accurate than the AMH algorithm. Therefore, the MCDWIS method is a more reliable technique for finite element model updating as it is more accurate and computationally efficient than the AMH technique, while the AMH algorithm is more computationally expensive due to the number

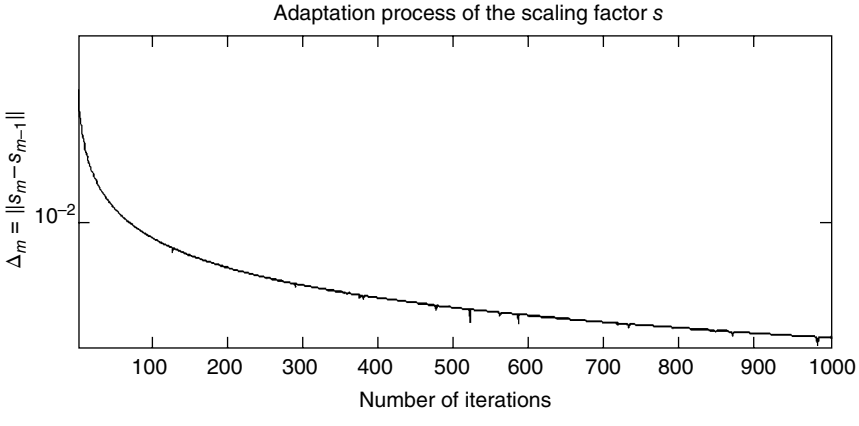

Figure 6.1 Adaptation of scaling parameter

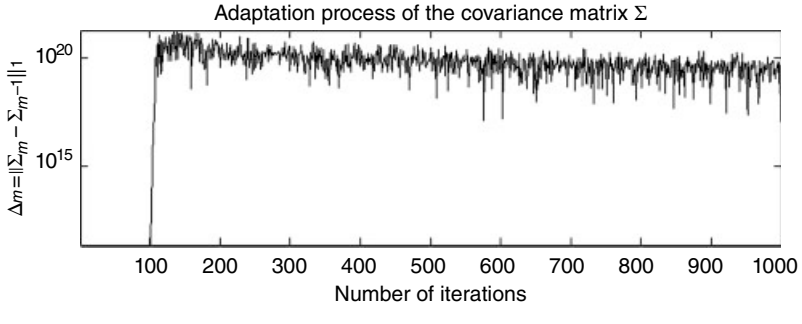

Figure 6.2 Adaptation of the covariance matrix

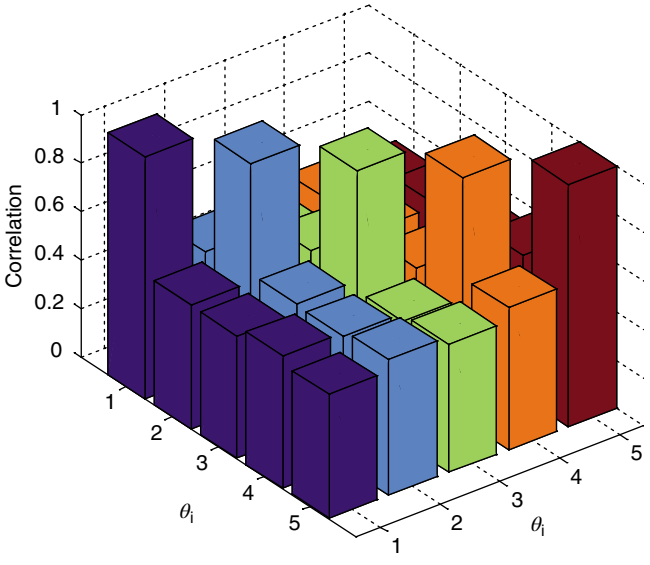

Figure 6.3 Parameter correlation

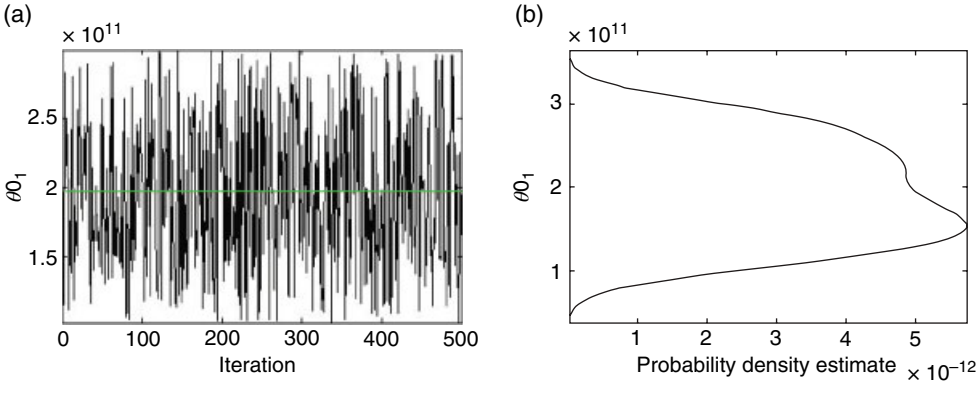

Figure 6.4 Adaptation of (a) first updating parameter and (b) probability density estimation

of iterations needed to update the covariance matrix for convergence to the target distribution and the fact that the scaling coefficient adapts during the pre-run and burn-in stages of the algorithm whereas the covariance updating occurs in the burn-in stage. Furthermore, in the AMH algorithm the runtime depends on the number of samples drawn; however, it was observed that with fewer burn-in iterations the algorithm performs poorly. Accordingly, the total number of iterations used in implementation of the AMH algorithm was 2100: 100 for the pre-run step, 1000 for burn-in step and 1000 in the final run step.

Table 6.1 Cantilever frequency results

Parameters	Initial $\boldsymbol{\theta}_0$	$\boldsymbol{\theta}$ vector MCDWIS method	$\boldsymbol{\theta}$ vector AMH method
E_1	2.4×10^{11}	2.1524×10^{11}	2.0993×10^{11}
E_2	2.4×10^{11}	2.1091×10^{11}	2.0997×10^{11}
E_3	2.4×10^{11}	2.0430×10^{11}	2.1006×10^{11}
E_4	2.4×10^{11}	2.0342×10^{11}	2.1001×10^{11}
E_5	2.4×10^{11}	2.0336×10^{11}	2.0994×10^{11}

Table 6.2 Cantilevered beam frequency results

Mode	Measured frequency (Hz)	Initial frequency (Hz)	MCDWIS frequency (Hz)	AMH frequency (Hz)
1	31.9	32.7	30.77	30.59
2	197.9	209.4	194.81	195.91
3	553	594.8	553.24	556.34
4	1082.2	1177.8	1094.4	1101.7
5	1781.5	1961.7	1824.00	1834.9

6.4 Application 2: Asymmetrical H-Shaped Beam

An asymmetrical H-shaped aluminium structure was modelled using the finite element model technique in the MATLAB environment and modal frequencies were obtained by using modal analysis (Marwala, 1997). This structure was excited using an electromagnetic shaker and modal hammer, and the response was measured using a roving accelerometer. Data acquisition was carried out and data were recorded; the uncertainty parameters in the finite element model were the corresponding areas and moments of inertia, $\boldsymbol{\theta} = [A_{x1}, A_{x2}, A_{x3}, I_{x1}, I_{x2}, I_{x3}]$.

The AMH and the MCDWIS techniques were employed to update a finite element model of the asymmetrical H-shaped beam. The statistical input parameters (i.e. covariance) and population control parameters were selected to meet the model needs in the MCDWIS algorithm. The results of the AMH method were compared to the results the MCDWIS algorithm. The measured natural frequencies were: 53.9, 117, 208.4 and 445 Hz. Physical properties not updated in the finite element model included Young's modulus, the material density and dimensions of the structure. Young's modulus was fixed at 7.2×10^{10} N/m^2 density at 2785 kg/m^2.

The covariance matrix $\boldsymbol{\Sigma}$ was started with a value of 5×10^{-8} on the diagonals of a 6×6 matrix and the proposal was truncated between $\theta_{\min} = [1.7265\times10^{-8}, 1.7265\times10^{-8}, 1.7265\times10^{-8}, 2.1556\times10^{-4}, 2.1556\times10^{-4}, 2.1556\times10^{-4}]$ and $\theta_{\max} = [3.7265\times10^{-8}, 3.7265\times10^{-8}, 3.7265\times10^{-8}, 4.1556\times10^{-4}, 4.1556\times10^{-4}, 4.1556\times10^{-4}]$. The methods were initialised at $\boldsymbol{\theta}_0 = [2.7265\times10^{-8}, 2.7265\times10^{-8}, 2.7265\times10^{-8}, 3.1556\times10^{-4}, 3.1556\times10^{-4}, 3.1556\times10^{-4}]$.

Figures 6.5 and 6.6 demonstrate the adaptation of the scaling factor and covariance matrix, respectively. Owing to the adaptation procedure of this algorithm it is vital to confirm that the covariance matrix is positive definite. This is because the covariance matrix should be

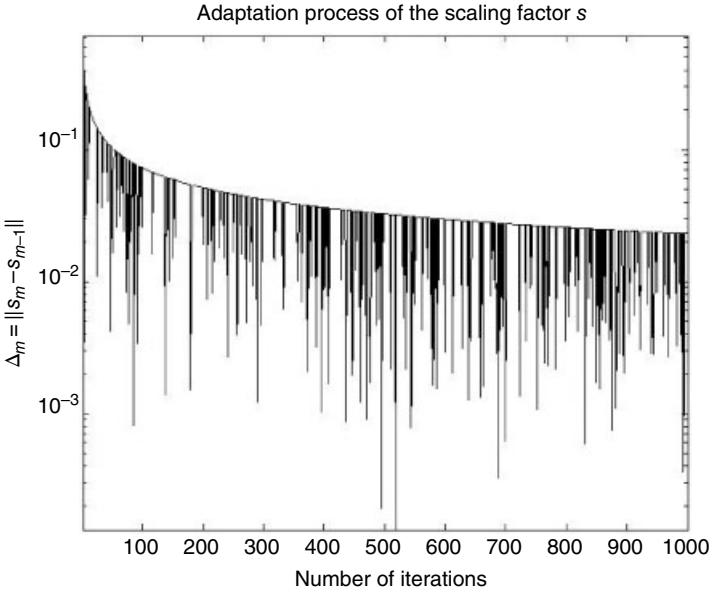

Figure 6.5 Adaption of scaling factor

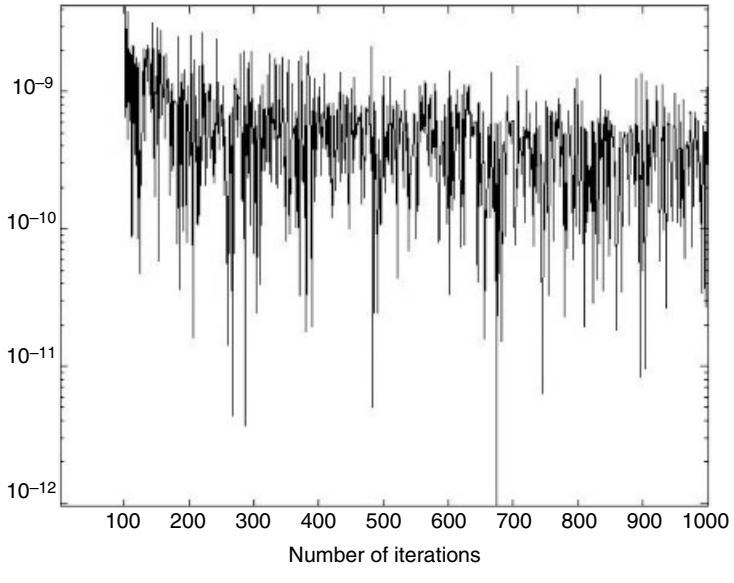

Figure 6.6 Adaption of covariance matrix

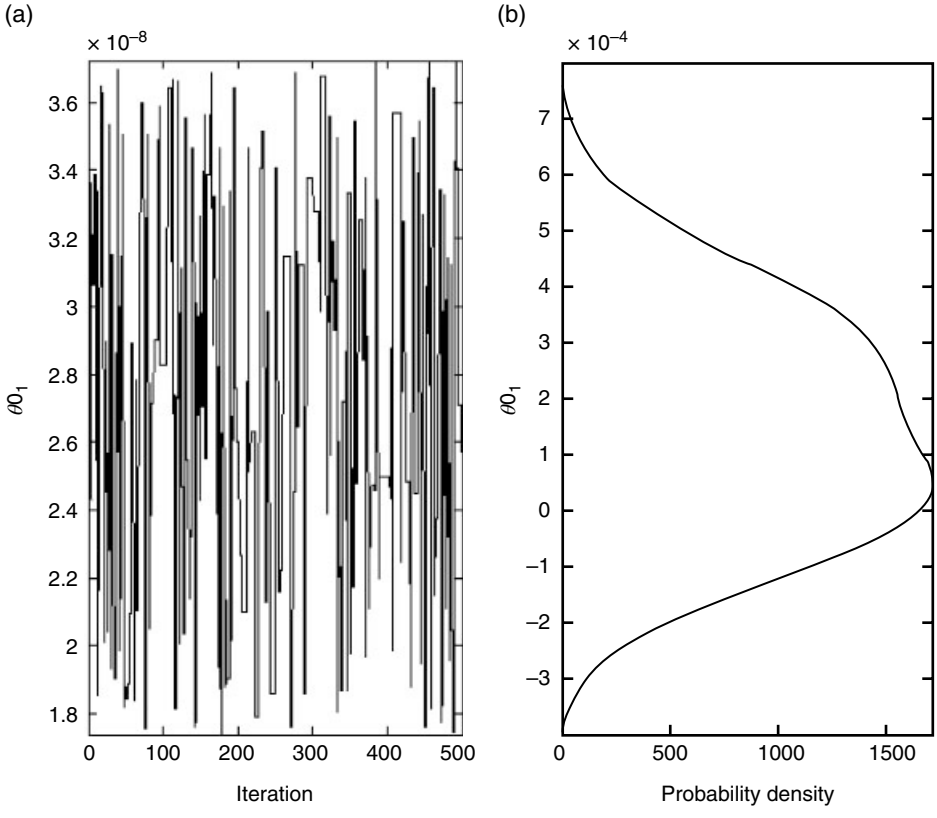

Figure 6.7 Adaption of (a) first parameter and (b) the kernel density estimate

symmetric, $\Sigma_{ij} = \Sigma_{ji}$, which is derived from the general definition $\Sigma_{ij} = E_{x\sim p}\left[\left(x_i - \mu_i\right)\left(x_j - \mu_j\right)\right]$ where i and j range from 1 to d. An additional prerequisite is to use the MATLAB Cholesky positive definite test. Figure 6.7 demonstrates the adaptation of a sample parameter of the first uncertain parameter θ_1 with its corresponding kernel density approximation which is fundamentally an approximation of the probability density. Figure 6.8 demonstrates the correlation between the sampled parameters. The algorithm was once more specified with 100 pre-run iterations, 1000 burn-in samples and in the final run 500 samples 20 times.

Tables 6.3 and 6.4 show the results for the parameter vector $\boldsymbol{\theta}$ and the resulting vector of computed natural frequencies, $\boldsymbol{f} = \{f_1, f_2, f_3, f_4, f_5\}$. From the results it can be determined that MCDWIS computes the most accurate sample estimates and the MCDWIS is 0.86% more accurate than AMH method.

6.5 Application 3: Aircraft GARTEUR Structure

The aircraft GARTEUR model is a familiar apparatus in the structural dynamics research arena, offering researchers shared infrastructure for testing different hypotheses in finite element model updating. This chapter applies the AMH and the MCDWIS procedures with higher-

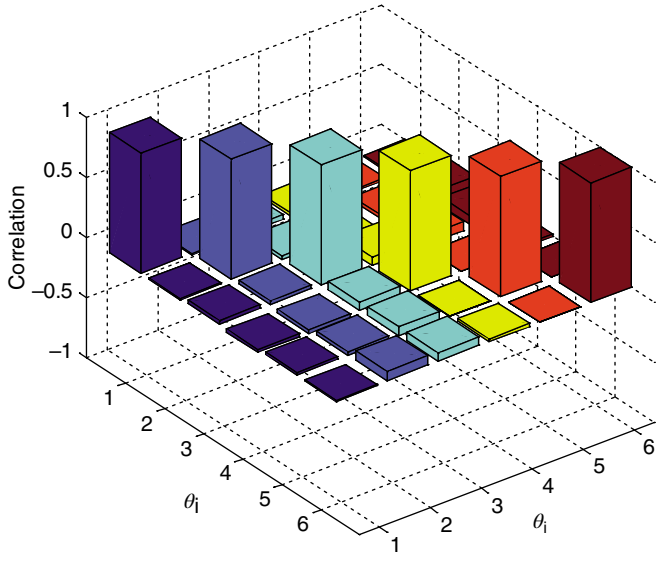

Figure 6.8 3D bar graph of the parameter correlation

Table 6.3 H-beam variable parameter vector results

Parameters	Initial $\boldsymbol{\theta}_0$	$\boldsymbol{\theta}$ vector MCDWIS method	$\boldsymbol{\theta}$ vector AMH method
A_{x1}	2.7×10^{-8}	2.2366×10^{-8}	2.7485×10^{-8}
A_{x2}	2.7×10^{-8}	2.4680×10^{-8}	2.6251×10^{-8}
A_{x3}	2.7×10^{-8}	2.7389×10^{-8}	3.0174×10^{-8}
I_{x1}	3.1×10^{-4}	3.9876×10^{-4}	3.4330×10^{-4}
I_{x2}	3.1×10^{-4}	2.3254×10^{-4}	2.5934×10^{-4}
I_{x3}	3.1×10^{-4}	2.1756×10^{-4}	2.3990×10^{-4}

Table 6.4 Beam frequency results

Mode	Measured frequency (Hz)	Initial frequency (Hz)	MCDWIS frequency (Hz)	AMH frequency (Hz)
1	53.9	51.0389	53.05	53.4294
2	117.3	115.7929	118.68	119.6341
3	208.4	199.8772	208.15	211.4322
4	254.0	246.0752	256.08	255.6062
5	445	389.1767	446.96	429.7475

dimensional uncertainty in the model to evaluate the usefulness of the procedures as well as their integrity in the finite model updating problem. The conventional M-H method was observed to give poor results for the higher-dimensional model (Joubert, 2015). Consequently, it is inferred that the ordinary M-H is devoid of computational efficiency and estimation precision for more complex finite element models such as the GARTEUR (Degener and Hermes, 1996; Friswell, 2003). The nominal length and width of the aircraft structure were 1.5 and 3 m respectively, it was made of aluminium with a total mass of 44 kg and with a visco-elastic layer bonded to the wings to stimulate increased damping effect. The Euler–Bernoulli beam elements with standard isotropic properties were used to create a finite element model.

As reported by Friswell (2003), the natural frequencies were determined to be 6.38, 16.10, 33.13, 33.53, 35.65, 48.38, 49.43, 55.08, 63.04, 66.52 Hz. The updating vector involved the right wing moments of inertia and torsional stiffness ($RI_{\min}$, $RI_{\max}$, RI_{tors}), the left wing moments of inertia and torsional stiffness ($LI_{\min}$, $LI_{\max}$, LI_{tors}), the vertical tail moment of inertia $(VTP_{I_{\min}})$ and the overall density of the structure ρ, and this is written as follows:

$\boldsymbol{\theta} = [\rho, VTP_{I_{\min}}, LI_{\min}, LI_{\max}, RI_{\min}, RI_{\max}, LI_{\text{tors}}, RI_{\text{tors}}]$ (Joubert, 2015). The vector used for finite element model updating was initiated at $\boldsymbol{\theta}_0 = [2785, 8.34\times10^{-9}, 8.34\times10^{-9}, 8.34\times10^{-9}, 8.34\times10^{-9}, 8.34\times10^{-7}, 4\times10^{-8}, 4\times10^{-8}]$ and confined between $\boldsymbol{\theta}_{\max} = [3500, 12\times10^{-9}, 10.2\times10^{-9}, 12\times10^{-9}, 10.2\times10^{-9}, 12\times10^{-7}, 6\times10^{-8}, 6\times10^{-8}]$ and $\boldsymbol{\theta}_{\min} = [2500, 6\times10^{-9}, 8\times10^{-9}, 6\times10^{-9}, 8\times10^{-9}, 6\times10^{-7}, 3\times10^{-8}, 3\times10^{-8}]$, and the diagonals of the covariance matrix were initialised at $\boldsymbol{\sigma} = [5\times10^{2}, 5\times10^{-9}, 5\times10^{-9}, 5\times10^{-9}, 5\times10^{-9}, 5\times10^{-7}, 5\times10^{-8}, 5\times10^{-8}]$.

On implementing the AMH, the first 1000 samples were deemed the pre-run stage of the simulation and the covariance matrix was proactively adjusted. The following 1000 samples were deemed the burn-in stage, and the scaling factor was modified and the covariance matrix was scaled to further increase the accuracy of the approximation. Figures 6.9 and 6.10 show the characteristics of the adaptation of the scaling factor and covariance, respectively. Figure 6.11 shows the second parameter adaptation and kernel density, while Figure 6.12 demonstrates the correlation between samples. The results of the parameter values of the updating vector as well as the natural frequencies are given in later in the chapter.

On implementing the MCDWIS a fast run of the M-H procedure was conducted which offered the initial set of samples for the MCDWIS to begin with. Using the 1000 iterations,

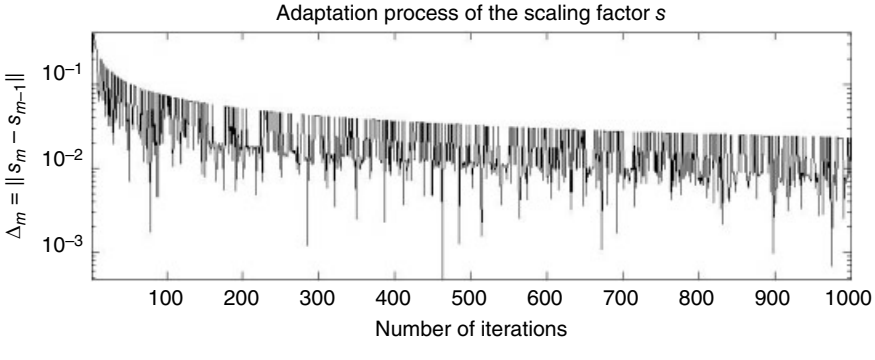

Figure 6.9 Adaptation of the scaling factor

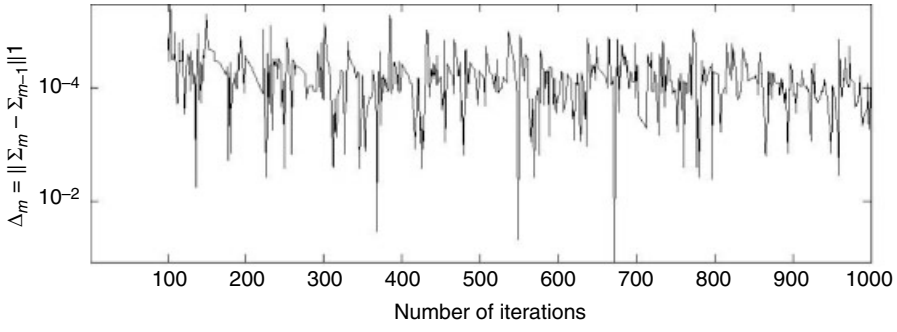

Figure 6.10 Adaptation of the covariance matrix

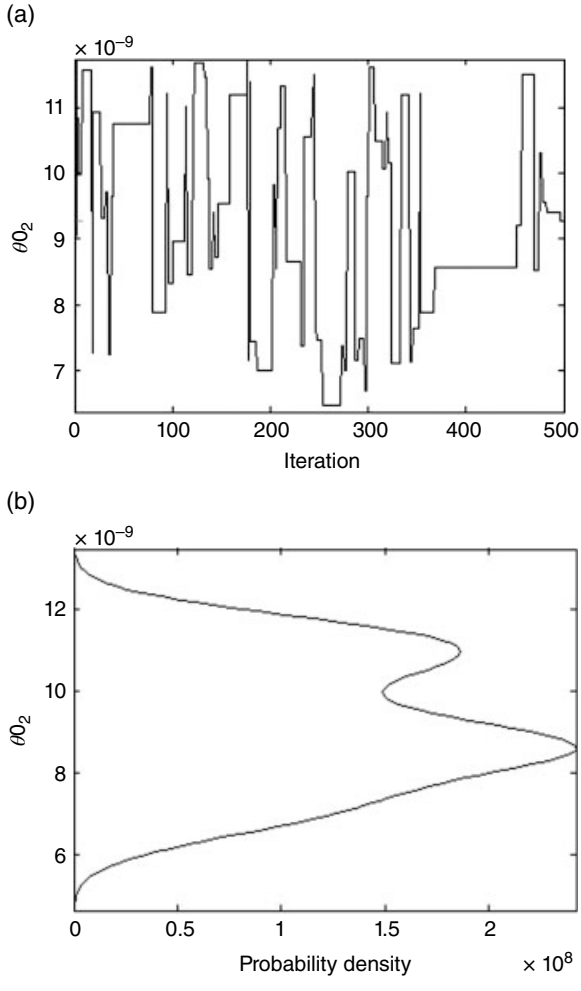

Figure 6.11 Adaptation of (a) second updated parameter and (b) the kernel density

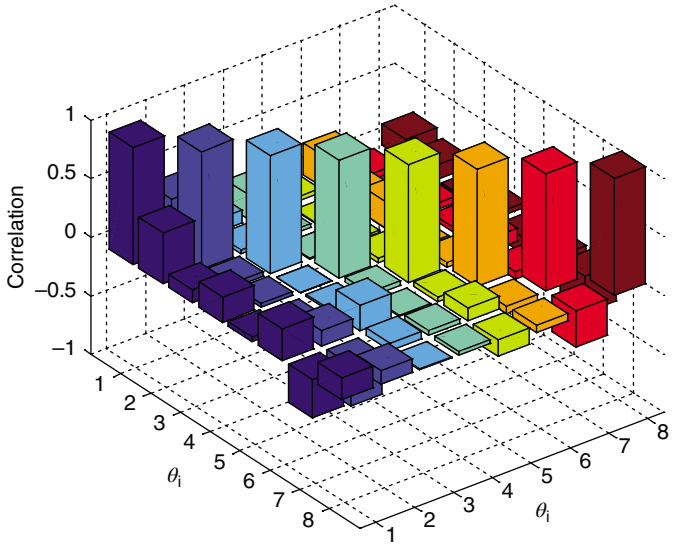

Figure 6.12 Correlation between different parameters

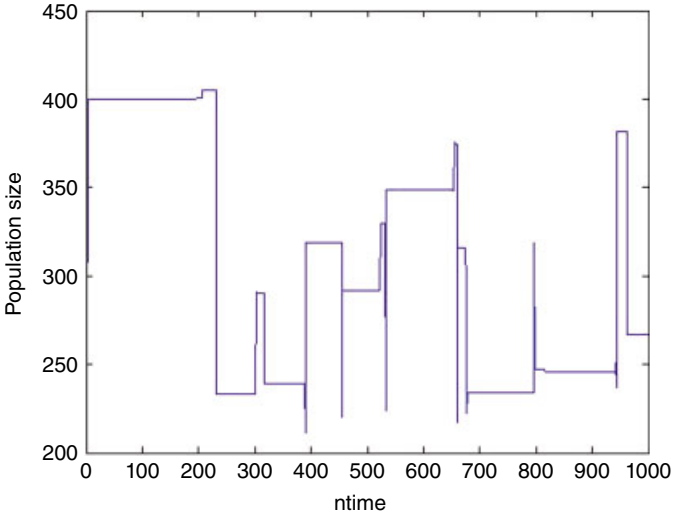

Figure 6.13 Adaptation of population size

the first 200 samples were rejected as burn-in and 200 samples were uniformly chosen from the remaining 800. The first samples were allotted a weight of 1 and the parameters for the algorithm parameters were $N = 1000$, burn-in 250, $W_c = e^7$, $n_{\min} = 100$, $n_{\max} = 1000$, $n_{\text{low}} = 200$, $n_{\text{up}} = 500$, $\lambda = 2$ and $k = \log_{10} W_c$. The adaptation of the population size is shown in Figures 6.13–6.15 and demonstrates an apparent Gaussian shape of the log weight at iterations 6 and 10, respectively. Because of the higher-dimensional vector, the procedure allocates

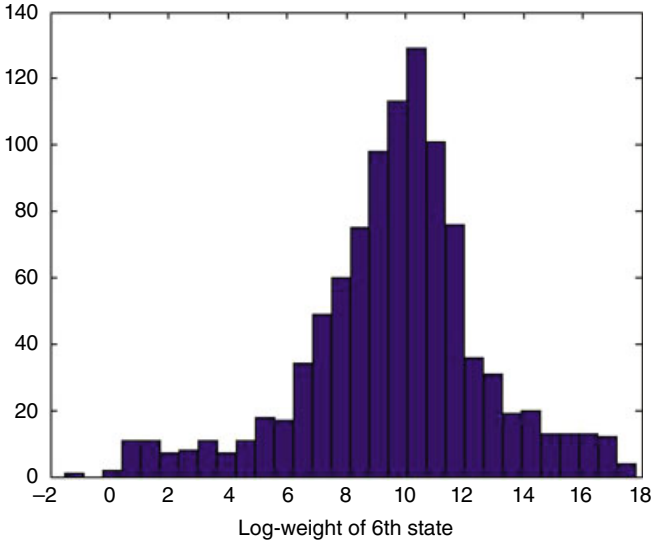

Figure 6.14 The sixth state of the log weight

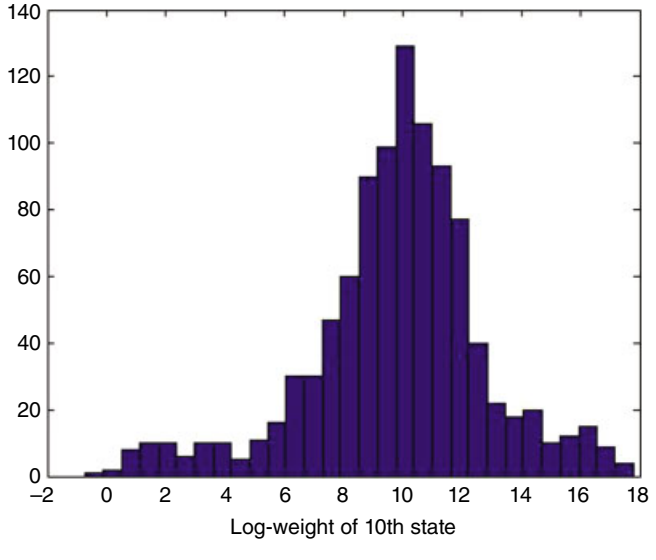

Figure 6.15 The 10th state of the log weight

unusually high weights in the first iteration and exponentially decreases and then varies. The results of the updating procedure are shown in Tables 6.5 and 6.6.

Table 6.5 gives the values of the elements in the updated vector after the simulation. Table 6.6 gives the resultant natural frequencies corresponding to each respective mode shape $\boldsymbol{\theta} = [\rho, VTP_{I_{\min}}, LI_{\min}, LI_{\max}, RI_{\min}, RI_{\max}, LI_{\text{tors}}, RI_{\text{tors}}]$. The results demonstrate that the MCDWIS method was substantially more accurate when compared to the AMH technique.

Table 6.5 Updating vector results

Parameters	Initial $\boldsymbol{\theta}_0$	$\boldsymbol{\theta}$ vector MCDWIS method	$\boldsymbol{\theta}$ vector AMH method
ρ	2785	2735	2847
$VTP_{I_{\min}}$	8.34×10^{-9}	7.1213×10^{-9}	9.1275×10^{-9}
LI_{min}	8.34×10^{-9}	1.02×10^{-9}	9.2278×10^{-9}
LI_{max}	8.34×10^{-7}	7.9716×10^{-7}	7.6652×10^{-7}
RI_{min}	8.34×10^{-9}	1.0202×10^{-9}	9.6244×10^{-9}
RI_{max}	8.34×10^{-7}	6.0834×10^{-7}	8.3867×10^{-7}
LI_{tors}	4×10^{-8}	4.1102×10^{-8}	4.4388×10^{-8}
RI_{tors}	4×10^{-8}	3.6212×10^{-8}	3.5720×10^{-8}

Table 6.6 Natural frequency results

Mode	Measured frequency	Initial frequency	AMH frequency	MCDWIS frequency
1	6.38	5.71	6.0099	6.3021
2	16.10	15.29	16.0666	15.9326
3	33.13	32.53	32.7091	32.3035
4	33.53	34.95	34.3994	34.0557
5	35.65	35.65	37.0980	35.8587
6	48.38	45.14	46.8730	48.5070
7	49.43	54.69	52.9663	49.5125
8	55.08	55.60	54.8028	54.2495
9	63.04	60.15	62.0316	63.5583
10	66.52	67.56	67.4895	67.4334

6.6 Conclusion

This chapter has presented the AMH procedure and Bayesian statistics for finite element model updating. In the AMH technique the Gaussian proposal distribution was adapted using the full information assembled previously; because of the adaptive features of the system, the AMH method is non-Markovian but also possesses full ergodic properties. The AMH technique was found to be simple to implement. The AMH method was implemented to update a finite element model of a cantilevered beam, an asymmetrical H-shaped structure and aircraft structure, and the results were compared to those from the MCDWIS method. The results obtained showed that the MCDWIS performed better than the AMH.

References

Albert J (2009) *Bayesian Computation with R* (2nd edition). New York: Springer.
Anderson DR (2008). *Model Based Inference in the Life Sciences*. London: Springer.
Ando T (2010) *Bayesian Model Selection and Statistical Modeling*. Boca Raton, FL: CRC Press.

Asmussen S, Glynn PW (2007) *Stochastic Simulation: Algorithms and Analysis*. Stochastic Modelling and Applied Probability, Vol. 57. New York: Springer.

Behmanesh I, Moaveni B (2013) Probabilistic damage identification of the Dowling Hall footbridge using Bayesian FE model updating. *Topics in Model Validation and Uncertainty Quantification* **5**:43–51.

Berg BA (2004) *Markov Chain Monte Carlo Simulations and Their Statistical Analysis*. Singapore: World Scientific.

Berlinet A, Thomas C (2004) *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Boston: Kluwer Academic Publishers.

Bhatia R (2007) *Positive Definite Matrices*. Princeton, NJ: Princeton University Press.

Breiman L (2001) Statistical modeling: the two cultures. *Statistical Science* **16**:199–231.

Brooks S, Gelman A, Jones GL, Meng X-L (2011) *Handbook of Markov Chain Monte Carlo*. Boca Raton, FL: Chapman & Hall /CRC.

Burnham KP, Anderson DR (2002) *Model Selection and Multi-model Inference: A Practical Information-Theoretic Approach* (2nd edition). New York: Springer-Verlag.

Cai B, Meyer R, Perron F (2008) Metropolis–Hastings algorithms with adaptive proposals. *Statistics and Computing* **18**:421–433.

Calderhead B (2014) A general construction for parallelizing Metropolis–Hastings algorithms. *Proceedings of the National Academy of Sciences of the United States of America* **111**:17408–17413.

Casella G, George EI (1992) Explaining the Gibbs sampler. *American Statistician* **46**:167–174.

Cui T, Fox C, O'Sullivan MJ (2011) Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm. *Water Resources Research*. **47**, article no. W10521.

Degener M, Hermes M (1996) Ground vibration test and finite element analysis of the GARTEUR SM-AG19 testbed. Report IB 232-96J 08, Deutsche Forschungsanstalt für Luft und Raumfahrt e.V. Institut für Aeroelastik.

Eberle A (2014) Error bounds for Metropolis–Hastings algorithms applied to perturbations of Gaussian measures in high dimensions. *Annals of Applied Probability* **24**:337–377.

Friswell MI (2003) Generation of validated structural dynamic models – results of a benchmark study utilizing the GARTEUR SM-AG19 testbed. *Mechanical Systems and Signal Processing*, **17**: 9–20.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) *Bayesian Data Analysis* (3rd edn). Boca Raton, FL: Chapman & Hall/CRC.

Gelman A, Roberts RO, Gilks WR (1996) Efficient Metropolis jumping rules. In: *Bayesian Statistics 5* (ed. Bernardo JM, Berger JO, Dawid AP, Smith AFM). Oxford: Oxford University Press.

Giordani P, Kohn R (2010) Adaptive independent Metropolis–Hastings by fast estimation of mixtures of normal. *Journal of Computational and Graphical Statistics* **19**:243–259.

Griffin JE, Walker SG (2013) On adaptive Metropolis–Hastings methods. *Statistics and Computing* **23**:123–134.

Gubernatis JE (2005) Marshall Rosenbluth and the Metropolis algorithm. *Physics of Plasmas* **12**, 057303.

Haario H, Saksman E, Tamminen J (2001) An adaptive Metropolis algorithm. *Bernoulli* **7**:223–242.

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**:97–109.

Holden L, Hauge R, Holden M (2009) Adaptive independent Metropolis–Hastings. *Annals of Applied Probability* **19**:395–413.

Horn RA, Johnson CR (1990) *Matrix Analysis*. Cambridge: Cambridge University Press.

Joubert, DJ (2015) Markov chain Monte Carlo method for finite element model updating. MEng thesis, University of Johannesburg.

Luengo, D. and Martino, L. (2013) Fully adaptive Gaussian mixture Metropolis–Hastings algorithm. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 6148–6152.

Marwala, T. (1997) A multiple criterion updating method for damage detection on structures. MEng thesis, University of Pretoria.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**:1087–1092.

Monroe S, Cai L (2014) Estimation of a Ramsay-curve item response theory model by the Metropolis–Hastings Robbins–Monro algorithm. *Educational and Psychological Measurement* **74**:343–369.

Newman MEJ, Barkema GT (1999) *Monte Carlo Methods in Statistical Physics*. New York: Oxford University Press.

Robert C, Casella G (2004) *Monte Carlo Statistical Methods*. New York: Springer.

Roberts GO, Gelman A, Gilks WR (1997) Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability* **7**:110–120.

Sejdinovic D, Strathmann H, Garcia ML *et al.* (2014) Kernel adaptive Metropolis–Hastings. Proceedings of the 31st International Conference on Machine Learning, pp. 3640–3653. Beijing: International Machine Learning Society.

Teller E (2001) *Memoirs: A Twentieth-Century Journey in Science and Politics*. New York: Perseus Publishing.

Van Dyk DA, Jiao X (2015) Metropolis–Hastings within partially collapsed Gibbs samplers. *Journal of Computational and Graphical Statistics* **24**:301–327.

Vu T, Vo B-N, Evans R (2014) A particle marginal Metropolis–Hastings multi-target tracker. *IEEE Transactions on Signal Processing* **62**:3953–3964.

Wolpert DH, Lee CF (2006) An adaptive Metropolis–Hastings scheme: sampling and optimization. *Europhysics Letters* **76**:353–359.

7

Hybrid Monte Carlo Technique for Finite Element Model Updating

7.1 Introduction

The hybrid Monte Carlo (HMC) technique, also known as the Hamiltonian Monte Carlo method, was first introduced in physics by Duane *et al.* (1987) to solve higher-dimensional complex engineering problems (Gupta *et al*., 1988; Alfaki, 2008; Cheung and Beck, 2009; Hanson, 2001; Beskos *et al.*, 2013; Boulkaibet, 2014; Boulkaibet *et al.*, 2012, 2014a). Duane *et al.* (1987) applied the HMC method to lattice field theory simulations of quantum chromodynamics for simulating molecular dynamics. The HMC algorithm was first used for statistical learning in neural network models by Neal (1998). It has been observed that the HMC technique performs better than the traditional Markov chain Monte Carlo (MCMC) method in correlated spaces and/or problems with high-dimensional spaces (Chen *et al*., 2001; Neal, 2011).

Cheung and Beck (2009) applied the HMC method to update a linear structural dynamic model with 31 uncertain parameters. This approach, formulated probabilistically using a Bayesian approach, was able to model the uncertainties associated with the underlying structural system. Boulkaibet (2014) compared the performance of the HMC algorithm with two other sampling techniques, the Metropolis–Hastings (M-H) method and slice sampling, by updating a structural beam model, and the HMC algorithm gave more accurate results and a faster convergence rate.

In the HMC method, the trajectory of the algorithm is governed by the derivative of the target log-density probability which moves towards areas of high probability in a short space of time during the search process (Beskos *et al.*, 2013). Through this approach, the updated parameter vector is handled as a system displacement and an auxiliary variable, called the *momentum* vector, introduced to construct a new molecular dynamic system. The total energy of the system,

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

generally referred to as the *Hamiltonian* function, is evaluated using the Velocity Verlet algorithm (also known as the leapfrog integrator). When using the HMC algorithm, the move step depends on the time step used by the leapfrog integrator. This means that a relatively large time step helps ensure fast convergence, in a reasonable number of iterations. In this chapter, the HMC algorithm is presented in Sections 7.2 and 7.3. Sections 7.4 and 7.5 present the results obtained when the HMC algorithm is employed to update structural beams: the cantilever and the asymmetrical H-shaped beams, respectively. Section 7.6 concludes the chapter.

7.2 Hybrid Monte Carlo Method

The HMC combines a molecular dynamic trajectory with a Monte Carlo (MC) rejection step (Akhmatskaya *et al*., 2009; Boulkaibet *et al.*, 2012; Heckman and Leamer, 2001; Marwala, 2012). This MC technique uses the gradient of the error, and this ensures that the simulation does sample through the regions of higher probabilities and, therefore, increases the time it takes to converge to a stationary probability density function (PDF). This technique is considered a type of Markov chain with transition between states achieved by changing between the 'stochastic' and 'dynamic' moves. The 'stochastic' moves allow the method to sample states with different total energy, while the 'dynamic' moves are achieved by applying the Hamiltonian dynamics and allowing the technique to search for states with a total energy that is nearly constant. Basically, the HMC method can be considered as a type of Monte Carlo sampling, which is driven by the gradient of the PDF at each state.

Aleksandrov *et al*. (2010) used the HMC method to study the vapour–liquid equilibria of copper, while Ghoufi and Maurin (2010) used the HMC method to estimate the structural transitions of a porous metal-organic framework material, and confirmed that combining the hybrid osmotic Monte Carlo method with a 'phase mixture' model is an effective procedure to approximate the adsorption behaviour. Zhang *et al*. (2010) successfully used the HMC method to simulate stress-induced texture evolution, while Rei *et al*. (2010) used the HMC method successfully in a single vehicle routeing problem. Wendt *et al.* (2011) used the HMC technique in graphics, while Bogaerts (2009) used the HMC technique to study the effects of oxygen addition to argon glow discharges. Qian *et al*. (2011) used the HMC technique to calculate the animal population affected by an environmental catastrophe, while Kulak (2009) used the HMC method to simulate fluorescence anisotropy decay. Other successful applications of the HMC method are in fluoride ion–water clusters (Suzuki *et al*., 2010), modelling structural heterogeneity in polyprolines (Hoefling *et al.*, 2011), spacecraft thermal models (Cheng *et al.*, 2011) and modelling the probability distributions in Riemannian space (Paquet and Viktor, 2011).

Through the use of the HMC method, a dynamic system is analysed by introducing an auxiliary variable, known as the momentum $p \in \mathbb{R}^d$. The updating vector $\boldsymbol{\theta}$, representing the variables in the finite element model requiring updating, is treated as displacement within the Hamiltonian dynamics context. Thus, the total energy (or Hamiltonian function) of the dynamic system is defined by Duane *et al.* (1987), Boulkaibet (2014) and Boulkaibet *et al.* (2014a, 2014b) as

$$H(\boldsymbol{\theta}, \mathbf{p}) = V(\boldsymbol{\theta}) + W(\mathbf{p}), \tag{7.1}$$

where $V(\boldsymbol{\theta})$ represents the potential energy which can be defined by $V(\boldsymbol{\theta}) = -\ln(P(\boldsymbol{\theta}|\mathcal{D}))$, $P(\boldsymbol{\theta}|\mathcal{D})$ being the posterior PDF, and where $W(\mathbf{p})$ is the kinetic energy and is given by $W(\mathbf{p}) = \mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}/2$ which depends only on $\mathbf{p}$ and a selected positive definite matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$. The partial derivative of Hamiltonian function, which determines the variation of the pair $(\boldsymbol{\theta}, \mathbf{p})$ over time, is given by Duane *et al.* (1987) and Boulkaibet (2014) as

$$\frac{d\boldsymbol{\theta}}{dt} = \frac{\partial H}{\partial p} = \mathbf{M}^{-1}\mathbf{p}(t), \tag{7.2}$$

$$\frac{dp}{dt} = \frac{\partial H}{\partial \boldsymbol{\theta}} = -\nabla V(\boldsymbol{\theta}(t)). \tag{7.3}$$

The HMC has several properties, described in the next section, that need to be satisfied in order to construct the Markov chain, and these properties can be proved easily.

7.3 Properties of the HMC Method

The Hamiltonian dynamics has several properties that are important for constructing the MCMC method: time reversibility, volume preservation and energy conservation (Duane *et al.*, 1987; Boulkaibet, 2014; Boulkaibet *et al.*, 2014a).

7.3.1 Time Reversibility

The Hamiltonian dynamics is invariant under the following transformations: $\dot{\boldsymbol{\theta}} = -\boldsymbol{\theta}, \dot{\boldsymbol{p}} = \boldsymbol{p}$ and $\dot{\boldsymbol{t}} = -\boldsymbol{t}$.This demonstrates that the Hamiltonian dynamics is invariant to the direction of time, that is, time reversibility. The propriety of time reversibility is used to demonstrate that Equations 7.2 and 7.3 converge to a desired invariant distribution.

7.3.2 Volume Preservation

The Hamiltonian dynamics preserves volume in the space $(\boldsymbol{\theta}, \mathbf{p})$, and this is known as Liouville's theorem (Rossberg, 1983; Goldstein, 1980). The mapping T_S from the state at time t, to the state at time $t+s$ is defined by Equations 7.2 and 7.3. This is applied to points in a certain region $\mathbb{R}$ of the space $(\boldsymbol{\theta}, \mathbf{p})$ with a volume v. The image of this region, under the mapping T_S, also has the same volume v (Neal, 2011). The volume preservation of the Hamiltonian dynamics can be proved by showing that the divergence of the vector field defined by Equations 7.2 and 7.3 is zero (Arnold, 1989):

$$\sum_{i=1}^{d}\left[\frac{\partial}{\partial\theta_i}\frac{d\theta_i}{dt} + \frac{\partial}{\partial p_i}\frac{dp_i}{dt}\right] = \sum_{i=1}^{d}\left[\frac{\partial}{\partial\theta_i}\frac{dH}{dp_i} - \frac{\partial}{\partial p_i}\frac{dH}{d\theta_i}\right] = \sum_{i=1}^{d}\left[\frac{\partial^2 H}{\partial\theta_i\partial p_i} - \frac{\partial^2 H}{\partial p_i\partial\theta_i}\right] = 0 \tag{7.4}$$

7.3.3 Energy Conservation

An additional important property of the Hamiltonian dynamics is that the Hamiltonian function (total energy) is invariant (conserved). This can be simply proven from Equations 7.2 and 7.3 as follows (Neal, 2011):

$$\frac{dH}{dt} = \sum_{i=1}^{d} \left[\frac{d\theta_i}{dt}\frac{\partial H}{\partial \theta_i} + \frac{dp_i}{dt}\frac{\partial H}{\partial p_i} \right] = \sum_{i=1}^{d} \left[\frac{\partial H}{\partial p_i}\frac{\partial H}{\partial \theta_i} - \frac{\partial H}{\partial \theta_i}\frac{\partial H}{\partial p_i} \right] = 0. \tag{7.5}$$

7.4 The Molecular Dynamics Algorithm

In this chapter, molecular dynamic simulations are accomplished under the conditions of the canonical ensemble. The density function $\rho(\boldsymbol{\theta}, \mathbf{p})$ of the canonical ensemble follows a Boltzmann distribution (Boulkaibet, 2014; Boulkaibet *et al.*, 2014a, 2015). This ensemble is a good illustration of the distribution of the Hamiltonian system, where a positive feature of this ensemble is that the position $\boldsymbol{\theta}$ and momentum $\mathbf{p}$ are independent for separable Hamiltonian functions (Skeel and Tupper, 2005).

The joint distribution function derived from the Hamiltonian function can be written as (Boulkaibet, 2014)

$$\rho(\boldsymbol{\theta}, \mathbf{p}) \propto \exp(-\beta_B H(\boldsymbol{\theta}, \mathbf{p})), \tag{7.6}$$

where $\beta_B = 1/K_B T$, K_B is the Boltzmann constant and T is a constant temperature. The expression $\rho(\boldsymbol{\theta}, \mathbf{p})$ can be rewritten as follows (Boulkaibet, 2014)

$$\rho(\boldsymbol{\theta}, \mathbf{p}) \propto \exp(-\beta_B V(\boldsymbol{\theta}))\exp(-\beta_B W(\mathbf{p})) \tag{7.7}$$

or

$$\rho(\boldsymbol{\theta}, \mathbf{p}) \propto P(\boldsymbol{\theta}|D)\exp\left(-\beta_B \mathbf{p}^T \mathbf{M}^{-1}\mathbf{p}/2\right). \tag{7.8}$$

Sampling $\boldsymbol{\theta}$ from the posterior distribution can also be done by sampling the pair $(\boldsymbol{\theta}, \mathbf{p})$ from the joint distribution $\rho(\boldsymbol{\theta}, \mathbf{p})$, where $\boldsymbol{\theta}$ and $\mathbf{p}$ are independent, using $H(\boldsymbol{\theta}, \mathbf{p})$, which is a separable Hamiltonian function.

The evolution of $(\boldsymbol{\theta}, \mathbf{p})$ through time t can be achieved, numerically, by using the velocity Verlet (leapfrog) scheme (Cheung and Beck, 2009; Boulkaibet, 2014; Boulkaibet *et al.*, 2014a):

$$\mathbf{p}\left(t + \frac{\delta t}{2}\right) = \mathbf{p}(t) - \frac{\delta t}{2}\nabla V(\boldsymbol{\theta}(t)), \tag{7.9}$$

$$\boldsymbol{\theta}(t + \delta t) = \boldsymbol{\theta}(t) + \delta t \mathbf{M}^{-1}\mathbf{p}\left(t + \frac{\delta t}{2}\right), \tag{7.10}$$

$$\mathbf{p}(t+\delta t)=\mathbf{p}\left(t+\frac{\delta t}{2}\right)-\frac{\delta t}{2}\nabla V(\boldsymbol{\theta}(t+\delta t)), \tag{7.11}$$

where δt is the time step and the gradient ∇V is achieved numerically using the finite difference approach as follows (Boulkaibet, 2014):

$$\frac{\partial V}{\partial \theta_i}\cong\frac{V(\boldsymbol{\theta}+\boldsymbol{\Delta}h)-V(\boldsymbol{\theta}-\boldsymbol{\Delta}h)}{2h\Delta_i}. \tag{7.12}$$

Here, $\boldsymbol{\Delta}=[\Delta_1,\Delta_2,\ldots,\Delta_N]$ is the perturbation vector and h is a scalar which determines the size of the perturbation of $\boldsymbol{\theta}$.

Practically, the leapfrog algorithm does not preserve the property of Equations 7.2 and 7.3, where the PDF is proportional to $\exp(-\beta_B H(\boldsymbol{\theta},\mathbf{p}))$ (Beskos *et al.*, 2013). With the intention of satisfying the property of Equations 7.2 and 7.3, the MC accept–reject step must be added. In such a case, after each iteration of Equations 7.9–7.11, the resulting candidate state is accepted or rejected according to the Metropolis criterion based on the value of the Hamiltonian $H(\boldsymbol{\theta},\mathbf{p})$ (Metropolis *et al.*, 1953).

Thus, if the pair $(\boldsymbol{\theta},\mathbf{p})$ is the initial state and $(\boldsymbol{\theta}^*,\mathbf{p}^*)$ is the state after Equations 7.9–7.11 have been updated, then the new candidate $(\boldsymbol{\theta}^*,\mathbf{p}^*)$ is accepted with probability $\min(1,\exp\{-\beta_B\Delta H\})$ where $\Delta H=H(\boldsymbol{\theta}^*,\mathbf{p}^*)-H(\boldsymbol{\theta},\mathbf{p})$. The vector $\boldsymbol{\theta}^*$ obtained is then used for the next iteration. The criterion used to stop this algorithm is defined by the number of $\boldsymbol{\theta}$ samples (N_S). The HMC algorithm can be summarised as follows (Boulkaibet, 2014):

1. Initialise the algorithm by specifying $\boldsymbol{\theta}_0$.
2. Initialise $\mathbf{p}_0$ in such a way that $\mathbf{p}_0\sim\mathcal{N}(0,\boldsymbol{M})$.
3. Initialise the leapfrog algorithm with $(\boldsymbol{\theta},\mathbf{p})$ and run the algorithm for L time steps to obtain $(\boldsymbol{\theta}^*,\mathbf{p}^*)$.
4. Update the finite element model to obtain the new analytical frequencies and then compute $H(\boldsymbol{\theta}^*,\mathbf{p}^*)$.
5. Accept $(\boldsymbol{\theta}^*,\mathbf{p}^*)$ with probability $\min(1,\exp\{-\beta_B\Delta H\})$.
6. Repeat steps 3–5 for N_S samples.

To improve the HMC algorithm, the leapfrog algorithm is evaluated L steps at every iteration where L is the trajectory length. This ensures that large moves of the algorithm during the search process and the final time step used at each iteration are equal to $\delta t\times L$. Generally, the value of L is uniformly selected from $\{1,L_{\max}\}$. This can assist in avoiding the non-ergodicity problem and ensures good performance of the algorithm (Cheung and Beck, 2009). Also, the value of $L_{\max}$ should not be large to reduce the computational cost of the algorithm ($L_{\max}\approx 30$ for small systems and 15 for large systems).

In practice, the time step of the HMC technique is bounded ($\delta t_{\min}<\delta t<\delta t_{\max}$). The performance of the HMC method degrades when the time step used by the leapfrog algorithm and/or the system size (number of updated parameters) is large (Boulkaibet, 2014; Izaguirre and Hampton, 2004; Sweet *et al.*, 2009); a large time step decreases the accuracy of the leapfrog algorithm (relatively large errors will be expected in the in the Hamiltonian function value) and thus the Hamiltonian function is not conserved. A significantly large time step increases the

numerical errors caused by the integrator used to evaluate the Hamiltonian function. The error caused by the velocity Verlet (leapfrog) integrator can cause the Hamiltonian function to fluctuate and thus to increase the rejection rate of the algorithm, and therefore no samples are accepted in the case where $\delta t \geq \delta t_{\max}$. To avoid a large rejection rate, the time step of this algorithm is set to less than $\delta t_{\max}$ and in the case where δt is too small, a high acceptance rate of the HMC algorithm is obtained. Nevertheless, a large number of samples and/or a large trajectory length are needed to cover more space during the search, particularly when $\delta t \leq \delta t_{\min}$. Alternatively, a relatively large δt facilitates a significant jump from the existing samples and therefore ensures a better exploration of the phase space. This chapter uses certain modifications to the HMC to improve its efficiency.

7.5 Improving the HMC

The HMC algorithm can be improved to efficiently sample from complex distributions. This can be achieved as follows (Boulkaibet, 2014):

1. Choose a good time step to optimise the move step and the acceptance rate of the algorithm.
2. Suppress the random walk when the momentum is sampled.
3. Set a better HMC gradient precision to improve the accuracy of the algorithm.

Next, the improvements of the HMC algorithm based on the above-mentioned points as well as the modification of the gradient when very large amounts of data are involved are discussed.

7.5.1 Choosing an Efficient Time Step

The acceptance rate of a candidate sample at the end of the $(\boldsymbol{\theta}, \mathbf{p})$ trajectory for the Hamiltonian dynamics of Equations 7.9–7.11 is affected by the discretisation errors introduced by the integration algorithm. The distance vector $\mathcal{L}(\delta t)$, which is also called the move step, moves in the $(\boldsymbol{\theta}, \mathbf{p})$ space after one evolution depends on the time step δt. Cheung and Beck (2009) maximised the distance vector $\mathcal{L}(\delta t)$ by using a small number of samples and empirically explored different δt to achieve maximum $\mathcal{L}(\delta t)$. However, this method is not efficient, especially when the algorithms reach the search space boundaries. Another way to deal with the time step problem is to use a variable step size or adaptive time step method.

Huang and Leimkuhler (1997) proposed an adaptive Verlet method which is based on time reparameterisation. A variable, τ, was introduced into the leapfrog scheme and this variable is related to a smooth, scalar-valued function $\psi(\boldsymbol{\theta}, \mathbf{p})$. The adaptive leapfrog scheme obtained is explicit if ψ depends only on $\boldsymbol{\theta}$, whereas it is a semi-explicit if ψ depends on $\mathbf{p}$. The function ψ is given by $\psi(\boldsymbol{\theta}, \mathbf{p}) = \left(\boldsymbol{p}^T \boldsymbol{M}^{-2} \boldsymbol{p} + \nabla_p V\right)^{1/2}$ (Alfaki, 2008; Huang and Leimkuhler, 1997). The scheme is semi-explicit, symmetric and time-reversible where $\psi(\boldsymbol{\theta}, \mathbf{p}) = \psi(\boldsymbol{\theta}, -\mathbf{p})$. The adaptive Verlet scheme is given by (Boulkaibet, 2014):

$$\mathbf{p}\left(t + \frac{\delta t}{2}\right) = \mathbf{p}\left(t - \frac{\delta t}{2}\right) - \frac{\delta t}{2} \cdot \left(\frac{1}{\tau(t + \delta t/2)} + \frac{1}{\tau(t - \delta t/2)}\right) \nabla V(\boldsymbol{\theta}(t)), \tag{7.13}$$

$$\tau\left(t+\frac{\delta t}{2}\right)=-\tau\left(t-\frac{\delta t}{2}\right)+\psi\left(\boldsymbol{\theta},\mathbf{p}\left(t+\frac{\delta t}{2}\right)\right)+\psi\left(\boldsymbol{\theta},\mathbf{p}\left(t-\frac{\delta t}{2}\right)\right), \tag{7.14}$$

$$\boldsymbol{\theta}(t+\delta t)=\boldsymbol{\theta}(t)+\frac{\delta t}{\tau(t+\delta t/2)}\mathbf{M}^{-1}\mathbf{p}\left(t+\frac{\delta t}{2}\right). \tag{7.15}$$

7.5.2 Suppressing the Random Walk in the Momentum

In a number of cases, the efficiency of MCMC method is limited by a strong dependency between components of the state. This might force the MCMC algorithms to move around the target distribution in small steps. This might also happen with the HMC algorithm when the momentum is drawn, and an effective way to deal with the dependencies between components, which are unavoidable, is by using an ordered over-relaxation approach to suppress the random walk of the momentum sampling process (Neal, 1998).

The ordered over-relaxation method can be used to sample from any type of distribution. The most common implementation used for the ordered over-relaxation is by employing the cumulative distribution function. In this chapter, the ordered over-relaxation method is used to obtain samples from the momentum distribution, which is built from the kinetic energy, given by $\Pi(\boldsymbol{p})=K^{-1}\exp(-\beta_B W(\boldsymbol{p}))$, where K^{-1} is the normalisation constant and $\mathbf{p}\in\mathbb{R}^d$. Supposing that $F(\boldsymbol{p}^i)$ is the cumulative distribution for the conditional distribution $\Pi\left(\boldsymbol{p}_i|\{\boldsymbol{p}_j\}_{i\neq j}\right)$, while $F^{-1}(\mathbf{p}^i)$ represents the inverse of the $F(\boldsymbol{p}^i)$, then the algorithm can be summarised as follows (Neal, 1996, 1998, 2011; Alfaki, 2008; Boulkaibet, 2014):

1. Compute $u=F(\boldsymbol{p})$.
2. Draw r from a Binomial$(k,\boldsymbol{u})$ distribution.
3. If $r>k-r$, randomly draw ϑ from $Beta(k-r+1,2r-k)$, where the new $u'=u\vartheta$.
4. If $r<k-r$, randomly draw ϑ from $Beta(r+1+1,k-2r)$, where the new $u'=1-(1-u)\vartheta$.
5. If $r=k-r$, the new $u'=u$.
6. The new momentum $\mathbf{p}=F^{-1}(u')$.

7.5.3 Gradient Computation

In general and due to the complexity of the modelled systems, the gradient of $V(\boldsymbol{\theta})$ is not analytically available. In this case numerical methods must be employed to identify its value. The most common technique uses the finite difference approximations which are based on forward and backward Taylor series expansion of $V(\boldsymbol{\theta})$. In this chapter, the first central difference approximation is used to compute the gradient, which in this work was more accurate than the forward/backward difference approximation.

However, in the case where the dimension of the uncertain parameters is high, forward difference approximation could be more practical since it requires d evaluations of V to compute the gradient (where d is the dimension of the uncertain parameters) while the central difference approximation requires $2d$ iterations.

Another problem could arise when huge or streaming data are involved, where computing the gradient in the presence of huge data is computationally expensive. Instead, a subset of the data can be used to compute a noisy gradient. This kind of gradient is called a stochastic gradient (Chen *et al*., 2014; Welling and Teh, 2011).

The idea behind the stochastic gradient is that instead of computing the gradient $\nabla V(\boldsymbol{\theta}(t))$ using the original data D, a subset of the data $\tilde{D}$, which is sampled randomly from the total data D, is used to compute a noisy estimation of the gradient $\nabla \tilde{V}(\boldsymbol{\theta}(t))$.The stochastic gradient is given by (Welling and Teh, 1997; Boulkaibet, 2014)

$$\nabla \tilde{V}(\boldsymbol{\theta}(t)) = -\frac{|\tilde{D}|}{|D|}\sum_{x\in\tilde{D}} \nabla P(x|D) - \nabla \log(P(\boldsymbol{\theta})). \tag{7.16}$$

It is assumed that the observations x are independent and, appealing to the central limit theorem, the noisy gradient is approximated as follows (Welling and Teh, 1997; Boulkaibet, 2014):

$$\nabla \tilde{V}(\boldsymbol{\theta}(t)) \cong \nabla V(\boldsymbol{\theta}(t)) + \mathcal{N}(0, \boldsymbol{V}(\boldsymbol{\theta}(t))), \tag{7.17}$$

where $V(\boldsymbol{\theta}(t))$ is the covariance matrix of the stochastic gradient noise and is modelled as multivariate Gaussian with zero mean and $\mathcal{N}$ represents a distribution. This covariance depends on the model uncertain parameters and the increase in the size of the subset data $\tilde{D}$ decreases the noise and makes the gradient more accurate. In this chapter, the data D used in all examples are limited and therefore the stochastic gradient approach will not be used.

The benefits of using Bayesian methods in the finite element model updating are not only to obtain accurate updated parameters, but also to represent the uncertainty of the obtained parameters and to measure the correlations between the uncertain parameters. In this chapter the mean estimated value $\hat{\boldsymbol{\theta}}$ and variance are used. In the next section, the HMC algorithm is applied for finite element model updating.

7.6 Application 1: Cantilever Beam

An experimental cantilever steel beam is updated based on the measurements of Kraaij (2006). The beam has the following dimensions: length 500 mm, width 60 mm, thickness 10 mm, $E = 2.1 \times 10^{11}\,\mathrm{N/m^2}$, $\upsilon = 0.3$ and $\rho = 7850\,\mathrm{kg/m^3}$.

Three accelerometers were used in the experiment, all positioned 490 mm from the clamped end. This position was selected because the response at this point is large (Kraaij, 2006). Each accelerometer has a mass of 40 g; the middle accelerometer is of type 303*A*3, the outer accelerometers are of type 303*A*2 (for more information on the experimental set-up, see Kraaij, 2006). To test the updating procedures, the beam was modelled using Version 6.3 of the Structural Dynamics Toolbox SDT® for MATLAB.

The beam was divided into 50 Euler–Bernoulli beam elements and excited at various positions (Boulkaibet, 2014; Kraaij, 2006). The measured natural frequencies of interest of this structure are 31.9, 197.9, 553, 1082.2 and 1781.5 Hz. The Young's modulus of the beam elements was used as an updating parameter, where for every 10 elements a different Young's

modulus is allocated. Thus, the parameters to be updated can be represented by a vector $\boldsymbol{\theta} = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$.

In this section, $\boldsymbol{\theta}$ is updated using the Bayesian approach trained using the HMC. The reason for using a large number of updating parameters is to determine the performance and the convergence speed of each sampling technique when a relatively large number of variables is introduced into the updating process.

N_S samples of $\boldsymbol{\theta}$ were generated from the posterior distribution function $P(\theta|D)$. The constant β_c was set equal to 1, and the coefficients α_i were set equal to $1/\sigma_i^2$ where σ_i^2 is the variance of the parameter θ_i which means that the parameters $\boldsymbol{\theta}$ are uncertain with a standard deviation σ. Since the updating parameter vector contains only the Young's modulus, all σ_i were set equal to 2×10^{11} and a large value of σ_i is used to give the algorithm more freedom and also to allow the algorithm to focus more on the likelihood term during the search. The updating parameters θ_i were bounded with a maximum value of 2.5×10^{11} and a minimum value of 1.7×10^{11}. The number of samples N_S was 1000 for all techniques and the initial $\boldsymbol{\theta}$ vector was

$$\boldsymbol{\theta}_0 = \left\{2.4 \times 10^{11}, 2.4 \times 10^{11}, 2.4 \times 10^{11}, 2.4 \times 10^{11}, 2.4 \times 10^{11}\right\}.$$

Instead of using the mean value of $\boldsymbol{\theta}$ for steel as an initial value, a large value of the initial parameter vector was selected to highlight the updating process. The initial time step used in the HMC algorithm was 1.25×10^{-3} seconds, and L was uniformly distributed in the interval $\{1,30\}$.

The HMC algorithm was implemented over 30 independent runs, and the results given in Table 7.1 are the average of these 30 runs as the randomness which exists in these three algorithms gives slightly different results. Figures 7.1 and 7.2 show the scatter plots with marginal histograms for four uncertain parameters using the HMC algorithm. The plots show that the HMC found the area of high probability.

The updated parameters are also presented in Table 7.2, and the results indicate that the HMC techniques update all vector parameters simultaneously and give results which are close to the mean value for steel of $2.5 \times 10^{11}\,\mathrm{N/m^2}$. The coefficient of variation (c.o.v.) represents the errors associated with the updated parameters and is estimated by dividing the standard deviation vector by the estimated $\boldsymbol{\theta}$ vector. Table 7.1 shows that the c.o.v. values are small and are less than 7% for the HMC. The obtained c.o.v. values indicate that the HMC algorithm estimated the uncertain parameters in an efficient way, and the time step used for the HMC was small, where the acceptance rate of this algorithm was equal to 99.9%. The c.o.v. values of the

Table 7.1 The updated vector of Young's modulus using HMC technique

	Initial	HMC method	$\frac{\sigma_i}{\mu_i}$ (%)
E_1	2.4×10^{11}	2.257×10^{11}	4.43
E_2	2.4×10^{11}	2.123×10^{11}	5.44
E_3	2.4×10^{11}	2.049×10^{11}	5.27
E_4	2.4×10^{11}	1.985×10^{11}	5.58
E_5	2.4×10^{11}	2.007×10^{11}	6.88

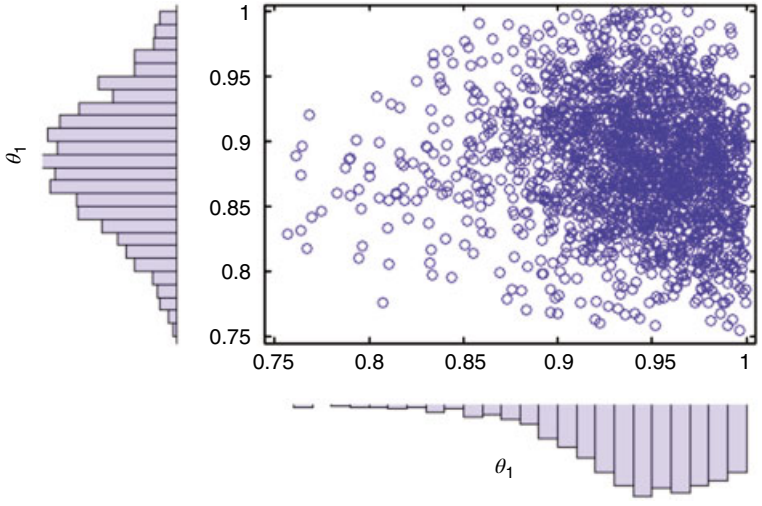

Figure 7.1 Scatter plots with marginal histograms using the HMC method. Scatter plot of θ_1 versus θ_2

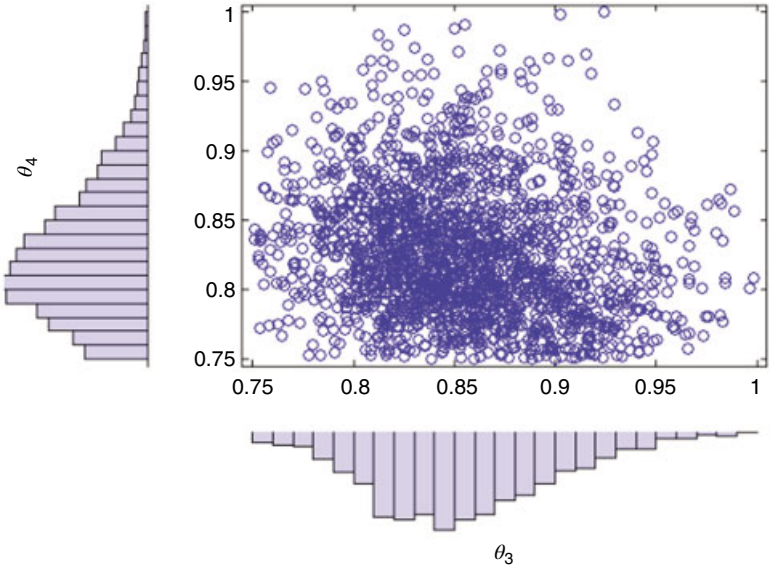

Figure 7.2 Scatter plots with marginal histograms using the HMC method. Scatter plot of θ_3 versus θ_4

HMC algorithm could be improved even more if a relatively large time step were used to increase the algorithm move step during the search process.

Figure 7.3 shows the correlation between all updated parameters for the HMC algorithms. Small values indicate that both parameters are weakly correlated (<0.3), large values (>0.7) indicate that the parameters are highly correlated, while a zero value indicates that the

Table 7.2 Results when the HMC technique is used to update Young's modulus

Modes	Measured frequency (Hz)	Initial frequency (Hz)	Error (%)	Frequencies HMC method (Hz)	Error (%)
1	31.9	32.7	2.51	31.2	2.15
2	197.9	209.4	5.83	195.7	1.14
3	553.0	594.8	7.55	553.4	0.07
4	1082.2	1177.8	8.84	1096.8	1.35
5	1781.50	1961.7	10.12	1828.0	2.61

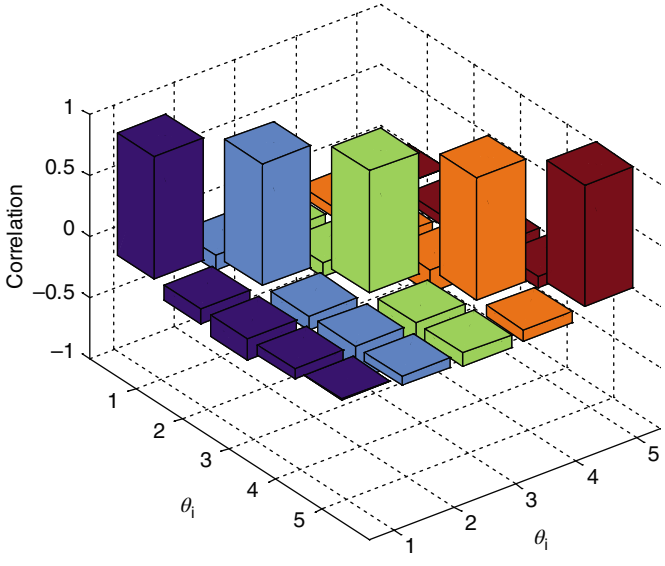

Figure 7.3 The correlation between the updated parameters (HMC algorithm)

parameters are not correlated. A positive correlation indicates that the variables are positively related, while a negative correlation indicates the opposite. Figure 7.3 indicates that for the HMC algorithm, all parameters are weakly correlated.

The results for the updated modes are given in Table 7.2 and show that the HMC algorithm gives results which are, on average, better than the initial finite element model. The HMC technique used additional parameters to evaluate the sampling and have the ability to suppress the random walk. The absolute mode errors and the final model percentage error for the HMC algorithm are also presented in Table 7.2 (Boulkaibet, 2014).

7.7 Application 2: Asymmetrical H-Shaped Structure

In this example, the asymmetrical H-shaped aluminium structure was updated using the HMC algorithm (Marwala, 1997). The structure was divided into 12 elements, each element modelled

as an Euler–Bernoulli beam. The beams were modelled using Version 6.3 of the Structural Dynamics Toolbox SDT® for MATLAB. The structure was excited and the acceleration was measured at 15 different positions. The structure was excited using an electromagnetic shaker while a roving accelerometer was used to measure the response (Marwala, 1997, 2010; Welling and Teh, 1997; Boulkaibet, 2014). A set of 15 frequency-response functions were calculated.

The measured natural frequencies are 53.9, 117.3, 208.4, 254.0 and 445.0 Hz. In this example, the moments of inertia and the cross sectional areas of the left, middle and right subsections of the beam were updated. The updating parameter vector is thus $\boldsymbol{\theta} = \{I_{x1}, I_{x2}, I_{x3}, A_{x1}, A_{x2}, A_{x3}\}$. The Young's modulus for the beam is set at $7.2 \times 10^{10} \mathrm{N/m^2}$, and the density is set to 2785 $\mathrm{kg/m^3}$. The updating parameters θ_i are bounded by maximum values equal to $\{3.73 \times 10^{-8}, 3.73 \times 10^{-8}, 3.73 \times 10^{-8}, 4.16 \times 10^{-4}, 4.16 \times 10^{-4}, 4.16 \times 10^{-4}\}$, and the minimum values are $\{1.73 \times 10^{-8}, 1.73 \times 10^{-8}, 1.73 \times 10^{-8}, 2.16 \times 10^{-4}, 2.16 \times 10^{-4}, 2.16 \times 10^{-4}\}$. The boundaries help to keep the updated vector physically realistic and the constant β_c of the posterior distribution was set equal 10, the coefficient α_i was set equal to $1/\sigma_i^2$, where σ_i^2 is the variance of the ith parameter, and the variance vector is defined as $\sigma = \{5 \times 10^{-8}, 5 \times 10^{-8}, 5 \times 10^{-8}, 5 \times 10^{-4}, 5 \times 10^{-4}, 5 \times 10^{-4}\}$. The constant β_c and the variance σ were chosen so that the weight of the likelihood terms was greater than the second term in the posterior PDF. The number of samples N_S was set to 1000, the initial time step was set to 0.0045 s for the HMC method, while L was uniformly distributed on the interval {1,30}.

The Bayesian simulation results are presented and evaluated in terms of the mean values of the samples obtained for each method. Also, the updated natural frequencies and the prediction percentage error are presented, where the percentage error is defined by the absolute differences between the updated value of the natural frequency and its experimental value divided by the experimental value (as in the previous section). Again, the HMC algorithm was implemented over 20 independent runs, and the results are tabulated in Tables 7.3 and 7.4.

Figures 7.4 and 7.5 show scatter plots with marginal histograms for four uncertain parameters using HMC algorithms.

The rest of the updated parameters are presented in Table 7.3. The acceptance rate for the HMC algorithm is 99.9%. Table 7.3 presents the initial and updated values of the updating vector obtained by HMC method along with the c.o.v. The HMC algorithm successfully updated the $\boldsymbol{\theta}$ vector, and the updated values are different than the initial $\boldsymbol{\theta}_0$. The updated vector obtained by the HMC algorithm is physically realistic.

Table 7.3 Initial and updated parameters using the HMC algorithm

Parameter	Initial $\boldsymbol{\theta}_0$	$\boldsymbol{\theta}$ vector HMC method	$\frac{\sigma_i}{\mu_i}$ (%)
I_{x1}	2.73×10^{-8}	2.21×10^{-8}	12.7
I_{x2}	2.73×10^{-8}	2.21×10^{-8}	1.37
I_{x3}	2.73×10^{-8}	2.90×10^{-8}	16.5
A_{x1}	3.16×10^{-4}	4.00×10^{-4}	1.39
A_{x2}	3.16×10^{-4}	2.30×10^{-4}	1.10
A_{x3}	3.16×10^{-4}	2.40×10^{-4}	1.95

Table 7.4 Results when the HMC algorithm was used to update the parameters

Mode	Measured frequency (Hz)	Initial frequency (Hz)	Frequency HMC method (Hz)
1	53.90	51.40	52.93
2	117.30	116.61	118.82
3	208.40	201.27	208.81
4	254.00	247.42	254.41
5	445.00	390.33	444.13

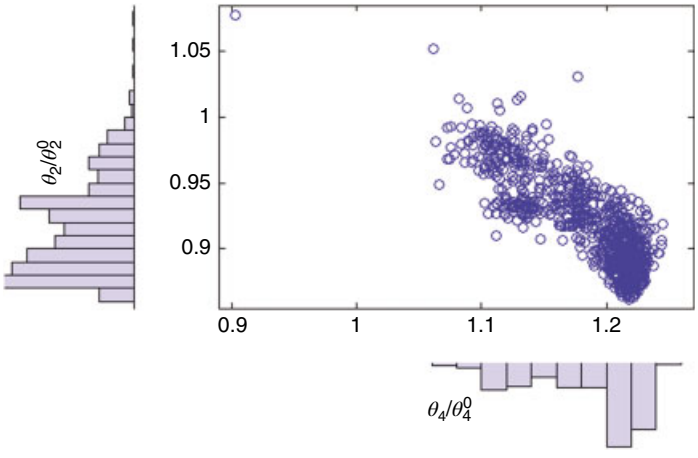

Figure 7.4 Scatter plots with marginal histograms using the HMC method. Scatter plot of θ_4 versus θ_2

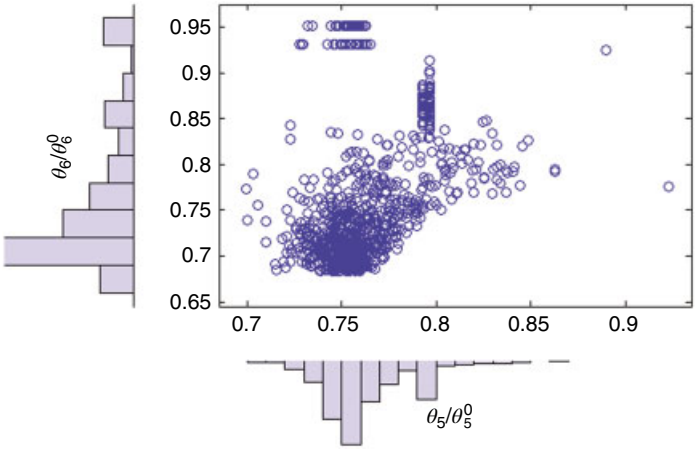

Figure 7.5 Scatter plots with marginal histograms using the HMC method. Scatter plot of θ_5 versus θ_6

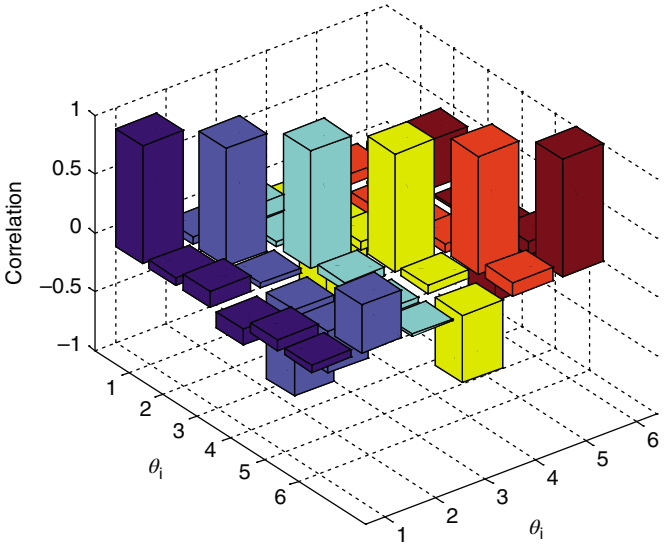

Figure 7.6 The correlation between the updated parameters (HMC algorithm)

Figure 7.6 shows the correlation between all updated parameters for the HMC algorithm. The figure indicates that all parameters are correlated for all algorithms (all values are non-zero), where most of the parameters are weakly correlated, except the pair (A_{x1}, A_{x3}) which is found to be highly correlated by the HMC algorithm.

Table 7.4 shows that the error between the first measured natural frequency and that of the initial model was 4.63% and when applying HMC method this error was reduced to 1.8%, whereas overall the HMC method improved the error from 4.7% to below than 0.8%.

7.8 Conclusion

This chapter has presented and described the HMC method along with its pseudo-code, and then certain improvements to the original algorithm were presented. The method was then implemented in finite element model updating of two structures and found to be satisfactory.

References

Akhmatskaya E, Bou-Rabee N, Reich S (2009) A comparison of generalized hybrid Monte Carlo methods with and without momentum flip. *Journal of Computational Physics* **228**: 2256–2265.

Aleksandrov T, Desgranges C, Delhommelle J (2010) Vapor–liquid equilibria of copper using hybrid Monte Carlo Wang–Landau simulations. *Fluid Phase Equilibria* **287**: 79–83.

Alfaki M (2008) Improving efficiency in parameter estimation using the Hamiltonian Monte Carlo algorithm. PhD thesis, University of Bergen.

Arnold VI (1989) *Mathematical Methods of Classical Mechanics* (2nd edition). London: Springer.

Beskos A, Pillai N, Roberts G, Sanz-Serna JM, Stuart A (2013) Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli* **19**: 1501–1534.

Bogaerts A (2009) Effects of oxygen addition to argon glow discharges: a hybrid Monte Carlo-fluid modeling investigation. *Spectrochimica Acta Part B: Atomic Spectroscopy* **64**: 1266–1279.

Boulkaibet, I (2014) Finite element model updating using Markov chain Monte Carlo techniques. PhD thesis, University of Johannesburg.

Boulkaibet I, Marwala T, Mthembu L, Friswell MI, Adhikari S (2012) Sampling techniques in Bayesian finite element model updating. *Proceedings of the Society for Experimental Mechanics* **29**: 75–83.

Boulkaibet I, Mthembu L, Marwala T, Friswell MI, Adhikari S (2014a) Finite element model updating using the shadow hybrid Monte Carlo technique. *Mechanical System and Signal Processing* **52–53**: 115–132.

Boulkaibet I, Mthembu L, Marwala T, Friswell MI,Adhikari S (2014b) Finite element model updating using the separable shadow hybrid Monte Carlo technique. *Topics in Modal Analysis II, Volume 8*. Cham: Springer, pp. 267–275.

Boulkaibet I, Mthembu L, Marwala T, Friswell MI, Adhikari S (2015) Finite element model updating using Hamiltonian Monte Carlo techniques. *Inverse Problems in Science and Engineering (IPSE)*, submitted for publication.

Chen L, Qin Z, Liu JS (2001) Exploring hybrid Monte Carlo in Bayesian computation. *Sigma* **2**: 2–5.

Chen T, Fox EB, and Guestrin, C. (2014) Stochastic gradient Hamiltonian Monte Carlo. Proceedings of the 31st International Conference on Machine Learning 32, arXiv:1402.4102.

Cheng WL, Liu N, Li Z, Zhong Q, Wang AM, Zhang ZM, He ZB (2011) Application study of a correction method for a spacecraft thermal model with a Monte-Carlo hybrid algorithm. *Chinese Science Bulletin* **56**: 1407–1412.

Cheung SH, Beck JL (2009) Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters. *Journal of Engineering Mechanics* **135**: 243–255.

Duane S, Kennedy AD, Pendleton BJ, Roweth D (1987) Hybrid Monte Carlo. *Physics Letters B* **195**: 216–222.

Ghoufi A, Maurin G (2010) Hybrid Monte Carlo simulations combined with a phase mixture model to predict the structural transitions of a porous metal-organic framework material upon adsorption of guest molecules. *Journal of Physical Chemistry C* **114**: 6496–6502.

Goldstein H (1980) *Classical Mechanics*. Reading, MA:Addison Wesley.

Gupta R, Kilcup GW, Sharpe SR (1988) Tuning the hybrid Monte Carlo algorithm. *Physical Review D* **38**: 1278.

Hanson KM (2001) Markov vhain Monte Carlo posterior sampling with the Hamiltonian method. *Proceedings of SPIE* **4322**: 456–467.

Heckman JJ, Leamer E (2001) *Handbook of Econometrics*, Vol. 5. Amsterdam: Elsevier.

Hoefling M, Lima N, Haenni D, Seidel CAM, Schuler B, Grubmuller H (2011) Structural heterogeneity and quantitative FRET efficiency distributions of polyprolines through a hybrid atomistic simulation and Monte Carlo approach. *PLoS ONE*, **6**, e19791.

Huang W, Leimkuhler B (1997) The adaptive Verlet method. *SIAM Journal on Scientific Computing* **18**: 239–256.

Izaguirre JA, Hampton SS (2004) Shadow hybrid Monte Carlo: an efficient propagator in phase space of macromolecules. *Journal of Computational Physics* **200**: 581–604.

Kraaij, C.S. (2006) Model updating of a 'clamped' free beam system using FEM tools. Technical Report, Technische Universiteit Eindhoven.

Kulak L (2009) Hybrid Monte-Carlo simulations of fluorescence anisotropy decay in three-component donor-mediator-acceptor systems in the presence of energy transfer. *Chemical Physics Letters* **467**: 435–438.

Marwala, T. (1997) *A multiple criterion updating method for damage detection on structures*. MEng thesis, University of Pretoria.

Marwala, T. (2010) *Finite Element Model Updating Using Computational Intelligence Techniques*. London: Springer-Verlag.

Marwala T (2012) *Condition Monitoring Using Computational Intelligence Methods*. London: Springer.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**: 1087–1092.

Neal RM (1996) *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics, Vol. **118**. New York: Springer-Verlag.

Neal RM (1998) Suppressing random walks in Markov Chain Monte Carlo using ordered overrelaxation. In: *Learning in Graphical Models* (ed. Jordan MI), pp. 205–228. Dordrecht: Kluwer Academic Publishers.

Neal RM (2011) MCMC using Hamiltonian dynamics. In: *Handbook of Markov Chain Monte Carlo* (ed. Brooks S, Gelman A, Jones GL, Meng X-L), pp. 113–162). Boca Raton, FL: Chapman & Hall/CRC.

Paquet E, Viktor HL (2011) Probability distributions from Riemannian geometry, generalized hybrid Monte Carlo sampling and path integrals. *Proceedings of SPIE*, **7864**, article no. 78640X.

Qian G, Li N, Huggins R (2011) Using capture–recapture data and hybrid Monte Carlo sampling to estimate an animal population affected by an environmental catastrophe. *Statistics and Data Analysis* **55**: 655–666.

Rei W, Gendreau M, Soriano P (2010) A hybrid Monte Carlo local branching algorithm for the single vehicle routing problem with stochastic demands. *Transportation Sciences* **44**: 136–146.

Rossberg K (1983) *A First Course in Analytical Mechanics*. New York: John Wiley & Sons, Inc.

Skeel RD, Tupper PF (2005) Mathematical issues in molecular dynamics. *Banff International Research Station Reports*.

Suzuki K, Tachikawa M, Shiga M (2010) Efficient ab initio path integral hybrid Monte Carlo based on the fourth-order Trotter expansion: application to fluoride ion-water cluster. *Journal of Chemical Physics* **132**: 144108.

Sweet CR, Hampton SS, Skeel RD, Izaguirre JA (2009) A separable shadow Hamiltonian hybrid Monte Carlo method. *Journal of Chemical Physics* **131**: 174106.

Welling M, Teh, YW (2011) Bayesian learning via stochastic gradient Langevin dynamics. Proceedings of the 28th International Conference on Machine Learning, pp. 681–688.

Wendt KA, Drut JE, Lähde TA (2011) Toward large-scale Hybrid Monte Carlo simulations of the Hubbard model on graphics processing units. *Computer Physics Communications* **182**: 1651–1656.

Zhang L, Bartel T, Lusk MT (2010) Parallelized hybrid Monte Carlo simulation of stress-induced texture evolution. *Computational Material Sciences* **48**: 419–425.

8

Shadow Hybrid Monte Carlo Technique for Finite Element Model Updating

8.1 Introduction

The application of the hybrid Monte Carlo (HMC) algorithm for finite element model updating has demonstrated pleasing results in terms of convergence speed and precision (Mares *et al*., 2000; Boulkaibet *et al.*, 2012; Boulkaibet, 2014). The HMC technique is a sampling technique which combines Markov chain Monte Carlo simulation with the gradient descent algorithm. It has been widely used in engineering, political science and economics problems (Marwala, 2010, 2012, 2013, 2014, 2015; Marwala and Lagazio, 2011). The results on finite element model updating obtained in Boulkaibet (2014) demonstrate that the HMC algorithm can offer improved results when compared to both the Metropolis–Hastings and slice sampling algorithms and without increasing the time step. Nevertheless, the results obtained from the implementation of the HMC technique can be improved even more by increasing the move step which can be achieved by increasing the time step. Regrettably, a large time step has a negative consequence for the sampling procedure because the resulting effect is that the acceptance rate tends to decrease exponentially as the time step increases.

In this chapter, a modified version of the HMC technique called the shadow hybrid Monte Carlo (SHMC) algorithm, which has a higher acceptance rate of samples and a better accuracy on estimating the probability distribution function, is applied for finite element model updating (Izaguirre and Hampton, 2004; Boulkaibet, 2014). This technique is basically a modified version of the HMC technique and is built to efficiently draw samples with relatively large systems and time steps.

This chapter covers the following aspects: Section 8.2 discusses the effect of the time step on the HMC algorithm. Section 8.3 examines the SHMC algorithm and Section 8.4 the shadow Hamiltonian function. Section 8.5 discusses the results obtained when an aircraft structure's

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

finite element model is updated using both the HMC and SHMC algorithms. Section 8.6 concludes the chapter.

8.2 Effect of Time Step in the Hybrid Monte Carlo Method

This section studies the effect of the choice of the time step on the performance of the HMC algorithm where the leapfrog integrator was applied to evaluate a Hamiltonian function. It was observed by Boulkaibet (2014) that a small time step conserves the Hamiltonian function, whereas a large time step oscillates the Hamiltonian function and this decreases the acceptance rate of the HMC algorithm. A straightforward technique to identify an appropriate time step that can maintain the high acceptance rate and thereby ensure relatively large moves in the search space is by running the HMC algorithm with varying time steps for a limited number of iterations and then selecting a time step that offers a high acceptance rate with acceptable moves in the search space. In this chapter, the Hamiltonian function is modified by ensuring that an accurate approximation of the Hamiltonian function is applied instead of the original function and this ensures a high acceptance rate when the time step is large because a large time step ensures large moves in the search space.

8.3 The Shadow Hybrid Monte Carlo Method

The SHMC technique is a generalisation of the HMC where the central notion is to apply a modified Hamiltonian estimate $\widetilde{H}(\boldsymbol{\theta},\boldsymbol{p})$ to sample from the extended phase space of the shadow Hamiltonian rather than sampling from the configuration space alone (Boulkaibet, 2014; Escribano *et al.*, 2014).

Fernández-Pendás *et al.* (2014) successfully applied a generalised SHMC for simulation of molecular dynamics (MD). Escribano *et al*. (2013) successfully applied a generalised SHMC method to simulate complex biological processes. Clark *et al.* (2011) proposed an integrator that not only approximately conserves Hamiltonian functions but also accurately conserves the nearby shadow Hamiltonian, whereas Sweet *et al.* (2009) proposed a separable shadow Hamiltonian HMC method that efficiently generates momenta. The SHMC method solves the problem of the decrease in the acceptance rate of the HMC technique when the system size d and/or time step δt is relatively large because the SHMC technique expands the HMC time step. The formulation of the SHMC method requires the introduction of a constant parameter c as well as the function $\widetilde{\rho}(\boldsymbol{\theta},\boldsymbol{p})$ which is a target density function for the SHMC where (Boulkaibet, 2014)

$$\widetilde{\rho}(\boldsymbol{\theta},\boldsymbol{p}) \propto \exp\left(-\beta_B \widetilde{H}(\boldsymbol{\theta},\boldsymbol{p})\right). \tag{8.1}$$

Here,

$$\widetilde{H}(\boldsymbol{\theta},\boldsymbol{p}) = \max\left(H(\boldsymbol{\theta},\boldsymbol{p}), H_{[2k]}(\boldsymbol{\theta},\boldsymbol{p}) - c\right). \tag{8.2}$$

$\widetilde{H}(\boldsymbol{\theta},\boldsymbol{p})$ is an accurate estimate of the shadow Hamiltonian defined in Section 8.4, and the constant c permits $H_{[2k]}(\boldsymbol{\theta},\boldsymbol{p})$ to differ from $H(\boldsymbol{\theta},\boldsymbol{p})$ because $H_{[2k]}(\boldsymbol{\theta},\boldsymbol{p})$ can have a significant separation from the original Hamiltonian $H(\boldsymbol{\theta},\boldsymbol{p})$.

On implementing the SHMC algorithm, a new momentum vector $\boldsymbol{p}$ is generated from a Gaussian $N(0, \boldsymbol{M})$ probability distribution function. Nevertheless, this momentum vector is accepted or rejected using the Metropolis algorithm which is a non-separable Hamiltonian function. The complexity of drawing new momenta from the non-separable Hamiltonian function is solved by using the von Neumann technique (von Neumann, 1951; Fishman, 2000).

This technique is implemented by allowing for a complicated target distribution $f(z)$ from which the samples are sampled. The Metropolis algorithm generates a random number with a probability distribution function $f(z)$ and then the probability distribution function is divided as follows: $f(z) = Cg(z)h(z)$ where $h(z)$ is a simple probability distribution function, C is a constant and $0 \le g(z) \le 1$. Thereafter a random variable Z with a probability distribution function of $h(z)$ is generated and then a uniform random number U from (0, 1) is generated. As a final point, if $U \le g(z)$, then Z has the probability distribution function of $f(z)$ or else the process is repeated and, in the case of $f(z) = \widetilde{\rho}(\boldsymbol{\theta}, \boldsymbol{p})$, then (Boulkaibet, 2014)

$$\begin{aligned}\widetilde{\rho}(\boldsymbol{\theta},\boldsymbol{p}) &= \exp\left(-\beta_B \max\left(H(\boldsymbol{\theta},\boldsymbol{p}), H_{[2k]}(\boldsymbol{\theta},\boldsymbol{p}) - c\right)\right) \\ &= \exp(-\beta_B H(\boldsymbol{\theta},\boldsymbol{p}))\min\left(1, \exp\left\{-\beta_B\left(H_{[2k]}(\boldsymbol{\theta},\boldsymbol{p}) - c - H(\boldsymbol{\theta},\mathbf{p})\right)\right\}\right).\end{aligned} \tag{8.3}$$

Then Equation 8.3 can be written as follows (Boulkaibet, 2014):

$$\widetilde{\rho}(\boldsymbol{\theta},\boldsymbol{p}) = \exp(-\beta_B V(\boldsymbol{\theta}))\exp(-\beta_B W(\boldsymbol{p}))\min\left(1, \exp\left\{-\beta_B\left(H_{[2k]}(\boldsymbol{\theta},\boldsymbol{p}) - c - H(\boldsymbol{\theta},\mathbf{p})\right)\right\}\right). \tag{8.4}$$

In the case where $C = \exp(-\beta_B V(\boldsymbol{\theta}))$, $h(\boldsymbol{p}) = \exp(-\beta_B W(\boldsymbol{p}))$ and $g(p) = \min(1, \exp\{-\beta_B \left(H_{[2k]}(\boldsymbol{\theta},\boldsymbol{p}) - c - H(\boldsymbol{\theta},\mathbf{p})\right)\})$, the vector $\boldsymbol{p}$ is generated from the Gaussian distribution $h(\boldsymbol{p})$ and then the sample is then accepted or rejected using the criterion

$$\min\left(1, \exp\left\{-\beta_B\left(H_{[2k]}(\boldsymbol{\theta},\boldsymbol{p}) - c - H(\boldsymbol{\theta},\mathbf{p})\right)\right\}\right). \tag{8.5}$$

The acceptance criterion is repeated until a new momentum vector is accepted. Selecting the correct parameter c can improve the efficiency of this technique by decreasing the effort required to produce the new momentum vector. The system is then assembled using an MD procedure. The SHMC algorithm can be summarised as follows (Boulkaibet *et al.*, 2014; Izaguirre and Hampton, 2004).

1. Set initial value $\boldsymbol{\theta}_0$.
2. Repeat for N_s samples.Monte Carlo (MC) step (Izaguirre and Hampton, 2004; Nakano, 2015; Zhang and You, 2015):
   a. Produce $\boldsymbol{p}$ such that $\boldsymbol{p} \sim \mathcal{N}(0, \boldsymbol{M})$.
   b. Accept with probability $\min\left(1, \exp\left\{-\beta_B\left(H_{[2k]}(\boldsymbol{\theta},\boldsymbol{p}) - c - H(\boldsymbol{\theta},\mathbf{p})\right)\right\}\right)$.
   c. Repeat until a new $\boldsymbol{p}$ is accepted

   MD step (Izaguirre and Hampton, 2004; Xie *et al.*, 2015; Abe and Tasaki, 2015):

   a. Initialise the extended leapfrog algorithm with $(\boldsymbol{\theta}, \boldsymbol{p})$ and run the algorithm for L time steps to obtain $(\boldsymbol{\theta}^*, \boldsymbol{p}^*)$.

b. Update the the finite element model to obtain the new analytical frequencies and then compute $\widetilde{H}(\boldsymbol{\theta}^*,\boldsymbol{p}^*)$.
c. Accept $(\boldsymbol{\theta}^*,\boldsymbol{p}^*)$ with probability $\min\left(1,\exp\left\{-\beta_B\Delta\widetilde{H}\right\}\right)$.

The MC and MD steps depend on the parameter c, and this parameter has a significant effect on the simulation. When c is positive and large, the SHMC algorithm is equivalent to the HMC algorithm with different momentum, and this decreases the MD step acceptance rate in the case where δt and/or the system size are large; otherwise the MC step acceptance rate increases. In contrast, a large negative c value increases the acceptance rate of the MD step and decreases the acceptance rate of the MC step for the case where δt and/or the system size is large. In this chapter, the algorithm is modified so that the value of c is chosen to be proportional to the average difference between the Hamiltonian and the shadow Hamiltonian, which can be done off-line as follows (Boulkaibet, 2014):

1. Implement the SHMC between 50 and 100 iterations and save $\Delta H = H_{[2k]}(\boldsymbol{\theta},\boldsymbol{p}) - H(\boldsymbol{\theta},\mathbf{p})$ in a vector for all iterations.
2. Calculate the expected value $\overline{\Delta H}$ and standard deviation $\sigma_{\Delta H}$ for the obtained vector.
3. Select c using the following formula: $c = \overline{\Delta H} - 1.2 \times \sigma_{\Delta H}^2$.

As a final point, in order to compute the balanced values of the mean, the results are reweighted using the formula $\rho(\boldsymbol{\theta},\boldsymbol{p})/\widetilde{\rho}(\boldsymbol{\theta},\boldsymbol{p})$ before estimating the averages. The average of an observable B is given by (Izaguirre and Hampton, 2004; Boulkaibet, 2014):

$$\langle B\rangle = \frac{\sum_{i=1}^{N_s} Ba_i}{\sum_{i=1}^{N_s} a_i}, \quad \text{where } a_i = \frac{\exp(-\beta_B H(\boldsymbol{\theta},\boldsymbol{p}))}{\exp\left(-\beta_B \widetilde{H}(\boldsymbol{\theta},\boldsymbol{p})\right)}. \tag{8.6}$$

If the weighted vector parameter is given by $\widetilde{\boldsymbol{\theta}}$, the mean value of the estimated parameter is given by (Izaguirre and Hampton, 2004; Boulkaibet, 2014):

$$\hat{\widetilde{\boldsymbol{\theta}}} = E\left(\widetilde{\boldsymbol{\theta}}\right) \cong \frac{1}{N_s}\sum_{i=1}^{N_s} \widetilde{\boldsymbol{\theta}}^i, \tag{8.7}$$

where $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta} a_i / \sum_{i=1}^{N_s} a_i$ and the mean value can be calculated as follows (Izaguirre and Hampton, 2004; Boulkaibet, 2014):

$$\hat{\widetilde{\boldsymbol{\theta}}} \cong \frac{1}{N_s}\sum_{i=1}^{N_s} \frac{a_i}{\sum_{j=1}^{N_s} a_j}\widetilde{\boldsymbol{\theta}}^i = \frac{1}{\sum_{j=1}^{N_s} a_j}\frac{1}{N_s}\sum_{i=1}^{N_s} a_i\boldsymbol{\theta}^i = \frac{1}{\sum_{j=1}^{N_s} a_j} E\left(\boldsymbol{\theta}\boldsymbol{a}^T\right). \tag{8.8}$$

By tracing the same thinking, the variance of the weighted estimated parameter is (Sweet *et al.*, 2006; Izaguirre and Hampton, 2004; Boulkaibet, 2014):

$$V\left(\tilde{\boldsymbol{\theta}}\right) = E\left(\left(\tilde{\boldsymbol{\theta}} - \hat{\tilde{\boldsymbol{\theta}}}\right)^2\right) \cong V(\boldsymbol{\theta}) \frac{\sum_{i=1}^{N_s} a_i^2}{\sum_{j=1}^{N_s} a_j^2}, \tag{8.9}$$

where $V(\boldsymbol{\theta})$ is the unweighted variance and the standard deviation (the error) is given by $\boldsymbol{\sigma}_{\tilde{\theta}} = \sqrt{V\left(\tilde{\boldsymbol{\theta}}\right)}$.

8.4 The Shadow Hamiltonian

On studying the Hamiltonian systems, the consequence of the discretisation error can be understood by investigating the 'modified equations' of this system (Skeel and Hardy, 2001; Engle *et al*., 2005). These equations are exactly reproducible by using the approximate discrete solution and can be represented by an asymptotic expansion in powers of the discretisation parameter (Creutz, 1988; Kennedy and Pendleton, 1991; Hairer *et al*., 2006). The modified integrator is Hamiltonian if and only if $\partial_y \boldsymbol{\varphi}(\boldsymbol{y})^{\boldsymbol{T}} \boldsymbol{J} \partial_y \boldsymbol{\varphi}(\boldsymbol{y}) \boldsymbol{J}$, where $\boldsymbol{y} = \boldsymbol{\varphi}(\boldsymbol{y})$ is a numerical integrator, $\boldsymbol{J} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I} \\ -\boldsymbol{I} & \boldsymbol{0} \end{bmatrix}$ and $\boldsymbol{I}$ is an identity matrix (Skeel and Hardy, 2001), and this property is known as symplectic. The numerical solution of a symplectic integrator remains approximately equal to the solution of a modified Hamiltonian $H^{\delta t}(\boldsymbol{\theta}, \boldsymbol{p})$ for an extended amount of time (Skeel and Hardy, 2001).

The leapfrog integrator is symplectic and therefore its modified differential equation is Hamiltonian, and this modified Hamiltonian of this integrator is expressed as follows (Skeel and Hardy, 2001; Sweet *et al*., 2006; Boulkaibet, 2014):

$$\begin{aligned} H^{\delta t} = H &+ \delta t^2 \left(\frac{1}{12}\{W,\{W,V\}\} - \frac{1}{24}\{V,\{V,W\}\} \right) \\ &+ \delta t^4 \left(\frac{7}{5760}\{V,\{V,\{V,\{V,W\}\}\}\} - \frac{1}{720}\{W,\{W,\{W,\{W,V\}\}\}\} \right. \\ &+ \frac{1}{360}\{V,\{W,\{W,\{W,V\}\}\}\} + \frac{1}{360}\{W,\{V,\{V,\{V,W\}\}\}\} \\ &\left. - \frac{1}{480}\{V,\{V,\{W,\{W,V\}\}\}\} + \frac{1}{120}\{W,\{W,\{V,\{V,W\}\}\}\} \right) + \ldots \end{aligned} \tag{8.10}$$

Here, the expression $\{A,B\} = \nabla_\theta A \nabla_p B - \nabla_p A \nabla_\theta B$ denotes the Poisson bracket of two functions which depends on $\boldsymbol{\theta}$ and $\boldsymbol{p}$. This formula is derived from the symmetric Baker–Campbell–Hausdorff formula and demonstrates how to calculate a modified Hamiltonian using a splitting technique, and the objective is to calculate (Hausdorff, 1906; Hairer *et al*., 2006; Skeel and Hardy, 2001; Sweet *et al*.; Boulkaibet, 2014)

$$H_{[2k]}(\boldsymbol{\theta}, \boldsymbol{p}) = H^{\delta t}(\boldsymbol{\theta}, \boldsymbol{p}) + O\left(\delta t^{2k}\right), \tag{8.11}$$

where $H_{[2k]}(\boldsymbol{\theta},\boldsymbol{p})$ is a shadow Hamiltonian of order $2k$, and this structure augments a new position variable and a conjugate momentum variable $\beta(t)$ to attain a modified Hamiltonian $\bar{H}(\boldsymbol{y}) = \frac{1}{2}\dot{\boldsymbol{y}}^T\boldsymbol{J}\boldsymbol{y}^T$, with $\boldsymbol{y} = \left[\boldsymbol{\theta}^T, \alpha, \boldsymbol{p}^T, \beta\right]^T$ and $\alpha = 1$. This modified Hamiltonian is homogeneous of order 2, and $\boldsymbol{y}$ is created using a numerical solution of the modified Hamiltonian system, the resulting solution satisfying Equation 8.11 (Skeel and Hardy, 2001). The expressions for the fourth and eighth shadow Hamiltonians, $k = 2$ and $k = 4$, are respectively given by (Sweet *et al.*, 2006; Boulkaibet, 2014)

$$H_{[4]}(\boldsymbol{\theta},\boldsymbol{p}) = \boldsymbol{A}_{10} - \frac{1}{6}\boldsymbol{A}_{12}, \tag{8.12}$$

$$H_{[8]}(\boldsymbol{\theta},\boldsymbol{p}) = \boldsymbol{A}_{10} - \frac{2}{7}\boldsymbol{A}_{12} - \frac{19}{210}\boldsymbol{A}_{14} + \frac{5}{42}\boldsymbol{A}_{30} + \frac{13}{105}\boldsymbol{A}_{32} - \frac{1}{140}\boldsymbol{A}_{34}, \tag{8.13}$$

where the $\mathbf{A}_{ij}$ are defined as (Sweet *et al.*, 2006; Boulkaibet, 2014)

$$\boldsymbol{A}_{ij} = \begin{cases} \mu\delta^i\boldsymbol{\theta}\delta^j\boldsymbol{p} - \delta^i\boldsymbol{\theta}\mu\delta^i\boldsymbol{p} - \mu\delta^i\beta, & j = 0, \\ \mu\delta^i\boldsymbol{\theta}\delta^j\boldsymbol{p} - \delta^i\boldsymbol{\theta}\mu\delta^i\boldsymbol{p}, & j \neq 0, \end{cases} \tag{8.14}$$

in which $\delta\boldsymbol{\theta}$ denotes the central difference of vector $\boldsymbol{\theta}$, $\boldsymbol{\theta} = \boldsymbol{\theta}^{1/2} - \boldsymbol{\theta}^{-1/2}$, and the averaging operator $\mu\boldsymbol{\theta}$ is likewise given by $\mu\boldsymbol{\theta} = \frac{1}{2}\left(\boldsymbol{\theta}^{1/2} + \boldsymbol{\theta}^{-1/2}\right)$, and it should be noted that $\boldsymbol{\theta}^{1/2} = \boldsymbol{\theta}(t + \delta t/2)$ and $\boldsymbol{\theta}^{-1/2} = \boldsymbol{\theta}(t - \delta t/2)$. To assess the leapfrog algorithm, Equations 8.7, 8.9 and 8.10 are implemented, and the term $\beta(t+1)$ is computed from $\beta(t)$ as follows (Sweet *et al.*, 2006; Boulkaibet, 2014):

$$\beta(t+1) = \beta(t) + \delta t\boldsymbol{\theta}(t)\nabla V(\boldsymbol{\theta}(t)) - 2V(\boldsymbol{\theta}(t)). \tag{8.15}$$

In the next section the SHMC technique is implemented for updating the finite element model of the GARTEUR SM-AG19 (or an aircraft) structure and the results are compared to those obtained by using an HMC technique.

8.5 Application: GARTEUR SM-AG19 Structure

The GARTEUR SM-AG19 structure, described in Appendix A, was used to test the finite element model updating procedure based on the SHMC method in a benchmark study by 12 members of the GARTEUR Structures and Materials Action Group 19 (Degener and Hermes, 1996; Balmes, 1998; Datta, 2002; Guyon and Elisseeff, 2003; Link and Friswell, 2003). The aeroplane has length 1.5 m, width 3 m, fuselage depth 15 cm and thickness 5 cm. It is made from aluminium with an overall mass of 44 kg. In order to increase the damping, a $1.1 \times 76.2 \times 1700$ m^3 visco-elastic constraining layer was bonded to the wings. Further details on this structure can also be found in Degener and Hermes (1996), Datta (2002), Guyon and Elisseeff (2003) and Link and Friswell (2003). In this study, all element materials are considered to be standard isotropic. The beam elements of the model were modelled as

Euler–Bernoulli elements. The experimental data used in this work was obtained from DLR Göttingen, Germany (Link and Friswell, 2003). The measured natural frequency data are 6.38, 16.10, 33.13, 33.53, 35.65, 48.38, 49.43, 55.08, 63.04 and 66.52 Hz. The updated parameters were the right wing moments of inertia and torsional stiffness ($RI_{\min}$, $RI_{\max}$, RI_{tors}), the left wing moments of inertia and torsional stiffness ($LI_{\min}$, $LI_{\max}$, LI_{tors}), the vertical tail moment of inertia ($VTP_{I_{\min}}$) and the overall structure's density ρ. The temperature was supposed to be $T = 300$ K, and $\rho_B = 1/(300K_B)$ where $K_B = 0.001\ 987\ 19$ kcal mol^{-1} K^{-1} and the updated vector is thus $\boldsymbol{\theta} = [\rho, VTP_{I_{\min}}, LI_{\min}, LI_{\max}, LI_{\min}, RI_{\max}, LI_{\text{tors}}, RI_{\text{tors}}]$.

The Young's modulus for the structure was set to $7.2 \times 10^{10}\text{N/m}^2$, the constant β_c of the posterior distribution was set to 100, all α_i coefficients were set equal to $1/\sigma_i^2$ where σ_i^2 was the variance of the ith parameter, and $\boldsymbol{\sigma} = [5 \times 10^2, 5 \times 10^{-9}, 5 \times 10^{-9}, 5 \times 10^{-7} 5 \times 10^{-9}, 5 \times 10^{-7} 5 \times 10^{-8}, 5 \times 10^{-8}]$.

The mean values of the updated parameters and their bounds are shown in Tables 8.1 and 8.2, respectively. The time step was set to $\delta t = 3$ ms and L was assumed to be uniformly distributed on the interval (1, 2), and the number of samples was set to $N_s = 1000$. The constant $c = 0.01$ was found to offer good results for the SHMC algorithms ($c = 0.009\ 789\ 12$ obtained from the off-line procedure). Each algorithm was run over 10 independent simulations and the results for the average of these 10 runs are shown in Tables 8.3 and 8.4 (Boulkaibet, 2014).

Several studies have been conducted to update the GARTEUR SM-AG19 structure. Mares *et al*. (2000) updated a finite element model to give an error of 0.66% using the sensitivity technique. Link and Friswell (2003) described the results attained by seven participants where each participant updated a different set of parameters using different updating approaches. The average errors ranged from 0.69% to 2.03% for all the results. An adjustment was made to the wing

Table 8.1 The initial values of the updating parameters for the GARTEUR example

Parameter	P (kg/m^3)	$VTP_{I_{\min}}\left(10^{-9}\ \text{m}^4\right)$	$LI_{\min}\left(10^{-9}\text{m}^4\right)$	$LI_{\max}\left(10^{-7}\text{m}^4\right)$
	2785.00	8.34	8.34	8.34
Parameter	$LI_{\text{tors}}\left(10^{-8}\ \text{m}^4\right)$	$RI_{\min}\left(10^{-9}\ \text{m}^4\right)$	$\text{RI}_{\max}\left(10^{-7}\ \text{m}^4\right)$	$RI_{\text{tors}}\left(10^{-8}\ \text{m}^4\right)$
	4.00	8.34	8.34	4.00

Table 8.2 The bounds of the updating parameters for the GARTEUR example

	Max	Min
ρ	3500	2500
$VTP_{I_{\min}}$	12×10^{-9}	5×10^{-9}
$LI_{\min}$	12×10^{-9}	5×10^{-9}
$LI_{\max}$	12×10^{-7}	5×10^{-7}
$RI_{\min}$	12×10^{-9}	5×10^{-9}
$RI_{\max}$	12×10^{-7}	5×10^{-7}
LI_{tors}	6×10^{-8}	3×10^{-8}
RI_{tors}	6×10^{-8}	3×10^{-8}

Table 8.3 Initial and updated parameter values for HMC and SHMC[4,8] algorithms at $\delta t = 3$ ms

	Initial (mean vector) θ_0	HMC method $\delta t = 3$ ms θ	SHMC[4] method $\delta t = 3$ ms θ	SHMC[8] method $\delta t = 3$ ms θ
P	2785.00	2667.33	2666.85	2686.97
$VTP_{I_{\min}}$	8.34×10^{-9}	6.94×10^{-9}	6.96×10^{-9}	7.15×10^{-9}
$LI_{\min}$	8.34×10^{-9}	10.12×10^{-9}	10.12×10^{-9}	10.13×10^{-9}
$LI_{\max}$	8.34×10^{-7}	7.90×10^{-7}	7.92×10^{-7}	8.02×10^{-7}
$RI_{\min}$	8.34×10^{-9}	10.15×10^{-9}	10.13×10^{-9}	10.12×10^{-9}
$RI_{\max}$	8.34×10^{-7}	6.11×10^{-9}	6.10×10^{-9}	6.11×10^{-9}
LI_{tors}	4.00×10^{-8}	4.04×10^{-8}	4.04×10^{-8}	4.02×10^{-8}
RI_{tors}	4.00×10^{-8}	3.57×10^{-8}	3.56×10^{-8}	3.57×10^{-8}

Table 8.4 Initial and updated parameter values for HMC and SHMC[4,8] algorithms at the $\delta t = 3$ ms

	Initial mean vector $\boldsymbol{\theta}_0$	HMC c.o.v. (%)	SHMC[4] c.o.v. (%)	SHMC[8] c.o.v. (%)
P	2785.00	1.97	2.27	2.98
$VTP_{I_{\min}}$	8.34×10^{-9}	5.62	5.50	6.92
$LI_{\min}$	8.34×10^{-9}	2.34	2.27	3.38
$LI_{\max}$	8.34×10^{-7}	2.61	2.75	3.34
$RI_{\min}$	8.34×10^{-9}	2.13	2.21	3.38
$RI_{\max}$	8.34×10^{-7}	3.14	4.02	4.21
LI_{tors}	4.00×10^{-8}	1.95	2.09	2.61
RI_{tors}	4.00×10^{-8}	2.17	2.49	3.18

of the original structure, and this time the average error between the participants ranged from 1.02% to 1.50%. In this chapter the HMC and SHMC techniques were used to update the finite element models.

Table 8.3 shows the initial (mean material or geometric) values of the vector $\boldsymbol{\theta}$ and the corresponding updated values achieved using the HMC and the SHMC[4,8] approaches for the time step of $\delta t = 3$ ms. Table 8.4 shows the coefficient of variation (c.o.v.) achieved using the HMC and SHMC[4,8] methods. There is a significant difference between the final updated values obtained by the HMC and the two versions of the SHMC because of the manner in which the SHMC algorithms apply the non-separable shadow Hamiltonian function for sampling. Additionally, the SHMC algorithms apply two extra parameters when calculating the MC and MD steps and there are the constants c and $\beta(t)$.

It can be observed that the c.o.v. is small for all techniques – less than 6% for both HMC and SHMC[4] and less than 7% for SHMC[8]. The time step applied gives a good acceptance sampling rate of 99.9% for all algorithms. The supplementary parameter c of the SHMC algorithm was used to control the difference between the modified and the true Hamiltonian function. In this instance, c is acquired from the offline algorithm where the offline technique gave $c = 0.009\ 789\ 12$. The value of c obtained ensures that the modified Hamiltonian function can differ from the true Hamiltonian function; however, different choices of c give different errors.

Figures 8.1–8.3 present Gaussian probability plots for the first updated parameter $\theta_1 = \rho$ using the three algorithms (SHMC[4], SHMC[8] and HMC; Boulkaibet, 2014). These graphs are an effective way to verify whether the updated parameters follow a Gaussian distribution, which is difficult to establish from histogram plots. Similar plots for other updating parameters can be created.

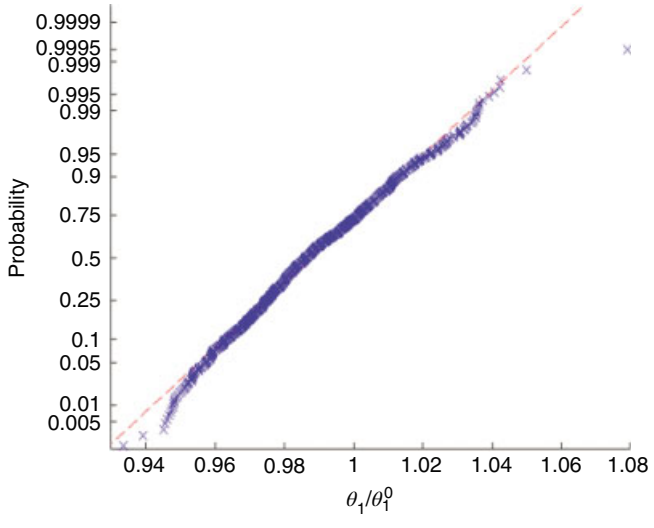

Figure 8.1 Normal probability plot for ρ (HMC)

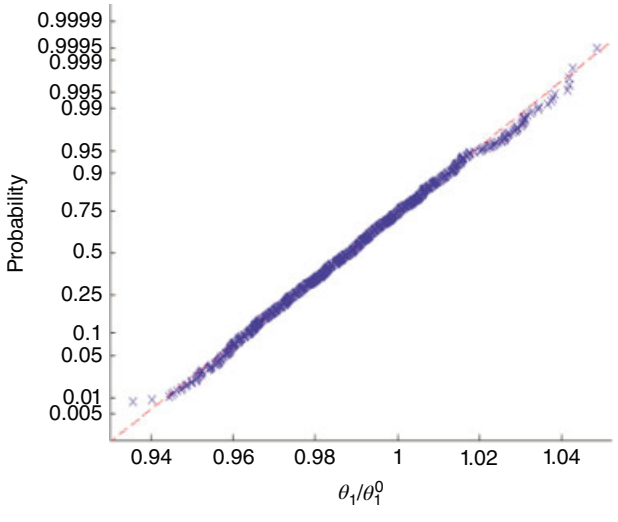

Figure 8.2 Normal probability plot for ρ (SHMC[4])

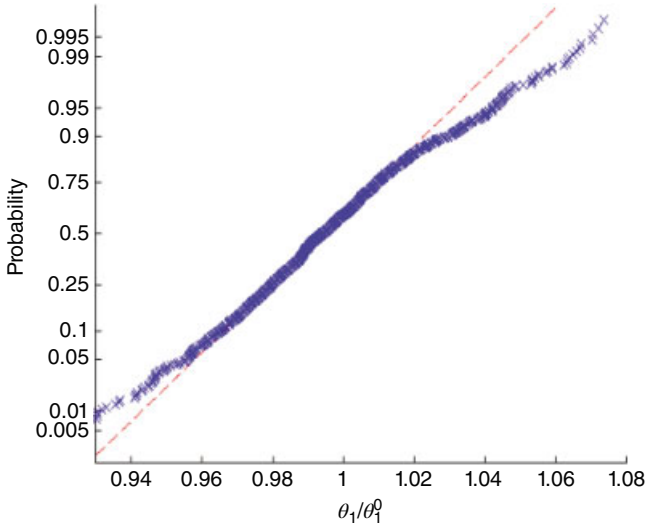

Figure 8.3 Normal probability plot for ρ (SHMC[8])

The distributions of the density ρ obtained from both HMC and SHMC[4] methods have a form close to the Gaussian. However, the distribution of ρ exhibits some non-Gaussian behaviour in the tails. The distribution of ρ obtained from the SHMC[8] method is relatively far from having a Gaussian form since a strong non-Gaussian behaviour in the tails was obtained. The same remarks can be made about the $LI_{\max}$ distribution obtained by the SHMC[4] algorithm. The HMC algorithm produces distributions close to the Gaussian for both LI_{tors} and RI_{tors}. In general, the distributions obtained by the three algorithms showed non-Gaussian behaviours.

Figures 8.4–8.6 provide the two-dimensional histogram plots of ρ versus $LI_{\max}$ for the HMC and SHMC[4,8] algorithms. The graphs exhibit the region of most probable values of the updated parameters where HMC and SHMC[4,8] algorithms were able to identify the region of high probability (shown in red). Additionally, the HMC and SHMC[4,8] give different histograms because of the reweighting step and the degree of the shadow Hamiltonian function in the SHMC method.

Figures 8.7 and 8.8 demonstrate the correlation between all updated parameters for both HMC and SHMC[8] algorithms; the correlation for the SHMC[4] algorithm is not displayed because it is almost identical to that for the SHMC[4]. For both HMC and SHMC[8], all parameters are correlated. Table 8.5 provides the modal results for the various sampling algorithms and the results obtained suggest that the updated finite element model natural frequencies are improvements on the initial finite element model in all cases. The error between the second measured natural frequency and that of the initial model was 5.01%; using the HMC method for finite element model updating, this error is reduced to 1.45%, and using the SHMC[4] and SHMC[8] method, it was reduced to 1.40% and 0.98%, respectively. A comparable observation was made on the fourth, sixth, seventh, eighth and ninth natural frequencies. Both SHMC[4] and SHMC[8] algorithms give a smaller final total average error when compared to that from the HMC algorithm for the same initial time step. The initial total average error was 4.6%; applying the HMC,

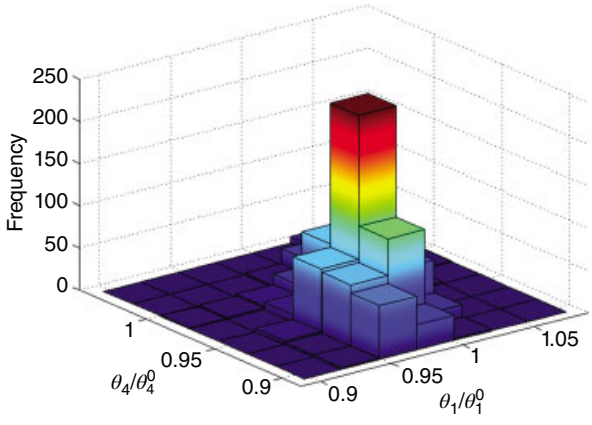

Figure 8.4 A two-dimensional histogram of ρ versus $LI_{\max}$ (HMC)

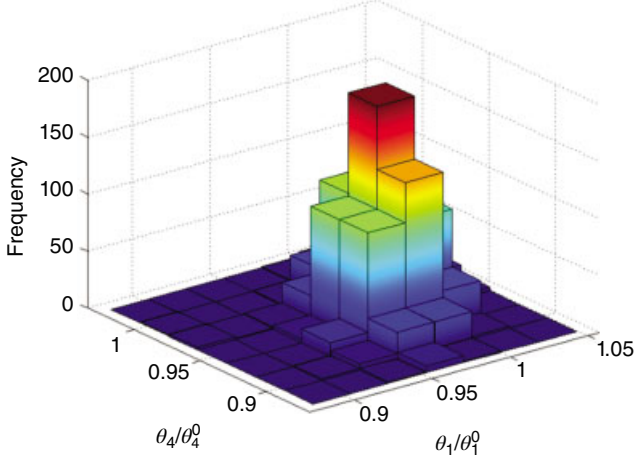

Figure 8.5 A two-dimensional histogram of ρ versus $LI_{\max}$ (SHMC^4)

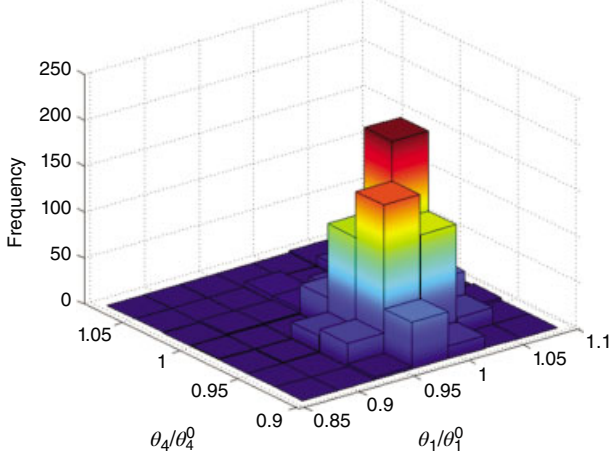

Figure 8.6 A two-dimensional histogram of ρ versus $LI_{\max}$ (SHMC^8)

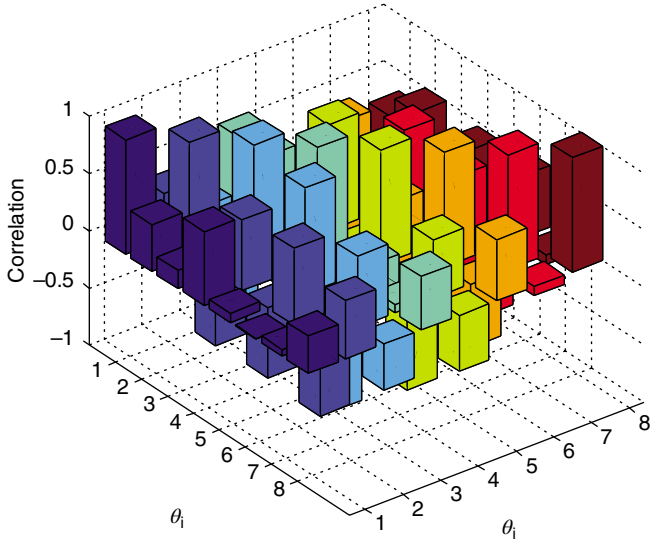

Figure 8.7 The correlation between the updated parameters (HMC)

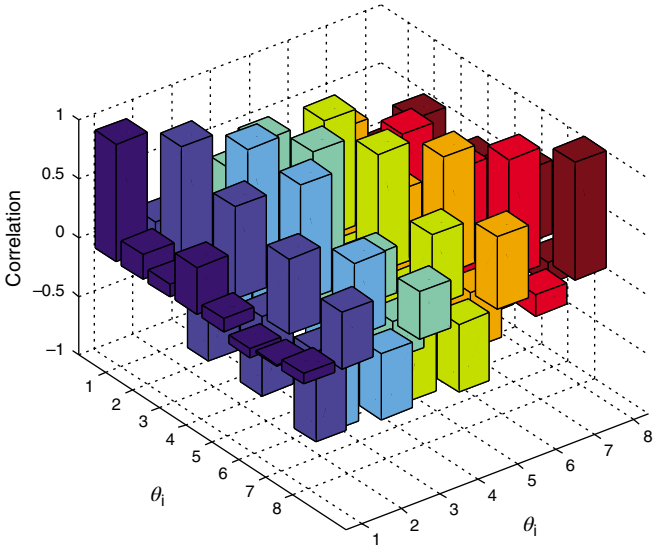

Figure 8.8 The correlation between the updated parameters (SHMC[8])

SHMC[4] and SHMC[8] techniques for finite element model updating, the error is reduced to 1.22%, 1.20%, and 1.09%, respectively. The additional observation indicates that all algorithms give robust estimation of the updated parameters which can be verified by the relatively low values of c.o.v. for all modes (less than 3%). The time step, $\delta t = 3$ ms, provides a good acceptance sampling rate for all methods, with the HMC and SHMC giving 99.9%.

Table 8.5 Modal results for the HMC and SHMC^{4,8} algorithms at $\delta t = 3$ ms

Mode	Measured frequency (Hz)	Initial FEM frequencies (Hz)	Initial error (%)	HMC frequencies (Hz)	SHMC^4 frequencies (Hz)	SHMC^8 frequencies (Hz)
1	6.380	5.71	10.47	6.313	6.312	6.301
2	16.100	15.29	5.01	15.866	15.875	15.930
3	33.130	32.53	1.82	32.236	32.238	32.330
4	33.530	34.95	4.23	33.900	33.880	33.930
5	35.650	35.65	0.012	35.643	35.620	35.589
6	48.380	45.14	6.69	48.840	48.800	48.783
7	49.430	54.69	10.65	49.871	49.860	49.702
8	55.080	55.60	0.94	54.364	54.418	54.591
9	63.040	60.15	4.59	63.888	63.896	63.828
10	66.520	67.56	1.57	67.446	67.447	67.449

FEM = finite element model.

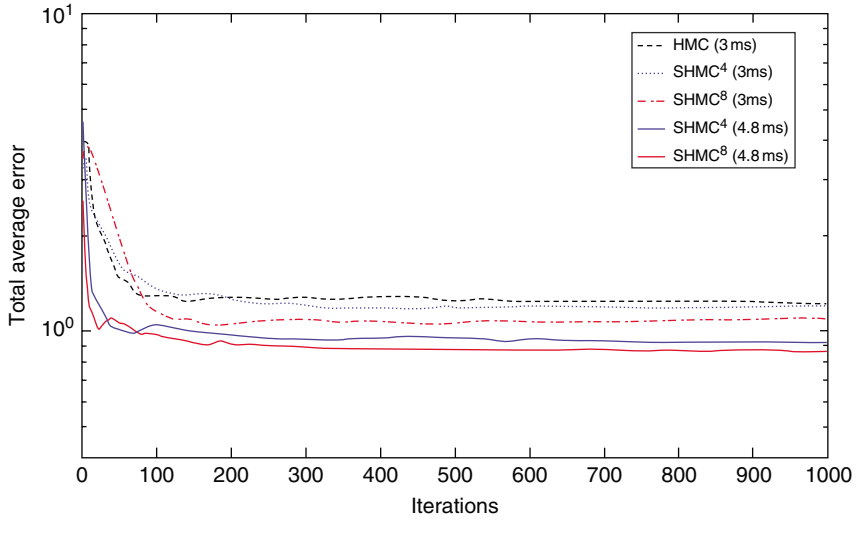

Figure 8.9 The total average error for HMC and SHMC^{4,8} for different time steps

Figure 8.9 demonstrates the total average error versus number of iterations for the three algorithms and convergence results for two time steps $\delta t = 3$ and 4.8 ms are plotted on the same graph, where $c = 0.01$ for both cases. Figure 8.9 shows the efficiency of these algorithms where 150–200 iterations (samples) are sufficient in order to obtain a good average error when time step $\delta t = 3$ ms. The three algorithms give almost the same convergence rate when the time step $\delta t = 3$ ms.

Nonetheless, the error decreases faster for both SHMC^{4,8} algorithms when the time step is $\delta t = 4.8$ ms, and this indicates that the SHMC algorithm is more efficient that the HMC

Table 8.6 Modal results and errors for HMC and SHMC^{4,8} algorithms for a time step of 4.8 ms

Mode	Measured frequency (Hz)	Initial error (%)	HMC error (%)	SHMC^4 error (%)	SHMC^8 error (%)
1	6.380	10.47	10.47	1.50	1.40
2	16.100	5.01	5.01	0.35	0.32
3	33.130	1.82	1.82	2.04	2.04
4	33.530	4.23	4.23	1.38	1.38
5	35.650	0.012	0.012	0.37	0.29
6	48.380	6.69	6.69	1.03	0.37
7	49.430	10.65	10.65	0.13	0.04
8	55.080	0.94	0.94	0.02	0.35
9	63.040	4.59	4.59	0.93	1.07
10	66.520	1.57	1.57	1.41	1.40
Total average errors	—	4.60	4.60	0.92	0.87

algorithm for large time steps (Table 8.6). When the time step is sufficiently large to permit significant jumps of the algorithm during the search process, the HMC technique offers poor updating parameters where no updating happened. This is because of the relatively large time step $\delta t = 4.8$ ms which does not conserve the Hamiltonian function, leading to a high rejection rate in the algorithm. The implementation of this time step caused significant numerical errors of the integrator used to assess the pair $(\boldsymbol{\theta}, \boldsymbol{p})$. The Hamiltonian function begins fluctuating with time, resulting in a sudden decrease in the acceptance rate: that of the HMC algorithm falls to less than 1% when the time step is $\delta t = 4.8$ ms. The acceptance rate for SHMC^4 was 70.8% and that of the SHMC^8 algorithm was 78.6%, and these values are sensible and more acceptable compared to those observed in the HMC method. The estimation error (c.o.v.) was relatively small for the SHMC algorithms, where the errors were less than 11% for the SHMC^4 and less than 6% for the SHMC^8.

For the case where the time step was increased ($\delta t = 4.8$ ms), the performance of the SHMC method improves the most (see Table 8.6) and the total average error was reduced to 0.92% for the SHMC^4 and 0.87% for the SHMC^8 . This is not the case for the HMC, where the acceptance rate decreases to less than 1% and the updated vector obtained from the HMC did not improve the finite element model results. Figure 8.10 with $c = 0.01$ shows the acceptance rate against the time step; the acceptance rate for the three methods was 99.9% when the time step was 3 ms. The acceptance rate started decreasing when the time step increased for the three methods, but this decrease was faster and more significant in the case of the HMC method. When the time step was $\delta t = 3.4$ ms, the acceptance rate for the HMC method decreased slightly to 98.7% but stayed the same for the SHMC methods (99.9% for both). When the time step was 3.8 ms the SHMC^8 acceptance rate reduced slightly to 99.8% while that of the SHMC^4 decreased to 88.2%; however, that of the HMC fell significantly to 53.2%. When the time step reached 4.8 ms, the acceptance rate decreased to 78.6% for the SHMC^8 and to 70.8% for the SHMC^4, both acceptable compared to that obtained for the HMC method (less than 1%).

In spite of the detailed mathematical formulation of the SHMC, this algorithm was easier to program, even though extra memory compared to the HMC algorithm was required for the extra parameters as well as the reweighting step during the execution of the program.

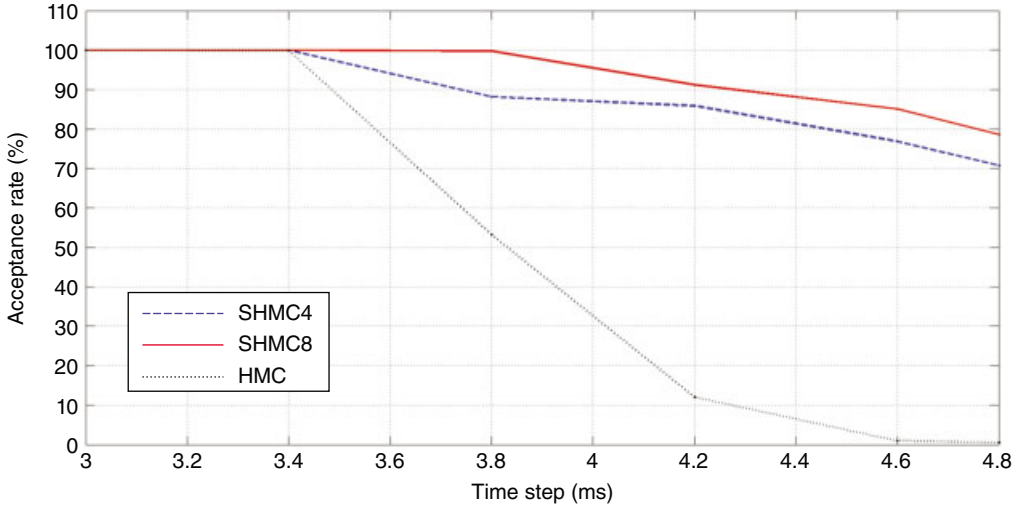

Figure 8.10 The acceptance rate obtained for HMC and SHMC^{4,8} for different time steps

Nevertheless, with a proper selection of the constant c, the SHMC algorithms performed satisfactorily. Sweet *et al*. (2006) proposed a technique to optimise the SHMC algorithm. This technique was based on identifying the mean and variance of the difference between the shadow Hamiltonian and the Hamiltonian, which can be useful to obtain the optimal values for both constant c and the time step δt. Furthermore, the SHMC technique successfully improved the time step upper bound which allowed large jumps during the search. Updating large structures with several updated parameters using the SHMC algorithms should be investigated because the number of the parameters in this method has lower effect than that of the HMC algorithm.

Different methods have been proposed to improve the efficiency of the HMC algorithm. Neal (1994) considered a transition between windows of several states (instead of a single state) at the beginning and end of the trajectory. The state within the selected window is selected according to the Boltzmann probability. The acceptance procedure using windows helped improve the acceptance rate of the HMC algorithm. Fischer *et al*. (1999) introduced an extra parameter to generalise the HMC algorithm by enhancing sampling at low temperatures, which is accomplished by sampling from a mixed canonical ensemble. Combining the SHMC algorithm with some of these modifications should be considered for finite element model updating. Additionally, the differences between the above techniques and other variations of the Hamiltonian Monte Carlo method should be studied.

8.6 Conclusion

In this chapter, the HMC technique was used to approximate the posterior distribution function, and the influence of the time step was studied. The results and limitations of the HMC in complicated finite element model updating problems were highlighted when the time step was increased. To overcome some of these limitations, a modified version based on a shadow

Hamiltonian – called the shadow hybrid Monte Carlo – was implemented. The SHMC used a modified Hamiltonian approximation to sample from the extended phase space of the shadow Hamiltonian rather than from the configuration space alone. The GARTEUR SM-AG19 structure was used to test this method and the efficiency of the algorithms was studied. Both these methods were observed to converge fast for a given time step, but the SHMC method was more efficient than HMC as it provided good samples even at larger time steps, which was not the case with the HMC technique. The sampling acceptance rate of the HMC algorithm decreased at larger time steps. The $SHMC^8$ gave more accurate results than the $SHMC^4$ since its shadow Hamiltonian consisted of more approximation terms. The technique of sampling the momentum and the molecular dynamic procedure depended on the constant c, which might be computationally expensive in cases where c was not optimised.

References

Abe Y, Tasaki S (2015) Molecular dynamics analysis of incoherent neutron scattering from light water via the Van Hove space-time self-correlation function with a new quantum correction. *Annals of Nuclear Energy* **83**: 302–308.

Balmes E (1998) Predicted variability and differences between tests of a single structure. Proceedings of the 16th International Modal Analysis Conference, Bethel, CT: SPIE, pp. 558–564.

Boulkaibet, I. (2014) Finite element model updating using Markov chain Monte Carlo techniques. PhD thesis, University of Johannesburg.

Boulkaibet I, Marwala T, Mthembu L, Friswell MI, Adhikari S (2012) Sampling techniques in Bayesian finite element model updating. *Topics in Model Validation and Uncertainty Quantification* **4**: 75–83.

Boulkaibet I, Mthembu L, Marwala T, Friswell MI, Adhikari S (2014) Finite element model updating using the shadow hybrid Monte Carlo technique. *Mechanical System & Signal Processing* **52–53**: 115–132.

Clark MA, Joó B, Kennedy AD, Silva PJ (2011) Improving dynamical lattice QCD simulations through integrator tuning using Poisson brackets and a force-gradient integrator. *Physical Review D* **84**, article no. 071502.

Creutz M (1988) Global Monte Carlo algorithms for many-fermion systems. *Physics Review D* **38**: 1228–1238.

Datta BN (2002) Finite element model updating, eigenstructure assignment and eigenvalue embedding techniques for vibrating systems. *Mechanical Systems & Signal Processing* **16**: 83–96.

Degener M, Hermes M (1996) Ground vibration test and finite element analysis of the GARTEUR SM-AG19 testbed. Report IB 232-96J 08, Deutsche Forschungsanstalt für Luft und Raumfahrt e.V. Institut für Aeroelastik.

Engle RD, Skeel RD, Drees M (2005) Monitoring energy drift with shadow Hamiltonians. *Journal of Computational Physics* **206**: 432–452.

Escribano B, Akhmatskaya E, Mujika JI (2013) Combining stochastic and deterministic approaches within high efficiency molecular simulations. *Central European Journal of Mathematics* **11**:787–799.

Escribano B, Akhmatskaya E, Reich S, Azpiroz JM (2014) Multiple-time-stepping generalized hybrid Monte Carlo methods. *Journal of Computational Physics* **280**: 1–20.

Fernández-Pendás M, Escribano B, Radivojević T, Akhmatskaya E (2014) Constant pressure hybrid Monte Carlo simulations in GROMACS. *Journal of Molecular Modeling* **20**: 2487.

Fischer A, Cordes F, Schütte C (1999) Hybrid Monte Carlo with adaptive temperature in mixed-canonical ensemble: efficient conformational analysis of RNA. *Computer Physics Communications*: **121–122**: 37–39.

Fishman GS (2000) *Monte Carlo: Concepts, Algorithms, and Applications*. Springer Series in Operations Research. New York: Springer-Verlag.

Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *Journal of Machine Learning Research* **3**: 1157–1182.

Hairer E, Lubich C, Wanner G (2006) *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Berlin: Springer-Verlag.

Hausdorff F (1906) Die symbolische Exponentialformel in der Gruppentheorie. *Berichte über Verhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig* **58**: 19–48.

Izaguirre JA, Hampton SS (2004) Shadow hybrid Monte Carlo: an efficient propagator in phase space of macromolecules. *Journal of Computational Physics* **200**: 581–604.

Kennedy AD, Pendleton B (1991) Acceptances and autocorrelations in hybrid Monte Carlo. *Nuclear Physics B, Proceedings Supplements* **20**: 118–121.

Li H-J, Yi H-B, Xu J-J (2015) High-order Cu(II) chloro-complexes in LiCl brines: insights from density function theory and molecular dynamics. *Geochimica et Cosmochimica Acta* **165**: 1–13.

Li YF, Valla S, Zio E (2015) Reliability assessment of generic geared wind turbines by GTST-MLD model and Monte Carlo simulation. *Renewable Energy* **83**: 222–233.

Link M, Friswell MI (2003) Generation of validated structural dynamic models – results of a benchmark study utilizing the GARTEUR SM-AG19 testbed. *Mechanical Systems & Signal Processing* **17**: 9–20.

Mares C, Mottershead JE, Friswell MI (2000) Selection and updating of parameters for the GARTEUR SM-AG19 testbed. Proceedings of the International Conference of Noise and Vibration Engineering (ISMA25), pp. 635–640.

Marwala T (2010) *Finite Element Model Updating Using Computational Intelligence Techniques*. London: Springer-Verlag.

Marwala T (2012) *Condition Monitoring Using Computational Intelligence Methods*. Heidelberg: Springer.

Marwala T (2013) *Economic Modeling Using Artificial Intelligence Methods*. Heidelberg: Springer.

Marwala T (2014) *Artificial Intelligence Techniques for Rational Decision Making*. Heidelberg: Springer.

Marwala T (2015) *Causality, Correlation, and Artificial Intelligence for Rational Decision Making*. Singapore: World Scientific.

Marwala T, Lagazio M (2011) *Militarized Conflict Modeling Using Computational Intelligence*. Heidelberg: Springer.

Nakano Y (2015) Quasi-Monte Carlo methods for choquet integrals. *Journal of Computational & Applied Mathematics*. **287**, article no. 10081.

Neal RM (1994) An improved acceptance procedure for the hybrid Monte Carlo algorithm. *Journal of Computational Physics* **111**: 94–203.

Skeel RD, Hardy DJ (2001) Practical construction of modified Hamiltonians. *SIAM Journal of Scientific Computing* **23**: 1172–1188.

Sweet, C., Hampton, S. and Izaguirre, J. (2006) Optimal implementation of the shadow hybrid Monte Carlo method. Technical Report TR-2006-09, University of Notre Dame.

Sweet CR, Hampton SS, Skeel RD, Izaguirre JA (2009) A separable shadow Hamiltonian hybrid Monte Carlo method. *Journal of Chemical Physics* **131**: article no. 174106.

Von Neumann J (1951) Various techniques used in connection with random digits. In: *Monte Carlo Method* (ed. Householder AS, Forsythe GE, Germond HH), Applied Mathematics Series **12**, pp. 36–38. Washington, DC: US Government Printing Office.

Xie Z-C, Gao T-H, Guo X-T, Xie Q (2015) Molecular dynamics simulation of nanocrystal formation and deformation behavior of Ti_3Al alloy. *Computational Materials Science* **98**: 245–251.

Zhang W, You C (2015) Numerical approach to predict particle breakage in dense flows by coupling multiphase particle-in-cell and Monte Carlo methods. *Powder Technology* **283**: 128–136.

9

Separable Shadow Hybrid Monte Carlo in Finite Element Updating

9.1 Introduction

In the previous chapters, the hybrid Monte Carlo (HMC) and shadow hybrid Monte Carlo (SHMC) algorithms were used to update the uncertain parameters of finite element models. The SHMC algorithm was able to preserve the HMC algorithm's ability to converge fast and improved the performance of the basic HMC algorithm by drawing samples when the time step was large. The SHMC algorithm was able to increase the upper bound of the time step while maintaining the acceptance rate of the algorithm at sensible levels, which gives more accurate results than the original HMC algorithm. However, the SHMC algorithm's performance can be further enhanced by implementing a separable shadow Hamiltonian function so as to streamline the Monte Carlo (MC) step and to decrease the time it takes to run the algorithm. In this chapter, an alternative modified version of the HMC, the separable shadow hybrid Monte Carlo (S2HMC) algorithm, is applied in finite element model updating (Sweet *et al.*, 2009; Boulkaibet, 2014; Boulkaibet *et al.*, 2014, 2015). The procedure samples from a separable modified Hamiltonian function by altering the search space and does not encompass any additional tuning parameter.

The chapter discusses and provides theoretical justification for the S2HMC algorithm. It then presents the results gained when finite element models of the asymmetrical H-shaped structural beam and aircraft GARTEUR SM-AG19 structure are updated by using the S2HMC algorithm. The results obtained are compared to those obtained from using both the HMC and SHMC algorithms.

9.2 Separable Shadow Hybrid Monte Carlo

The S2HMC algorithm is a modified form of the HMC and SHMC algorithms where a separable shadow Hamiltonian function is used to create samples (Sweet *et al.*, 2009; Boulkaibet,

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

2014; Boulkaibet *et al.*, 2014, 2015). A shadow Hamiltonian function signifies an accurate estimate of the Hamiltonian function (total energy), which can be more conserved (remaining almost constant when the Verlet integrator is evaluated) than the original Hamiltonian function, which starts to fluctuate, when relatively large time steps are used for evolving a Verlet integrator. However, the use of the shadow Hamiltonian function might obfuscate the sampling process of the momentum vector. The S2HMC algorithm improves the sampling efficiency by sampling from a separable shadow Hamiltonian function by altering the configuration space. These transformations make it unnecessary to use additional parameters such as the constant c in the SHMC algorithm and circumvent the difficulty of using an augmented integrator. Additionally, the process of sampling a new momentum vector is the same as in the HMC algorithm, and these enhancements speed up the convergence of averages calculated and the transformations used advance the acceptance rate with a relatively insignificant further computational cost (Sweet *et al.*, 2009; Boulkaibet, 2014; Boulkaibet *et al.*, 2014, 2015).

The S2HMC algorithm uses the velocity Verlet (VV) integrator to increase the order of accuracy of the sampling technique and this is achieved by altering the configuration space by introducing the pre-processing and post-processing steps (Sweet *et al.*, 2009; Boulkaibet, 2014; Boulkaibet *et al.*, 2014, 2015). The modified Hamiltonian function applied in this procedure is conserved to $O(\delta t^4)$ by the treated technique, instead of merely $O(\delta t^2)$ by the untreated technique. Analogous to the SHMC procedure, the S2HMC procedure also necessitates a reweighting step to deal with the alteration of the potential energy. The shadow Hamiltonian function applied in the S2HMC is separable and is of fourth order (Sweet *et al.*, 2009; Boulkaibet *et al.*, 2014, 2015; Boulkaibet, 2014):

$$\tilde{H}(\boldsymbol{\theta},\mathbf{p}) = \frac{1}{2}\mathbf{p}^{\mathbf{T}}\mathbf{M}^{-1}\mathbf{p} + V(\boldsymbol{\theta}) + \frac{\delta t^2}{24}V_{\boldsymbol{\theta}}{}^{T}\mathbf{M}^{-1}V_{\boldsymbol{\theta}} + O\left(\delta t^4\right), \tag{9.1}$$

where $V_{\boldsymbol{\theta}}$ is the derivative of the potential energy V with respect to $\boldsymbol{\theta}$. The joint distribution derived from the separable shadow Hamiltonian function can be expressed as $\tilde{\rho}(\boldsymbol{\theta},\mathbf{p}) \propto \exp\left(-\beta_B \tilde{H}(\boldsymbol{\theta},\mathbf{p})\right)$, while the separable shadow Hamiltonian function is a result of applying backward error analysis to numerical integrators (Hairer *et al.*, 2002; Leimkuhler and Reich, 2004; Sweet and Izaguirre, 2006). The pre-processing step is expressed as follows (Sweet *et al.*, 2009; Boulkaibet, 2014; Boulkaibet *et al.*, 2014, 2015):

$$\hat{\mathbf{p}} = \mathbf{p} - \frac{\delta t}{24}\left(V_{\boldsymbol{\theta}}\left(\boldsymbol{\theta} + \delta t\mathbf{M}^{-1}\hat{\mathbf{p}}\right) - V_{\boldsymbol{\theta}}\left(\boldsymbol{\theta} - \delta t\mathbf{M}^{-1}\hat{\mathbf{p}}\right)\right), \tag{9.2}$$

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + \frac{\delta t^2}{24}\mathbf{M}^{-1}\left(V_{\boldsymbol{\theta}}\left(\boldsymbol{\theta} + \delta t\mathbf{M}^{-1}\hat{\mathbf{p}}\right) + V_{\boldsymbol{\theta}}\left(\boldsymbol{\theta} - \delta t\mathbf{M}^{-1}\hat{\mathbf{p}}\right)\right). \tag{9.3}$$

Equations 9.2 and 9.3 necessitate an iterative solution for $\hat{\mathbf{p}}$ and a direct calculation for $\hat{\boldsymbol{\theta}}$. The post-processing step is written as follows (Sweet *et al.*, 2009; Boulkaibet, 2014; Boulkaibet *et al.*, 2014, 2015):

$$\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} - \frac{\delta t^2}{24}\mathbf{M}^{-1}\left(V_{\boldsymbol{\theta}}\left(\boldsymbol{\theta} + \delta t\mathbf{M}^{-1}\hat{\mathbf{p}}\right) + V_{\boldsymbol{\theta}}\left(\boldsymbol{\theta} - \delta t\mathbf{M}^{-1}\hat{\mathbf{p}}\right)\right), \tag{9.4}$$

$$\mathbf{p} = \hat{\mathbf{p}} + \frac{\delta t}{24}\left(V_{\boldsymbol{\theta}}\left(\boldsymbol{\theta} + \delta t\mathbf{M}^{-1}\hat{\mathbf{p}}\right) - V_{\boldsymbol{\theta}}\left(\boldsymbol{\theta} - \delta t\mathbf{M}^{-1}\hat{\mathbf{p}}\right)\right). \tag{9.5}$$

Equations 9.4 and 9.5 necessitate an iterative solution for $\boldsymbol{\theta}$ and direct calculation for $\mathbf{p}$. To compute balanced values of the mean, the results need to be reweighted and the average of an observable A is written as follows (Sweet *et al.*, 2009; Boulkaibet, 2014; Boulkaibet *et al.*, 2014, 2015):

$$\langle B \rangle = \frac{\sum_{i=1}^{N_s} Ba_i}{\sum_{i=1}^{N_s} a_i}, \quad \text{where } a_i = \frac{\exp(-\beta_B H(\boldsymbol{\theta}, \mathbf{p}))}{\exp\left(-\beta_B \tilde{H}(\boldsymbol{\theta}, \mathbf{p})\right)}. \tag{9.6}$$

The S2HMC procedure can be abridged as follows (Sweet *et al.*, 2009; Boulkaibet, 2014; Boulkaibet *et al.*, 2014, 2015):

1. Initiate $\boldsymbol{\theta}_0$.
2. Initiate $\mathbf{p}_0$ so that $\mathbf{p}_0 \sim \mathcal{N}(0, \boldsymbol{M})$.
3. Calculate the initial shadow energy $\tilde{H}(\boldsymbol{\theta}, \mathbf{p})$ using Equation 9.1.
4. Pre-processing step: beginning with ($\boldsymbol{\theta}$, $\mathbf{p}$), iteratively calculate $\hat{\mathbf{p}}$ and directly estimate $\hat{\boldsymbol{\theta}}$ using Equations 9.2 and 9.3.
5. Start the leapfrog algorithm with $\left(\hat{\boldsymbol{\theta}}, \hat{\mathbf{p}}\right)$ and run for L time steps to obtain $\left(\hat{\boldsymbol{\theta}}^*, \hat{\mathbf{p}}^*\right)$.
6. Post-processing step: beginning with $\left(\hat{\boldsymbol{\theta}}^*, \hat{\mathbf{p}}^*\right)$, iteratively calculate $\boldsymbol{\theta}^*$ and directly estimate $\mathbf{p}^*$ using Equations 9.4 and 9.5.
7. Update the finite element model to get the new analytical frequencies and then compute H ($\boldsymbol{\theta}^*$, $\mathbf{p}^*$).
8. Accept ($\boldsymbol{\theta}^*$, $\mathbf{p}^*$) with probability $\min\left(1, \exp\left\{-\beta_B \Delta \tilde{H}\right\}\right)$.
9. Repeat steps 3–8 to obtain N_s samples.
10. Calculate the weight by using Equation 9.6 to calculate the averages of the quantity $\mathbf{A}(\boldsymbol{\theta})$.

The idea of the constructed processed integrator is to alter the phase space, that is, the pre-processing of the pair ($\boldsymbol{\theta}$, $\mathbf{p}$), where the propagation is conducted in a different space and by using another integrator which has a non-separable shadow Hamiltonian. The inverse of the pre-processing step, which is also called the post-processing step, is evaluated when the output is required and thus, compared to the SHMC method, the momentum sampling process is simplified and an accurate and faster simulation is assured (Blanes *et al.*, 2004; Sweet *et al.*, 2009; Boulkaibet, 2014; Boulkaibet *et al.*, 2015).

Particular changes can be made to improve the S2HMC technique, including sampling the momentum vector by avoiding the dependency between their components when the momentum is drawn. The best approach for handle the dependencies between components, which are unavoidable, is by applying an ordered over-relaxation technique to minimise the random walk in the momentum sampling process (Neal, 1998). Finite difference estimation is employed to calculate the gradient $V_{\boldsymbol{\theta}}$ and this gradient is based on forward and backward Taylor series expansion of the function. Nevertheless, in the situation where the dimension of the uncertain parameters is high, the forward/backward difference estimation could be more practical since it necessitates d, which is the dimension of the uncertain parameters evaluation of $V(\boldsymbol{\theta})$ to calculate the gradient, while the central difference approximation necessitates $2d$, and the forward difference approximation gradient, which is given by

$$\frac{\partial V}{\partial E_i} = \frac{V(\boldsymbol{\theta} + \Delta h) - V(\boldsymbol{\theta})}{h\Delta_i}.$$

In a situation where significant amounts or streaming data are involved when calculating the gradient it is computationally taxing and as an alternative a subset of the data is used to compute a noisy gradient. This type of gradient is called a stochastic gradient (Welling and Teh, 2011; Boulkaibet, 2014). As a final point, in order to compute balanced values of the mean, the results must be reweighted and this is achieved using $\rho(\boldsymbol{\theta},\boldsymbol{p})/\widetilde{\rho}(\boldsymbol{\theta},\boldsymbol{p})$ before evaluating the averages. The average of an observable A is given by Equation 9.6. If the weighted vector parameter is given by $\widetilde{\boldsymbol{\theta}}$, the mean value of the estimated parameters is given by (Boulkaibet, 2014)

$$\hat{\widetilde{\boldsymbol{\theta}}} = E\left(\widetilde{\boldsymbol{\theta}}\right) \cong \frac{1}{N_s}\sum_{i=1}^{N_s}\widetilde{\boldsymbol{\theta}}^i, \tag{9.7}$$

where $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta} a_i \Big/ \sum_{i=1}^{Ns} a_i$ and $\boldsymbol{\theta}$ is the unweighted uncertain vector. The mean value can be written as (Boulkaibet, 2014)

$$\hat{\widetilde{\boldsymbol{\theta}}} \cong \frac{1}{N_s}\sum_{i=1}^{N_s}\frac{a_i}{\sum_{j=1}^{Ns} a_j}\widetilde{\boldsymbol{\theta}}^i = \frac{1}{\sum_{j=1}^{Ns} a_j}\frac{1}{N_s}\sum_{i=1}^{N_s} a_i\widetilde{\boldsymbol{\theta}}^i = \frac{1}{\sum_{j=1}^{Ns} a_j}E\left(\boldsymbol{\theta}\boldsymbol{a}^{\boldsymbol{T}}\right). \tag{9.8}$$

By the same logic, the variance of the weighted estimated parameters can be written as (Boulkaibet, 2014)

$$V\left(\widetilde{\boldsymbol{\theta}}\right) = E\left(\left(\widetilde{\boldsymbol{\theta}} - \hat{\widetilde{\boldsymbol{\theta}}}\right)^2\right) \cong V(\boldsymbol{\theta})\frac{\sum_{i=1}^{N_s} a_i{}^2}{\sum_{j=1}^{N_s} a_j{}^2}, \tag{9.9}$$

where $V(\boldsymbol{\theta})$ is the unweighted variance, which is similar to the value estimated by the HMC algorithm, or any version of the Markov chain Monte Carlo algorithm. The standard deviation is then given by $\boldsymbol{\sigma}_{\widetilde{\boldsymbol{\theta}}} = \sqrt{V\left(\widetilde{\boldsymbol{\theta}}\right)}$.

9.3 Theoretical Justifications of the S2HMC Method

In the HMC algorithm, the molecular dynamic (MD) integrator ensures reversibility and preservation of volume to satisfy the detailed balance (Duane *et al.*, 1987; Gupta *et al.*, 1988; Hanson, 2001; Alfaki, 2008; Boulkaibet *et al.*, 2012, 2014; Beskos *et al.*, 2013; Boulkaibet, 2014). In this section, it is demonstrated that the processed leapfrog is symplectic and reversible and that the shadow Hamiltonian function $\tilde{H}$ of the S2HMC is conserved to fourth order for the processed leapfrog. In the S2HMC algorithm, the objective is to apply a processed leapfrog integrator where its fourth order shadow Hamiltonian is separable (Sweet *et al.*, 2009):

$$\tilde{H}(\boldsymbol{\theta},\mathbf{p}) = \frac{1}{2}\mathbf{p}^{\mathbf{T}}\mathbf{M}^{-\mathbf{1}}\mathbf{p} + \tilde{V}(\boldsymbol{\theta}), \tag{9.10}$$

with $\tilde{V}(\boldsymbol{\theta})$ signifying the modified potential energy. The fourth-order shadow Hamiltonian, which is a result of applying backward error to the numerical integrator, is conserved to $O(\delta t^4)$ (Skeel and Hardy, 2001; Engle *et al.*, 2005). The processed integrator is created by changing the search space variables (pre-processing step) for better propagation by another processed integrator. The inverse mapping (post-processor step) can be then evaluated when the output is needed. The processed integrator has more advantages in terms of sampling accuracy and less complexity of the MC step (Blanes *et al.*, 2004). The mapping function used by the S2HMC algorithm preserves the phase space volume and confirms the reversibility principle of momentum, and these fulfil the detailed balance. The following generation function is used to obtain a symplectic mapping function (Sweet *et al.*, 2009):

$$S(\boldsymbol{\theta},\hat{\mathbf{p}}) = \boldsymbol{\theta}^T\hat{\mathbf{p}} + \frac{\delta t}{24}\left(V_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}+\delta t\mathbf{M}^{-1}\hat{\mathbf{p}}\right) - V_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}-\delta t\mathbf{M}^{-1}\hat{\mathbf{p}}\right)\right), \tag{9.11}$$

where the mapping function $(\boldsymbol{\theta},\hat{\mathbf{p}}) = \chi\left(\hat{\boldsymbol{\theta}},\hat{\mathbf{p}}\right)$ can be obtained as follows:

$$\hat{\boldsymbol{\theta}} = \frac{\partial S}{\partial \hat{\mathbf{p}}}, \quad \hat{\mathbf{p}} = \frac{\partial S}{\partial \boldsymbol{\theta}}. \tag{9.12}$$

The mapping function in Equation 9.11, which is attained by pre-processing steps in Equations 9.2 and 9.3, preserves the symplectic property of the processed integrator. The inverse mapping χ^{-1} is given by the post-processing step (Equations 9.4 and 9.5). The reversibility of the processed integrator can be proven by showing that $\chi(\boldsymbol{\theta},-\mathbf{p}) = \mathrm{diag}(\mathbf{I},-\mathbf{I})\chi(\boldsymbol{\theta},\mathbf{p})$, which demonstrates that the processed integrator is both symplectic and reversible. Therefore, the S2HMC algorithm satisfies the detailed balance relations.

It is vital to demonstrate the relationship between the unprocessed leapfrog integrator and the processed leapfrog integrator. The fourth-order shadow Hamiltonian function for the processed leapfrog integrator was shown in Equation 9.1. Nevertheless, the numerical solution of the unprocessed leapfrog is known to be approximately equal to the analytical solution of the following fourth-order shadow Hamiltonian function (Skeel and Hardy, 2001; Engle *et al.*, 2005):

$$\tilde{H}_0\left(\hat{\boldsymbol{\theta}},\hat{\mathbf{p}}\right) = \frac{1}{2}\hat{\mathbf{p}}^{\mathbf{T}}\mathbf{M}^{-\mathbf{1}}\hat{\mathbf{p}} + V\left(\hat{\boldsymbol{\theta}}\right) + \frac{\delta t}{12}\mathbf{K}_{\mathbf{p}}^T V_{\boldsymbol{\theta}\boldsymbol{\theta}}\mathbf{K}_{\mathbf{p}} - \frac{\delta t^2}{24}V_{\boldsymbol{\theta}}{}^T\mathbf{M}^{-\mathbf{1}}V_{\boldsymbol{\theta}} + O\left(\delta t^4\right), \tag{9.13}$$

where $V_{\boldsymbol{\theta}\boldsymbol{\theta}}$ is the derivative of the potential energy $V_{\boldsymbol{\theta}}$ with respect to $\boldsymbol{\theta}$, and $\mathbf{K}_{\mathbf{p}}$ is the derivative of kinetic energy scalar value K($\mathbf{p}$) with respect to $\mathbf{p}$. The shadow Hamiltonian function of the transformed variables $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{p}}$ is shown in Equation 9.13. The mapping function $\left(\hat{\boldsymbol{\theta}},\hat{\mathbf{p}}\right) = \chi(\boldsymbol{\theta},\mathbf{p})$ connects the original to the transformed variables and because the transformation is symplectic, the shadow Hamiltonian function of the processed integrator can be expressed as $\tilde{H}(\boldsymbol{\theta},\mathbf{p}) = \tilde{H}_0(\chi(\boldsymbol{\theta},\mathbf{p}))$. The power expansions of the transformed variables $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{p}}$ are defined as (Skeel and Hardy, 2001; Engle *et al.*, 2005):

$$\hat{\mathbf{p}} = \mathbf{p} - \frac{\delta t^2}{12}V_{\boldsymbol{\theta}}\mathbf{M}^{-\mathbf{1}}\mathbf{p} + O\left(\delta t^4\right), \tag{9.14}$$

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + \frac{\delta t^2}{12}\mathbf{M}^{-\mathbf{1}}V_{\boldsymbol{\theta}} + O\left(\delta t^4\right). \tag{9.15}$$

Table 9.1 The momentum, MD and reweighting steps for HMC, SHMC and S2HMC algorithms

Methods	Sampling new momentums (MC) step	MD step	Reweighting step
HMC	Given $\boldsymbol{\theta}$, generate $\mathbf{p}$ such that $\mathbf{p}\sim\mathcal{N}(0,\mathbf{M})$	Accept $(\boldsymbol{\theta}^*, p^*)$ with probability $\min(1,\exp\{-\beta_B\Delta H\})$	—
SHMC	Given $\boldsymbol{\theta}$, generate $\mathbf{p}$ such that $\mathbf{p}\sim\mathcal{N}(0,\boldsymbol{M})$ Accept with probability $\left(1,\exp\left\{-\beta_B\left(H_{[2k]}(\boldsymbol{\theta},\boldsymbol{p})-c-H(\boldsymbol{\theta},\mathbf{p})\right)\right\}\right)$	Accept $(\boldsymbol{\theta}^*, p^*)$ with probability $\min\left(1,\exp\left\{-\beta_B\Delta\tilde{H}\right\}\right)$ where $\tilde{H}=H_{[2k]}(\boldsymbol{\theta},\boldsymbol{p})-c$	The observable B is given by: $\langle B\rangle=\frac{\sum_{i=1}^{N_s}Ba_i}{\sum_{i=1}^{N_s}a_i}$, where $a_i=\frac{\exp(-\beta_B H(\boldsymbol{\theta},\mathbf{p}))}{\exp\left(-\beta_B\tilde{H}(\boldsymbol{\theta},\mathbf{p})\right)}$
S2HMC	Given $\boldsymbol{\theta}$, generate $\boldsymbol{p}$ such that $\mathbf{p}\sim\mathcal{N}(0,\boldsymbol{M})$ Then solve $\hat{\mathbf{p}}=\mathbf{p}-\frac{\delta t}{24}\left(V_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}+\delta t\mathbf{M}^{-1}\hat{\mathbf{p}}\right)-V_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}-\delta t\mathbf{M}^{-1}\hat{\mathbf{p}}\right)\right)$	Accept $(\boldsymbol{\theta}^*, p^*)$ with probability $\min\left(1,\exp\left\{-\beta_B\Delta\tilde{H}\right\}\right)$	The observable B is given by: $\langle B\rangle=\frac{\sum_{i=1}^{N_s}Ba_i}{\sum_{i=1}^{N_s}a_i}$, where $a_i=\frac{\exp(-\beta_B H(\boldsymbol{\theta},\mathbf{p}))}{\exp\left(-\beta_B\tilde{H}(\boldsymbol{\theta},\mathbf{p})\right)}$

Then, the shadow Hamiltonian function of the processed integrator can be attained by substituting Equations 9.14 and 9.15 into Equation 9.13 and identifying the power expansion. Table 9.1 describes the S2HMC, HMC and SHMC algorithms (Boulkaibet, 2014). The differences between the procedures of sampling momentum, along with the MD step, are explained for the three algorithms. The function $H_{[2k]}(\boldsymbol{\theta}, \mathbf{p})$ characterises the accurate shadow Hamiltonian of order k.

In the next section the S2HMC algorithm is applied to update an asymmetrical H-shaped aluminium structure.

9.4 Application 1: Asymmetrical H-Shaped Structure

The asymmetrical H-shaped aluminium structure is now investigated using the S2HMC technique to update its finite element model (Marwala, 1997, 2010). The structure is modelled using finite element modelling software. The results obtained from finite element model updating using the S2HMC algorithm are compared to those obtained using HMC and SHMC algorithms. Because the S2HMC algorithm uses a fourth-order Hamiltonian function, the results obtained by the SHMC[4] algorithm are compared with those obtained by the S2HMC algorithm.

The measured natural frequencies of this structure are 53.9, 117.3, 2041.4, 254.0 and 445.0 Hz. The moments of inertia and the cross-sectional areas of the left, middle and right subsections of the beam are updated. The updating parameter vector is $\boldsymbol{\theta}=\{I_{x1},I_{x2},I_{x3},A_{x1},A_{x2},A_{x3}\}$. The posterior constants, beam parameters and boundaries of these

Table 9.2 Initial and updated parameters using HMC, SHMC and S2HMC

	Initial	HMC method	$\frac{\sigma_i}{\theta_i}$ (%)	SHMC method	$\frac{\sigma_i}{\theta_i}$ (%)	S2HMC method	$\frac{\sigma_i}{\theta_i}$ (%)
I_{x1}	2.73×10^{-8}	2.21×10^{-8}	12.67	2.18×10^{-8}	12.60	2.19×10^{-8}	13.44
I_{x2}	2.73×10^{-8}	2.60×10^{-8}	1.37	2.49×10^{-8}	3.99	2.44×10^{-8}	2.70
I_{x3}	2.73×10^{-8}	2.90×10^{-8}	16.50	2.96×10^{-8}	14.93	2.95×10^{-8}	13.31
A_{x1}	3.16×10^{-4}	4.00×10^{-4}	1.39	4.05×10^{-4}	2.19	3.91×10^{-4}	2.93
A_{x2}	3.16×10^{-4}	2.30×10^{-4}	1.10	4.05×10^{-4}	2.10	2.34×10^{-4}	2.51
A_{x3}	3.16×10^{-4}	2.40×10^{-4}	1.95	2.25×10^{-4}	3.20	2.21×10^{-4}	7.44

Table 9.3 Natural frequencies and errors when HMC, SHMC and S2HMC are used to update the parameters

Measured frequency (Hz)	Initial frequency (Hz)	Frequencies HMC method (Hz)	Frequencies SHMC method (Hz)	Frequencies S2HMC method (Hz)
53.90	51.40	52.93 (0.17%)	52.94 (0.04%)	53.16 (0.83%)
117.30	116.61	118.82 (0.21%)	118.23 (0.15%)	118.78 (0.87%)
208.40	201.27	208.81 (0.24%)	207.91 (0.27%)	208.56 (0.99%)
254.00	247.42	254.41 (0.22%)	253.84 (0.17%)	254.04 (1.20%)
445.00	390.33	444.13 (0.41%)	443.00 (0.20%)	444.16 (1.56%)

parameters are given in Marwala (1997). The number of samples is $N_s = 1000$, and the time step is set to 0.0045 s, and L is uniformly distributed on the interval $\{1, 30\}$. The constant is $c = 0.001$ for the SHMC algorithm. Each algorithm was implemented over 20 independent runs and the results are shown in Tables 9.2 and 9.3 (Boulkaibet, 2014).

Figure 9.1 shows the scatter plots with marginal histograms for four uncertain parameters using S2HMC algorithm. These graphs show concentration in a specific area, which means that the algorithm found the area of high probability (almost the same as HMC and SHMC algorithms). The expectation values of the rest of the parameters along with their coefficient of variation (c.o.v.) are given in Table 9.2.

Table 9.2 presents the initial, the updated values of the uncertain parameters, and the c.o.v. of parameters for the S2HMC algorithm. The c.o.v. represents the percentage of estimated standard deviation divided by the estimated $\boldsymbol{\theta}$ for each algorithm, $100\sigma_i/\theta_i$. The estimations of the middle beam parameters are more accurate than the left and the right beam parameters, which can be seen from the values of the c.o.v. in Table 9.2. The c.o.v. values of the middle beam parameters are lower than those obtained for both the left and right beams. This could be attributed to the structure being excited at the position indicated by the double arrow (located in the middle beam), and thus more information on the middle beam was used in the updating process. In general, the S2HMC algorithm updated the $\boldsymbol{\theta}$ vector where the updated vector obtained by the algorithm was physically realistic. The time step used provides a very good sampling acceptance rate for all algorithms (99.9%). The updated values and the c.o.v. for the HMC and SHMC algorithms are found to be close to those obtained by the S2HMC algorithm (Figure 9.2).

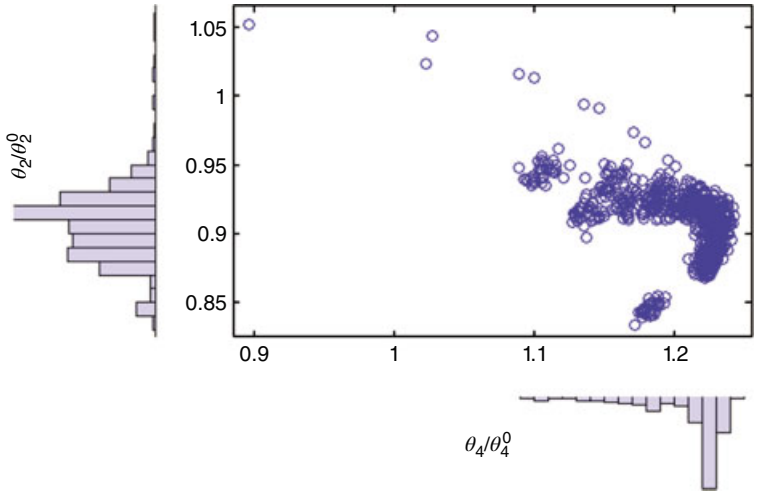

Figure 9.1 Scatter plot of A_{x1} versus I_{x2} with marginal histograms using S2HMC methods

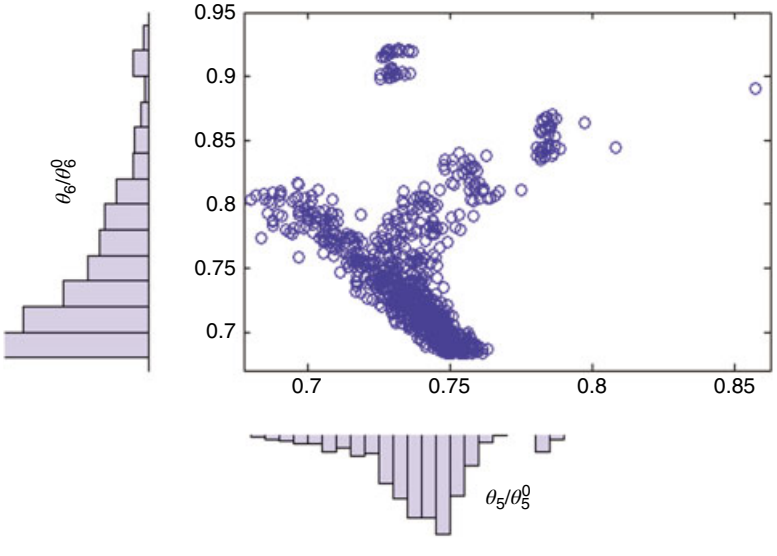

Figure 9.2 Scatter plot of A_{x2} versus A_{x3} with marginal histograms using S2HMC methods

Table 9.3 shows the updated natural frequencies for each mode, and the coefficient of variation in brackets represents the estimated standard deviation of the frequency divided by the estimated frequency.

The HMC algorithm reduced the error to 0.73%, while the fourth-order SHMC and S2HMC algorithms gave errors of 0.66% and 0.58%, respectively. From Table 9.3, the error between the first measured natural frequency and that of the initial model was 4.63%. When the S2HMC algorithm was applied, the error decreased to 1.37%. The same comments can be made for the

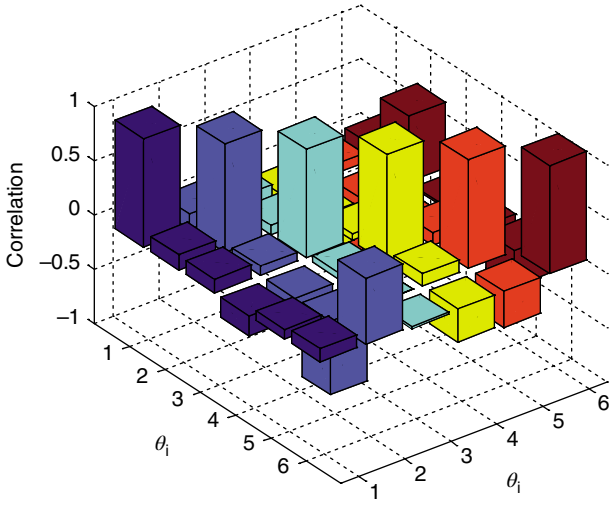

Figure 9.3 The correlation between the updated parameters (S2HMC algorithm)

third, fourth, and fifth natural frequencies. Overall, the updated finite element model natural frequencies for the S2HMC algorithm were better than the initial .

When the results obtained by the S2HMC algorithm were compared with those calculated by both the HMC and SHMC algorithms, the S2HMC algorithm gave a slightly better total average error than both the HMC and SHMC algorithms. As can be seen in Table 9.3, finite element model updating using the S2HMC method improved the error from 4.7% to 0.58%, which is an acceptable percentage. The c.o.v. (in parentheses) for the S2HMC algorithms is very small, suggesting that the updating process using the Bayesian approach while predicting the posterior distribution function by the separable shadow Hamiltonian algorithms produced more accurate results compared to the other algorithms. The c.o.v. values obtained by the S2HMC are slightly higher than those obtained by HMC algorithm (but still less than 1.6%). This is to be expected since the S2HMC algorithm uses an approximation of the Hamiltonian function for the sampling procedure and the original Hamiltonian function is almost conserved for lower time steps. However, the separable shadow Hamiltonian function has advantages when the time step is relatively large. Figure 9.3 shows the correlation between all updated parameters. All parameters are weakly correlated, except for the pair (A_{x1}, A_{x3}) which is highly correlated (the correlation is equal to 0.71).

The total average error is plotted on a logarithmic scale against the number of iterations in Figure 9.4, and this figure can be used to determine the convergence of the S2HMC algorithm in terms of iterations (samples) needed to converge. The results obtained by the S2HMC are compared in the same plot with those from the HMC and SHMC algorithms. The results show that the S2HMC algorithm converges faster and within the first 100 iterations along with the other two algorithms when the time step is relatively small. This means that 200 iterations (or 200 samples) could be sufficient to obtain good updated parameters.

The time step used in Hamiltonian algorithms determines the accuracy and speed of these algorithms. Figure 9.5 shows the sampling acceptance rate (AR) of the S2HMC algorithm when the time step varies between 0.006 and 0.01 s. Moreover, Figure 9.5 shows both the

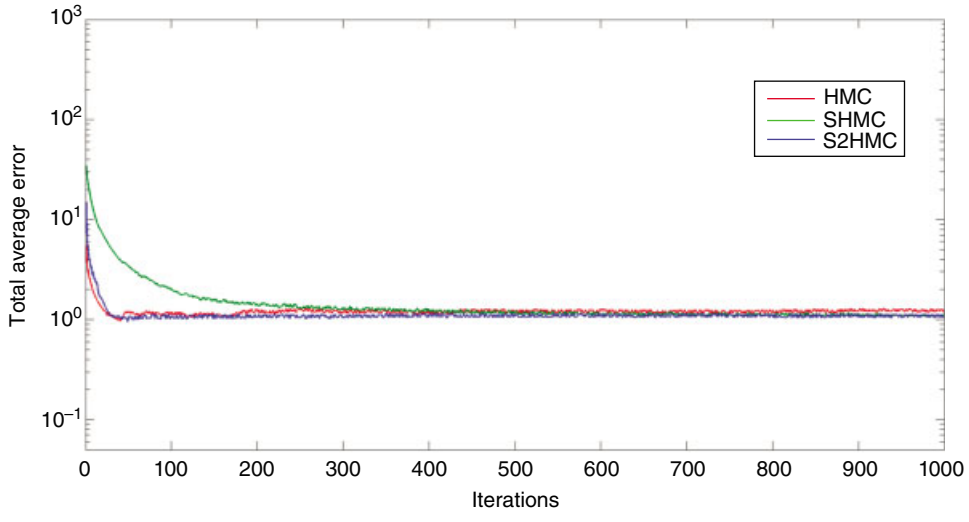

Figure 9.4 The total average error using the HMC, SHMC and S2HMC methods

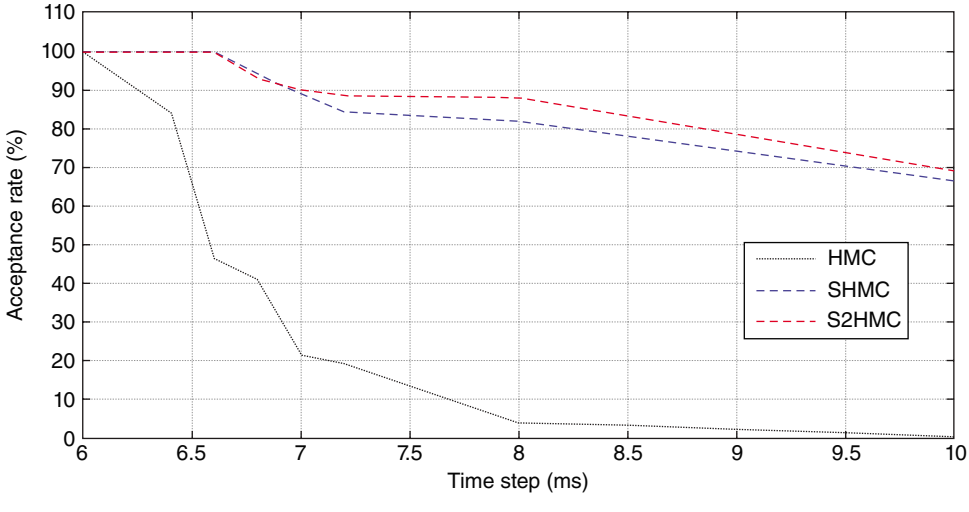

Figure 9.5 The acceptance rate obtained for different time steps using the HMC, SHMC and S2HMC methods

HMC and SHMC acceptance rates. It shows that the S2HMC algorithm extended the upper boundary of the HMC time step. The S2HMC method maintains a good AR when the time step is increased. At a time step of 0.006 s, the S2HMC algorithm has an AR equal to 99.9%. It preserves the same AR value until $\delta t = 0.0066$ s, beyond which the AR of the algorithm starts decreasing to reach an acceptance rate of 69.7% at time step $\delta t = 0.01$ s.

The S2HMC algorithm gives an AR slightly higher than that of the SHMC (66% for the latter). The AR of the HMC algorithm at time step $\delta t = 0.01$ s is less than 1% (no samples were accepted).

Figures 9.4 and 9.5 show that the results of the S2HMC and SHMC algorithms are close in terms of convergence or acceptance rates. However, Table 9.1 indicates that the SHMC has a complicated procedure for obtaining new momentums, where an acceptance–rejection procedure (which depends on the constant c) is needed to accept new momentum vectors. Comparing the MC and MD procedures of the SHMC algorithm when the constant c varies between 0.001 and 1.0, it is observed that c plays a very important role since the AR of both MC and MD steps can be balanced using this constant. The constant c should be small to ensure a good AR in the MD step. At the same time, the value of c should be large in order to ensure a good AR when the momentum is sampled (MC step). The MC acceptance rate started at a very low 28% when $c = 0.001$, since the algorithm needs more time to produce new momentums. It reached 76% when $c = 1.0$. In contrast, the MC acceptance rate started very high at 98.1% when $c = 0.001$ and decreased to 29.6% when $c = 1.2$.

To explore the S2HMC algorithm further, a complex structure is considered in the next section.

9.5 Application 2: GARTEUR SM-AG19 Structure

In this second example, the S2HMC technique is used to update the finite element model of the GARTEUR SM-AG19 structure (Degener and Hermes, 1996; Balmes, 1998; Datta, 2002; Guyon and Elisseeff, 2003; Link and Friswell, 2003; Carvalho *et al.*, 2007). The measured frequencies are 6.38, 16.10, 33.13, 33.53, 35.65, 48.38, 49.43, 55.08, 63.04 and 66.52 Hz. The parameters of the structure to be updated are the right wing moments of inertia and torsional stiffness ($RI_{\min}, RI_{\max}, RI_{\text{tors}}$), the left wing moments of inertia and torsional stiffness ($LI_{\min}, LI_{\max}, LI_{\text{tors}}$), the vertical tail moment of inertia ($VTP_{I_{\min}}$), and the overall structure density ρ.

The mean values of the updated vector and their bounds are given in Tables 9.3 and 9.4 respectively. The time step is set to $\delta t = 3$ ms for the first scenario and $\delta t = 4.8$ ms for the second scenario, L is uniformly distributed on the interval $\{1, 20\}$, and the number of samples is $N_s = 1000$. Each algorithm was run for over 10 independent simulations. The final results presented in Tables 9.4 and 9.5 are the average of these 10 runs. The S2HMC algorithm was implemented to update the finite element model and the results obtained are compared with those obtained by the SHMC and HMC methods. Table 9.4 presents the initial value of the updated vector $\boldsymbol{\theta}$, the updated values and the corresponding c.o.v. obtained by S2HMC method for two scenarios of time steps.

When the time step was $\delta t = 3$ ms, the S2HMC algorithm updated all parameters with reasonable precision, as indicated by the small c.o.v. values for all parameters. The time step used is relatively small, which might mean that the HMC algorithm has some advantages over the SHMC and S2HMC algorithms. This can be seen in the c.o.v. values, which are smaller than for both SHMC and S2HMC. However, the SHMC and S2HMC algorithms give better total average error than the HMC algorithm.

In general, the c.o.v. values are small for the updated parameters when the Hamiltonian algorithms are implemented to update the structure (less than 6% for both the HMC and SHMC algorithms, and lower than 6.5% for the S2HMC algorithm). In the second scenario, when

Table 9.4 Initial and updated parameter values for the HMC, SHMC and S2HMC algorithms at two different time steps

	Initial (mean vector) $\boldsymbol{\theta}_0$	HMC method ($\delta t = 3$ ms)	$\frac{\sigma_i}{\mu_i}$ (%) c.o.v.	SHMC[4] method ($\delta t = 3$ ms)	$\frac{\sigma_i}{\mu_i}$ (%) c.o.v.	S2HMC method ($\delta t = 3$ ms)	$\frac{\sigma_i}{\mu_i}$ (%) c.o.v.	S2HMC method ($\delta t = 4.8$ ms)	$\frac{\sigma_i}{\mu_i}$ (%) c.o.v.
ρ	2785.00	2667.33	1.97	2666.85	2.27	2732.4	2.39	2764.20	2.24
$VTP_{I_{\min}}$	8.34×10^{-9}	6.94×10^{-9}	5.62	6.96×10^{-9}	5.50	7.18×10^{-9}	6.20	7.25×10^{-9}	4.49
$LI_{\min}$	8.34×10^{-9}	10.12×10^{-9}	2.34	10.12×10^{-9}	2.27	10.42×10^{-9}	3.12	10.52×10^{-9}	1.96
$LI_{\max}$	8.34×10^{-7}	7.90×10^{-7}	2.61	7.92×10^{-7}	2.75	8.13×10^{-7}	2.29	8.21×10^{-7}	2.48
$RI_{\min}$	8.34×10^{-9}	10.15×10^{-9}	2.13	10.13×10^{-9}	2.21	10.13×10^{-9}	2.50	10.17×10^{-9}	1.37
$RI_{\max}$	8.34×10^{-7}	6.11×10^{-7}	3.14	6.09×10^{-7}	4.02	6.10×10^{-7}	4.34	6.09×10^{-7}	2.30
LI_{tors}	4.0×10^{-8}	4.04×10^{-8}	1.95	4.04×10^{-8}	2.09	4.04×10^{-8}	1.62	4.04×10^{-8}	1.79
RI_{tors}	4.0×10^{-8}	3.57×10^{-8}	2.17	3.56×10^{-8}	2.49	3.57×10^{-8}	2.09	3.57×10^{-8}	1.80

Table 9.5 Modal results and errors for HMC, SHMC and S2HMC at two different time steps

Measured frequency (Hz)	Initial FEM frequency (Hz)	Frequency HMC method (Hz) $\delta t =$ 3 ms	Frequency SHMC method (Hz) $\delta t = 3$ ms	Frequency S2HMC method (Hz) $\delta t = 3$ ms	Frequency S2HMC method (Hz) $\delta t = 4.8$ ms
6.38	5.7	6.3 (0.95%)	6.31 (1.30%)	06.326 (1.34%)	06.334 (0.92%)
16.10	15.29	15.87 (0.81%)	15.87 (1.38%)	15.995 (0.84%)	16.037 (0.93%)
33.13	32.53	32.24 (0.75%)	32.24 (1.43%)	32.290 (0.72%)	32.307 (0.75%)
33.53	34.95	33.90 (0.76%)	33.88 (1.52%)	33.906 (0.63%)	33.929 (0.75%)
35.65	35.65	35.64 (0.31%)	35.62 (1.45%)	35.609 (0.42%)	35.606 (0.67%)
48.38	45.14	48.84 (0.61%)	48.80 (1.36%)	48.599 (0.71%)	48.434 (0.73%)
49.43	54.69	49.87 (1.45%)	49.86 (1.61%)	49.706 (1.47%)	49.630 (0.94%)
55.08	55.60	54.36 (0.83%)	54.42 (1.47%)	54.684 (0.76%)	54.728 (0.93%)
63.04	60.150	63.888 (0.68%)	63.896 (1.39%)	63.827 (0.77%)	63.773 (0.76%)
66.52	67.560	67.446 (0.03%)	67.447 (1.48%)	67.444 (0.26%)	67.441 (0.54%)

a relatively large time step ($\delta t = 4.8$ ms) is used, the updated parameters obtained by the S2HMC method are much closer to the mean value than those obtained earlier by using a 3 ms time step. In this setting, the HMC method gives poor updating parameters. The same initial values were obtained since no new samples are accepted. The reason is that the Hamiltonian function fluctuated very rapidly with time, causing a sudden decrease in the AR (which dropped to less than 1% when the time step was 4.8 ms). The AR for the S2HMC was 71.3%, which is acceptable compared to that of the HMC method. However, the S2HMC algorithm has an less implementation time than that of the SHMC algorithm. The constant $c = 0.01$ used earlier reduces the SHMC acceptance rate of the MC step to 61.2%. (The SHMC algorithm uses an acceptance–rejection procedure to sample new momentums.) This indicates that the SHMC algorithm takes longer to sample momentums, and more time is needed to run the algorithm. However, the S2HMC uses a similar method to the HMC to sample new momentum vectors (sampling directly from a Gaussian probability distribution function. Then the pre-processing step transfers the momentum to be evaluated to a different search space.

A 4.8 ms time step allows the S2HMC algorithm to give slightly more precise results than those obtained by a 3 ms time step. This can be verified from Table 9.4, where the c.o.v. values obtained by the S2HMC algorithm when $\delta t = 4.8$ ms are smaller than those obtained by the same algorithm when $\delta t = 3$ ms. Also, the c.o.v. values obtained by the S2HMC algorithm are smaller than those obtained using the HMC and SHMC algorithms. This means that the S2HMC algorithm produced more accurate results than those obtained by the other two algorithms when the time step is relatively large (the c.o.v. values are less than 4.5% for S2HMC algorithm, while exceeding 10% for some parameters using the SHMC algorithm). Figures 9.6–9.9 present the histograms of two updating parameters using the S2HMC method. Again, θ_i refers to the sequential numbering of the updating parameters in the updating vector, that is, $\theta_1 = \rho$ (density) and θ_2 refers to $VTP_{I_{\min}}$, while the normalisation constants θ_i^0 are the initial (mean) values of the updated parameters presented in Table 9.4. The obtained results show that the S2HMC algorithm was successfully able to identify the high probability region (the density of the parameters varies in a very small region, especially those of θ_5 and θ_6).

The parameters ρ, $VTP_{I_{\min}}$, $LI_{\min}$, $LI_{\max}$ (shown in Figures 9.6–9.9), LI_{tors} and RI_{tors} (not shown) have forms similar (or close) to normal distribution function forms, while the densities of the rest of the parameters have different forms.

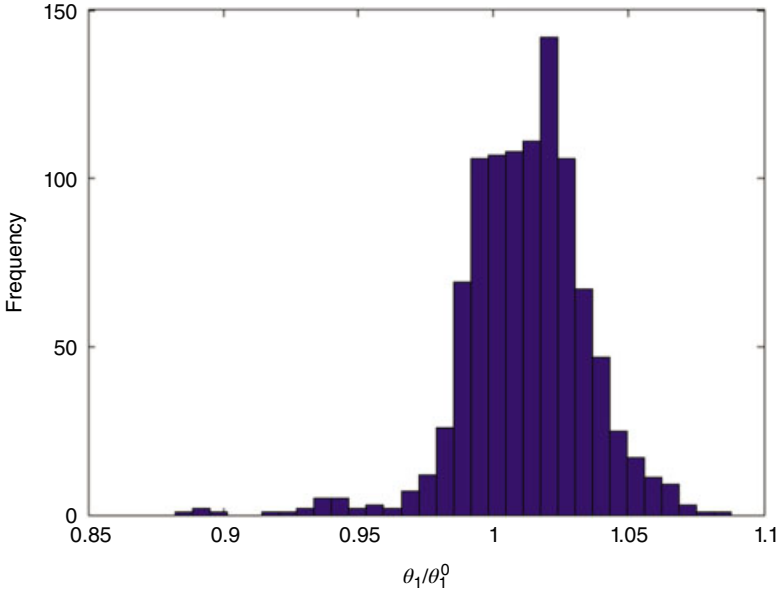

Figure 9.6 Histograms of updating model parameters ρ using the S2HMC method. The normalisation constant ${\theta_1}^0$ is the initial (mean) values

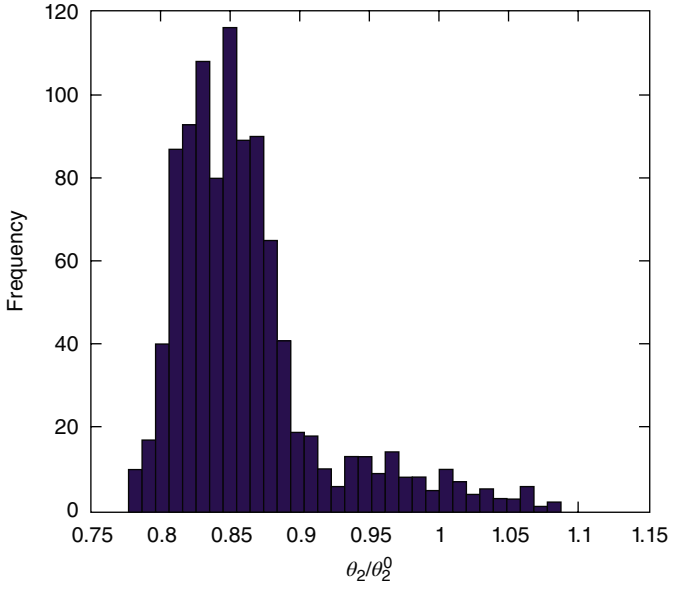

Figure 9.7 Histograms of updating model parameters $VTP_{I_{\min}}$ using the S2HMC method

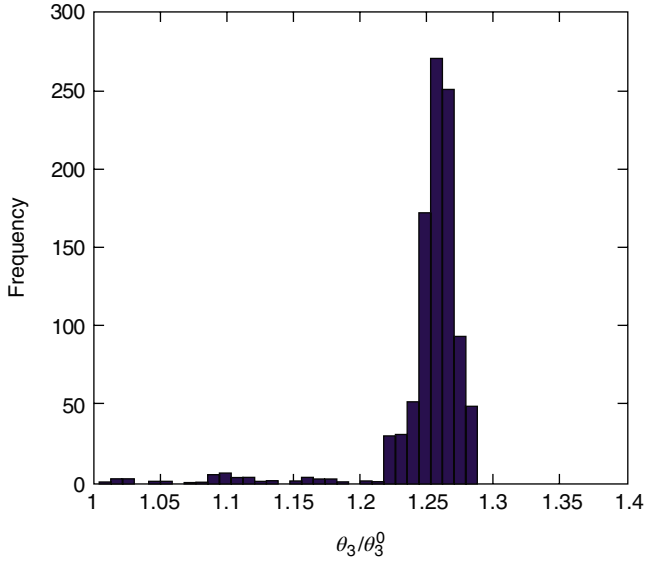

Figure 9.8 Histograms of updating model parameters $LI_{\min}$ using the S2HMC method

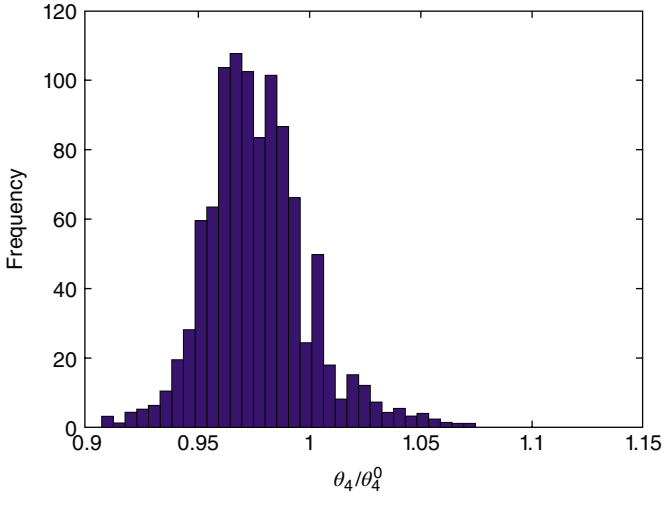

Figure 9.9 Histograms of updating model parameters $LI_{\max}$ using the S2HMC method

Figure 9.10 shows the correlations between all updated parameters for the S2HMC algorithm. All parameters are correlated: $(LI_{\min}, RI_{\min})$, $(LI_{\min}, RI_{\max})$ and $(RI_{\max}, RI_{\min})$ are strongly so, $(\rho, RI_{\max})$, $(\rho, RI_{\min})$ and (ρ, LI_{tors}) weakly so. Table 9.5 contains the updated natural frequencies and output errors when the S2HMC algorithm is employed. The results show

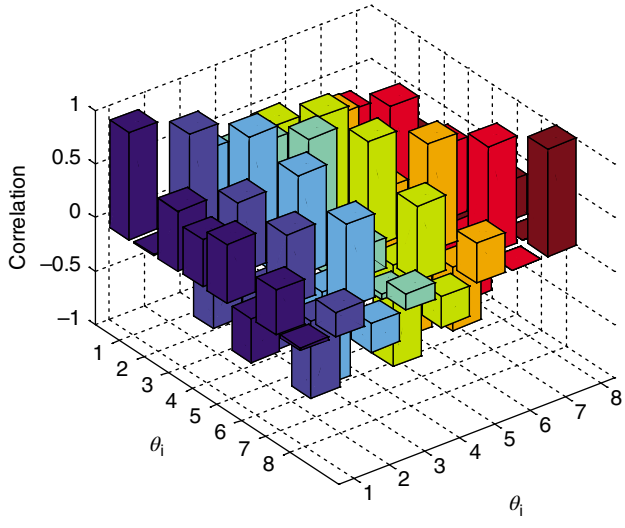

Figure 9.10 The correlation between the updated parameters (S2HMC algorithm)

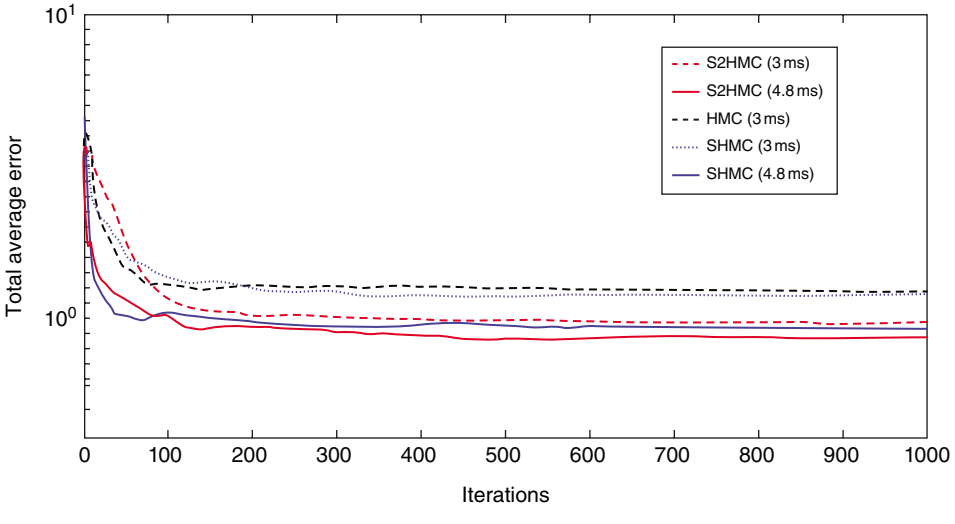

Figure 9.11 The total average error using the HMC, SHMC and S2HMC methods

that the updated finite element model natural frequencies are better than the initial finite element model for an S2HMC algorithm in both time steps. In the case of $\delta t = 3$ ms, the error between the first measured natural frequency and that of the initial model is 10.47%; the S2HMC reduced the error to 0.85%. A similar observation can be made for the fourth, fifth, sixth, eighth and ninth natural frequencies. The initial total average error was 4.6%, but after using the S2HMC method, it fell to 0.964%.

Figure 9.11 shows the total average error (plotted on a log scale) of the S2HMC, SHMC and HMC algorithms for both scenarios and for 1000 iterations. The S2HMC algorithm converges fast and has almost the same convergence rate for both scenarios (the algorithms start

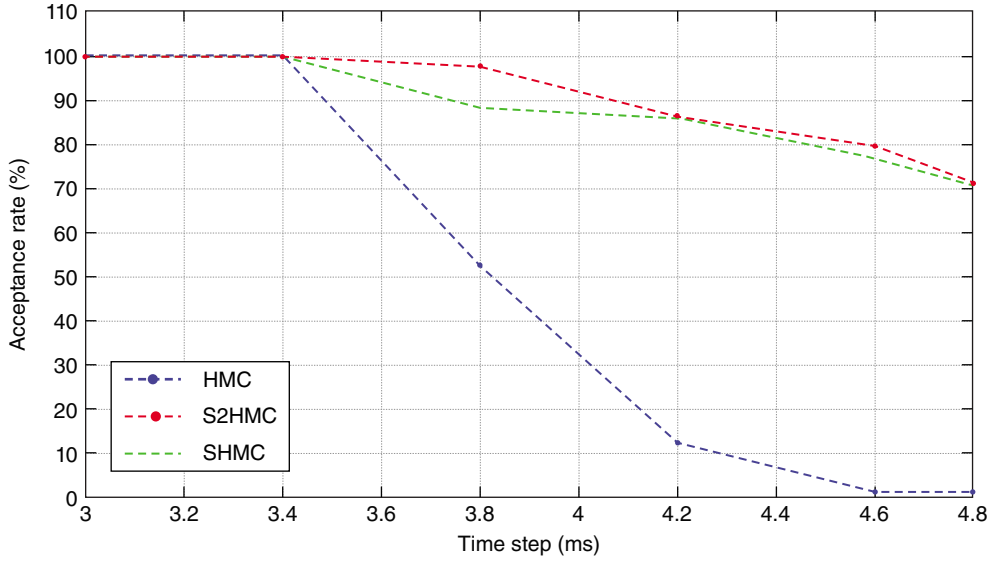

Figure 9.12 The acceptance rate obtained for different time steps using HMC, SHMC and S2HMC methods.

converging in the first 100–150 iterations). Also, the S2HMC algorithm has very similar convergence rate to those of the SHMC and HMC algorithms. The time step, $\delta t = 3$ ms, provides a good sampling AR for the S2HMC algorithm (99.9 %). A different time step for the Hamiltonian algorithms may reduce the sampling AR for these methods. However, when the time step increases the results obtained are significantly affected, as well as the convergence rate, particularly for those algorithms that use the original Hamiltonian function. The time step changes can also provide a good convergence rate for the S2HMC method, since the S2HMC provides samples when the time step is relatively large. In the case where the time step $\delta t = 4.8$ ms (the second scenario), the S2HMC method improves the results and reduces the c.o.v. values, which indicates that the results are more accurate than the first scenario. This can be seen in Table 9.5 where the total average error is reduced to 0.86% with an AR of 71.3%. However, this is not the case for HMC, where the acceptance rate was found to be less than 1 and the updated vector obtained from the HMC does not improve the finite element modelling results.

Figure 9.12 shows the AR when the time step varies between 3 and 4.8 ms. The AR for the S2HMC algorithm is 99.9% when the time step is 3 ms. It starts decreasing when the time step increases, but this decrease is faster and more significant in the case of the HMC method. When the time step is 3.4 ms, the AR for the S2HMC is 99.9%, reducing slightly to 97.8% when the time step reaches 3.8 ms. Finally, when the time step is 4.8 ms, the AR for the S2HMC reduces to 71.3%, while SHMC algorithm has an AR equal to 70.8%, which is acceptable compared to that obtained by the HMC method (less than 1%).

9.6 Conclusions

This chapter has illustrated the use of the separable shadow hybrid Monte Carlo algorithm in finite element model updating problems. The S2HMC technique has the basics of the SHMC technique – both methods use a modified Hamiltonian function. However, the Hamiltonian

function used by the S2HMC algorithm is a separable modified function which is easy to program, and the sampling procedure of the momentum vector is similar to the basic HMC algorithm. Two different real structures, with the same set of updating parameters introduced, are used as test cases for this sampling technique. The structures used are the asymmetrical H-shaped aluminium structure and the GARTEUR SM-AG19 structure. The S2HMC algorithm is more efficient than both HMC and SHMC algorithm – the S2HMC has a simple procedure to sample new momentums which the SHMC algorithm does not. Also, it can provide samples with large time step. The S2HMC algorithm has the ability to produce samples with relatively large time and without using extra parameters. In both implementations, the S2HMC method gave better results than both the SHMC and HMC for the examples in different scenarios when the time step is equal to 3 and 4.8 ms, respectively.

References

Alfaki M (2008) Improving efficiency in parameter estimation using the Hamiltonian Monte Carlo algorithm. PhD thesis, University of Bergen.

Balmes E (1998) Predicted variability and differences between tests of a single structure. Proceedings of the 16th International Modal Analysis Conference, Bethel, CT: SPIE, pp. 558–564.

Beskos A, Pillai N, Roberts G, Sanz-Serna JM, Stuart A (2013) Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli* **19**: 1501–1534.

Blanes S, Casas S, Murua A (2004) On the numerical integration of ordinary differential equations by processed methods. *SIAM Journal of Numerical Analysis* **42**: 531–552.

Boulkaibet I (2014) Finite element model updating using the Markov chain Monte Carlo technique. PhD thesis, University of Johannesburg.

Boulkaibet I, Marwala T, Mthembu L, Friswell MI,Adhikari S (2012) Sampling techniques in Bayesian finite element model updating. *Proceedings of the Society for Experimental Mechanics* **29**: 75–83.

Boulkaibet I, Mthembu L, Marwala T, Friswell MI, Adhikari S (2014) Finite element model updating using the shadow hybrid Monte Carlo technique. *Mechanical System & Signal Processing* **52**: 115–132.

Boulkaibet I, Mthembu L, Marwala T, Friswell MI, Adhikari S (2015) Finite element model updating using Hamiltonian Monte Carlo techniques. *Inverse Problems in Science and Engineering (IPSE)*, submitted for publication.

Carvalho J, Datta BN, Gupta A, Lagadapati M (2007) A direct method for model updating with incomplete measured data and without spurious modes. *Mechanical Systems and Signal Processing* **21**: 2715–2731.

Datta BN (2002) Finite element model updating, eigenstructure assignment and eigenvalue embedding techniques for vibrating systems. *Mechanical Systems and Signal Processing* **16**: 83–96.

Degener M, Hermes M (1996) Ground vibration test and finite element analysis of the GARTEUR SM-AG19 testbed. Report IB 232-96J 08, Deutsche Forschungsanstalt für Luft und Raumfahrt e.V. Institut für Aeroelastik.

Duane S, Kennedy AD, Pendleton BJ,Roweth D (1987) Hybrid Monte Carlo. *Physics Letters* **B195**: 216–222.

Engle RD, Skeel RD, Drees M (2005) Monitoring energy drift with shadow Hamiltonians. *Journal of Computational Physics* **206**: 432–452.

Gupta R, Kilcup GW, Sharpe SR (1988) Tuning the hybrid Monte Carlo algorithm. *Physical Review* **D38**: 1278.

Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *Journal of Machine Learning Research* **3**: 1157–1182.

Hairer E, Lubich C, Wanner G (2002) *Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Berlin: Springer-Verlag.

Hanson KM (2001) Markov chain Monte Carlo posterior sampling with the Hamiltonian method. *Proceedings of SPIE* **4322**: 456–467.

Leimkuhler B and Reich S (2004) *Simulating Molecular Dynamics*. Cambridge: Cambridge University Press.

Link M, Friswell MI (2003) Generation of validated structural dynamic models – results of a benchmark study utilizing the GARTEUR SM-AG19 testbed. *Mechanical Systems & Signal Processing* **17**: 9–20.

Marwala T (1997) A multiple criterion updating method for damage detection on structures. MEng thesis, University of Pretoria.

Marwala T (2010) *Finite Element Model Updating Using Computational Intelligence Techniques*. London: Springer-Verlag.

Neal RM (1998) Suppressing random walks in Markov chain Monte Carlo using ordered over-relaxation. In Jordan MI (ed.) *Learning in Graphical Models*. Dordrecht: Kluwer Academic Publishers, pp. 205–228.

Skeel RD,Hardy DJ (2001) Practical construction of modified Hamiltonians. *SIAM Journal of Scientific Computing* **23**: 1172–1188.

Sweet CR, Hampton SS, Skeel RD, Izaguirre JA (2009) A separable shadow Hamiltonian hybrid Monte Carlo method. *Journal of Chemical Physics* **131**, article no. 174106.

Sweet CR, Izaguirre JA (2006). Backward error analysis of multiscale symplectic integrators and propagators. Proceedings of the Third International Conference Multiscale Materials Modeling MMM2006.

Welling M and Teh YW (2011). Bayesian learning via stochastic gradient Langevin dynamics. Proceedings of the 28th International Conference on Machine Learning, pp. 681–688.

10

Evolutionary Approach to Finite Element Model Updating

10.1 Introduction

Finite element modelling is a computational estimation of a real system by breaking the system into substructures called elements, and the process of connecting these elements (Marwala, 2010; Friswell and Mottershead, 1995). This chapter is essentially about constructing mathematical models based on evolutionary Markov chain Monte Carlo (MCMC) to estimate the parameters that are uncertain in a finite element model (Boulkaibet *et al.*, 2015a). If it is apparent which variables are uncertain, one tactic is to describe their probability distributions. This is conceivable for finite element models of real systems for the reason that the uncertain variable values have to be of realistic magnitudes. The sampling methods can be used to compute these uncertain parameters. Nevertheless, the difficulty is that – occasionally – the distribution of these uncertain variables is not known *a priori,* and then the updated finite element models have multiple optimal solutions.

One of the sampling methods is the MCMC (Bishop, 2006; Marwala, 2010; Boulkaibet *et al.*, 2012). MCMC methods have the benefit that the sampling processes draw samples with an element of randomness, even though being steered by their performance on the problem objective function also known as the posterior function. Consequently, some samples are rejected while others are accepted, and this essentially creates a chain of samples. The learning part of the sampling technique uses the paradigm of evolution-based algorithms (Holland, 1975; Mitchell, 1996; Goldberg, 1989; Hukushima and Nemoto, 1996).

These procedures are based on the concept that, in evolution, new individuals are an enhancement based on their parents' performance in a precise problem. Accordingly, new individuals can be created if we can model the variables that gave advantage to their parents. The

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

similarity of individuals in evolution algorithms to the finite element model context is the particular combination of uncertain variable values that constitute a possibly correct solution to the finite element model problem. Therefore, individuals in evolution procedures are all the probable solutions to the finite element model. The individual's performance on the problem is assessed by the fitness function and, in this chapter, the fitness function is defined by the logarithm of the posterior density distribution function.

There are two basic instruments for developing new individuals in evolution algorithms: mutation and crossover. Mutation is when a variable, randomly chosen, alters its value. The crossover is the procedure where two individuals interchange their variable values at certain positions, also called crossover points, in the uncertain variable vector. The choice of the locations can either be conducted using methods such as roulette wheel selection or random selection (Davis, 1991; Liang and Wong, 2000). The algorithm described in this chapter, proposed by Boulkaibet *et al.* (2015a), has an extra operator called the exchange which permits the mixing of individuals within a population to eradicate early convergence. With the intention of characterising the uncertainties of the uncertain variables, the problem can be formulated in the Bayesian context. In this chapter, Bayesian statistics, the MCMC, evolutionary Markov chain Monte Carlo (EMCMC) algorithms and the application of the EMCMC on an asymmetrical H-shaped structure are studied.

10.2 The Bayesian Formulation

The EMCMC described in this chapter is essentially used to solve the finite element updating problem, which is formulated in the Bayesian framework (Marwala and Sibisi, 2005). The Bayesian method of solving the finite element updating problem is based on Bayes' rule, proposed by Reverend Thomas Bayes. Bayes' rule has been applied extensively in applied mechanics by Marwala and Sibisi (2005), Marwala (2010), Sundar and Manohar (2013), Yan (2012, 2014) and Yan *et al*. (2015). It is expressed as follows (Bishop, 1995; Cheung and Beck, 2009; Yuen, 2010; Boulkaibet *et al.*, 2015a, 2015b):

$$P(\boldsymbol{\theta}|\mathcal{D},\boldsymbol{\mathcal{M}}) \propto P(\mathcal{D}|\boldsymbol{\theta},\boldsymbol{\mathcal{M}})P(\boldsymbol{\theta}|\boldsymbol{\mathcal{M}}), \tag{10.1}$$

where $\boldsymbol{\mathcal{M}}$ denotes the model type for the target model, which is described by the parameters of the updated model $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^d$; $\mathcal{D}$ is the measured data, in this chapter the natural frequencies f_i^m and mode shapes ϕ_i^m; and the expression $P(\boldsymbol{\theta}|\boldsymbol{\mathcal{M}})$ is the prior probability distribution function. The prior probability function indicates prior knowledge of the uncertain variables given a model without the knowledge of the measured data $\mathcal{D}$. The expression is the posterior density distribution function of the variables given the observed data and the presumed model type is the likelihood function (Marwala, 2010; Marwala and Sibisi, 2005; Cheung and Beck, 2009), which measures the difference between the measured data and the finite element model results for given variables and assumed model. The dependence on $\boldsymbol{\mathcal{M}}$ is only relevant in cases where more than one model type is studied. In this chapter, only one model is studied, and the term $\boldsymbol{\mathcal{M}}$ is ignored.

In the finite element model setting, the likelihood density distribution function is given by (Cheung and Beck, 2009; Boulkaibet *et al.*, 2015a):

$$P(\mathcal{D}|\boldsymbol{\theta}) = \frac{1}{(2\pi/\beta_c)^{N_m/2}\prod_{i=1}^{N_m} f_i^m}\exp\left(-\frac{\beta_c}{2}\sum_{i}^{N_m}\left(\frac{f_i^m - f_i}{f_i^m}\right)^2\right), \tag{10.2}$$

where β_c is a constant, N_m is the number of measured modes, and f_i^m and f_i are respectively the ith analytical natural frequency and the ith measured natural frequency.

The prior density distribution function is the prior knowledge of the updating variables $\boldsymbol{\theta}$ and it is chosen to be a Gaussian distribution (Cheung and Beck, 2009):

$$\begin{aligned} P(\boldsymbol{\theta}) &= \frac{1}{(2\pi)^{Q/2}\prod_{i=1}^{Q}\left(1/\sqrt{\alpha_i}\right)}\exp\left(-\sum_{i=1}^{Q}\frac{\alpha_i}{2}\left\|\boldsymbol{\theta}^i - \boldsymbol{\theta}_0^i\right\|^2\right) \\ &= \frac{1}{(2\pi)^{Q/2}\prod_{i=1}^{Q}\left(1/\sqrt{\alpha_i}\right)}\exp\left(-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right), \end{aligned} \tag{10.3}$$

where Q is the number of variables to be updated, $\boldsymbol{\theta}_0$ represents the mean value of the updated vector, α_i is the coefficient of the prior density distribution function for the ith updating variables and $\|.\|$ denotes the Euclidean norm.

The posterior density distribution function of the variables $\boldsymbol{\theta}$ given the observed data $\mathcal{D}$ is written as $P(\boldsymbol{\theta}|\mathcal{D})$ and is obtained by applying Bayes' theorem in Equation 10.1. The distribution $P(\boldsymbol{\theta}|\mathcal{D})$ is estimated by substituting Equations 10.2 and 10.3 into Equation 10.1 to give (Cheung and Beck, 2009)

$$P(\boldsymbol{\theta}|\mathcal{D}) \propto \frac{1}{Z_s(\alpha,\beta_c)}\exp\left(-\frac{\beta_c}{2}\sum_{i}^{N_m}\left(\frac{f_i^m - f_i}{f_i^m}\right)^2 - \sum_{i}^{Q}\frac{\alpha_i}{2}\left\|\boldsymbol{\theta}^i - \boldsymbol{\theta}_0^i\right\|^2\right), \tag{10.4}$$

where

$$Z_s(\alpha,\beta_c) = \left(\frac{2\pi}{\beta_c}\right)^{N_m/2}\prod_{i=1}^{N_m} f_i^m (2\pi)^{Q/2}\prod_{i=1}^{Q}\frac{1}{\sqrt{\alpha_i}}. \tag{10.5}$$

In systems with a large number of uncertain variables, calculating a posterior density distribution function analytically is impossible. Sampling methods such as MCMC can give a numerical estimation of this function (Cheung and Beck, 2009; Ching and Leu, 2009).

If Y is a sample of specific variables at different discrete time periods, then the estimation of the future responses of this parameter Y at another time period can be obtained by using the total probability theorem (Cheung and Beck, 2009):

$$P(Y|\mathcal{D}) = \int_{\boldsymbol{\theta}} P(Y|\boldsymbol{\theta})P(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}. \tag{10.6}$$

Equation 10.6 depends on the posterior density distribution function and consequently, given a set of N_s random variable vectors sampled from a probability distribution function $P(\boldsymbol{\theta}|\mathcal{D})$, the expected value of any observed function Y can be approximated easily.

The integral in Equation 10.6 can be calculated using sampling techniques (Bishop, 2006; Marwala, 2010; Boulkaibet *et al.*, 2015a, 2015b). These methods are applied to estimate a series of vectors $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{N_s}\}$ that can be applied to create a Markov chain. This approximated vector is then applied to approximate the expression of the posterior distribution function $P(\boldsymbol{\theta}|\mathcal{D})$ and the integral in Equation 10.6 can written as (Cheung and Beck, 2009)

$$\tilde{Y} \cong \frac{1}{N_s} \sum_{i=1}^{N_s} G(\boldsymbol{\theta_i}), \tag{10.7}$$

where G is a function that depends on the updated variables $\boldsymbol{\theta}_i$. For instance, if $G(\boldsymbol{\theta}) = \boldsymbol{\theta}$ then $\tilde{Y}$ equals the expected value of $\boldsymbol{\theta}$. Normally, $\tilde{Y}$ is the vector that has the measured data and N_s is the number of retained states. In this chapter, the EMCMC technique is used to sample from the posterior density distribution function.

10.3 The Evolutionary MCMC Algorithm

The EMCMC method combines evolutionary programming techniques with the MCMC method (Casanellas and Kedzierska, 2013; Liang and Wong, 2000, 2001; Geyer, 1991). Theofilatos *et al.* (2015) proposed an evolutionary based Markov clustering for predicting protein complexes, while Ponce-de-Leon-Senti and Diaz-Diaz (2012) proposed an evolutionary method based on Gibbs sampling and the Markov model. Goh *et al.* (2012) successfully applied an evolutionary hidden Markov model for medical image analysis, while Ruiz-Cárdenas *et al.* (2012) successfully applied EMCMC algorithms to monitor network designs.

In order to understand the EMCMC technique, it is important to first understand the MCMC. The MCMC method produces a chain of states through a random walk process which obeys the Markov process. Each sample depends only on the previous sample and nothing else. The MCMC is a hybrid of two methods: the Markov process and Monte Carlo simulation (Markov, 1971; Doob, 1953; Malve *et al.*, 2007; Liesenfeld and Richard, 2008; Jing and Vadakkepat, 2010).

In the MCMC technique a system is considered which is represented by a stochastic process comprising random variables $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_i\}$ where a variable $\boldsymbol{x}_i$ inhabits a state $\mathbf{x}$ at discrete time i. If the probability that the system occupies state $\boldsymbol{x}_{i+1}$ at time $i+1$ depends entirely on the state $\boldsymbol{x}_i$ at time i, then the random variables $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_i\}$ form a Markov chain. The transition between states in the MCMC is attained by adding a random component ($\boldsymbol{\varepsilon}$) to the present state as follows (Bishop, 2006; Marwala, 2010):

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \boldsymbol{\varepsilon}. \tag{10.8}$$

The current state is either accepted or rejected, normally using the Metropolis algorithm (Metropolis *et al.*, 1953; Wang *et al.*, 2015). There are several types of Metropolis algorithms, including a generalised multiple-point Metropolis (Kobayashi and Kozumi, 2015), an adaptive Metropolis algorithm (Mbalawata *et al.*, 2015), a pseudo-marginal random walk Metropolis algorithm (Sherlock *et al.*, 2015), a parallel Metropolis–Hastings (M-H) algorithm (Calderhead, 2014) and a differential evolution adaptive Metropolis algorithm (Zhou *et al.*, 2015).

When applying the Metropolis algorithm for sampling a stochastic process with random variables $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_i\}$, random changes to $\boldsymbol{x}$ are evaluated and are either accepted or rejected according to the following criterion (Metropolis *et al.*, 1953):

$$\begin{cases} \text{If } F_{\text{new}} < F_{\text{old}} \text{ accept state}(\boldsymbol{x}_{i+1}) \\ \text{Else} \\ \text{Accept state}(\boldsymbol{x}_{i+1}) \text{ with probability } \exp\{-(F_{\text{new}} - F_{\text{old}})\}. \end{cases} \tag{10.9}$$

Here F_{new} and F_{old} are the fitness function (posterior distribution function) values corresponding to $\boldsymbol{x}_{i+1}$ and $\boldsymbol{x}_i$, respectively.

The other aspect of the EMCMC is genetic programming. The genetic algorithm is based on the theory of evolution and consists of two operators: crossover and mutation (Chatterjee *et al.*, 1996). The genetic algorithm has been applied successfully to optimise a laser-mixing scheme (Kou *et al.*, 2015), for virtual topology design (Din, 2015), for moving object detection (Lee *et al.*, 2015), for capacitor allocation (Gholami *et al.*, 2015), to evaluate discharge (Dunca *et al.*, 2015), for container packing (Feng *et al.*, 2015), for finite element model updating (Boulkaibet *et al.*, 2015c) and in decision support (Bukharov and Bogolyubov, 2015).

The EMCMC algorithm combines the genetic algorithm operators with the dynamics of MCMC algorithms for sampling and learning. In this section, the basics of the EMCMC methods are described.

Let $\boldsymbol{\theta} = \{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^i, \ldots, \boldsymbol{\theta}^N\}$ denote a population of samples, where $\boldsymbol{\theta}^i = \{\boldsymbol{\theta}_1^i, \boldsymbol{\theta}_2^i, \ldots, \boldsymbol{\theta}_d^i\}$ is a d-dimensional vector called an individual, or a chromosome in the genetic algorithm, and N is the population size. In Bayesian statistics, $\boldsymbol{\theta}^i$ is a vector of variables, while the fitness function $F(\boldsymbol{\theta}^i)$ is the negative of the log-posterior of $\boldsymbol{\theta}^i$. In EMCMC, a different temperature T_i is attached to each individual $\boldsymbol{\theta}^i$, and the temperatures form a schedule with the ordering $T_1 > T_1 > T_2 > \ldots > T_N = 1$. The concept of temperature is derived from annealing, which is a natural process in which objects are heated and then cooled in such a way that the system achieves a stable orientation. This physical process has been used to design an optimisation technique called simulated annealing, and this technique has been successfully applied to optimise complex processes such as a window assembly line (García-Villoria *et al.*, 2015), for single-row equidistant facility layout (Palubeckis, 2015), solving a multi-objective facility layout problem (Matai, 2015) and optimal placement of synchrophasor measurement units in smart power grids (Gopakumar *et al.*, 2015). A Boltzmann distribution for each individual $\boldsymbol{\theta}^i$ can be defined as (Li, 2015; Mo *et al.*, 2015)

$$Z_i(T_i) = \sum\nolimits_{\{\boldsymbol{\theta}^i\}} \exp\left(-\frac{F(\boldsymbol{\theta}^i)}{T_i}\right) g(\boldsymbol{\theta}^i) = \frac{1}{Z_i(T_i)} \exp\left(-\frac{F(\boldsymbol{\theta}^i)}{T_i}\right), \tag{10.10}$$

where $Z_i(T_i)$ is the normalising constant.

10.3.1 Mutation

A mutation operator selects a binary digit of a chromosome at random and changes it. This introduces new information into the population, and thus prevents the genetic algorithm

simulation from converging to a local optimum solution. Mutation happens with a particular probability, and in numerous physical systems the probability of mutation is very low. An example of mutation is binary mutation (Goldberg, 1989). When binary mutation is applied, a number expressed in binary format is selected and one bit value is changed. For instance, the chromosome 11001011 is mutated to the chromosome 11000011.

Another example of mutation is non-uniform mutation, which operates by increasing the probability of mutation so that is the amount of mutation will be close to 0 as the generation number increases. This inhibits the population from stagnating in the initial stages of the evolution process, and then allows the algorithm to improve the solution in the terminal stages of the evolution.

Ulivi *et al.* (2015) applied gene mutation analysis to patient selection, while Diaz-Uriarte (2015) applied mutation analysis to study tumour progression. Ok *et al.* (2015) applied TP53 mutation in therapy-related myelodysplastic syndromes and acute myeloid leukaemia.

In the mutation operator, an individual – say $\boldsymbol{\theta}^k$ – is randomly chosen from the present population $\boldsymbol{\theta} = \left\{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^k, \ldots, \boldsymbol{\theta}^N\right\}$. It is then mutated to form a new individual $\widetilde{\boldsymbol{\theta}}^k$ by altering the values of some bits which are also selected randomly. A new population $\widetilde{\boldsymbol{\theta}} = \left\{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \widetilde{\boldsymbol{\theta}}^k, \ldots, \boldsymbol{\theta}^N\right\}$ is therefore formed, and it is accepted with probability $\min(1, r_m)$ using the Metropolis criterion, where r_m is the M-H ratio (Metropolis *et al.*, 1953; Hastings, 1970) given by (Liang and Wong, 2000)

$$r_m = \frac{g\left(\widetilde{\boldsymbol{\theta}}^k\right)}{g\left(\boldsymbol{\theta}^k\right)} \cdot \frac{\mathrm{Tr}\left(\boldsymbol{\theta}^k \middle| \widetilde{\boldsymbol{\theta}}^k\right)}{\mathrm{Tr}\left(\widetilde{\boldsymbol{\theta}}^k \middle| \boldsymbol{\theta}^k\right)} = \exp\left(-\frac{\left(F\left(\widetilde{\boldsymbol{\theta}}^k\right) - F\left(\boldsymbol{\theta}^k\right)\right)}{T_k}\right) \cdot \frac{\mathrm{Tr}\left(\boldsymbol{\theta}^k \middle| \widetilde{\boldsymbol{\theta}}^k\right)}{\mathrm{Tr}\left(\widetilde{\boldsymbol{\theta}}^k \middle| \boldsymbol{\theta}^k\right)} \tag{10.11}$$

where $\mathrm{Tr}(\cdot|\cdot)$ indicates the transition probability between populations. If the proposal is accepted, the current population $\boldsymbol{\theta}$ is replaced by $\widetilde{\boldsymbol{\theta}}$, otherwise the population $\boldsymbol{\theta}$ is not changed. This paper uses the two-point mutation operator which indicates that $\mathrm{Tr}\left(\boldsymbol{\theta}^k \middle| \widetilde{\boldsymbol{\theta}}^k\right) = \mathrm{Tr}\left(\widetilde{\boldsymbol{\theta}}^k \middle| \boldsymbol{\theta}^k\right)$.

10.3.2 Crossover

In the genetic algorithm, the crossover operator fuses genetic information in the population by cutting pairs of chromosomes at random points along their length and exchanging the cut sections over. This combines successful operators together. Crossover is essentially an algorithmic operator used to change the programming of a potential solution. Crossover is executed with a particular probability and, in many instances, the probability of crossover happening is higher than the probability of mutation occurring (Goldberg, 1989). An example of crossover is the simple crossover where one crossover point is chosen and a binary string from the beginning of a chromosome to the crossover point is duplicated from one parent and the rest is duplicated from the second parent. For example, if two chromosomes in binary space $a = 11101011$ and $b = 11111111$ undergo a one-point crossover at the midpoint, then the resulting offspring is $c = 11101111$. Another example is the arithmetic crossover, which is a mathematical operator

which operates by combining two solutions $a = 11001011$ and $b = 11011111$ to form an offspring $c = 11001011$. There are many other types of crossover, including genetic crossover (Sakae *et al.*, 2015) and multi-parent crossover (Moin *et al.*, 2015). Other studies on crossover and its applications are by Šprogar (2015), Rowan *et al.* (2015), and Vannucci and Colla (2015).

As described by Boulkaibet *et al.* (2015a), in the crossover operator, two individuals – say $\boldsymbol{\theta}^a$ and $\boldsymbol{\theta}^b (a \neq b)$ – are chosen from the current population $\boldsymbol{\theta}$ using some selection process (Liang and Wong, 2000). If we assume $F(\boldsymbol{\theta}^a) \geq F\left(\boldsymbol{\theta}^b\right)$ then from these two individuals, known as parents, two new individuals are created in accordance with the crossover operator, and this is done by randomly choosing a crossover point in the uncertain variable vector. For example, if position p is chosen to be the crossover position in the size d element variable vector, then the elements from position $p+1$ in $\boldsymbol{\theta}^a$ will be moved to individual $\boldsymbol{\theta}^b$ and vice versa. The offspring with smaller fitness value is denoted by $\widetilde{\boldsymbol{\theta}}^b$ and the other by $\widetilde{\boldsymbol{\theta}}^a$. A new population of individuals then becomes $\widetilde{\boldsymbol{\theta}} = \left\{\boldsymbol{\theta}^1, \cdots, \widetilde{\boldsymbol{\theta}}^a, \cdots, \widetilde{\boldsymbol{\theta}}^b, \cdots, \boldsymbol{\theta}^N\right\}$. According to the Metropolis algorithm, the new population is accepted with probability $\min(1, r_c)$ (Liang and Wong, 2000):

$$r_c = \frac{g\left(\widetilde{\boldsymbol{\theta}}^a\right).g\left(\widetilde{\boldsymbol{\theta}}^b\right)}{g(\boldsymbol{\theta}^a).g\left(\widetilde{\boldsymbol{\theta}}^b\right)} \cdot \frac{\mathrm{Tr}\left(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}}\right)}{\mathrm{Tr}\left(\widetilde{\boldsymbol{\theta}}|\boldsymbol{\theta}\right)} = \exp\left(-\frac{\left(F\left(\widetilde{\boldsymbol{\theta}}^a\right) - F(\boldsymbol{\theta}^a)\right)}{T_a} - \frac{\left(F\left(\widetilde{\boldsymbol{\theta}}^a\right) - F\left(\boldsymbol{\theta}^b\right)\right)}{T_b} \right) \cdot \frac{\mathrm{Tr}\left(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}}\right)}{\mathrm{Tr}\left(\widetilde{\boldsymbol{\theta}}|\boldsymbol{\theta}\right)}, \tag{10.12}$$

where $\mathrm{Tr}\left(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}}\right) = P\left(\left(\boldsymbol{\theta}^a, \boldsymbol{\theta}^b\right)|\boldsymbol{\theta}\right) \cdot P\left(\left(\widetilde{\boldsymbol{\theta}}^a, \widetilde{\boldsymbol{\theta}}^b\right)|\left(\boldsymbol{\theta}^a, \boldsymbol{\theta}^b\right)\right)$ in which $P\left(\left(\boldsymbol{\theta}^a, \boldsymbol{\theta}^b\right)|\boldsymbol{\theta}\right)$ denotes the selection probability of individuals $(\boldsymbol{\theta}^a, \boldsymbol{\theta}^b)$ from the population $\boldsymbol{\theta}$, and $P\left(\left(\widetilde{\boldsymbol{\theta}}^a, \widetilde{\boldsymbol{\theta}}^b\right)|\left(\boldsymbol{\theta}^a, \boldsymbol{\theta}^b\right)\right)$ denotes the generating probability of individuals $\left(\widetilde{\boldsymbol{\theta}}^a, \widetilde{\boldsymbol{\theta}}^b\right)$ from the parents $(\boldsymbol{\theta}^a, \boldsymbol{\theta}^b)$.

Boulkaibet *et al.* (2015a) chose the parental individuals as follows. The first individual $\boldsymbol{\theta}^a$ is selected according to a roulette wheel technique with Boltzmann weights, meaning that $\boldsymbol{\theta}^a$ is selected with probability given by Equation 10.10. The second individual $\boldsymbol{\theta}^b$ is selected randomly from the rest of the population. The selection probability of $\boldsymbol{\theta}^b$ is then (Liang and Wong, 2000)

$$P\left(\left(\boldsymbol{\theta}^a, \boldsymbol{\theta}^b\right)|\boldsymbol{\theta}\right) = \frac{1}{(N-1) \cdot Z(\theta)} \cdot \left(\exp\left(-\frac{H(\boldsymbol{\theta}^a)}{T_a} \right) + \exp\left(-\frac{H\left(\boldsymbol{\theta}^b\right)}{T_b} \right) \right), \tag{10.13}$$

where

$$Z(\boldsymbol{\theta}) = \sum_{\{\boldsymbol{\theta}^a\}} \exp\left(-\frac{H(\boldsymbol{\theta}^a)}{T_a} \right) P\left(\left(\widetilde{\boldsymbol{\theta}}^a, \widetilde{\boldsymbol{\theta}}^b\right)|\widetilde{\boldsymbol{\theta}}\right).$$

The crossover operator used is a two-point crossover.

10.3.3 Exchange

This operation was first proposed in the parallel tempering sampling technique by Geyer (1991) and in the exchange Monte Carlo sampling technique by Hukushima and Nemoto (1996). Using the current population $\boldsymbol{\theta}$ and the temperature schedule $\mathbf{T}$, where $(\boldsymbol{\theta},\boldsymbol{T}) = \{\boldsymbol{\theta}^1, T_1, \boldsymbol{\theta}^2, T_2, \ldots, \boldsymbol{\theta}^N, T_N\}$, an exchange between individuals $\boldsymbol{\theta}^a$ and $\boldsymbol{\theta}^b$ can be made without changing their temperatures, meaning we attempt to change $(\boldsymbol{\theta},\boldsymbol{T}) = \{\boldsymbol{\theta}^1, T_1, \boldsymbol{\theta}^a, T_a, \ldots, \boldsymbol{\theta}^b, T_b, \ldots, \boldsymbol{\theta}^N, T_N\}$ to $(\overline{\boldsymbol{\theta}},\boldsymbol{T}) = \{\boldsymbol{\theta}^1, T_1, \boldsymbol{\theta}^b, T_a, \ldots, \boldsymbol{\theta}^a, T_b, \ldots, \boldsymbol{\theta}^N, T_N\}$. The new population is accepted with probability $\min(1, r_e)$ using the Metropolis criterion, where (Liang and Wong, 2000)

$$r_e = \frac{g(\overline{\boldsymbol{\theta}})}{g(\boldsymbol{\theta})} \cdot \frac{\mathrm{Tr}(\boldsymbol{\theta}|\overline{\boldsymbol{\theta}})}{\mathrm{Tr}(\overline{\boldsymbol{\theta}}|\boldsymbol{\theta})} = \exp\left(-\left(H(\boldsymbol{\theta}^a) - H(\boldsymbol{\theta}^b)\right)\left(\frac{1}{T_a} - \frac{1}{T_b}\right)\right) \frac{\mathrm{Tr}(\boldsymbol{\theta}|\overline{\boldsymbol{\theta}})}{\mathrm{Tr}(\overline{\boldsymbol{\theta}}|\boldsymbol{\theta})} \tag{10.14}$$

and $\mathrm{Tr}(.|.)$ denotes the transition probability between populations. Normally, the exchange is only performed on states with neighbouring temperature values, that is, $|a-b| = 1$ where $\mathrm{Tr}(\boldsymbol{\theta}|\overline{\boldsymbol{\theta}}) = \mathrm{Tr}(\overline{\boldsymbol{\theta}}|\boldsymbol{\theta})$ (Liang and Wong, 2000). Thus the EMCMC algorithm is summarised as follows (Boulkaibet *et al.*, 2015a):

1. *Initialisation*. Create N individuals, initiate the temperature vector $\mathbf{T}$, and calculate the posterior (fitness) of each individual.
2. *Selection*. Choose particular individuals from the current population.
3. *Crossover*. Generate offspring by a recombination of the vector elements in the mating individuals, where this operation is accepted by $\min(1, r_c)$.
4. *Mutation*. Generate offspring by random changes in individuals in the population, where these offspring are accepted by $\min(1, r_m)$.
5. *Exchange*. Perform the exchange for each set of neighbouring individuals in the population and accept using $\min(1, r_e)$.
6. Repeat steps 2–5 until N_s samples are obtained.

The finite element model updating of an asymmetrical H-shaped beam structure (Marwala, 1997) is carried out using the EMCMC and the M-H method in Section 10.5.

10.4 Metropolis–Hastings Method

In this chapter, we compare the EMCMC to the M-H algorithm. The M-H algorithm is a type of MCMC method in which samples are drawn from multivariable densities (Metropolis *et al.*, 1953; Hastings, 1970; Chib and Greenberg, 1995; Roberts and Smith, 1994; Gilks, 2005; Boulkaibet, 2014; Heckman and Leamer, 2001; Hairer *et al.*, 2014; Amin *et al.*, 2014). Kennedy *et al.* (2014) combined principal component analysis and line searches to enhance the performance of Metropolis–Hastings MCMC, while Vu *et al.* (2014) applied the marginal M-H for a multi-target tracker. Monroe and Cai (2014) applied the M-H algorithm to Ramsay curve item response theory estimation, while Eberle (2014) applied the M-H algorithm to perturbations of Gaussian measures. This algorithm is similar to rejection and importance

sampling and is implemented by proposing a probability distribution function and using it to create proposed values (Marwala, 2010; Boulkaibet, 2014). The proposed distribution is also utilised to obtain the move probability, which is applied to establish whether the drawn value is accepted as part of the Markov chain or not. The move probability is defined by the ratio of the target density times the ratio of the proposed density. This means that a normalising constant of the target density distribution function is not required in this algorithm.

In order to sample from the posterior distribution function $p(\boldsymbol{\theta}|\mathcal{D})$, which is the 'target' distribution function, where $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_d\}$ is a d-dimensional parameter vector, the proposal density distribution $q(\boldsymbol{\theta}|\boldsymbol{\theta}_{t-1})$ is introduced in order to generate a random vector θ given the value at the previous iteration of the algorithm. The M-H algorithm consists of two basic steps: draw from the proposed density stage and accept or reject the sample. The M-H algorithm is summarised as follows (Boulkaibet, 2014):

1. Initialise $\boldsymbol{\theta}_0$.
2. At iteration t, $\boldsymbol{\theta}^*$ is drawn from the proposed density $q(\boldsymbol{\theta}|\boldsymbol{\theta}_{t-1})$, where $\boldsymbol{\theta}_{t-1}$ is the parameter value at the previous step.
3. Update the finite element model to obtain the new analytical frequencies, then calculate the acceptance probability given by

$$a(\boldsymbol{\theta}^*, \boldsymbol{\theta}_{t-1}) = \min\left\{1, \frac{P(\boldsymbol{\theta}^*|\mathcal{D})q(\boldsymbol{\theta}_{t-1}|\boldsymbol{\theta}^*)}{P(\boldsymbol{\theta}_{t-1}|\mathcal{D})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_{t-1})}\right\}.$$

4. Draw u from the uniform distribution $U(0, 1)$.
5. If $u \le a(\boldsymbol{\theta}^*, \boldsymbol{\theta}_{t-1})$ accept $\boldsymbol{\theta}^*$. Otherwise, reject $\boldsymbol{\theta}^*$.
6. Return to step 2 until N_s samples are obtained.

10.5 Application: Asymmetrical H-Shaped Structure

An asymmetrical H-shaped aluminium structure is updated using the EMCMC algorithm. The structure was divided into 12 elements, each modelled as an Euler–Bernoulli beam. The structure was excited using an electromagnetic shaker, and the acceleration was measured at 15 different positions using a roving accelerometer. A set of 15 frequency-response functions were calculated.

The measured frequencies were 53.9, 117.3, 208.4, 254.0 and 445.0 Hz. The moments of inertia and the cross-section areas of the left, middle and right subsections of the beam were selected to be updated. The updating parameter vector was $\boldsymbol{\theta} = \{I_{x1}, I_{x2}, I_{x3}, A_{x1}, A_{x2}, A_{x3}\}$. The Young's modulus for the structure was $7.2 \times 10^{10}\,\mathrm{N/m^2}$, the density 2785 kg/m^3, and the structure dimensions are given in Marwala (1997). The updating parameters $\boldsymbol{\theta}_i$ were bounded by maximum values equal to $[3.73 \times 10^{-8}, 3.73 \times 10^{-8}, 3.73 \times 10^{-8}, 4.16 \times 10^{-4}, 4.16 \times 10^{-4}, 4.16 \times 10^{-4}]$ and minimum values are equal to $[1.73 \times 10^{-8}, 1.73 \times 10^{-8}, 1.73 \times 10^{-8}, 2 \times 10^{-4}, 2.16 \times 10^{-4}, 2.16 \times 10^{-4}]$ and these bounds keep the updated vector physically realistic.

The constant β_c of the posterior distribution function was equal to 1 for the EMCMC algorithm and 10 for the M-H algorithm. In this book, the EMCMC results are compared with those

obtained from the M-H algorithm (Cheung and Beck, 2009). The coefficient α_i was equal to $1/\sigma_i^2$, where σ_i^2 is the variance of the ith parameter, and the variance vector was defined as $\sigma = \left[5\times10^{-8}, 5\times10^{-8}, 5\times10^{-8}, 5\times10^{-4}, 5\times10^{-4}, 5\times10^{-4}\right]$. The number of samples N_s was equal to 1000, the population size for the EMCMC algorithm was to be equal to 12, and the mutation rate was equal to 0.2 while the selection rate was equal to 0.5. The temperatures used in this implementation were $\mathbf{T} = [1557.8, 1277.4, 1054.8, 915.4, 880.7, 815.8, 651, 308.2, 212.2, 189.4, 138.5, 1]$ where $T = 1$ corresponds the initial posterior distribution.

The Bayesian simulation results were analysed in terms of the mean values of the samples obtained for each method as shown Table 10.1. The updated natural frequencies and the prediction error percentage are shown in Table 10.2. The acceptance rate (AR) for the EMCMC algorithm is 61.9%, while theAR for the M-H algorithm was 46.9%. The M-H algorithm gave a bad AR because the move step used for this algorithm was chosen to be relatively large in order to have acceptable results with fast convergence as the variance of the proposal distribution was relatively large. The EMCMC algorithm has an acceptable AR, the reason being that the exchange probabilities used in this implementation were set to 0.8 when the selected chromosomes (or individuals) were the first and second chromosomes in the population list ($\boldsymbol{\theta}^1, \boldsymbol{\theta}^2$), or the last two chromosomes in the population list ($\boldsymbol{\theta}^{N-1}, \boldsymbol{\theta}^N$), otherwise the probability was set to 0.5.

Table 10.1 shows the initial values of the parameters and the updated values for both EMCMC and M-H algorithms. Both the algorithms updated the $\boldsymbol{\theta}$ vector and the updated values were different from the initial $\boldsymbol{\theta}_0$. The updated vectors obtained by both algorithms were physically realistic. The coefficients of variance obtained by M-H algorithm were large when

Table 10.1 Initial and updated parameters using EMCMC and M-H algorithms

	Initial $\boldsymbol{\theta}_0$	$\boldsymbol{\theta}$ vector EMCMC method	$\frac{\sigma_i}{\mu_i}$ (%)	$\boldsymbol{\theta}$ vector M-H method	$\frac{\sigma_i}{\mu_i}$ (%)
I_{x1}	2.73×10^{-8}	2.578×10^{-8}	0.75	2.31×10^{-8}	22.59
I_{x2}	2.73×10^{-8}	2.576×10^{-8}	0.78	2.68×10^{-8}	15.25
I_{x3}	2.73×10^{-8}	2.533×10^{-8}	3.84	2.17×10^{-8}	13.96
A_{x1}	3.16×10^{-4}	3.693×10^{-4}	0.04	2.85×10^{-4}	14.36
A_{x2}	3.16×10^{-4}	2.157×10^{-4}	0.001	2.83×10^{-4}	14.36
A_{x3}	3.16×10^{-4}	3.006×10^{-4}	0.003	2.77×10^{-4}	13.08

Table 10.2 Natural frequencies when using EMCMC and M-H algorithms

Mode	Measured frequency (Hz)	Initial frequency (Hz)	Frequencies EMCMC method (Hz)	Frequencies M-H method (Hz)
1	53.9	51.40	53.1	53.92
2	117.3	116.61	119.99	122.05
3	208.4	201.27	210.23	210.93
4	254.0	247.42	249.01	258.94
5	445	390.33	431.84	410.33

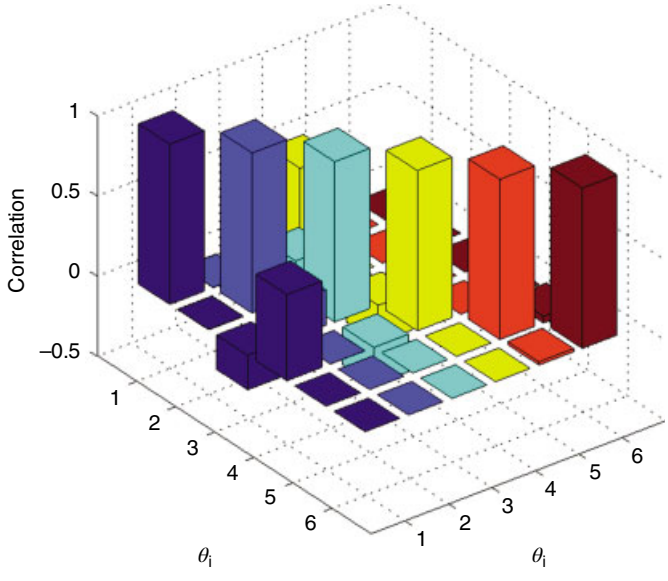

Figure 10.1 The correlation between the updated parameters (EMCMC algorithm)

compared with those obtained by from the implementation of the EMCMC algorithm. The reason for this is that large move steps were used for the M-H algorithm to ensure a fast convergence and to improve the total average errors, while the EMCMC algorithm used the mutation and the crossover operators to move from one sample to the next.

Figures 10.1 and 10.2 show the correlation between all updated parameters for both algorithms and these results indicate that all parameters were correlated for all algorithms (all values are non-zero) and most of the parameters were weakly correlated except the pair (A_{x1}, A_{x3}) which was found to have a correlation of 0.51 when using the EMCMC algorithm.

Table 10.2 shows the finite element model updated frequencies when both EMCMC and M-H algorithms are implemented to update the structure. The error between the first measured natural frequency and that of the initial model was 4.63%. When applying the EMCMC algorithm this error was reduced to 1.49%; however, the M-H algorithm reduced the error to 0.04%. The overall updated finite element model natural frequencies for both algorithms are better than the initial finite element model. However, the EMCMC algorithms produced better total average error results than the M-H algorithm. The updating procedure using the EMCMC method improved the error from 4.7 % to 1.92%. The coefficient of variance values obtained by the EMCMC algorithm are very small compared to those obtained by the M-H algorithm, which indicates that the EMCMC algorithm produces more accurate results than the M-H algorithm, and these can be seen in Table 10.3.

The results obtained show that both algorithms converge fast using around 100 iterations for the M-H algorithm, while the EMCMC algorithm requires less than 45 iterations to start to converge. The reason why the M-H algorithm has a high convergence rate is that large move steps are used. Mutation and the crossover procedures assists the EMCMC algorithm to converge fast. The error obtained by the EMCMC algorithm is smaller than that obtained by the M-H algorithm.

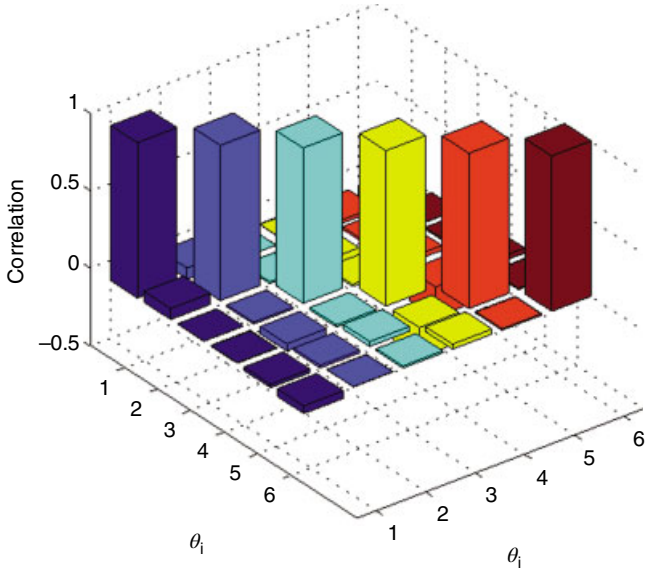

Figure 10.2 The correlation between the updated parameters (M-H algorithm)

Table 10.3 Errors when using EMCMC and M-H algorithms

Mode	Error between initial and measured natural frequencies (%)	Coeffients of variance for EMCMC (%)	Error for EMCMC (%)	Coeffients of variance for M-H (%)	Error for M-H (%)
1	4.63	0.013	1.49	3.96	0.04
2	0.59	0.041	2.29	4.28	4.05
3	3.42	0.172	0.88	4.95	1.22
4	2.59	0.091	1.97	4.81	1.94
5	12.28	0.044	2.96	4.74	7.79
Total average error	4.70	—	1.92	—	3.01

10.6 Conclusion

This chapter presented the applicability and performance of the EMCMC algorithm on the finite element model updating problem. The EMCMC algorithm was tested on a real beam structure with a number of uncertain variables and was compared to the M-H method. The results showed better results of the EMCMC approach as compared to the M-H algorithm. The performance of the latter decreases with the complexity of the system and the size of the uncertain vector and its error was large compared to that obtained from the EMCMC algorithm.

References

Amin NAM, Adam MB, Ibrahim NA (2014) Multiple-try Metropolis Hastings for modeling extreme PM10 data. *AIP Conference Proceeedings*, **1605**: 949–954.

Bishop CM (1995) *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.

Bishop CM (2006) *Pattern Recognition and Machine Learning*. New York: Springer-Verlag.

Boulkaibet, I. (2014) Finite element model updating using Markov chain Monte Carlo techniques. PhD thesis, University of Johannesburg.

Boulkaibet I, Marwala T, Mthembu L, Friswell MI, Adhikari S (2012) Sampling techniques in Bayesian finite element model updating. *Proceedings of the Society for Experimental Mechanics* **29**: 75–83.

Boulkaibet I, Mthembu L, Marwala T, Friswell MI, Adhikari S (2015a) Finite element model updating using the shadow hybrid Monte Carlo technique. *Mechanical Systems & Signal Processing* **52–53**: 115–132.

Boulkaibet I, Mthembu L, Marwala T, Friswell MI, Adhikari S (2015b) Finite element model updating using an evolutionary Markov chain Monte Carlo algorithm. *Dynamics of Civil Structures* **2**: 245–253.

Boulkaibet I, Mthembu L, De Lima Neto F and Marwala T (2015c) Finite element model updating using fish school search and volitive particle swarm optimization. *Integrated Computer-Aided Engineering* **22**(4): 361–376.

Bukharov OE, Bogolyubov DP (2015) Development of a decision support system based on neural networks and a genetic algorithm. *Expert Systems with Applications* **42**: 6177–6183.

Calderhead B (2014) A general construction for parallelizing Metropolis–Hastings algorithms. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 17408–17413.

Casanellas M, Kedzierska AM (2013) Generating Markov evolutionary matrices for a given branch length. *Linear Algebra and Its Applications* **438**: 2484–2499.

Chatterjee S, Carrera C, Lynch LA (1996) Genetic algorithms and traveling salesman problems. *European Journal of Operational Research* **93**: 490–510.

Cheung SH, Beck JL (2009) Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters. *Journal of Engineering Mechanics* **135**: 243–255.

Chib S, Greenberg E (1995) Understanding the Metropolis–Hastings algorithm. *American Statistician* **49**: 327–335.

Ching J, Leu SS (2009) Bayesian updating of reliability of civil infrastructure facilities based on condition-state data and fault-tree model. *Reliability Engineering & System Safety* **94**: 1962–1974.

Davis L (1991) *Handbook of Genetic Algorithms*. New York: Van Nostrand Reinhold.

Diaz-Uriarte, R (2015) Identifying restrictions in the order of accumulation of mutations during tumor progression: effects of passengers, evolutionary models, and sampling. *BMC Bioinformatics* **16**, article no. 41.

Din D-R (2015) Genetic algorithm for virtual topology design on MLR WDM networks. *Optical Switching and Networking* **18**: 20–34.

Doob JL (1953) *Stochastic Processes*. New York: John Wiley & Sons, Inc.

Dunca G, Bucur DM, Cervantes MJ, Popa R (2015) Discharge evaluation from pressure measurements by a genetic algorithm based method. *Flow Measurement and Instrumentation* **45**: 49–55.

Eberle A (2014) Error bounds for Metropolis–Hastings algorithms applied to perturbations of Gaussian measures in high dimensions. *Annals of Applied Probability* **24**: 337–377.

Feng X, Moon I, Shin J (2015) Hybrid genetic algorithms for the three-dimensional multiple container packing problem. *Flexible Services and Manufacturing Journal* **27**: 451–477.

Friswell MI, Mottershead JE (1995) *Finite Element Model Updating in Structural Dynamics*. Heidelberg: Kluwer Academic Publishers.

García-Villoria A, Corominas A, Pastor R (2015) Heuristics and simulated annealing procedures for the accessibility windows assembly line problem level 1 (AWALBP-L1). *Computers and Operations Research* **62**: 1–11.

Geyer CJ (1991) Markov chain Monte Carlo maximum likelihood. Proceedings of the 23rd Symposium on the Interface, (Keramidas EM, ed.). Fairfax Station: Interface Foundation, pp. 156–163.

Gholami R, Shahabi M, Haghifam M-R (2015) An efficient optimal capacitor allocation in DG embedded distribution networks with islanding operation capability of micro-grid using a new genetic based algorithm. *International Journal of Electrical Power and Energy Systems* **71**: 335–343.

Gilks WR (2005) Markov chain Monte Carlo. In *Encyclopedia of Biostatistics*. Chichester: John Wiley & Sons, Ltd.

Goh, J., Tang, H.L., Peto, T. and Saleh, G. (2012) An evolutionary approach for determining hidden Markov model for medical image analysis. Proceedings of the IEEE Congress on Evolutionary Computation. Brisbane: IEEE.

Goldberg DE (1989) *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston: Addison-Wesley.

Gopakumar P, Jaya Bharata Reddy M, Mohanta DK (2015) Pragmatic multi-stage simulated annealing for optimal placement of synchrophasor measurement units in smart power grids. *Frontiers in Energy* **9**: 148–161.

Hairer M, Stuart AM, Vollmer SJ (2014) Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *Annals of Applied Probability* **24**: 2455–2490.

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 97–109.

Heckman JJ, Leamer E (2001) *Handbook of Econometrics*, Vol. **5**. Amsterdam: Elsevier.

Holland JH (1975) *Adaptation in Natural and Artificial Systems*. Cambridge, MA: MIT Press.

Hukushima K, Nemoto K (1996) Exchange Monte Carlo method and application to spin glass simulations. *Journal of the Physics Society of Japan* **65**: 1604–1608.

Jing L, Vadakkepat P (2010) Interacting MCMC particle filter for tracking maneuvering target. *Digital Signal Processing* **20**: 561–574.

Kennedy DA, Dukic V, Dwyer G (2014) Combining principal component analysis with parameter line-searches to improve the efficacy of Metropolis–Hastings MCMC. *Environmental and Ecological Statistics* **28**: 247–274.

Kobayashi G, Kozumi H (2015) Generalized multiple-point Metropolis algorithms for approximate Bayesian computation. *Journal of Statistical Computation and Simulation* **85**: 675–692.

Kou K, Li X, Li L, Li H, Wu T (2015) Absolute distance estimation with improved genetic algorithm in laser self-mixing scheme. *Optics and Laser Technology* **68**: 113–119.

Lee G, Mallipeddi R, Jang G-J, Lee M (2015) A genetic algorithm-based moving object detection for real-time traffic surveillance. *IEEE Signal Processing Letters* **22**: 1619–1622.

Li H (2015) Cauchy problem for linearized non-cutoff Boltzmann equation with distribution initial datum. *Acta Mathematica Scientia* **35**: 459–476.

Liang F, Wong WH (2000) Evolutionary Monte Carlo: application to C_p model sampling and change point problem. *Statistica Sinica* **10**: 317–342.

Liang F, Wong WH (2001) Real parameter evolutionary Monte Carlo with applications in Bayesian mixture models. *Journal of the American Statistical Association* **96**: 653–666.

Liesenfeld R, Richard J (2008) Improving MCMC, using efficient importance sampling. *Computational Statistics and Data Analysis* **53**: 272–288.

Malve O, Laine M, Haario H, Kirkkala T, Sarvala J (2007) Bayesian modelling of algal mass occurrences – using adaptive MCMC methods with a lake water quality model. *Environmental Modelling and Software* **22**: 966–977.

Markov AA (1971) Extension of the limit theorems of probability theory to a sum of variables connected in a Chain, reprinted in Appendix B of *Dynamic Probabilistic Systems, Vol. 1: Markov Chains* (ed. Howard R). New York: John Wiley & Sons, Inc.

Marwala, T. (1997) Multi-criteria method for determining damage on structures. MEng thesis, University of Pretoria.

Marwala T (2010) *Finite Element Model Updating Using Computational Intelligence Techniques*. London: Springer-Verlag.

Marwala T, Sibisi S (2005) Finite element model updating using Bayesian framework and modal properties. *Journal of Aircraft* **42**: 275–278.

Matai R (2015) Solving multi objective facility layout problem by modified simulated annealing. *Applied Mathematics and Computation* **261**: 302–311.

Mbalawata IS, Särkkä S, Vihola M, Haario H (2015) Adaptive Metropolis algorithm using variational Bayesian adaptive Kalman filter. *Computational Statistics and Data Analysis* **83**: 101–115.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**: 1087–1092.

Mitchell M (1996) *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press.

Mo J, Simha A, Kheifets S, Raizen MG (2015) Testing the Maxwell–Boltzmann distribution using Brownian particles. *Optics Express* **23**: 1888–1893.

Moin NH, Chung Sin O, Omar M (2015) Hybrid genetic algorithm with multiparents crossover for job shop scheduling problems. *Mathematical Problems in Engineering*, article no. 210680.

Monroe S, Cai L (2014) Estimation of a Ramsay-curve item response theory model by the Metropolis–Hastings Robbin–Monro algorithm. *Educational and Psychological Measurement* **74**: 343–369.

Ok CY, Patel KP, Garcia-Manero G, Routbort MJ, Peng J, Tang G, Goswami M, Young KH, Singh R, Medeiros LJ, Kantarjian HM, Luthra R, Wang SA (2015) TP53 mutation characteristics in therapy-related myelodysplastic syndromes and acute myeloid leukemia is similar to de novo diseases. *Journal of Hematology and Oncology* **8**, article no. 139.

Palubeckis G (2015) Fast simulated annealing for single-row equidistant facility layout. *Applied Mathematics and Computation* **263**: 287–301.

Ponce-de-Leon-Senti EE, Diaz-Diaz E (2012) Adaptive evolutionary algorithm based on a cliqued Gibbs sampling over graphical Markov model structure. *Adaptation, Learning, and Optimization* **14**: 109–123.

Roberts GO, Smith AF (1994) Simple conditions for the convergence of the Gibbs sampler and Metropolis–Hastings algorithms. *Stochastic Processes and Their Applications* **49**: 207–216.

Rowan BA, Patel V, Weigel D, Schneeberger K (2015) Rapid and inexpensive whole-genome genotyping-by-sequencing for crossover localization and fine-scale genetic mapping, *G3: Genes, Genomes, Genetics* **5**: 385–398.

Ruiz-Cárdenas R, Ferreira MAR, Schmidt AM (2012) Evolutionary Markov chain Monte Carlo algorithms for optimal monitoring network designs. *Statistical Methodology* **9**: 185–194.

Sakae Y, Hiroyasu T, Miki M, Ishii K, Okamoto Y (2015) Conformational search simulations of Trp-cage using genetic crossover. *Molecular Simulation* **41**(10–12): 1045–1049.

Sherlock C, Thiery AH, Roberts GO, Rosenthal JS (2015) On the efficiency of pseudo-marginal random walk metropolis algorithms. *Annals of Statistics* **43**: 238–275.

Šprogar M (2015) Prudent alignment and crossover of decision trees in genetic programming. *Genetic Programming and Evolvable Machines* **16**(4): 499–530.

Sundar VS, Manohar CS (2013) Updating reliability models of statically loaded instrumented structures. *Structural Safety* **40**: 21–30.

Theofilatos K, Pavlopoulou N, Papasavvas C, Likothanassis S, Dimitrakopoulos C, Georgopoulos E, Moschopoulos C, Mavroudi S (2015) Predicting protein complexes from weighted protein–protein interaction graphs with a novel unsupervised methodology: evolutionary enhanced Markov clustering. *Artificial Intelligence in Medicine* **63**: 181–189.

Ulivi P, Delmonte A, Chiadini E, Calistri D, Papi M, Mariotti M, Verlicchi A, Ragazzini A, Capelli L, Gamboni A, Puccetti M, Dubini A, Burgio MA, Casanova C, Crinò L, Amadori D, Dazzi C (2015) Gene mutation analysis in EGFR wild type NSCLC responsive to erlotinib: are there features to guide patient selection? *International Journal of Molecular Sciences* **16**: 747–757.

Vannucci M, Colla V (2015) Fuzzy adaptation of crossover and mutation rates in genetic algorithms based on population performance. *Journal of Intelligent and Fuzzy Systems* **28**: 1805–1818.

Vu T, Vo B-N, Evans R (2014) A particle marginal Metropolis–Hastings multi-target tracker. *IEEE Transactions on Signal Processing* **62**: 3953–3964.

Wang B, Sun R, Yin X, Zhang G (2015) Nonlinear inversion based on Metropolis sampling algorithm. *Oil Geophysical Prospecting* **50**: 111–117.

Yan, G. (2012) A Bayesian approach for identification of structural crack using strain measurements. Proceedings of the 6th European Workshop – Structural Health Monitoring, pp. 763–770.

Yan G (2014) A Bayesian approach for impact load identification of stiffened composite panel. *Inverse Problems in Science and Engineering* **22**: 940–965.

Yan G, Sun H, Waisman H (2015) A guided Bayesian inference approach for detection of multiple flaws in structures using the extended finite element method. *Computers and Structures* **152**: 27–44.

Yuen KV (2010)*Bayesian Methods for Structural Dynamics and Civil Engineering*. Hoboken, NJ: John Wiley & Sons, Inc.

Zhou J, Mita A, Mei L (2015) Posterior density estimation for structural parameters using improved differential evolution adaptive Metropolis algorithm. *Smart Structures and Systems* **15**: 735–749.

11

Adaptive Markov Chain Monte Carlo Method for Finite Element Model Updating

11.1 Introduction

Finite element modelling is one of the most widely implemented numerical techniques and has been applied in many engineering problems from various domains, including mechanical, civil and electrical engineering (Rao, 2004; Bhatti, 2005; Oñate, 2009; Li *et al*., 2016; Fabbri and Cevoli, 2016). It produces numerical models for real systems (or structures) and these models can be fairly accurate for simple structures or systems. However, the estimation worsens when the modelled system is sufficiently complex, and the results achieved from the finite element models are often different from those obtained from the experimental investigation. A number of factors can cause the degradation of the accuracy of finite element models, including the variability of model parameters implemented in the models and the errors resulting from the general modelling assumptions such as non-linearity or modelling damping. To increase the accuracy of finite element models, some of the uncertain model parameters can be adjusted or updated to decrease the error between the measured data and the numerical model (Friswell and Mottershead, 1995; Marwala, 2010; Shan *et al.*, 2015; Sanayei *et al.*, 2015; Seifi and Abbasi, 2015). This procedure of adjusting or updating these uncertain parameters, to identify the most probable parameters that accurately characterise the structure using the measured responses of the system, is what we have referred to throughout this book as finite element model updating (Friswell and Mottershead, 1995; Marwala, 2010).

There are three main approaches to conducting finite element model updating: direct, indirect (also called iterative updating) and probabilistic approaches (Friswell and Mottershead, 1995; Marwala, 2010). Calling this approach the iterative approach is now outdated because probabilistic models which are the subject of this book are not necessarily all iterative. To implement the direct finite element model updating process, the model output is directly equated to the measured data. However, in the iterative approach the differences between

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

the measured data and the finite element model output are minimised by adjusting some uncertain variables (Friswell and Mottershead, 1995).

The most widely implemented iterative algorithms are the sensitivity-based procedures and the metaheuristic algorithms (Friswell and Mottershead, 1995; Marwala, 2010; Boulkaibet *et al.*, 2012, 2013; Boulkaibet, 2014). These approaches are characteristically optimisation problems and their objective functions are constructed by the error between the analytical and experimental data. The objective functions of an iterative updating strategy are then minimised by identifying the optimal solution of the updated parameters of the finite element model. The strategy of minimising some fitness function is called the maximum likelihood or frequentist approach (Neyman, 1937).

Additional methodologies implemented in finite element model updating are statistical approaches (Marwala and Sibisi, 2005; Yuen, 2010; Khodaparast, 2010; Marwala, 2010; Rathmanner and Hutter, 2011). Statistical approaches, also called uncertainty quantification techniques, are beneficial mathematical approaches that can be implemented to update finite element models and to offer additional information on the uncertainties of the updated parameters.

The statistical approaches that have been implemented for finite element model updating are of two types: non-probabilistic or possibilistic methods, such as rough sets and fuzzy logic, which can be applied to approximate the uncertainty of the model parameters where the uncertain parameters are characterised by membership functions (Pawlak, 1982; Moens and Vandepitte, 2006; Bazan *et al.*, 2004; Biacino and Gerla, 2002; Arabacioglu, 2010); and probabilistic methods which treat the uncertain parameter as a random distribution with joint probability density function (Yuen, 2010; Khodaparast, 2010). The Bayesian technique, which is the most common probabilistic procedure together with the perturbation methods, is widely used in system identification and uncertainty quantification (Boulkaibet, 2014; Boulkaibet *et al.,* 2015a, 2015c; Green *et al*., 2015). This method has been applied in finite element model updating problems, and good results have been obtained for uncertainty quantification (Marwala and Sibisi, 2005; Marwala, 2010; Mthembu *et al.*, 2011b; Boulkaibet, 2014; Jensen *et al.*, 2014).

In the Bayesian technique the uncertain parameters are modelled as random parameters with joint probability density functions, referred to as *posterior* probability density functions, estimated by multiplying the likelihood function by the prior probability and dividing by the evidence. In situations where the posterior probability density function is analytically incalculable, sampling techniques can be implemented to obtain numerical solutions for it. The most widely implemented sampling procedure is the Markov chain Monte Carlo (MCMC) method, which is a random walk procedure that is based on the Markov chain and the Metropolis acceptance procedure (Ginting *et al*., 2015). MCMC methods have been widely implemented for finite element model updating (Lam *et al*., 2015; Lee and Sohn, 2015; Wu and Chu, 2015).

The most widely implemented MCMC algorithm is the Metropolis–Hastings (M-H) method which has been implemented for finite element model updating problems and uncertainty quantification (Mares *et al*., 2006; Boulkaibet, 2014). Additionally, a number of MCMC algorithms have been implemented to solve Bayesian model updating problems. A Monte Carlo inverse technique was implemented by Mares *et al*. (2006) for Bayesian finite element model updating, while Nichols *et al.* (2010) implemented the MCMC technique to sample from the posterior probability distribution function of a non-linear system. The Gibbs sampling method was implemented by Ching *et al*. (2006) for finite element model updating, while Ching and Chen (2007) proposed a modified form of the M-H algorithm called the transitional Markov chain

Monte Carlo (TMCMC) algorithm which was implemented by Muto and Beck (2008) to update hysteretic structural models. Cheung and Beck (2009) applied the hybrid Monte Carlo (HMC) technique to update a structural dynamic linear system with 31 uncertain parameters. The updating procedure was able to depict uncertainties related to the underlying structural system.

The HMC technique has demonstrated promising results in solving complex higher-dimensional problems (Chen and Roux, 2015; Venditto *et al.*, 2015; Ito *et al.*, 2015). The trajectory of the HMC method is directed by the derivative of the posterior log-density which enables convergence to areas with high probability during the search process (Duane *et al.*, 1987; Kennedy and Pendleton, 1991; Beskos *et al.*, 2011). In the HMC technique, a molecular dynamic system is created and its total energy, which is called the Hamiltonian function, is implemented for sampling. The Hamiltonian is calculated numerically by implementing the leapfrog integrator. However, this integrator does not conserve the Hamiltonian function, particularly when a moderately large time step is required to accelerate the convergence process or when the system size is comparatively large, and this is in conflict with the principle of conservation of energy. Boulkaibet (2014) attempted to solve this problem by implementing two modified versions of the HMC algorithm – the shadow hybrid Monte Carlo (SHMC; Boulkaibet *et al.*, 2015a) and separable shadow hybrid Monte Carlo (S2HMC; Boulkaibet *et al.*, 2014) algorithms – and applied these for Bayesian finite element model updating. Both of these algorithms produced samples with a relatively large time step and gave more accurate results than the HMC algorithm.

In this chapter, additional modification of the HMC algorithm is discussed, and the idea is to deal with the acceptance rate (AR) degradation and improve the accuracy of the results. Initially, the procedure adaptively selects the length of the trajectory to obtain a better AR without wasting computation time. This can be done by adjusting the trajectory length at every iteration (or several iterations) to control the AR, and with comparatively large length of the trajectory. Furthermore, to handle region separation problems, two high-probability spaces are isolated by regions of low probability. Many MCMC techniques, including HMC, have difficulties in moving from one search space to another when these two regions are separated by other regions with low probability. This can be a problem when the attained samples are only attained from the regions with local minima. To overcome this dilemma, the trajectory of the adaptive hybrid Monte Carlo (AHMC) algorithm is moderated where the samples are attained from a sequence of distributions that are more diffused than the original posterior probability distribution function. In this chapter, the AHMC algorithm is studied to sample the posterior probability distribution function. This technique is examined by updating the finite element models of two structures: a linear system with three degrees of freedom and an asymmetrical H-shaped structure. The advantages and disadvantage of the AHMC technique will be discussed. The chapter begins by examining the posterior distribution function of the uncertain parameters.

11.2 Bayesian Theory

The Bayesian theory is a mechanism for estimating the posterior distribution function from the likelihood function, the prior distribution function and the evidence. The Bayesian method is expressed in terms of Bayes' rule, given by (Bishop, 2006; Hasegawa *et al.*, 2016; Zhang *et al.*, 2016)

$$P(\boldsymbol{\theta}|\mathcal{D},\boldsymbol{\mathcal{M}}) \propto P(\mathcal{D}|\boldsymbol{\theta},\boldsymbol{\mathcal{M}})P(\boldsymbol{\theta}|\boldsymbol{\mathcal{M}}). \tag{11.1}$$

Here, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^d$ is the vector of uncertain parameters representing the variables in the finite element model that need to be updated. The parameter $\boldsymbol{\mathcal{M}}$ designates which model class of the target system is to be updated. Typically, the model classes are disjointed by the model updated vector $\boldsymbol{\theta}$. The parameter $\mathcal{D}$ denotes the experimental measurements, which are the frequencies f_i^m and the mode shapes ϕ_i^m. The expression $P(\boldsymbol{\theta}|\boldsymbol{\mathcal{M}})$ is the prior probability distribution function which represents the knowledge of the uncertain parameters when the model class is known. The expression $P(\mathcal{D}|\boldsymbol{\theta},\boldsymbol{\mathcal{M}})$ represents the likelihood function, which is obtained by the difference between the measurements and the finite element model data when both $\boldsymbol{\theta}$ and $\boldsymbol{\mathcal{M}}$ are given. The term $P(\boldsymbol{\theta}|\mathcal{D},\boldsymbol{\mathcal{M}})$ is the posterior probability distribution function which characterises the probability of the update parameters when both $\mathcal{D}$ and $\boldsymbol{\mathcal{M}}$ are given. Since only a single model class is considered in this chapter, the expression $\boldsymbol{\mathcal{M}}$ will henceforth be suppressed.

The posterior probability distribution function $P(\boldsymbol{\theta}|\mathcal{D})$ implemented in this chapter is the same as the one implemented by Boulkaibet (2014) and Boulkaibet *et al.* (2016) and is mathematically expressed as follows:

$$P(\boldsymbol{\theta}|\mathcal{D}) = \frac{1}{Z_s(\alpha,\beta_c)} \exp\left(-\frac{\beta_c}{2} \sum_{i}^{N_m} \left(\frac{f_i^m - f_i}{f_i^m} \right)^2 - \sum_{i}^{Q} \frac{\alpha_i}{2} \left\| \boldsymbol{\theta}^i - \boldsymbol{\theta}_0^i \right\|^2 \right), \tag{11.2}$$

where $Z_s(\alpha, \beta_c)$ is a normalising constant given by (Boulkaibet, 2014; Boulkaibet *et al.*, 2016)

$$Z_s(\alpha,\beta_c) = \left(\frac{2\pi}{\beta_c} \right)^{N_m/2} \prod_{i=1^{\cdot}}^{N_m} f_i^m (2\pi)^{Q/2} \prod_{i=1}^{Q} \frac{1}{\sqrt{\alpha_i}}. \tag{11.3}$$

The parameter β_c is a constant that can be implemented to give more weight to the likelihood terms. f_i^m and f_i are the ith analytical natural frequency and the ith measured natural frequency. The parameter N_m is the number of measured modes implemented in the finite element model updating procedure. The parameter Q is the size of the updated vector, and the parameter $\boldsymbol{\theta}_0$ is the initial value of the updated vector, which is typically the mean value. The parameter α_i is the ith coefficient of the ith updating variable, and these coefficients can be implemented to weight the prior probability distribution function. The expression $\|.\|$ denotes the Euclidean norm. In complex structures, it is not possible to obtain an analytical solution from the posterior probability distribution function and one must turn to sampling for a numerical solution of the probability distribution function in Equation 11.1 (Boulkaibet, 2014; Boulkaibet *et al.*, 2016). In this chapter, the AHMC algorithm is implemented to sample from the posterior probability distribution function.

11.3 Adaptive Hybrid Monte Carlo

The central notion of the AHMC algorithm is to improve the HMC trajectory by providing an adaptive trajectory length as well as a tempered trajectory (Boulkaibet *et al.*, 2016). Many methods that have been proposed to achieve this. Wang *et al*. (2013) proposed a method using

adaptive Hamiltonian and Riemann manifold Monte Carlo samplers. In that work, a Bayesian optimisation is used to adapt the HMC parameters (time step δt and number of steps L). Burda and Maheu (2011) applied an AHMC method to Baba–Engle–Kraft–Kroner generalised autoregressive conditional heteroscedasticity (GARCH) models. Fischer *et al*. (1998) applied the HMC with adaptive temperature for conformational analysis of RNA, while Wang *et al.* (2014) applied the AHMC method for motion tracking.

The AHMC algorithm is based on the original HMC algorithm (Duane *et al.*, 1987; Neal, 2011; Andrieu and Thoms, 2008). The HMC algorithm has demonstrated good results in solving higher-dimensional complex engineering problems (Duane *et al.*, 1987; Neal, 2011; Andrieu and Thoms, 2008). It operates by combining the molecular dynamic (MD) trajectory (Alder and Wainwright, 1959), the Monte Carlo algorithm (Kolokoltsov, 2010) as well as the Metropolis algorithm (Metropolis *et al.*, 1953) for accepting or rejecting the proposal state. The same conceptions are implemented for the AHMC algorithm where a new dynamical system is constructed by introducing a new auxiliary variable, called the momentum, $\mathbf{p} \in \mathbb{R}^d$. The uncertain vector $\boldsymbol{\theta}$ is treated as the system displacement, while the total energy or the Hamiltonian function of the new dynamical system can be defined as $H(\boldsymbol{\theta}, \mathbf{p}) = V(\boldsymbol{\theta}) + W(\mathbf{p})$ (Resnick and Eisberg, 1985). The potential energy is defined as $V(\boldsymbol{\theta}) = -\ln(P(\boldsymbol{\theta}|D))$, while the kinetic energy of the dynamic system is expressed as $W(\mathbf{p}) = \mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}/2$ and depends on the matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$ which is a positive definite matrix (Jain, 2009). The Hamiltonian dynamics is then governed by (Bransden and Joachain, 1983; Neal, 2001)

$$\frac{d\boldsymbol{\theta}}{dt} = \mathbf{M}^{-1}\mathbf{p}(t), \quad \frac{d\boldsymbol{p}}{dt} = -\nabla V(\boldsymbol{\theta}(t)). \tag{11.4}$$

In this chapter, the joint density function $\rho(\boldsymbol{\theta}, \mathbf{p}) \propto \exp(-\beta_B H(\boldsymbol{\theta}, \mathbf{p}))$ follows a Boltzmann distribution (Landau and Lifshitz, 1980), where $\beta_B = 1/T$, T is a constant temperature, and here the Boltzmann constant is neglected. It is evident that $\rho(\boldsymbol{\theta}, \mathbf{p})$ can be mathematically rewritten as

$$\rho(\boldsymbol{\theta}, \mathbf{p}) \propto \exp\left(-\frac{V(\boldsymbol{\theta})}{T}\right)\exp\left(-\frac{W(\boldsymbol{p})}{T}\right)$$

or (Neal, 2001; Boulkaibet *et al.*, 2016)

$$\rho(\boldsymbol{\theta}, \mathbf{p}) \propto P(\boldsymbol{\theta}|\mathcal{D}) \cdot \exp\left(-\frac{1}{T}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}/2\right) \tag{11.5}$$

It is obvious from Equation 11.5 that sampling the vector $\boldsymbol{\theta}$ from the posterior probability distribution function can also be done by sampling the pair $(\boldsymbol{\theta}, \mathbf{p})$ from the joint probability distribution function $\rho(\boldsymbol{\theta}, \mathbf{p})$ and then discarding $\mathbf{p}$ at the end of the simulation. The pair $(\boldsymbol{\theta}, \mathbf{p})$ is assessed through time t by applying the following leapfrog integrator (Neal, 2001; Boulkaibet *et al.*, 2016):

$$\mathbf{p}\left(t + \frac{\delta t}{2}\right) = \mathbf{p}(t) - \frac{\delta t}{2}\nabla V(\boldsymbol{\theta}(t)), \tag{11.6}$$

$$\boldsymbol{\theta}(t+\delta t)=\boldsymbol{\theta}(t)+\delta t\mathbf{M}^{-1}\mathbf{p}\left(t+\frac{\delta t}{2}\right), \tag{11.7}$$

$$\mathbf{p}(t+\delta t)=\mathbf{p}\left(t+\frac{\delta t}{2}\right)-\frac{\delta t}{2}\nabla V(\boldsymbol{\theta}(t+\delta t)), \tag{11.8}$$

where δt is the time step and ∇V is the gradient which can be obtained numerically by using the finite difference technique which can be expressed as follows (Cheung and Beck, 2009; Boulkaibet *et al.*, 2014, 2016):

$$\frac{\partial V}{\partial \theta_i}\cong\frac{V(\boldsymbol{\theta}+\Delta h)-V(\boldsymbol{\theta}-\Delta h)}{2h\Delta_i}. \tag{11.9}$$

Here, the parameter h is a scalar that prescribes the size of the perturbation of $\boldsymbol{\theta}$, while $\boldsymbol{\Delta}=[\Delta_1,\Delta_2,\ldots,\Delta_N]$ is the perturbation vector. After the evaluation of Equations 11.6–11.8, a Monte Carlo accept–reject step is added to satisfy the property of the system in Equation 11.4. Therefore, if the pair $(\boldsymbol{\theta}, \mathbf{p})$ is the initial vector while the pair $(\boldsymbol{\theta}^*, \mathbf{p}^*)$ is the vector after running Equations 11.6–11.8, then the candidate $(\boldsymbol{\theta}^*, \mathbf{p}^*)$ is accepted with probability $\min(1,\exp\{-\Delta H/T\})$, where $\Delta H=H(\boldsymbol{\theta}^*,\mathbf{p}^*)-H(\boldsymbol{\theta},\mathbf{p})$.

Because the AHMC algorithm has the same fundamentals as the HMC algorithm, all the previous equations and properties are implemented by the AHMC algorithm. However, the AHMC algorithm has particular modifications that improve the sampling performance. To circumvent the non-ergodicity problem and to ensure good performance of the AHMC algorithm, the calculation of the leapfrog algorithm is performed for L steps during each iteration. This improves the algorithm trajectory and confirms large move steps. The value of L can be uniformly selected from the interval $\{1, L_{\max}\}$. Furthermore, because the time step implemented by the leapfrog integrator is bounded ($\delta t_{\min}<\delta t<\delta t_{\max}$), the time step is adjusted after a number of iterations. Adjusting the time step after a fixed number of iterations allows a large rejection rate to be avoided when the time step is large enough (Boulkaibet *et al.*, 2014, 2016). Also, the adaptation of the time step circumvents the use of a small trajectory length and consequently more iterations are required for convergence. To adapt the time step, an initial random value of the parameter δt is selected from the interval $[\delta t_{\min}, \delta t_{\max}]$ and then Equations 11.6–11.8 are evaluated for N_b iterations. The N_b samples attained are implemented to calculate the acceptance rate $\bar{\alpha}^b$ and used to decide if the time step is to be increased or decreased using the following equation (Boulkaibet *et al.*, 2016):

$$\delta t^{i+1}=\begin{cases}\delta t^i-\gamma^i\delta t^i, & \bar{\alpha}^b<\bar{\alpha},\\ \delta t^i+\gamma^i\delta t^i, & \bar{\alpha}^b\geq\bar{\alpha},\end{cases} \tag{11.10}$$

where γ^i is a random variable chosen from the interval [0.01, 0.05] and $\bar{\alpha}$ is the target AR. The value of the target AR can be chosen high to ensure that more different samples are involved in computing the mean values of the uncertain parameters. Thereafter, after every N_b samples or iterations the time step is adapted by increasing or decreasing the time step by a small value (between 1% and 5%). This strategy ensures that the time step does not produce a small trajectory move and that a relatively large or significant number of iterations are not wasted.

The second modification proposed in this algorithm is to sample from distributions that are more diffused than the original posterior probability distribution function (Neal, 2001; Boulkaibet *et al.*, 2016). This strategy can facilitate the movement between high-probability areas separated by regions of low probability. This can be achieved by increasing the temperate T which will eventually offer more a diffused distribution (Neal, 2001; Boulkaibet *et al.*, 2016); note that the posterior probability distribution function is when $T = 1$. In this chapter, the temperature is changed at each iteration by small value according to the expression $T(i+1) = \bar{\alpha}^T T(i)$, where $\bar{\alpha}^T > 1$. The AHMC algorithm is summarised as follows (Neal, 2001; Boulkaibet *et al.*, 2016):

1. A value $\boldsymbol{\theta}_0$ is used to start the algorithm.
2. Initiate $\mathbf{p}_0$ such that $\mathbf{p}_0 \sim \mathcal{N}(0, \boldsymbol{M})$.
3. Perform the following steps for N_b:
   a. Start the leapfrog integrator with the previously accepted pair $(\boldsymbol{\theta}, \mathbf{p})$ and implement the algorithm for L time steps to obtain $(\boldsymbol{\theta}^*, \mathbf{p}^*)$.
   b. Update the finite element model and use the analytical frequencies obtained to compute $H(\boldsymbol{\theta}^*, \mathbf{p}^*)$.
   c. Accept $(\boldsymbol{\theta}^*, \mathbf{p}^*)$ with probability $\min\{1, \exp(-\Delta H / T(i))\}$.
   d. Change the temperature according to $T(i+1) = \bar{\alpha}^T T(i)$.
4. Use the N_b samples to obtain the acceptance rate $\bar{\alpha}^b$.
5. Adjust the time step according to Equation 11.10.
6. Repeat steps 3–5 for N_s samples.

In the next sections, the performance of the AHMC algorithm is discussed when two different finite element models are updated.

11.4 Application 1: A Linear System with Three Degrees of Freedom

In this section a simulation of a mass-and-spring linear system with three degrees of freedom is studied (Boulkaibet *et al.*, 2016). The system is shown in Figure 11.1. The deterministic parameters of this system are masses $m_1 = 2$ kg, $m_2 = 1$ kg and $m_3 = 3$ kg. The nominal mean values of the uncertain parameters are $k_1 = 10\,\mathrm{N/m}$, $k_2 = k_3 = 6\,\mathrm{N/m}$ and $k_4 = 10\,\mathrm{N/m}$, and these values are used to calculate the natural frequencies of interest of this structure: $f_1 = 1.12$ Hz, $f_2 = 3.5$ Hz and $f_3 = 4.1$ Hz. However, the initial values of the uncertain parameters

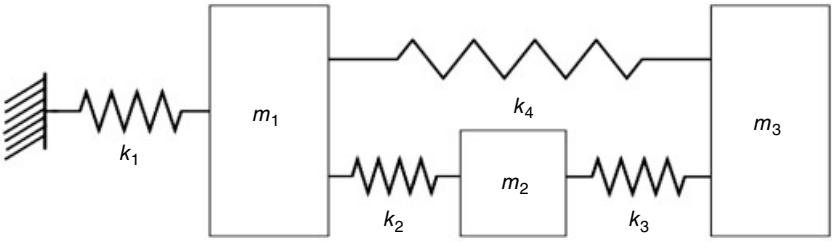

Figure 11.1 A mass-and-spring linear system with three degrees of freeedom

are $k_1 = 12$ N/m, $k_2 = k_3 = 4$ N/m and $k_4 = 7$ N/m. Thus, the parameters to be updated are k_1, k_2 and k_4, and can be represented by a vector of $d = 3$ variables $\boldsymbol{\theta} = \{\theta_1, \theta_2, \theta_3\}$.

11.4.1 Updating the Stiffness Parameters

In this subsection, the linear system is updated by adjusting a vector of three parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, \theta_3\}$ using the Bayesian method, and the AHMC algorithm is implemented to sample from the posterior probability distribution function. The number of samples is set to $N_s = 5000$, the coefficients α_i in Equation 11.3 were set equal to $1/\sigma_i^2$, where σ_i^2 is the variance of θ_i, and because only the stiffness parameters are updated, the σ_i $(i = 1, 2, 3)$ have equal value and are all set to 500. The constant β_c, the weight of the likelihood term, in Equation 11.3 was set equal to 1. The updating parameters θ_i were bounded by a maximum of 14 N/m and a minimum of 2 N/m.

The initial value of $\boldsymbol{\theta}$ was set to $\boldsymbol{\theta}_0 = \{12,4,7\}$. The initial time step implemented in the HMC algorithm was $\delta t^0 = 1 \times 10^{-3}$ s, while the time step was bounded by a minimum of $\delta t^0 = 0.001 \times 10^{-3}$ s and a maximum of $\delta t^0 = 15 \times 10^{-3}$ s. The parameter L was uniformly distributed on the interval $\{1, 30\}$, the target acceptance rate was set to $\bar{\alpha} = 0.95$ (95%), the initial value of the temperature was set to $T(1) = 1$, the parameter $\bar{\alpha}^T$ was set to 1.021 and the final results are tabulated in Tables 11.1 and 11.2.

Figure 11.2 shows the scatter plots for the three uncertain parameters using the AHMC algorithm. The uncertain parameters plotted in this figure are normalised by dividing their values by the initial value θ_i^0. In addition, error ellipses (or confidence ellipses) for the samples obtained are shown. The error ellipse represents a contour that allows one to visualise the confidence interval. This confidence interval describes the region that contains 95% of the parameter

Table 11.1 The updated vector of stiffness parameters using the AHMC technique

	Stiffness parameters (N/m)				
	Initial	Nominal values	Error (%)	AHMC algorithm (μ_i)	$\frac{\sigma_i}{\mu_i}$ (%)
θ_1	12.00	10.00	20	9.89	1.77
θ_2	4.00	6.00	50	5.99	2.82
θ_3	7.00	10.00	30	10.05	5.79

Table 11.2 Frequencies and errors when the AHMC technique is implemented to update the stiffness parameters

Modes	Nominal frequency (Hz)	Initial frequency (Hz)	Error (%)	Frequency AHMC method (Hz)	c.o.v. values (%)	Error (%)
1	1.120	1.122	0.16	1.119	0.43	0.05
2	3.500	2.932	16.23	3.498	0.48	0.07
3	4.100	3.649	11.01	4.100	0.45	0.01

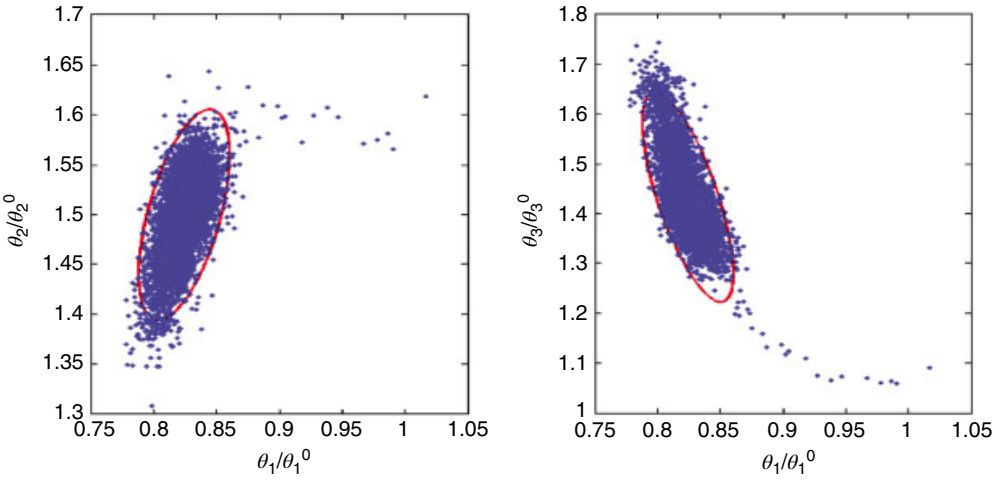

Figure 11.2 Scatter plots with error ellipses using the AHMC method

samples. The figure demonstrates that the AHMC algorithm found the high-probability areas after a few iterations.

The updated values of the uncertain parameters are presented in Table 11.1 along with their initial values, nominal values and coefficient of variation (c.o.v.) values. The c.o.v., which is the obtained standard deviation divided by the mean (the updated values θ_i are the mean values of the samples), can be used to describe the accuracy of the updated parameters. Table 11.1 shows that the c.o.v. values are small for the AHMC algorithm (less than 6%), which indicates that the AHMC algorithm has efficiently estimated the uncertain parameters. This can be seen from the updated vector which is nearly identical to the nominal values.

Table 11.2 contains the updated modes along with the initial modes, the absolute mode errors $\left|f_i^m - f_i\right|/f_i^m$, the total average error (TAE) as a percentage,

$$TAE = \frac{1}{N_m}\sum_{i=1}^{N_m}\frac{\left|f_i^m - f_i\right|}{f_i^m} \quad (N_m = 3),$$

and the c.o.v. The results obtained by implementing the AHMC algorithm are, on average, better than the initial modes. The initial error for the first frequency was 0.16%, and when the AHMC algorithm was applied to update the finite element model, the error was reduced to 0.05%. The same can be said for the rest of the frequencies. In general, using the AHMC algorithm to update the finite element model of this system reduced the TAE from 9.13% to 0.04%. In addition, the c.o.v. obtained by the AHMC algorithm shows that the error for all modes was very small.

Figure 11.3 plots the TAE against the iteration numbers within the first 5000 iterations. Figure 11.3 was obtained by computing the mean value of samples, which was evaluated at every iteration, according to $\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}) \cong (1/N_s)\sum_{j=1}^{i}\boldsymbol{\theta}^i$, and i represents the current iteration.

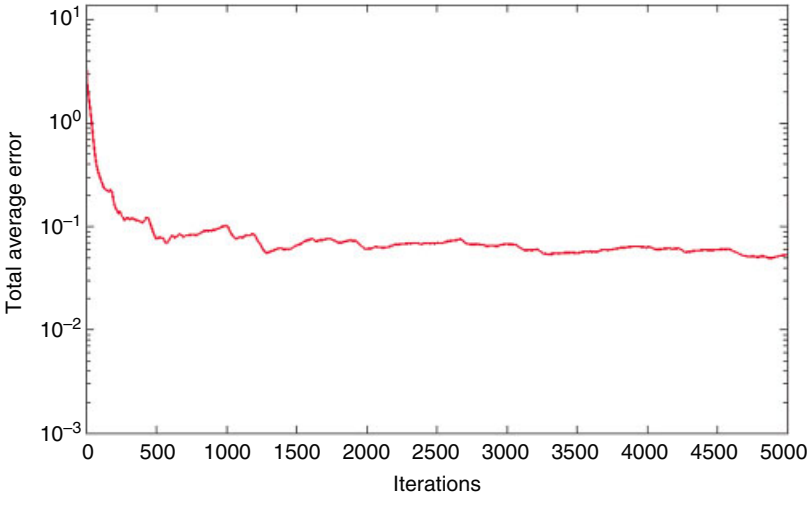

Figure 11.3 The total average error using the AHMC algorithm

Then the obtained mean value was implemented to obtain the new modes from the finite element model and the absolute TAE was computed as $TAE(i) = (1/N_m)\sum_{j=1}^{N_m} \left| f_j^m - f_j \right| / f_j^m$. The result plotted in Figure 11.2 demonstrates that the AHMC algorithm converges faster and within the first 300 iterations.

11.5 Application 2: Asymmetrical H-Shaped Structure

In this section, an asymmetrical H-shaped aluminium structure with a real measured data is updated using the AHMC algorithm (Marwala, 1997, 2010; Marwala and Heyns, 1998; Boulkaibet, 2014). The structure is described in Appendix A. The structure was modelled by assembling 12 beam elements, each modelled as an Euler–Bernoulli beam. An electromagnetic shaker was used to excite the structure, the response measured by an accelerometer, and a set of 15 frequency-response functions were calculated (Marwala, 1997, 2010; Marwala and Heyns, 1998; Boulkaibet, 2014). The measured modes are 53.9, 117.3, 208.4, 254.0 and 445.0 Hz. In this example, the uncertain parameters are the moments of inertia and the section areas of the three beams. The AHMC algorithm was implemented to obtain the updated vector $\boldsymbol{\theta} = \{I_{x1}, I_{x2}, I_{x3}, A_{x1}, A_{x2}, A_{x3}\}$.

11.5.1 H-Shaped Structure Simulation

The parameters of the structure are as follows: Young's modulus is set at 7.2×10^{10} N/m^2 and the density is set to 2785 kg/m^3. The same simulation sets and boundaries implemented previously are implemented for the AHMC algorithm (Boulkaibet, 2014). To help to keep the uncertain parameters physically realistic, the updated parameter vectors are bounded by maximum and minimum values of $[3.73 \times 10^{-8}, 3.73 \times 10^{-8}, 3.73 \times 10^{-8}, 4.16 \times 10^{-4},$

$4.16\times10^{-4}, 4.16\times10^{-4}]$ and $[1.73\times10^{-8}, 1.73\times10^{-8}, 1.73\times10^{-8}, 2.16\times10^{-4}, 2.16\times10^{-4}, 2.16\times10^{-4}]$, respectively.

The likelihood weight β_c in Equation 11.3 was set to 10. The coefficients α_i were set equal to $1/\sigma_i^2$ where $\boldsymbol{\sigma} = [5\times10^{-8}, 5\times10^{-8}, 5\times10^{-8}, 5\times10^{-4}, 5\times10^{-4}, 5\times10^{-4}]$. The number of samples N_s was set to 1000, the initial time step implemented in the AHMC algorithm was $\delta t^0 = 4.5\times10^{-3}$ s while the time step was bounded with a minimum value of $\delta t^0 = 0.001\times10^{-3}$ s and a maximum value of $\delta t^0 = 7\times10^{-3}$ s. The parameter L was uniformly distributed on the interval {1, 30}, the target AR was $\bar{\alpha} = 0.95$ (95%), the initial value of the temperature $T(1) = 1$, and the parameter $\bar{\alpha}^T$ was set to 1.0081. The results obtained using the AHMC algorithm, the updated parameters and the updated frequencies, are shown in Tables 11.3 and 11.4, respectively. Figure 11.4 shows the kernel smoothing density estimation of the updating parameters along with the updated values of the uncertain parameters.

The parameter θ_i denotes the sequential numbering of the updating parameters, while the normalisation constants θ_i^0 were the initial values of the updated parameters. The results obtained demonstrate that the AHMC algorithm identified the high-probability region. Furthermore, the shapes of the plotted density functions are not Gaussian. The updated parameters values, the initial values of these parameters and their c.o.v. values are shown in Table 11.3. The total AR for the AHMC algorithm was 96.3% which is very good. The AHMC algorithm successfully updated the uncertain parameters (the updated values are different than the initial $\boldsymbol{\theta}_0$). The coefficients of variation obtained by AHMC algorithm demonstrate that the middle beam parameters were better approximated than those on the left and right. This

Table 11.3 Initial and updated parameters using the AHMC algorithm

	$\boldsymbol{\theta_0}$ vector initial	$\boldsymbol{\theta}$ vector AHMC method	$\frac{\sigma_i}{\theta_i}$ (%)
I_{x1}	2.73×10^{-8}	3.51×10^{-8}	15.56
I_{x2}	2.73×10^{-8}	2.30×10^{-8}	2.80
I_{x3}	2.73×10^{-8}	3.12×10^{-8}	14.50
A_{x1}	3.16×10^{-4}	3.79×10^{-4}	1.35
A_{x2}	3.16×10^{-4}	2.45×10^{-4}	2.83
A_{x3}	3.16×10^{-4}	2.28×10^{-4}	3.98

Table 11.4 Natural frequencies and errors when the AHMC algorithm is implemented to update the structure

Mode	Measured frequency (Hz)	Initial frequency (Hz)	Error (%)	Frequencies AHMC method (Hz)	Error (%)
1	53.90	51.40	4.63	53.44 (0.87%)	0.85
2	117.30	116.61	0.59	118.96 (0.96%)	1.42
3	208.40	201.27	3.42	208.38 (1.04%)	0.01
4	254.00	247.42	2.59	254.30 (1.40%)	0.12
5	445.00	390.33	12.28	445.08 (1.21%)	0.02

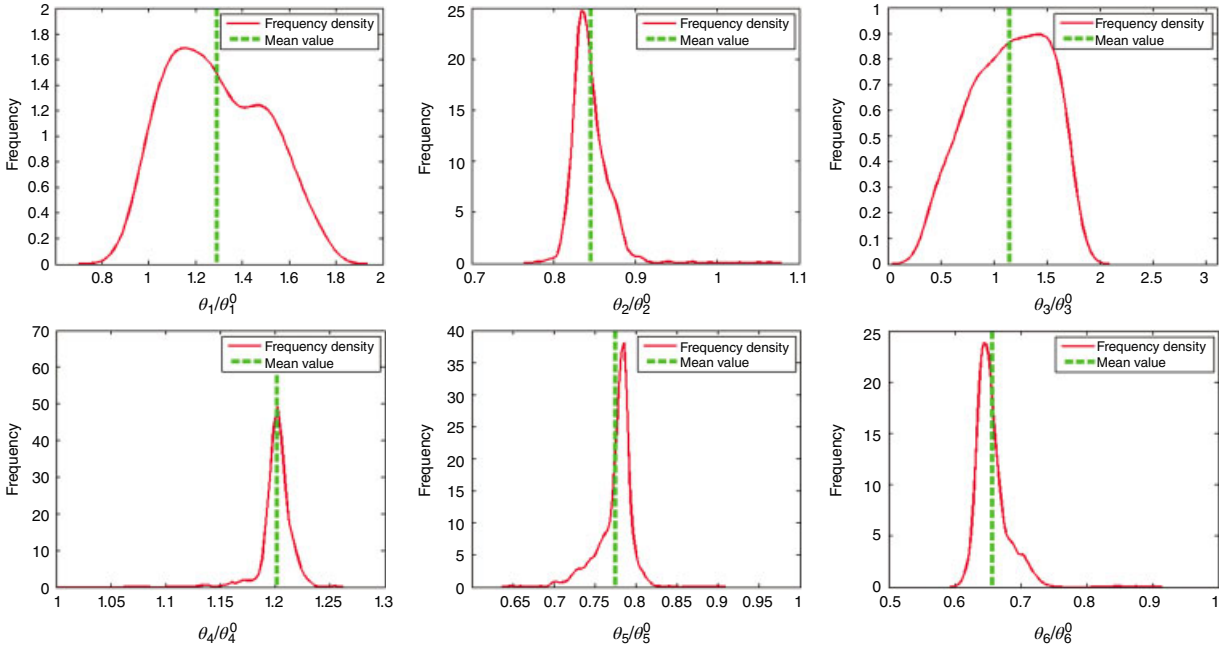

Figure 11.4 The kernel smoothing density estimation of updating model parameters using the AHMC method

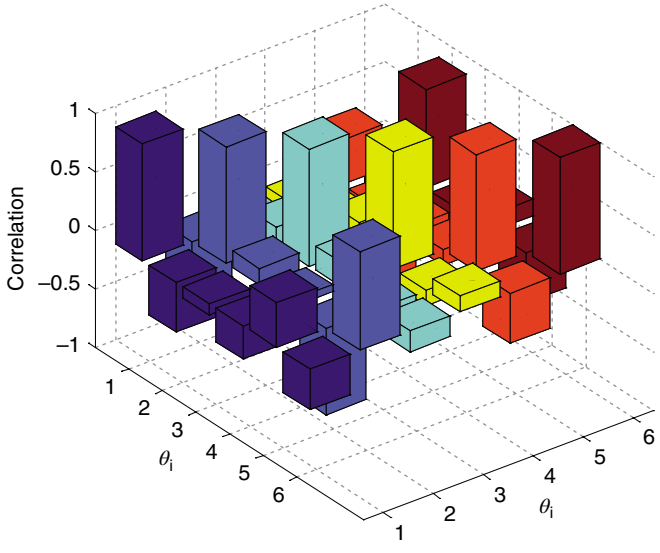

Figure 11.5 The correlation between uncertain parameters

is apparent because the structure was excited in the middle beam, and more information on the middle beam was used in the updating process. The AHMC algorithm successfully updated the uncertain parameters where the values obtained were different than the initial vector $\boldsymbol{\theta}_0$.

The correlations between all updated parameters are shown in Figure 11.5. It suggests that the uncertain parameters are weakly correlated except for the pair (I_{x2}, A_{x2}) which is strongly correlated (the correlation is equal to 0.73). The asymmetrical H-shaped aluminium structure has been updated many times by different (deterministic and statistic) methods (Marwala, 1997, 2010; Marwala and Heyns, 1998). One method that has been used for finite element model updating is the Nelder–Mead optimisation technique (Marwala, 2010), which is a simplex optimisation technique; more details on this procedure can be obtain in Nelder and Mead (1965) and Avriel (2003). Another method that has been used for finite element model updating is the genetic algorithm which is a heuristic optimisation method inspired by the way genes operate (Abdella, 2006; Crossingham and Marwala, 2008; Marwala, 2010). A third technique which has been used together with neural networks for finite element model updating is the response surface method (Marwala, 2004). This is an approximate optimisation method where the objective function is replaced by a simpler one. A final optimisation method to mention which has also been used for finite element model updating is the particle swarm optimisation method, a heuristic method based on the way animals such as birds flock to identify food sources (Kennedy and Eberhart, 1995; Mthembu *et al.*, 2011a).

The Nelder–Mead simplex method gave a TAE equal to 2.14%, while the genetic algorithm reduced the TAE to 1.1%. Between these two, the response surface method produced a TAE of 1.84%) (Marwala, 2010). The particle swarm optimisation algorithm gave a better result, reducing the error to 0.4% (Marwala, 2010). Three MCMC algorithms were also applied to this structure, the M-H, slice sampling and HMC algorithms (Boulkaibet, 2014). The results obtained were 3.01%, 2.98% and 0.73%, respectively. Two further modified versions of the

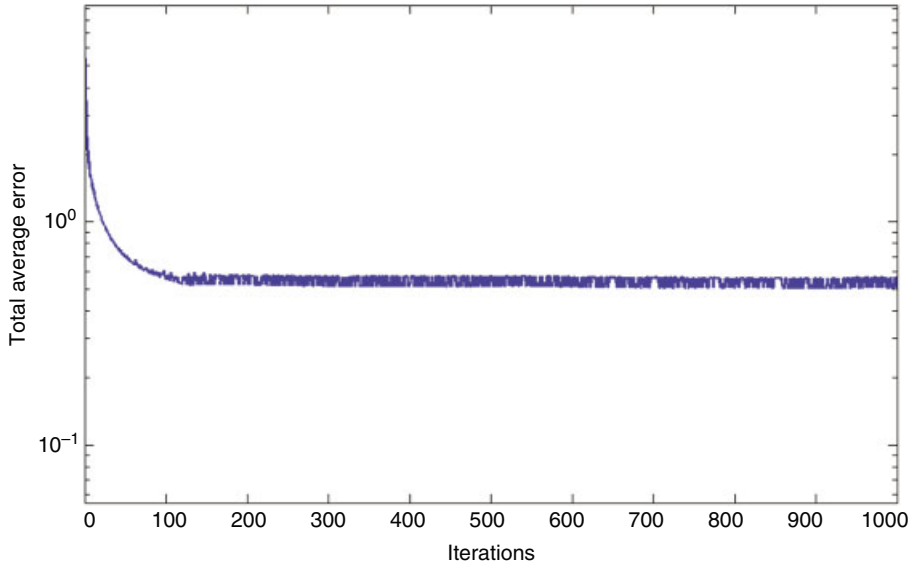

Figure 11.6 The total average error for the AHMC algorithm

HMC algorithm (SHMC and S2HMC algorithms) were also applied to update this structure, and the results obtained were 0.66% and 0.58%, respectively.

Table 11.4 shows the updated frequencies obtained by the AHMC algorithm. The error between the first measured mode and initial one was equal to 4.63%. When the AHMC algorithm was implemented to update the FEM of the structure the error was reduced to 0.85%. On the whole, the results obtained by using the AHMC algorithm were far better than the initial finite element model. The AHMC algorithm reduced the total error to 0.48%, which is a very good result compared to those obtained in previous work (better than all implemented algorithms except the particle swarm optimisation algorithm). Figure 11.6 shows the variation of the TAE with time (iterations). The strategy implemented in the first example to plot Figure 11.3 was implemented to plot Figure 11.6. The results demonstrate that the AHMC algorithm converges fast and within the first 100 iterations.

11.6 Conclusion

In this chapter, an adaptive MCMC algorithm, called the adaptive hybrid Monte Carlo algorithm, was studied to solve the Bayesian finite element model formulation. In this technique, the time step was adaptively designed to improve the trajectory length of the algorithm. This technique was tested by updating the models of two structural systems: a simulated linear system with three degrees of freedom and an asymmetrical H-shaped aluminium structure. For linear system, the AHMC method offered good results and reduced the error to less than 0.05%. For the asymmetrical H-shaped aluminium structure, the AHMC technique gave good results and reduced the total error to 0.48%. The results obtained using the AHMC algorithms were better than those obtained using the HMC, SHMC and S2HMC algorithms.

Notwithstanding the fact that the SHMC and S2HMC approaches use relatively larger time steps (i.e. large trajectory moves in the search space), the AHMC algorithm's ability to adjust its time step and temperature at each iteration gave the algorithm the advantage of escaping from local solutions.

References

Abdella M (2006) The use of genetic algorithms and neural networks to approximate missing data in database. MSc thesis, University of the Witwatersrand.

Alder BJ, Wainwright TE (1959) Studies in molecular dynamics I: General method. *Journal of Chemical Physics* **31**: 459.

Andrieu C, Thoms J (2008) A tutorial on adaptive MCMC. *Statistics and Computing* **18**: 343–373.

Arabacioglu BC (2010) Using fuzzy inference system for architectural space analysis. *Applied Soft Computing* **10**: 926–937.

Avriel M (2003) *Nonlinear Programming: Analysis and Methods*. Mineola, NY: Dover Publishing.

Bazan J, Szczuka M, Wojna A, Wojnarski M (2004) On the evolution of rough set exploration system. In Tsumoto S, Słowiński R, Komorowski J, Grzymała-Busse JW, eds, *Rough Sets and Current Trends in Computing*. Lecture Notes in Artificial Intelligence Vol. **3066**, pp. 592–601. Berlin: Springer-Verlag.

Beskos A, Pillai NS, Roberts GO, Sanz-Serna JM, Stuart AM (2011) Optimal tuning of the hybrid Monte-Carlo algorithm. *Bernoulli* **19**: 1501–1534.

Bhatti MA (2005) *Fundamental Finite Element Analysis and Applications*. Hoboken, NJ: John Wiley & Sons.

Biacino L, Gerla G (2002) Fuzzy logic, continuity and effectiveness. *Archive for Mathematical Logic* **41**: 643–667.

Bishop CM (2006) *Pattern Recognition and Machine Learning*. New York: Springer-Verlag.

Boulkaibet I (2014) Finite element model updating using Markov chain Monte Carlo techniques. PhD thesis. University of Johannesburg.

Boulkaibet I, Marwala T, Friswell MI, Adhikari S (2016) An adaptive Markov chain Monte Carlo method for Bayesian finite element model updating. In Allemang R, De Clerck J, Niezrecki C, Wicks A, eds, *Special Topics in Structural Dynamics 6*, pp. 55–65. New York: Springer-Verlag.

Boulkaibet I, Marwala T, Mthembu L, Friswell MI,Adhikari S (2012) Sampling techniques in Bayesian finite element model updating. *Topics in Model Validation and Uncertainty Quantification* **4**: 75–83.

Boulkaibet I, Marwala T, Mthembu L, Friswell MI, Adhikari S (2015a) Finite element model updating using an evolutionary Markov chain Monte Carlo algorithm, *Dynamics of Civil Structures* **2**: 245–253.

Boulkaibet I, Mthembu L, De Lima Neto F, Marwala T (2015b) Finite element model updating using fish school search and volitive particle swarm optimization. *Integrated Computer-Aided Engineering* **22**: 361–376.

Boulkaibet I, Mthembu L, Marwala T, Friswell MI, Adhikari S (2015c) Finite element model updating using the shadow hybrid Monte Carlo technique. *Mechanical Systems and Signal Processing* **52**: 115–132.

Boulkaibet I, Mthembu L, De Lima Neto F, Marwala T (2013) Finite element model updating using fish school search optimization method. 1st BRICS and 11th CBIC Brazilian Congress on Computational Intelligencear, arXiv:1308.2307.

Boulkaibet I, Mthembu L, Marwala T, Friswell MI, Adhikari S (2014) Finite element model updating using the separable shadow hybrid Monte Carlo technique. *Topics in Modal Analysis II, Volume 8*. Cham: Springer, pp. 267–275.

Bransden BH, Joachain CJ (1983) *Physics of Atoms and Molecules*. Harlow: Longman.

Burda, M and Maheu, J. (2011) Bayesian Adaptive Hamiltonian Monte Carlo with an Application to High-Dimensional BEKK GARCH Models. https://www.economics.utoronto.ca/public/workingPapers/tecipa-438.pdf (accessed 20 October 2015).

Chen Y, Roux B (2015) Enhanced sampling of an atomic model with hybrid non-equilibrium molecular dynamics – Monte Carlo simulations guided by a coarse-grained model. *Journal of Chemical Theory and Computation* **11**: 3572–3583.

Cheung SH, Beck JL (2009) Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters. *Journal of Engineering Mechanics* **135**: 243–255.

Ching J, Chen YJ (2007) Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. *Journal of Engineering Mechanics* **133**: 816–832.

Ching J, Muto M, Beck JL (2006) Structural model updating and health monitoring with incomplete modal data using Gibbs sampler. *Computer-Aided Civil and Infrastructure Engineering* **21**: 242–257.

Crossingham B, Marwala T (2008) Using genetic algorithms to optimise rough set partition sizes for HIV data analysis. *Advances in Intelligent and Distributed Computing, Studies in Computational Intelligence* **78**: 245–250.

Duane S, Kennedy AD, Pendleton BJ, Roweth D (1987) Hybrid Monte Carlo. *Physics Letters* **B195**: 216–222.

Fabbri A, Cevoli C (2016) Rheological parameters estimation of non-Newtonian food fluids by finite elements model inversion. *Journal of Food Engineering* **169**: 172–178.

Fischer A, Cordes F, Schütte C (1998) Hybrid Monte Carlo with adaptive temperature in mixed-canonical ensemble: efficient conformational analysis of RNA. *Journal of Computational Chemistry* **19**: 1689–1697.

Friswell MI, Mottershead JE (1995) *Finite Element Model Updating in Structural Dynamics*. Dordrecht: Kluwer Academic Publishers.

Ginting V, Pereira F, Rahunanthan A (2015) Multi-physics Markov chain Monte Carlo methods for subsurface flows. *Mathematics and Computers in Simulation* **118**: 224–238.

Green PL, Cross EJ, Worden K (2015) Bayesian system identification of dynamical systems using highly informative training data. *Mechanical Systems and Signal Processing* **56–57**: 109–122.

Hasegawa T, Niida A, Mori T, Shimamura T, Yamaguchi R, Miyano S, Akutsu T, Imoto S (2016) A likelihood-free filtering method via approximate Bayesian computation in evaluating biological simulation models. *Computational Statistics and Data Analysis* **94**: 63–74.

Ito AM, Takayama A, Oda Y, Tamura T, Kobayashi R, Hattori T, Ogata S, (2015) Molecular dynamics and Monte Carlo hybrid simulation for fuzzy tungsten nanostructure formation. *Nuclear Fusion* **55**, article no. 073013.

Jain MC (2009) *Textbook of Engineering Physics (Part I)*. New Delhi: PHI Learning.

Jensen HA, Millas E, Kusanovic D, Papadimitriou C (2014) Model-reduction techniques for Bayesian finite element model updating using dynamic response data. *Computer Methods in Applied Mechanics and Engineering* **279**: 301–324.

Kennedy AD, Pendleton B (1991) Acceptances and autocorrelations in hybrid Monte Carlo. *Nuclear Physics B, Proceedings Supplements* **20**: 118–121.

Kennedy J, Eberhart R (1995) Particle swarm optimization. Proceedings of the IEEE International Conference on Neural Networks, Piscataway, NJ: IEEE, pp. 1942–1948.

Khodaparast HH (2010) Stochastic finite element model updating and its application in aeroelasticity. PhD thesis, University of Liverpool.

Kolokoltsov V (2010) *Nonlinear Markov Processes*. Cambridge: Cambridge University Press.

Lam H-F, Yang J, Au S-K (2015) Bayesian model updating of a coupled-slab system using field test data utilizing an enhanced Markov chain Monte Carlo simulation algorithm. *Engineering Structures* **102**: 144–155.

Landau LD, Lifshitz EM (1980) *Statistical Physics: Course of Theoretical Physics*, Vol. 5 (3rd edn). Oxford: Pergamon Press.

Lee M, Sohn K (2015) Inferring the route-use patterns of metro passengers based only on travel-time data within a Bayesian framework using a reversible-jump Markov chain Monte Carlo (MCMC) simulation. *Transportation Research Part B: Methodological* **81**: 1–17.

Li J, Huang Y, Yang W (2016) Mathematical analysis and finite element simulation of a magnetized ferrite model. *Journal of Computational and Applied Mathematics* **292**: 279–291.

Mares C, Mottershead J, Friswell MI (2006) Stochastic model updating: part 1 – Theory and simulated example. *Mechanical Systems and Signal Processing* **20**: 1674–1695.

Marwala T (1997) Multi-criteria method for determining damage on structures. MEng thesis, University of Pretoria.

Marwala T (2004) Finite element model updating using response surface method. Proceedings of the 45th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, pp. 5165–5173.

Marwala T (2010) *Finite Element Model Updating Using Computational Intelligence Techniques*. London: Springer-Verlag.

Marwala T, Heyns PS (1998) A multiple criterion method for detecting damage on structures. *American Institute of Aeronautics and Astronautics Journal* **195**: 1494–1501.

Marwala T, Sibisi S (2005) Finite element model updating using Bayesian approach. *Proceedings of the IMAC XXIII: A Conference & Exposition on Structural Dynamics*. Bethel, CT: Society for Experimental Mechanics.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**: 1087–1092.

Moens D, Vandepitte D (2006) Recent advances in non-probabilistic approaches for non-deterministic dynamic finite element analysis. *Archives of Computational Methods in Engineering* **13**: 389–464.

Mthembu L, Marwala T, Friswell MI, Adhikari S (2011a) Finite element model selection using particle swarm optimization. In *Dynamics of Civil Structures, Volume 4*. Conference Proceedings of the Society for Experimental Mechanics Series, pp. 41–52. New York: Springer.

Mthembu L, Marwala T, Friswell MI, Adhikari S (2011b) Model selection in finite element model updating using the Bayesian evidence statistic. *Mechanical Systems and Signal Processing* **25**: 2399–2412.

Muto M, Beck JL (2008) Bayesian updating and model class selection for hysteretic structural models using stochastic simulation. *Journal of Vibration and Control* **14**:7–34.

Neal RM (2001) Annealed importance sampling. *Statistics and Computing* **11**: 125–139.

Neal RM (2011) MCMC using Hamiltonian dynamics. In Brooks S, Gelman A, Jones GL, and Meng X-L, eds, *Handbook of Markov Chain Monte Carlo*. Boca Raton, FL: Chapman & Hall/CRC, pp. 113–162.

Nelder JA, Mead R (1965) A simplex method for function minimization. *Computer Journal* **7**: 308–313.

Neyman J (1937) Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London* **A236**: 333–380.

Nichols J, Link W, Murphy K, Olson C (2010) A Bayesian approach to identifying structural nonlinearity using free-decay response: application to damage detection in composites. *Journal of Sound and Vibration* **329**: 2995–3007.

Oñate E (2009) *Structural Analysis with the Finite Element Method. Linear Statics. Vol. 1: Basis and Solids*. Dordrecht: Springer.

Pawlak Z (1982) Rough sets. *International Journal of Parallel Programming* **11**: 341–356.

Rao SS (2004) *The Finite Element Method in Engineering* (4th edn). Amsterdam: Elsevier/Butterworth-Heinemann.

Rathmanner S, Hutter M (2011) A philosophical treatise of universal induction. *Entropy* **13**: 1076–1136.

Resnick R, Eisberg R (1985) *Quantum Physics of Atoms, Molecules, Solids, Nuclei and Particles* (2nd edn). New York: John Wiley & Sons, Inc.

Sanayei M, Khaloo A, Gul M, NecatiCatbas F (2015) Automated finite element model updating of a scale bridge model using measured static and modal test data. *Engineering Structures* **102**: 66–79.

Seifi R, Abbasi K (2015) Friction coefficient estimation in shaft/bush interference using finite element model updating. *Engineering Failure Analysis* **57**: 310–322.

Shan D, Li Q, Khan I, Zhou X (2015) A novel finite element model updating method based on substructure and response surface model. *Engineering Structures* **103**: 147–156.

Venditto JG, Wolf S, Curotto E, Mella M (2015) Replica exchange hybrid Monte Carlo simulations of the ammonia dodecamer and hexadecamer. *Chemical Physics Letters* **635**: 127–133.

Wang F, Li X, Lu M, Xiao Z (2014) Robust abrupt motion tracking via adaptive Hamiltonian Monte Carlo sampling. In Pham D-N, Park S-B, eds, *PRICAI 2014: Trends in Artificial Intelligence*. Lecture Notes in Computer Science **8862**, pp. 52–63. Cham: Springer.

Wang Z, Mohamed S, De Freitas, N (2013) Adaptive Hamiltonian and Riemann manifold Monte Carlo samplers. 30th International Conference on Machine Learning (Part 3), pp. 2512–2520.

Wu S-J, Chu MT (2015) Constructing optimal transition matrix for Markov chain Monte Carlo. *Linear Algebra and its Applications* **487**: 184–202.

Yuen KV (2010) *Bayesian Methods for Structural Dynamics and Civil Engineering*. Hoboken, NJ: John Wiley & Sons, Inc.

Zhang JA, Chen Z, Cheng P, Huang X (2016) Multiple-measurement vector based implementation for single-measurement vector sparse Bayesian learning with reduced complexity. *Signal Processing* **118**: 153–158.

12

Conclusions and Further Work

12.1 Introduction

This book has studied model selection in finite element model updating. It has introduced various methods for selecting the finite element model that best reflects the measured data. The principle of Occam's razor principle states that the simplest model that explains the observed data is the most appropriate one. Additionally, this book has studied the criteria for model selection, including the Akaike information criterion, optimal design, statistical hypothesis testing, the Bayes factor, structural risk minimisation, cross-validation and the Bayesian information criterion, within the context of finite element model updating. Furthermore, this book has applied nested sampling, cross-validation and regularisation techniques for model selection in mechanical structures.

The book introduced Bayesian statistics within the context of structural mechanics for finite element model updating. In this regard, the distribution of the measured natural frequency and the mass were assumed to be known, and Bayesian statistics was applied to approximate the distribution of the stiffness. Furthermore, the book applied the Markov chain Monte Carlo (MCMC) method and Bayesian statistics for finite element model updating. MCMC was used for computationally sampling a probability distribution function based on the Markov process, random walk and Monte Carlo simulation. Two methods were used to update a finite element model of a mechanical structure: Metropolis–Hastings and slice sampling. The slice sampling technique was found to give an adaptive step size, which was automatically adjusted to match the characteristics of the posterior distribution function. The slice sampling method was operated by sampling uniformly from the area under the posterior distribution function.

This book also used a probabilistic technique to minimise the distance between the models and measurements and consequently to model uncertainties. Finite element model updating was conducted using the knowledge learnt using the observation of stochastic characteristics

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

of finite element model parameters, and this was realised through the use of Bayesian statistics. The Monte Carlo dynamically weighted importance sampling algorithm was used to estimate unbiased estimates which were regulated within a targeted reduced variance range from the drawn samples. By using population control techniques, the system was made adaptive to the requirements of the environment and fixing the posterior distribution function. Monte Carlo dynamically weighted importance sampling was compared to the Metropolis–Hastings method and found to be more computationally efficient and accurate.

The adaptive Metropolis–Hastings procedure and Bayesian statistics were also applied for finite element model updating. In the adaptive Metropolis–Hastings technique, the Gaussian proposal distribution was adapted using the full information collected hitherto; because of the adaptive features of the method, this technique is non-Markovian but possesses full ergodic properties. The technique was found to be simple to use, and it was used to update a finite element model of a cantilevered beam and H-shaped structure. The results obtained compared favourably to the results from the Monte Carlo dynamically weighted importance sampling method.

Furthermore, this book presented the hybrid Monte Carlo (HMC) technique for finite element model updating of a truss. HMC is a MCMC method which operates by making the transition between two states using the Metropolis algorithm, the gradient of the energy function and Monte Carlo simulation. This transition between two states, in the context of this book, is the transition between a finite element model and the next, where the difference between the models is the state of the parameters to be updated. In this book, it was observed that the HMC method gave satisfactory results when tested on a cantilevered beam and an asymmetrical H-shaped structure.

The MCMC is operated by moving from one state to another using the random walk technique, where the transition between one state and another uses the Markov chain, and the acceptance of a state is determined using the Metropolis–Hastings technique. The HMC method uses the gradient information to move from one state to another to increase the acceptance rate. The HMC is a hybrid of the Monte Carlo procedure and the gradient descent optimisation method, and it is implemented by approximating the Hamiltonian, which is the sum of the potential energy (position) and the kinetic energy (velocity) of a system. The acceptance rate of the HMC depends on the system size and the time step applied to sample the molecular dynamics trajectory. To handle this constraint, the shadow hybrid Monte Carlo (SHMC) algorithm was applied; this samples large system sizes and time steps by sampling from a modified Hamiltonian function, as an alternative to the normal Hamiltonian function. The SHMC was applied to update a finite element model of an aircraft structure and found to perform better than the HMC.

Next to be applied was the separable shadow Hamiltonian Monte Carlo (S2HMC) algorithm for finite element model updating problems. This is a simplified model in which the Hamiltonian function used is a separable modified function. Two examples were used to test the procedure on finite element model updating: the H-shaped beam and the aircraft structure. It was found that the S2HMC algorithm converged faster and was more efficient than the HMC and SHMC methods.

In finite element model updating and in the situations where the system has many degrees of freedom this requires concurrent and efficient sampling of the uncertain free parameters from multiple unknown probability distributions. The MCMC methods are approaches for sampling a complex probability distribution function. The MCMC can be improved by producing several

random models of a system and evolving them concurrently over time, rather than improving one model in the form of a chain. By applying ideas from Darwin's theory of evolution, finite element model updating can be constructed to converge to a globally optimal model. A number of evolution-based approaches for finite element model updating have been used and, in this book, an evolutionary Markov chain Monte Carlo (EMCMC) algorithm was used to update finite element models. This method combined concepts from genetic algorithms, simulated annealing and MCMC methods. The EMCMC is a sampling technique in which genetic operators, mutation and crossover, are used to design the Markov chain, which gives samples that approximate probability distributions. This method was used to update a beam structure, and the results compared favourably to the results from the Metropolis–Hastings technique.

12.2 Further Work

Further work is required on a number of techniques in finite element model updating; these include sequential Monte Carlo, reversible jump Monte Carlo, Markov chain quasi-Monte Carlo, multiple-try Metropolis–Hastings and dynamic programming. These methods are briefly described in this concluding section.

12.2.1 Reversible Jump Monte Carlo

The reversible jump MCMC method (Green, 1995) has been successfully applied in a number of complex problems (Fox *et al*., 2015; Lee and Sohn, 2015; Drovandi *et al.*, 2014). As a sampling procedure it is an extension of MCMC, where the number of parameters to be updated is variable. In this way, reversible jump MCMC is able to automatically identify the parameters of the finite element model that need to be updated. The acceptance probability is (Green, 1995)

$$a(q,q') = \min\left(1, \frac{p_{q'q}p_{q'}f_{q'}(q')}{p_{qq'}N_{qq'}p_q f_q(q)}\right)\left|\det\left(\frac{\partial h_{qq'}(q,u)}{\partial(q,u)}\right)\right|. \tag{12.1}$$

Here, q' is the proposal, $h_{qq'}$ is the mapping of q' and q, u is a random component of U drawn from a distribution $\mathcal{N}$, det is the determinant, |.| is the absolute value, $p_q f_q$ is the joint posterior probability $p_q f_q = c^{-1}p(y|q,r_q)p(r_q)$ and c is the normalisation constant.

12.2.2 Multiple-Try Metropolis–Hastings

In order to speed up the convergence of the updating process, a multiple-try Metropolis–Hastings algorithm can be used. This technique was proposed by Liu *et al*. (2000), and is a modified version of the Metropolis–Hastings technique, which accelerates the sampling trajectory by increasing both the step size and the acceptance rate. The acceptance criterion for this method can be written as (Liu *et al*., 2000)

$$r = \min\left(1, \frac{w(y_1,\mathbf{x}) + \ldots + w(y_k,\mathbf{x})}{w(x_1,\mathbf{y}) + \ldots + w(x_k,\mathbf{y})}\right), \tag{12.2}$$

where $\mathbf{x}$ is the state and $\mathbf{y}$ is the proposal.

12.2.3 Dynamic Programming

Dynamic programming is an optimisation technique for identifying optimal solutions to problems with features of overlapping sub-problems and optimal sub-structures, for finite element updating (Bellman, 1957). Dynamic programming has been successfully applied in energy scheduling (Xu *et al*., 2015), wind farm power system stability (Guo *et al.*, 2015), pattern recognition in a hybrid electric car (Zhang and Xiong, 2015) and missing data estimation (Nelwamondo *et al*., 2013). Usually, problems that exhibit optimal sub-structures can be solved by using a three-step method of dividing the finite element updating problem into sub-problems, solving each sub-problem optimally by using the three-step procedure iteratively and using the optimal solutions from these sub-problems to identify an optimal solution for the entire problem (Bertsekas, 2000). The sub-problems are solved by breaking them into sub-sub-problems until a simple problem is found that can easily be solved. The necessary condition for optimality of the global problem given the optimal solutions for sub-problems in dynamic programming is ensured by the use of the Bellman equation (Bellman, 1957):

$$V(x) = \max_{y \in \Im(x)} [F(x,y) + \beta V(y)], \quad \forall x \in X, \tag{12.3}$$

where V is the value function, x is the finite element updating design variable and $y(x)$ is the objective function.

12.2.4 Sequential Monte Carlo

Another technique which can be applied for finite element model updating is the sequential Monte Carlo method. This method is particularly applicable for situations where the system being modelled is changing, and there is a need to update a finite element model sequentially (Doucet *et al*., 2000). Sequential Monte Carlo techniques are Monte Carlo methods that are very similar to genetic algorithms. They have been successfully applied to problems for tracking (Cherry *et al.*, 2015), capital allocation (Targino *et al*., 2015) and automotive batteries (Li *et al.*, 2015).

References

Bellman R (1957) *Dynamic Programming*. Princeton, NJ: Princeton University Press.

Bertsekas DP (2000) *Dynamic Programming and Optimal Control*. Belmont, MA: Athena Scientific.

Cherry KM, Peplinski B, Kim L, Wang S, Lu L, Zhang W, Liu J, Wei Z, Summers RM (2015) Sequential Monte Carlo tracking of the marginal artery by multiple cue fusion and random forest regression. *Medical Image Analysis* **19**: 164–175.

Doucet A, Godsill S, Andrieu C (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* **10**: 197–208.

Drovandi CC, Pettitt AN, Henderson RD, McCombe PA (2014) Marginal reversible jump Markov chain Monte Carlo with application to motor unit number estimation. *Computational Statistics & Data Analysis* **72**: 128–146.

Fox M, Bodin T, Shuster DL (2015) Abrupt changes in the rate of Andean Plateau uplift from reversible jump Markov chain Monte Carlo inversion of river profiles. *Geomorphology* **238**: 1–14.

Green PJ (1995) Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**: 711–732.

Guo W, Liu F, Si J, He D, Harley R, Mei S (2015) Approximate dynamic programming based supplementary reactive power control for DFIG wind farm to enhance power system stability. *Neurocomputing* **170**: 417–427.

Lee M, Sohn K (2015) Inferring the route-use patterns of metro passengers based only on travel-time data within a Bayesian framework using a reversible-jump Markov chain Monte Carlo (MCMC) simulation. *Transportation Research Part B: Methodological* **81**: 1–17.

Li J, Barillas JK, Guenther C, Danzer MA (2015) Multi-cell state estimation using variation based sequential Monte Carlo filter for automotive battery packs. *Journal of Power Sources* **277**: 95–103.

Liu JS, Liang F, Wong WH (2000) The multiple-try method and local optimization in Metropolis sampling. *Journal of the American Statistical Association* **95**: 121–134.

Nelwamondo FV, Golding D, Marwala T (2013) A dynamic programming approach to missing data estimation using neural networks. *Information Sciences* **237**: 49–58.

Targino RS, Peters GW, Shevchenko PV (2015) Sequential Monte Carlo samplers for capital allocation under copula-dependent risk models. *Insurance: Mathematics and Economics* **61**: 206–226.

Xu Y, Liu D, Wei Q (2015) Action dependent heuristic dynamic programming based residential energy scheduling with home energy inter-exchange. *Energy Conversion and Management* **103**: 553–561.

Zhang S, Xiong R (2015) Adaptive energy management of a plug-in hybrid electric vehicle based on driving pattern recognition and dynamic programming. *Applied Energy* **155**: 68–78.

Appendix A

Experimental Examples

A.1 Cantilevered Beam

The experimental cantilevered steel beam was measured by Kraaij (2006) and was found to have the following characteristics: length 500 mm, width 60 mm, and thickness 10 mm, $E = 2.1 \times 10^{11}\,\mathrm{N/m^2}$, $v = 0.3$ and $\rho = 7850\,\mathrm{kg/m^3}$. Three accelerometers were used, located 490 mm from the clamped location, to take the measurements and this position was selected because of the large response (Kraaij, 2006; Boulkaibet *et al.*, 2012; Boulkaibet, 2014). Each accelerometer had a mass of 40 g; the middle accelerometer was of type 303*A*3 and the outer accelerometers were of type 303*A*2 (Kraaij, 2006). The cantilevered beam was modelled using the Structural Dynamics Toolbox SDT® for MATLAB, and the beam was segmented into 50 Euler–Bernoulli beam elements and excited at a number of positions (Boulkaibet *et al.*, 2012). The measured natural frequencies of interest were 31.9, 197.9, 553, 1082.2 and 1781.5 Hz. Figures A.1–A.5 show five modes of the modelled cantilevered beam.

A.2 H-Shaped Structure Simulation

In this example, the asymmetrical H-shaped aluminium structure shown in Figure A.6 was considered. The model of this structure was segmented into 12 elements, each modelled using the Euler–Bernoulli beam using the Structural Dynamics Toolbox SDT® for MATLAB (Marwala, 1997). The structure was excited at the location specified by the double arrow, and the acceleration was measured at 15 different locations. The structure was excited using an electromagnetic shaker, while a roving accelerometer was used to measure the response. A set of 15 frequency-response functions were computed, and the measured natural frequencies were

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

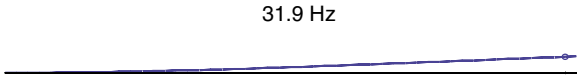

Figure A.1 Cantilevered beam mode 1

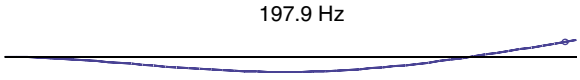

Figure A.2 Cantilevered beam mode 2

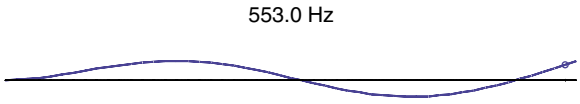

Figure A.3 Cantilevered beam mode 3

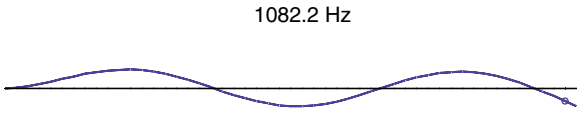

Figure A.4 Cantilevered beam mode 4

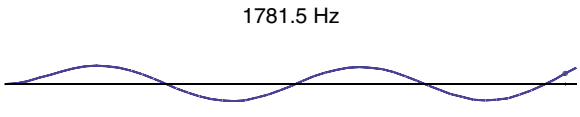

Figure A.5 Cantilevered beam mode 5

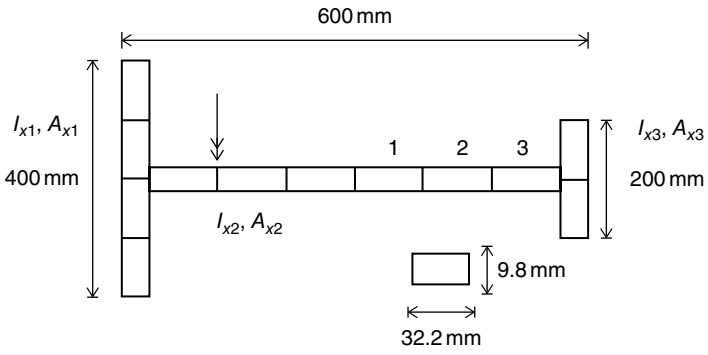

Figure A.6 The H-shaped aluminium structure

53.9, 117.3, 208.4, 254.0 and 445.0 Hz, with modes indicated in Figures A.7–A.11. In this example, the moments of inertia and the cross-sectional areas of the left, middle and right subsections of the beam were updated, and the updating parameter vector was $\boldsymbol{\theta} = \{I_{x1}, I_{x2}, I_{x3}, A_{x1}, A_{x2}, A_{x3}\}$. The Young's modulus for the beam was specified as $7.2 \times 10^{10}\,\mathrm{N/m^2}$, and the density was specified as 2785 $\mathrm{kg/m^3}$.

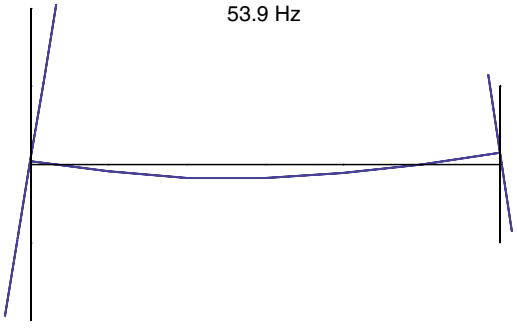

Figure A.7 H-shaped aluminium structure mode 1

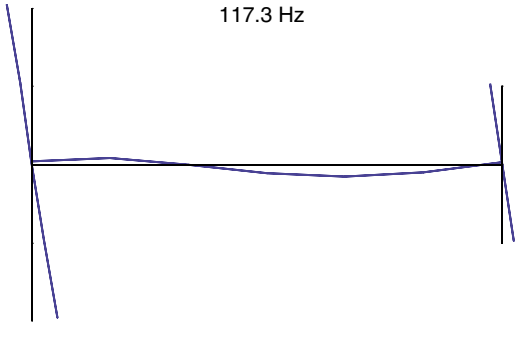

Figure A.8 H-shaped aluminium structure mode 2

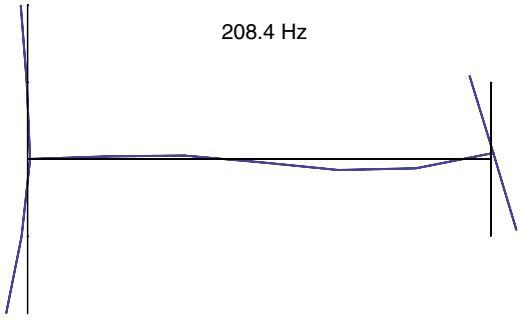

Figure A.9 H-shaped aluminium structure mode 3

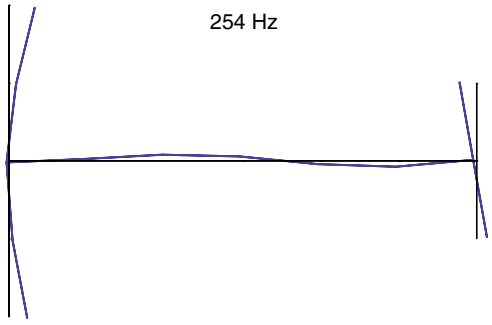

Figure A.10 H-shaped aluminium structure mode 4

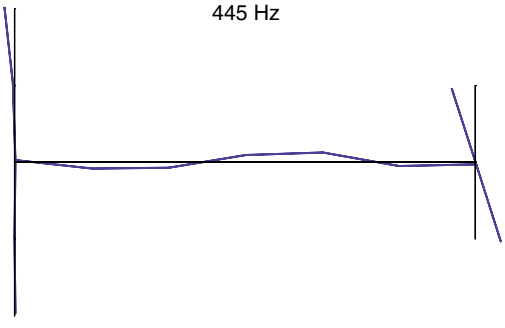

Figure A.11 H-shaped aluminium structure mode 5

A.3 GARTEUR SM-AG19 Structure

The GARTEUR SM-AG19 structure was used as a benchmark study by 12 members of the GARTEUR Structures and Materials Action Group 19 (Degener and Hermes, 1996; Balmes, 1998; Datta, 2002; Guyon and Elisseeff, 2003; Link and Friswell, 2003; Carvalho *et al.*, 2007). The aeroplane was 1.5 m long, 3 m wide, with a fuselage 15 cm deep and 5 cm thick. The material used was aluminium and the overall mass was 44 kg, and a $1.1 \times 76.2 \times 1700\,\text{mm}^3$ visco-elastic constraining layer was bonded to the wings to increase damping. Euler–Bernoulli elements were used and all element materials were deemed to be standard isotropic. The measured natural frequencies were 6.38, 16.10, 33.13, 33.53, 35.65, 48.38, 49.43, 55.08, 63.04 and 66.52 Hz, and the mode shapes are shown in Figures A.12–A.21. The parameters of the structure to be updated were the right wing moments of inertia and torsional stiffness ($RI_{\min}, RI_{\max}, RI_{\text{tors}}$), the left wing moments of inertia and torsional stiffness ($LI_{\min}, LI_{\max}, LI_{\text{tors}}$), the vertical tail moment of inertia ($VTP_{I_{\min}}$), and the overall structure density ρ. The temperature was set at $T = 300$ K and $\beta_B = 1/300K_B$, where $K_B = 0.00198719$ kcal mol^{-1} K^{-1}. The updated vector was $\boldsymbol{\theta} = [\rho, VTP_{I_{\min}}, LI_{\min}, LI_{\max}, RI_{\min}, RI_{\max}, LI_{\text{tors}}, RI_{\text{tors}}]$. The Young's modulus for the structure was set at 7.2×10^{10} N/m^2.

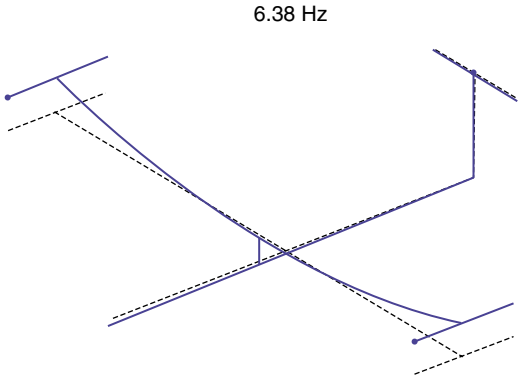

Figure A.12 Aeroplane structure mode 1

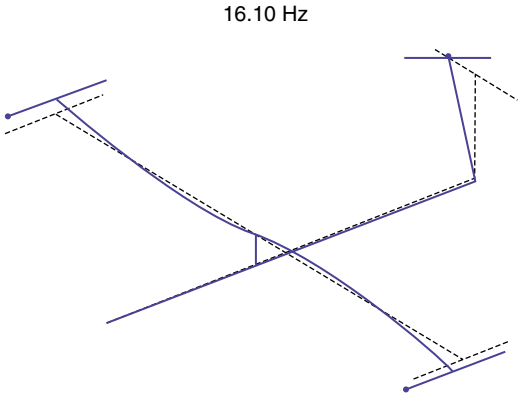

Figure A.13 Aeroplane structure mode 2

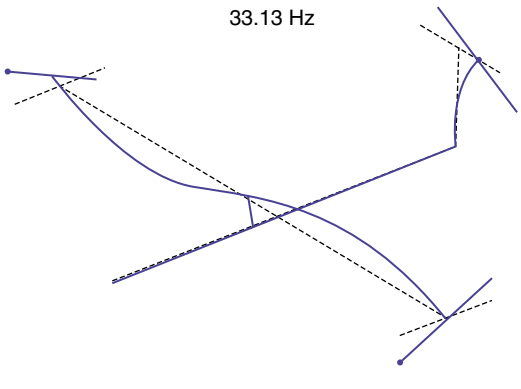

Figure A.14 Aeroplane structure mode 3

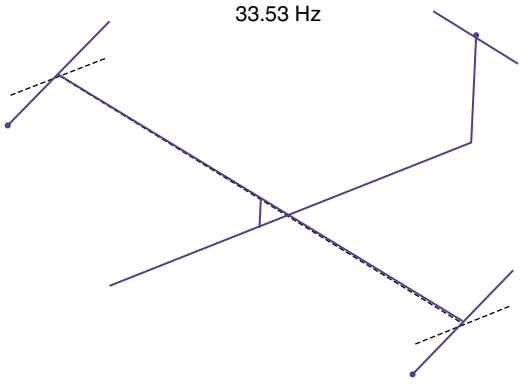

Figure A.15 Aeroplane structure mode 4

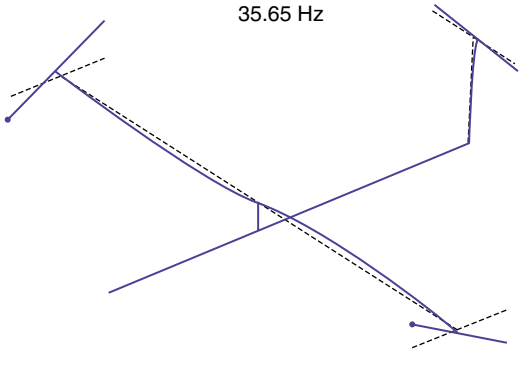

Figure A.16 Aeroplane structure mode 5

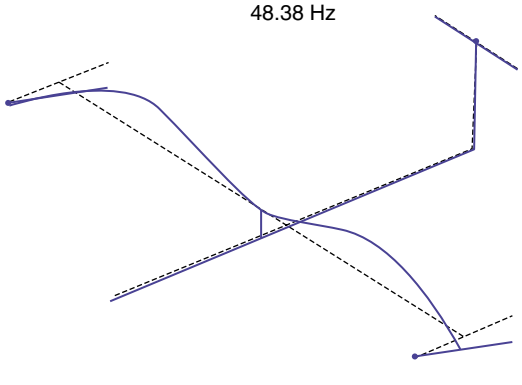

Figure A.17 Aeroplane structure mode 6

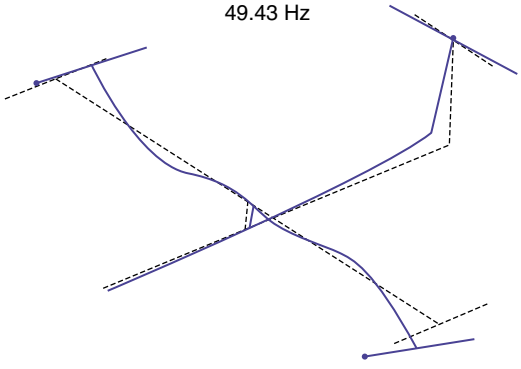

Figure A.18 Aeroplane structure mode 7

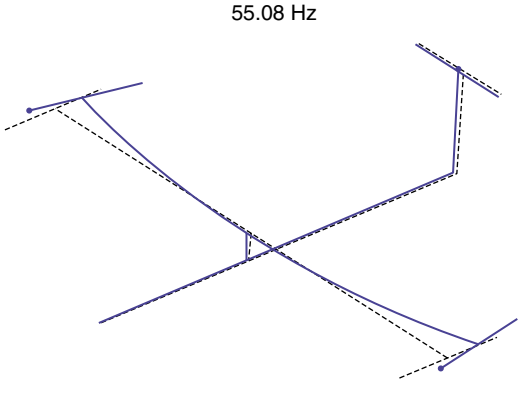

Figure A.19 Aeroplane structure mode 8

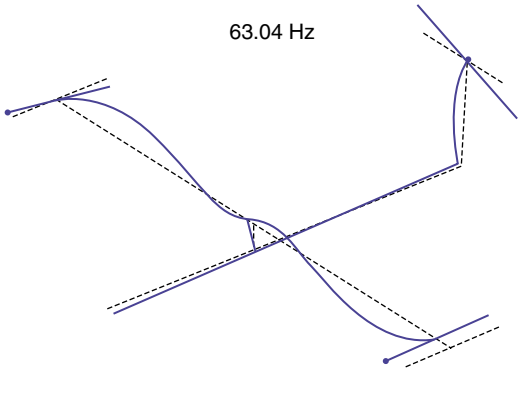

Figure A.20 Aeroplane structure mode 9

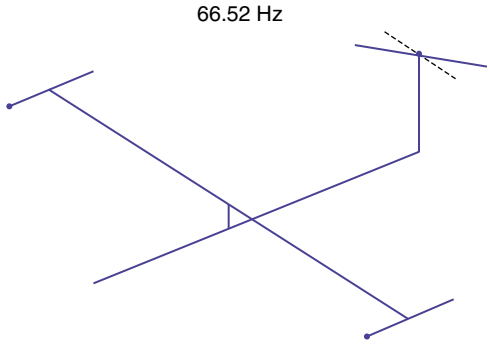

Figure A.21 Aeroplane structure mode 10

References

Balmes E (1998) Predicted variability and differences between tests of a single structure. Proceedings of the 16th International Modal Analysis Conference, Bethel, CT: SPIE, pp. 558–564.

Boulkaibet I (2014) Finite element model updating using Markov Chain Monte Carlo techniques. PhD thesis, University of Johannesburg.

Boulkaibet I, Marwala T, Mthembu L, Friswell MI, Adhikari S (2012) Sampling techniques in Bayesian finite element model updating. *Proceedings of the Society for Experimental Mechanics* **29**: 75–83.

Carvalho J, Datta BN, Gupta A, Lagadapati M (2007) A direct method for model updating with incomplete measured data and without spurious modes. *Mechanical Systems and Signal Processing* **21**: 2715–2731.

Datta BN (2002) Finite element model updating, eigenstructure assignment and eigenvalue embedding techniques for vibrating systems. *Mechanical Systems and Signal Processing* **16**: 83–96.

Degener M, Hermes M (1996) Ground vibration test and finite element analysis of the GARTEUR SM-AG19 testbed. Report IB 232-96J 08, Deutsche Forschungsanstalt für Luft und Raumfahrt e.V. Institut für Aeroelastik.

Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *Journal of Machine Learning Research* **3**: 1157–1182.

Kraaij CS (2006) Model updating of a 'clamped' free beam system using FEM tools. Technical Report DCT2006.128, Technische Universiteit Eindhoven.

Link M, Friswell MI (2003) Generation of validated structural dynamic models – results of a benchmark study utilizing the GARTEUR SM-AG19 testbed. *Mechanical Systems and Signal Processing* **17**: 9–20.

Marwala, T. (1997) A multiple criterion updating method for damage detection on structures. MEng thesis, University of Pretoria.

Appendix B

Markov Chain Monte Carlo

B.1 Introduction

Some types of sampling approaches are grounded in the notion of Markov chains, which tend to converge to a particular specified invariant distribution (Boulkaibet, 2014). Therefore, this approach can be applied to approximate expectations and correlations of this distribution. In this book, sampling procedures are founded on the concept of a Markov chain. Section B.2 explains the Markov chain theory, and Section B.3 deliberates on the resulting invariant probability distribution. Section B.4 describes the fundamentals of reversibility and ergodicity of a Markov chain.

B.2 Basic Definition of the Markov Chain

A series of random variables denoted by $z^{(1)}, \ldots, z^{(M)}$ is deemed to be a Markov chain if the following conditional independence characteristics are true for $m \in \{1, \ldots, M-1\}$ (Bishop, 2006; Boulkaibet, 2015):

$$p\left(z^{(m+1)}|z^{(1)}, \ldots, z^{(m)}\right) = p\left(z^{(m+1)}|z^{(m)}\right). \tag{B.1}$$

Equation B.1 can also be considered to be a directed graph expressed as a chain. The Markov chain can then be quantified by the probability distribution for the variable $p(z^{(0)})$ at time 0 and the conditional probabilities $T_m\left(z^{(m+1)}|z^{(m)}\right)p\left(z^{(m+1)}|z^{(m)}\right)$, which are also called the *transition* probabilities.

It should be borne in mind that if all transition probabilities do not change for all m then the Markov chain is considered to be *homogeneous* (Bishop, 2006). The marginal probability for a

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

specific variable can be characterised by all marginal probabilities for the preceding variables in the chain, and the new expression for this is as follows (Bishop, 2006; Boulkaibet, 2015):

$$p\left(z^{(m+1)}\right) = \sum_{z^{(m)}} p\left(z^{(m+1)} | z^{(m)}\right) \cdot p\left(z^{(m)}\right). \tag{B.2}$$

This equation can define the chain characteristics for all stages in a situation where the initial probabilities $p(z^{(0)})$ are specified.

B.3 Invariant Distribution

A probability distribution is considered *invariant* or *stationary,* in relation to a Markov chain, if it remains invariant at each iteration in the Markov chain. In the situation of a homogeneous Markov chain which is prescribed by the transition probabilities $T(z', z)$, the distribution $p^*(z)$ is considered invariant if (Bishop, 2006; Boulkaibet, 2015):

$$p^*(z) = \sum_{z'} T(z', z) \cdot p^*(z') \tag{B.3}$$

To ensure that the distribution $p^*(z)$ is invariant, the transition probabilities are chosen to satisfy the property of *detailed balance,* which is defined for a particular distribution $p^*(z)$ as follows (Bishop, 2006; Boulkaibet, 2015):

$$p^*(z) \cdot T(z, z') = p^*(z') \cdot T(z', z). \tag{B.4}$$

It should be borne in mind that a transition probability that obeys a detailed balance in relation to a specific distribution maintains an invariant distribution where (Bishop, 2006; Boulkaibet, 2015):

$$\sum_{z'} p^*(z') \cdot T(z', z) = \sum_{z'} p^*(z) \cdot T(z, z') = p^*(z) \sum_{z'} p(z' | z) = p^*(z). \tag{B.5}$$

B.4 Reversibility and Ergodicity

A Markov chain that fulfils the detailed balance is also considered to be *reversible.* From a specified distribution, a Markov chain is intended to randomly select samples, and this is done by creating a Markov chain in such a way that the sought distribution is invariant. However, it is essential that for $m \to \infty$, the sought distribution $p(z^{(m)})$ converges to $p^*(z)$, which is the desired invariant distribution, notwithstanding the selection of the initial distribution $p(z^{(0)})$. This characteristic is known as the *ergodicity* and, therefore, the invariant distribution is an *equilibrium* distribution (Bishop, 2006; Boulkaibet, 2015). An ergodic Markov chain only has one equilibrium distribution, and it can be demonstrated that a homogeneous Markov chain is ergodic as long as there exist weak limitations on the invariant distribution and the transition probabilities

(Neal, 2003). Actually, the transition probabilities are constructed from a class of 'base' transitions $B_1, \ldots, B_K$ using a mixture distribution that can be expressed as follows (Bishop, 2006; Boulkaibet, 2015):

$$T(z',z) = \sum_{k=1}^{K} \alpha_k B_K(z',z). \tag{B.6}$$

Here, a particular class of mixing coefficients $\alpha_1, \ldots, \alpha_K$ satisfy $\alpha_k \geq 0$ and $\sum_k \alpha_k = 1$; otherwise the supporting transitions may be hybridised using consecutive applications where (Bishop, 2006; Boulkaibet, 2015):

$$T(z',z) = \sum_{z_1} \ldots \sum_{z_{n-1}} B_1(z',z_1) \ldots B_{K-1}(z_{K-2},z_{K-1}) B_K(z_{K-1},z). \tag{B.7}$$

If it is true that a distribution is invariant in relation to each of the base transitions, then it is likewise invariant with respect to either $T(z', z)$ as expressed by Equations B.6 and B.7. For the situation of the mixture which is expressed in Equation B.7, the mixture transition T satisfies the detailed balance if each of these base transitions satisfies the detailed balance.

References

Bishop CM (2006) *Pattern Recognition and Machine Learning*. New York: Springer-Verlag.

Boulkaibet, I. (2014) Finite element model updating using the Markov Chain Monte Carlo technique. PhD thesis, University of Johannesburg.

Neal RM (2003) Slice sampling. *Annals of Statistics* **31**: 705–741.

Appendix C

Gaussian Distribution

C.1 Introduction

For a dynamic system, the uncertain parameters are considered as a probabilistic vector $\boldsymbol{\theta}$, and the uncertainty is expressed in terms of a joint density function $f(\boldsymbol{\theta})$ with d variables (Bishop, 2006; Boulkaibet, 2014). The basic parametric function class may be more useful for complex systems, for example the Gaussian or normal distribution. In consequence, the uncertainty of the parameters is characterised by the mean and the standard deviation values, while the dependence of these parameters can be expressed using a correlation or covariance matrix. Section C.2 briefly describes the Gaussian distribution, and Section C.3 describes particular properties of the Gaussian probability distribution function.

C.2 Gaussian Distribution

The Gaussian density distribution is a probabilistic continuous function which has been comprehensively implemented to model uncertain parameters in system dynamics (Bishop, 2006; Boulkaibet, 2015). The Gaussian distribution for a single variable θ can be written as

$$\mathcal{N}\left(\theta|\mu,\sigma^2\right) = \frac{1}{\left(2\pi\sigma^2\right)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(\theta-\mu)^2\right\}, \tag{C.1}$$

where μ is the mean value, σ^2 denotes the variance and σ is the standard deviation. In the situation of a multi-dimensional vector $\boldsymbol{\theta}$ (of dimension d), the multivariate Gaussian distribution can be expressed as

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.

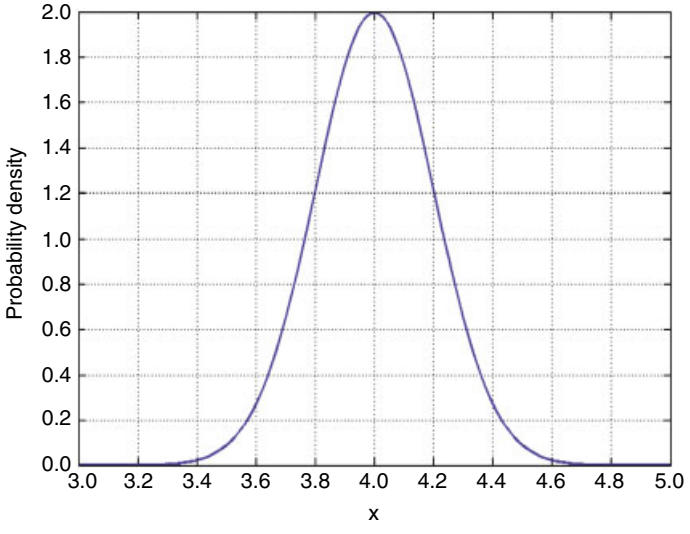

Figure C.1 Graph of a probability distribution function for a univariate Gaussian distribution function

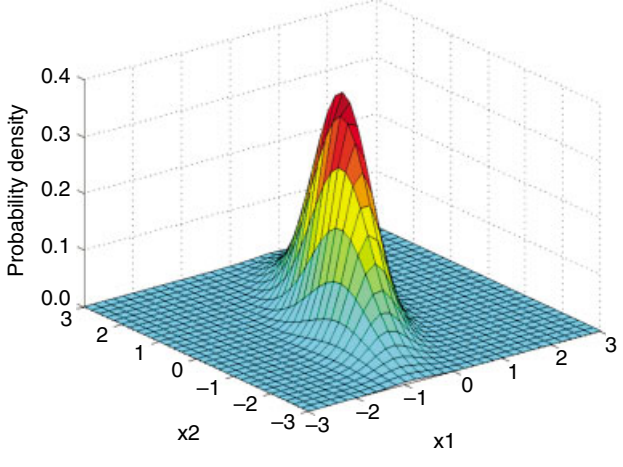

Figure C.2 Graph of a probability distribution function graph for a bivariate Gaussian distribution function

$$\mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\boldsymbol{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{\theta}-\boldsymbol{\mu})\right\}, \tag{C.2}$$

where now $\boldsymbol{\mu}$ is a d-dimensional mean vector, $\boldsymbol{\Sigma}$ is a $d \times d$ covariance matrix and $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$. Figure C.1 shows a univariate function with $\mu = 4$ and $\sigma = 0.2$, while Figure C.2 shows a bivariate normal distribution with $\boldsymbol{\mu} = [0, 0]$ and standard deviation $\boldsymbol{\Sigma} = \begin{bmatrix} 0.25 & 0.3 \\ 0.3 & 1 \end{bmatrix}$

(Boulkaibet, 2015). As demonstrated in these figures, the mean of the Gaussian distribution lies at the maximum of the probability, which is desirable for modelling uncertain parameters.

C.3 Properties of the Gaussian Distribution

The parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are significant properties of the Gaussian distribution. The expectation value or mean of the Gaussian distribution is mathematically expressed as follows (Neal, 2003; Bishop, 2006; Boulkaibet, 2015):

$$\begin{aligned} E(\boldsymbol{\theta}) &= \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \int \exp\left\{ -\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})^T \cdot \Sigma \cdot (\boldsymbol{\theta}-\boldsymbol{\mu}) \right\} \boldsymbol{\theta} d\boldsymbol{\theta} \\ &= \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \int \exp\left\{ -\frac{1}{2} z^T \cdot \boldsymbol{\Sigma} \cdot z \right\} (z+\boldsymbol{\mu}) dz. \end{aligned} \tag{C.3}$$

Equation C.3 is obtained by change of variables using the expression $z = \boldsymbol{\theta} - \boldsymbol{\mu}$. From Equation C.3, it is evident that the exponent is an even function of the components of z. However, the integral in Equation C.3 is calculated over the range $(-\infty, \infty)$, which signifies that the terms in z in the factor $z + \boldsymbol{\mu}$ cancel one another because of the symmetrical nature of the expression. Consequently, the resulting quantity of the expectation value is expressed as follows (Chib and Greenberg, 1995):

$$E(\boldsymbol{\theta}) = \boldsymbol{\mu}, \tag{C.4}$$

where $\boldsymbol{\mu}$ is the mean of the Gaussian distribution.

To derive the covariance matrix, the second-order moments of the Gaussian are used. If we assume that the distribution is univariate, then the second-order moment is expressed by $E(\theta^2)$. However, in the multivariate Gaussian case, there are d^2 second-order moments expressed as (θ_i, θ_j), which can be combined to form the matrix $E\left(\boldsymbol{\theta} \cdot \boldsymbol{\theta}^T\right)$. This matrix is expressed as follows (Bishop, 2006; Boulkaibet, 2015):

$$\begin{aligned} E\left(\boldsymbol{\theta} \cdot \boldsymbol{\theta}^T\right) &= \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \int \exp\left\{ -\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})^T \cdot \boldsymbol{\Sigma} \cdot (\boldsymbol{\theta}-\boldsymbol{\mu}) \right\} \boldsymbol{\theta} \cdot \boldsymbol{\theta}^T d\boldsymbol{\theta} \\ &= \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \int \exp\left\{ -\frac{1}{2} z^T \cdot \boldsymbol{\Sigma} \cdot z \right\} (z+\boldsymbol{\mu}) \cdot (z+\boldsymbol{\mu}) \, dz. \end{aligned} \tag{C.5}$$

Equation C.5 is similarly obtained by change of variables using the expression $z = \boldsymbol{\theta} - \boldsymbol{\mu}$. The cross-terms relating to $\boldsymbol{\mu} z^T$ and $z \boldsymbol{\mu}^T$ cancel each other due to the property of symmetry. The expression $\boldsymbol{\mu} \boldsymbol{\mu}^T$ is a constant and can be factored out of the integral, and its value is unity because the Gaussian distribution is normalised. The expression with $z z^T$ is written as follows (Bishop, 2006; Boulkaibet, 2015):

$$\frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \int \exp\left\{ -\frac{1}{2} z^T \cdot \boldsymbol{\Sigma} \cdot z \right\} z \cdot z dz = \boldsymbol{\Sigma}. \tag{C.6}$$

The result of Equation C.5 is expressed as follows (Bishop, 2006; Boulkaibet, 2015):

$$E\left(\boldsymbol{\theta}\cdot\boldsymbol{\theta}^{T}\right)=\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^{T}. \tag{C.7}$$

In the class of single random variables, the mean can be subtracted before calculating the second moments in order to simplify the process of obtaining the variance. Likewise, the mean can be subtracted in the multivariate case, and the covariance of a random vector $\boldsymbol{\theta}$ is then given by (Bishop, 2006; Boulkaibet, 2015)

$$\operatorname{cov}(\boldsymbol{\theta})=E\left[(\boldsymbol{\theta}-E(\boldsymbol{\theta}))(\boldsymbol{\theta}-E(\boldsymbol{\theta}))^{\boldsymbol{T}}\right]=\boldsymbol{\Sigma}. \tag{C.8}$$

The parameter matrix $\boldsymbol{\Sigma}$ is the covariance matrix.

References

Bishop CM (2006) *Pattern Recognition and Machine Learning*. New York: Springer-Verlag.

Boulkaibet I (2014) Finite element model updating using the Markov chain Monte Carlo technique. PhD thesis, University of Johannesburg.

Chib S,Greenberg E (1995) Understanding the Metropolis–Hastings algorithm. *American Statistician* **49**: 327–335.

Neal RM (2003) Slice sampling. *Annals of Statistics* **31**: 705–741.

Index

Probabilistic Finite Element Model Updating Using Bayesian Statistics: Applications to Aeronautical and Mechanical Engineering, First Edition. Tshilidzi Marwala, Ilyes Boulkaibet and Sondipon Adhikari.