Bayesian Methods for Structural Dynamics and Civil Engineering

Ka-Veng Yuen

WILEY

www.EngineeringBooksPdf.com

www.EngineeringBooksPdf.com

BAYESIAN METHODS FOR STRUCTURAL DYNAMICS AND CIVIL ENGINEERING

www.EngineeringBooksPdf.com

www.EngineeringBooksPdf.com

BAYESIAN METHODS FOR STRUCTURAL DYNAMICS AND CIVIL ENGINEERING

Ka-Veng Yuen

Department of Civil and Environmental Engineering
University of Macau

John Wiley & Sons (Asia) Pte Ltd

www.EngineeringBooksPdf.com

Copyright © 2010 John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop, # 02-01, Singapore 129809

Visit our Home Page on www.wiley.com

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as expressly permitted by law, without either the prior written permission of the Publisher, or authorization through payment of the appropriate photocopy fee to the Copyright Clearance Center. Requests for permission should be addressed to the Publisher, John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop, #02-01, Singapore 129809, tel: 65-64632400, fax: 65-64646912, email: enquiry@wiley.com.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The Publisher is not associated with any product or vendor mentioned in this book. All trademarks referred to in the text of this publication are the property of their respective owners.

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstrasse 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons Canada Ltd, 5353 Dundas Street West, Suite 400, Toronto, ONT, M9B 6H8, Canada

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Yuen, Ka-Veng.
Bayesian methods for structural dynamics and civil engineering / Ka-Veng Yuen.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-82454-2 (cloth)
1. Engineering–Statistical methods. 2. Structural engineering–Mathematics. 3. Bayesian statistical decision theory. I. Title.
TA340.Y84 2010
624.101'519542–dc22

2009043690

ISBN 978-0-470-82454-2 (HB)

Typeset in 10/12pt Times by Thomson Digital, Noida, India.
Printed and bound in Singapore by Markono Print Media Pte Ltd, Singapore.
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

www.EngineeringBooksPdf.com

To my family

To my wife, *Vanessa,*
my daughter, *Hoi-Man* and my son, *Kin-Hang*

www.EngineeringBooksPdf.com

www.EngineeringBooksPdf.com

Contents

Preface **xi**

Acknowledgements **xiii**

Nomenclature **xv**

1 Introduction **1**
1.1 Thomas Bayes and Bayesian Methods in Engineering 1
1.2 Purpose of Model Updating 3
1.3 Source of Uncertainty and Bayesian Updating 5
1.4 Organization of the Book 8

2 Basic Concepts and Bayesian Probabilistic Framework **11**
2.1 Conditional Probability and Basic Concepts 12
2.1.1 Bayes' Theorem for Discrete Events 13
2.1.2 Bayes' Theorem for Continuous-valued Parameters by Discrete Events 15
2.1.3 Bayes' Theorem for Discrete Events by Continuous-valued Parameters 17
2.1.4 Bayes' Theorem between Continuous-valued Parameters 18
2.1.5 Bayesian Inference 20
2.1.6 Examples of Bayesian Inference 24
2.2 Bayesian Model Updating with Input–output Measurements 33
2.2.1 Input–output Measurements 33
2.2.2 Bayesian Parametric Identification 34
2.2.3 Model Identifiability 35
2.3 Deterministic versus Probabilistic Methods 40
2.4 Regression Problems 43
2.4.1 Linear Regression Problems 43
2.4.2 Nonlinear Regression Problems 47
2.5 Numerical Representation of the Updated PDF 48
2.5.1 General Form of Reliability Integrals 48
2.5.2 Monte Carlo Simulation 49
2.5.3 Adaptive Markov Chain Monte Carlo Simulation 50
2.5.4 Illustrative Example 54

www.EngineeringBooksPdf.com

2.6 Application to Temperature Effects on Structural Behavior 61
2.6.1 Problem Description 61
2.6.2 Thermal Effects on Modal Frequencies of Buildings 61
2.6.3 Bayesian Regression Analysis 64
2.6.4 Analysis of the Measurements 66
2.6.5 Concluding Remarks 68
2.7 Application to Noise Parameters Selection for the Kalman Filter 68
2.7.1 Problem Description 68
2.7.2 Kalman Filter 68
2.7.3 Illustrative Examples 71
2.8 Application to Prediction of Particulate Matter Concentration 77
2.8.1 Introduction 77
2.8.2 Extended-Kalman-filter based Time-varying Statistical Models 80
2.8.3 Analysis with Monitoring Data 87
2.8.4 Conclusion 98

3 Bayesian Spectral Density Approach 99
3.1 Modal and Model Updating of Dynamical Systems 99
3.2 Random Vibration Analysis 101
3.2.1 Single-degree-of-freedom Systems 101
3.2.2 Multi-degree-of-freedom Systems 102
3.3 Bayesian Spectral Density Approach 104
3.3.1 Formulation for Single-channel Output Measurements 105
3.3.2 Formulation for Multiple-channel Output Measurements 110
3.3.3 Selection of the Frequency Index Set 115
3.3.4 Nonlinear Systems 116
3.4 Numerical Verifications 116
3.4.1 Aliasing and Leakage 117
3.4.2 Identification with the Spectral Density Approach 122
3.4.3 Identification with Small Amount of Data 126
3.4.4 Concluding Remarks 127
3.5 Optimal Sensor Placement 127
3.5.1 Information Entropy with Globally Identifiable Case 128
3.5.2 Optimal Sensor Configuration 129
3.5.3 Robust Information Entropy 130
3.5.4 Discrete Optimization Algorithm for Suboptimal Solution 131
3.6 Updating of a Nonlinear Oscillator 132
3.7 Application to Structural Behavior under Typhoons 138
3.7.1 Problem Description 138
3.7.2 Meteorological Information of the Two Typhoons 140
3.7.3 Analysis of Monitoring Data 142
3.7.4 Concluding Remarks 152
3.8 Application to Hydraulic Jump 152
3.8.1 Problem Description 152
3.8.2 Fundamentals of Hydraulic Jump 153
3.8.3 Roller Formation-advection Model 153

www.EngineeringBooksPdf.com

3.8.4 Statistical Modeling of the Surface Fluctuation 154
3.8.5 Experimental Setup and Results 155
3.8.6 Concluding Remarks 159

4 Bayesian Time-domain Approach 161
4.1 Introduction 161
4.2 Exact Bayesian Formulation and its Computational Difficulties 162
4.3 Random Vibration Analysis of Nonstationary Response 164
4.4 Bayesian Updating with Approximated PDF Expansion 167
4.4.1 Reduced-order Likelihood Function 172
4.4.2 Conditional PDFs 172
4.5 Numerical Verification 174
4.6 Application to Model Updating with Unmeasured Earthquake Ground Motion 179
4.6.1 Transient Response of a Linear Oscillator 179
4.6.2 Building Subjected to Nonstationary Ground Excitation 182
4.7 Concluding Remarks 186
4.8 Comparison of Spectral Density Approach and Time-domain Approach 187
4.9 Extended Readings 189

5 Model Updating Using Eigenvalue–Eigenvector Measurements 193
5.1 Introduction 193
5.2 Formulation 196
5.3 Linear Optimization Problems 198
5.3.1 Optimization for Mode Shapes 199
5.3.2 Optimization for Modal Frequencies 199
5.3.3 Optimization for Model Parameters 200
5.4 Iterative Algorithm 200
5.5 Uncertainty Estimation 201
5.6 Applications to Structural Health Monitoring 202
5.6.1 Twelve-story Shear Building 202
5.6.2 Three-dimensional Six-story Braced Frame 205
5.7 Concluding Remarks 210

6 Bayesian Model Class Selection 213
6.1 Introduction 213
6.1.1 Sensitivity, Data Fitness and Parametric Uncertainty 216
6.2 Bayesian Model Class Selection 219
6.2.1 Globally Identifiable Case 221
6.2.2 General Case 225
6.2.3 Computational Issues: Transitional Markov Chain Monte Carlo Method 228
6.3 Model Class Selection for Regression Problems 229
6.3.1 Linear Regression Problems 229
6.3.2 Nonlinear Regression Problems 234

www.EngineeringBooksPdf.com

6.4 Application to Modal Updating 235
6.5 Application to Seismic Attenuation Empirical Relationship 238
6.5.1 Problem Description 238
6.5.2 Selection of the Predictive Model Class 239
6.5.3 Analysis with Strong Ground Motion Measurements 241
6.5.4 Concluding Remarks 249
6.6 Prior Distributions – Revisited 250
6.7 Final Remarks 252

Appendix A: Relationship between the Hessian and Covariance Matrix for Gaussian Random Variables 257

Appendix B: Contours of Marginal PDFs for Gaussian Random Variables 263

Appendix C: Conditional PDF for Prediction 269
C.1 Two Random Variables 269
C.2 General Cases 273

References 279

Index 291

www.EngineeringBooksPdf.com

Preface

Bayesian inference is a statistical process that quantifies the degree of belief of hypothesis, events or values of parameters. Many Bayesian methods have been developed in various areas of science and engineering, especially in statistical physics, medical sciences, electrical engineering, and information sciences, etc. Since there are many types of modeling and parametric uncertainty in civil engineering problems, Bayesian probabilistic methods are useful in the estimation of uncertain parameters and quantification of the associated uncertainties. For example, loadings, such as earthquake ground motion or complete wind pressure profile, cannot be predetermined at the structural design stage. It is also difficult to determine to a very precise level the mechanical properties for some materials, e.g., concrete, rock and soil, etc. Hourly/daily emission rates by vehicles and factories are uncertain. It is also difficult to obtain the spatial distribution of the air quality information in the nearby region. Otherwise, the transportation of air pollutants can be modeled. Furthermore, the meteorological conditions including wind speed, wind direction and rainfall of the day for prediction are also uncertain. These are important factors for modeling the pollutant flow and also for determination of dam design capacity. Traffic loads are also uncertain. These are just some of the civil engineering examples for which Bayesian probabilistic methods are applicable.

Even though Bayesian inference is useful for uncertainty quantification in civil engineering applications, the literature shows that Bayesian research in civil engineering has great potential for exploration. This book introduces some recently developed Bayesian methods and applications to a variety of areas in civil engineering although structural dynamics is the main focus. These methods are developed for the identification of dynamical systems, but some of them are also applicable to static systems. There are two levels of system identification problems to be considered although they are strongly related. The first level is parametric identification with a specified model class. A number of methods are presented for different working conditions in different identification problems. The second level is on the selection of model class. In other words, it attempts to use measurement to infer not only the uncertain parameters but also a suitable model class for system identification. This book presents various applications in civil engineering, including air quality prediction, finite-element model updating, hydraulic jump, seismic attenuation relationship, and structural health monitoring, etc.

www.EngineeringBooksPdf.com

www.EngineeringBooksPdf.com

Acknowledgements

In the course of my research, I have benefited a lot from Professor James L. Beck (California Institute of Technology), Professor Lambros S. Katafygiotis (The Hong Kong University of Science and Technology) and Professor Costas Papadimitriou (University of Thessaly, Greece) since my graduate studies. They brought me to the area of Bayesian analysis and their philosophical thinking has had a great influence on my work throughout the years. I would also like to express my sincere gratitude to my long-term collaborators: Professor Siu-Kui Au (City University of Hong Kong), Professor Jianye Ching (National Taiwan University), Professor Heung-Fai Lam (City University of Hong Kong), Professor Kai-Meng Mok (University of Macao) and Professor Wai-Man Yan (The University of Hong Kong). There have been a lot of enjoyable moments in the collaborations and I have learned much from them, especially in the areas outside my expertise. Special thanks are due to Professor Siu-Kui Au for providing his computer code to prepare the numerical example provided in Chapter 2, Section 2.5.4. I would also like to thank my previous and current graduate students, Dr Y.F. Shi, Dr J. Wang, Dr K.I. Hoi, K.C. Kuok, K.K. Wong, S.C. Kuok, H. Zang and H.Q. Mu, for preparing part of the experimental and computational results of this book and for their interaction and stimulation throughout these years.

I gratefully acknowledge the continuous support from the Science and Technology Development Fund and the Meteorological and Geophysical Bureau of the Macao Government and the Research Committee of the University of Macau.

I would like to thank my parents and my little brother, Ka-Fai, for giving me the joy and stimulation in different dimensions of life. The writing process of this book overlapped with Vanessa's pregnancy of our second child, Kin-Hang, and I am in debt to them and our first child, Hoi-Man, for not being able to take perfect care of them during this important period of time.

Finally, I would like to express my special thanks to Mr James Murphy and Mr Roger Bullen (Editors, at John Wiley & Sons (Asia)) for their great encouragement and professional assistance in preparing the proposal and manuscript of this book.

www.EngineeringBooksPdf.com

www.EngineeringBooksPdf.com

Nomenclature

$\mathcal{C}$:	class of models
$\mathbf{C}$:	damping matrix
Δt:	sampling time interval
$\delta_{nn'}$:	Kronecker Delta
$\lvert\mathbf{A}\rvert$:	determinant of the matrix $\mathbf{A}$
$\mathcal{G}(\mathbf{x}; \mu, \boldsymbol{\Sigma})$:	Gaussian distribution of random vector $\mathbf{x}$ with mean μ and covariance matrix $\boldsymbol{\Sigma}$
$\mathbb{C}$:	set of complex numbers
$\mathcal{H}$:	Hessian matrix
H:	information entropy
J_g:	goodness-of-fit function
J:	objective function
$\mathbf{K}$:	stiffness matrix
$\mathcal{K}$:	frequency range index set
$\mathbf{L}_o$:	observation matrix
$\mathbf{M}$:	mass matrix
N:	number of observed points in time domain
N_C:	number of model classes
N_d:	number of degrees of freedom (DOFs)
N_o:	number of observed DOFs
N_m:	number of modes
N_j:	number of uncertain parameters of the model class $\mathcal{C}_j$
$\mathcal{O}_j$:	Ockham factor for model class $\mathcal{C}_j$
P:	plausibility/probability
p:	probability density function
$\mathbf{q}$:	modal coordinates
$\boldsymbol{\phi}^{(m)}$:	mode shape vector of the mth mode
$\boldsymbol{\Phi}$:	modal matrix, $\boldsymbol{\Phi} = [\boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}^{(2)}, \ldots, \boldsymbol{\phi}^{(N_m)}]$
$\mathbb{R}$:	set of real numbers
S_{f0}:	spectral intensity of white noise excitation
$S_{\epsilon 0}$:	spectral intensity of prediction error and measurement noise
$S(\omega)$:	spectral density function
$S_{y,N}$:	spectral density estimator for the process y
$\mathbf{T}_0$:	force distributing matrix
$\boldsymbol{\theta}$:	parameter vector for identification
$\tilde{\boldsymbol{\theta}}$:	actual parameter vector

www.EngineeringBooksPdf.com

$\boldsymbol{\theta}^{\star}$:	optimal/updated parameter vector
$\boldsymbol{\theta}^{\eta}$:	nominal parameter vector
$\mathcal{U}$:	user's preference
$\Omega^{(m)}$:	modal frequency of the mth mode
$\mathbf{x}$:	physical coordinates
$\mathbf{X}$:	state vector
$\zeta^{(m)}$:	damping ratio of the mth mode

www.EngineeringBooksPdf.com

1

Introduction

Keywords: Thomas Bayes; information; model updating; uncertainty

1.1 Thomas Bayes and Bayesian Methods in Engineering

The Reverend Thomas Bayes [1702–1761] was a British mathematician and Presbyterian minister. He is well known for his paper 'An essay towards solving a problem in the doctrine of chances' [14], which was submitted by Richard Price two years after Bayes' death. In this work, he interpreted *probability of any event* as 'the chance of the event expected upon its happening.' There were ten propositions in his essay and Proposition 3, 5 and 9 are particularly important. Proposition 3 stated that the probability of an event X conditional on another event Y is simply the ratio of the probability of both events to the probability of the event Y. This is the definition of *conditional probability*. In Proposition 5, he introduced the concept of conditional probability and showed that it can be expressed regardless of the order in which the events occur. Therefore, the concern in conditional probability and Bayes' theorem is on *correlation* but not *causality*. The consequence of Proposition 3 and 5 is the Bayes' theorem even though this was not what Bayes emphasized in his article. In Proposition 9, he used a billiard example to demonstrate his theory. The work was republished in modern notation by G. A. Barnard [13]. In 1774, Pierre-Simon Laplace extended the results by Bayes in his article 'Mémoire sur la probabilité des causes par les événements (in French).' He treated probability as a tool for filling up the gap of knowledge. The Bayes' theorem is one of the most frequently encountered eponyms in the literature of statistics.

Bayes is also well known for another article 'A Defence of the Mathematicians against the Objections of the Author of the Analyst (1736)', in which he defended the logical foundation of the calculus by Issac Newton. Thomas Bayes was elected as a fellow of The Royal Society of the United Kingdom in 1742.

In the area of statistical analysis, there are two main parties, namely *frequency probability* and *Bayesian inference*. *Frequentists* define probability (of an *event*) as the limit of its relative

Bayesian Methods for Structural Dynamics and Civil Engineering Ka-Veng Yuen
© 2010 John Wiley & Sons (Asia) Pte Ltd

www.EngineeringBooksPdf.com

frequency for a large number of trials:

$$P = \lim_{N \to \infty} \frac{N_1}{N} \tag{1.1}$$

where N_1 is the number of trials that the event occurs among the total of N trials. This definition of probability narrows down its applicability since probability can only be assigned to an event whose *random experiment* is available. On the contrary, Bayesian inference allows the definition for the probability of an event or a *statement*, in which a random experiment cannot be designed. This offers a different dimension of probability for *degree of belief*, which is regarded as *plausibility*.

Bayesian methods have been widely used in many areas due to the pioneering work by Thomas Bayes. The Bayesian theory has had great advancement after the significant work by Harold Jeffreys [121], Richard T. Cox [63] and Edwin T. Jaynes [115, 117–120]. Since then, different Bayesian methods have been developed and widely applied to many different disciplines of natural sciences, social sciences and engineering, especially in statistical physics [116, 119], electricity usage [42], engineering hydrology [165], econometrics [300], archeology [38], information sciences [123, 180], system reliability [167, 301], prediction of concrete strength [87], structural dynamics [199], medical sciences [144, 254, 274], fatigue [166], forensic science [259], marketing [223], mechanical engineering [269], computer science [152], engineering geology [220], aerospace engineering [40], finance [213], population migration [217] and many others [80], etc.

Bayesian inference is very useful for civil engineering applications because there are many types of modeling and parametric uncertainty in civil engineering problems. In structural engineering, excitation, such as earthquake ground motion or complete time-varying wind pressure profile, cannot be predetermined at the structural design stage. Material properties are difficult to determine to a precise level for some materials, e.g., concrete. The number and size of cracks in concrete beams are also uncertain. There are also modeling errors, such as rigid joint assumption, that can be treated as modeling uncertainty. In geotechnical engineering, the underground geological profile is determined by very limited information with ground investigation. Furthermore, even for carefully handled samples, laboratory test results of soil properties, e.g., plastic limit and liquid limit, exhibit high level of uncertainty. There are also modeling errors in constitutive relationships and they can be treated as modeling uncertainty. In environmental engineering, hourly/daily emission by vehicles and factories is also uncertain. It is also difficult to obtain the complete spatial distributions of the air quality information and meteorological conditions in the city of concern and other surrounding cities. In traffic engineering, the traffic loading and traffic flow are uncertain. These are only some of the examples of uncertainty appearing in civil engineering and the Bayesian probabilistic method is useful for explicit treatment of the modeling uncertainty and quantification of the parametric uncertainty.

Even though Bayesian inference is useful for uncertainty quantification that fulfills the need in civil engineering, the literature shows that developments and applications of this powerful tool in civil engineering are still at an early stage. Therefore, there is plenty of room to be explored for Bayesian applications in civil engineering. This book introduces some recently developed Bayesian methods and applications to a number of areas in civil engineering. The main concern here is on the identification of dynamical systems, but some of the methods are also applicable to static problems. Two types of problems in system identification are

www.EngineeringBooksPdf.com

considered. The first type is *parametric identification* for a prescribed mathematical model with uncertain parameters. The second type is on the selection of a class of parametric models based on system response measurements. In other words, the measurement is used to infer not only the uncertain parameters but also the structure of the mathematical model. Applications include different areas in civil engineering, such as air quality prediction, finite-element model updating, hydraulic jump, seismic attenuation empirical relationship, and structural health monitoring, etc.

The idea of Bayesian updating is similar to our thinking process but it provides also the basis for quantification. We have a perception of different people and matters based on our experience, i.e., data. When a new event happens (i.e., new data is obtained), it modifies our perception. In other words, our perception is not only determined by the latest piece of information but it also depends on the original perception. In Bayesian analysis, the original perception is regarded as the prior information and the new piece of information is utilized to update our perception or mathematical model.

1.2 Purpose of Model Updating

There are two main purposes of model updating or system identification. One common goal is to identify physical parameters, e.g., stiffness of a structural element or the diffusion rate of an air pollutant. These identified parameters can be further used as indicator for the status of the system or phenomenon. For example, the stiffness parameter of a structural member can be monitored from time to time and an abnormal reduction indicates possible damage of the member. However, reduction may be simply due to statistical uncertainty. Therefore, it is necessary to quantify the uncertainty of the estimation so that one can distinguish whether the parameter change is due to deterioration of the structural member. In this case, it is desirable to obtain a narrow distribution of the parameter so that small changes can be detected with a high level of confidence.

Another purpose of model updating is to obtain a mathematical model to represent the underlying system for future prediction. Even though there are also parameters to be identified as in the previous case, these parameters may not necessarily be physical, e.g., coefficients of auto-regressive models. In this situation, the identified parameters are not necessarily as important as the previous case provided that the *identified model* provides an accurate prediction for the system output. It will be shown in the following chapters that there is no direct relationship between satisfactory model predictions and small posterior uncertainty of the parameters. This point will be further elaborated in Chapter 6. Nevertheless, no matter for which purpose, quantification of the parametric uncertainty is useful for further processing. For example, it can be utilized for comparison of the identified parameter values at different stages or for uncertainty analysis of the output of the identified model. Furthermore, it will be demonstrated in Chapter 6 that quantification of the posterior uncertainty allows for the selection of a suitable class of models for parametric identification.

Using the quantified uncertainty obtained from Bayesian methods, there are two important types of applications. The first category is *robust reliability analysis*. Under severe earthquake excitations, buildings and bridges may exhibit significant nonlinear behavior. With a stochastic representation of the anticipated ground motions [96, 130, 237], one important reliability problem is to determine the first passage probability of some response quantities of interest in a

www.EngineeringBooksPdf.com

prescribed duration. The first passage problem involving nonlinear dynamics is a well-known challenging problem especially for multi-degree-of-freedom systems. Monte Carlo simulation methods offer a promising means regardless of the number of uncertain parameters or degrees of freedom [225]. However, it is not efficient for estimation of rare events because it requires a large number of samples in order to obtain a sufficient number of rare failed samples. In order to reduce the number of samples without sacrificing the accuracy of the estimates, many simulation techniques were developed, including numerical solutions of the Kolmogorov equation [221], *importance sampling* [12, 176, 231] and adaptive sampling [37]. The importance sampling and adaptive sampling methods attempt to increase computational efficiency by generating samples in the high-probability density regions of the failure domain. However, these methods encounter substantial difficulties when the number of random variables is large [11], e.g., in the case of reliability analysis for dynamical systems.

The controlled Monte Carlo simulation method [209], the subset simulation method [9] and the spherical subset simulation [135] were developed to solve problems with a high dimension, i.e., a large number of random variables. These methods are applicable to general nonlinear systems and are shown to be robust and significantly more efficient than standard Monte Carlo simulations. However, it is reasonable to expect that a general/robust simulation method might not be the most efficient for some special cases, e.g., for linear systems. For example, ISEE [10, 293] was developed for linear time-invariant systems due to the appreciation of the fact that 'the efficiency of a reliability method is often gained at the expense of generality' and it was proven to be extremely efficient. Another efficient algorithm for solving the reliability problem of a linear system is the domain decomposition method [134].

The aforementioned methods can be applied to evaluate the reliability of engineering systems subjected to stochastic input with a given mathematical model. On the other hand, if a parametric model of the underlying system is available and the probability density function of these parameters is obtained by Bayesian methods, the uncertain parameter vector can be augmented to include the model parameters and the uncertain input components. Then, robust reliability analysis can proceed for stochastic excitation with an uncertain mathematical model. This allows for more realistic reliability evaluation in practice so that the modeling error and other types of uncertainty of the mathematical model can be taken into account.

Another type of important applications is *structural vibration control*. This area has received great attention in the last several decades [112, 248]. Because complete information about a dynamical system and its environment is never available, the system and excitation can not be modeled exactly. Classical control design methods based on a single nominal model of the system may fail to create a controlled system that provides satisfactory and robust performance. Robust control methods, e.g., H_2, H_∞ and μ-synthesis, were therefore proposed so that the optimal controller provides robust performance and stability for a set of *possible models* of the system [71, 72]. In a probabilistic robust control approach, an additional dimension is introduced by using probabilistic descriptions of all the possible models when selecting the controller to achieve optimal performance. These probability distributions give a measure of how plausible the parameter values are and they may be obtained from engineering judgement or Bayesian system identification methods.

Over the last two decades, there has been increasing interest in probabilistic, or stochastic, robust control theory. Monte Carlo simulation methods have been used to synthesize and analyze controllers for uncertain systems [170, 255]. First- and second-order reliability methods were incorporated to compute the probable performance of linear-quadratic-regulator

www.EngineeringBooksPdf.com

controllers (LQRs) [78, 79, 252, 253]. By the probability density function of the uncertain parameters of a structural model, an efficient asymptotic expansion of Laplace type [197] was used to approximate the probability integrals that are needed to determine the optimal parameters for a passive tuned mass damper [200] and the optimal control gains for an active mass driver [172, 283] for robust structural control. In May and Beck [172], the proposed controller feeds back output measurements at the current time, where the output corresponds to certain response quantities that need not be the full state vector of the system. In Yuen and Beck [283], the method was extended to consider additional feedback information from the past output measurements. It improves the performance of the control system without requiring extra hardware cost. This method was also applied to semi-active controlled systems with magnetorheological (MR) dampers [298].

1.3 Source of Uncertainty and Bayesian Updating

There has been a lot of debate in the literature of Philosophy and Physics on whether natural phenomena are deterministic or random but the uncertainty considered in this book is not directly related to this issue. First, consider a 100-degree-of-freedom chain-like system depicted in Figure 1.1. The mass and spring constants are taken to be $M_1 = M_2 = \cdots = M_{100} = 1.0$ kg and $K_1 = K_2 = \cdots = K_{100} = 5000$ N/m. The governing equation of the system is:

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) = 0 \tag{1.2}$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_{100}]^T$ is the nodal displacement vector, the mass matrix is $\mathbf{M} = M_1\mathbf{I}_{100}$, $\mathbf{I}_{100}$ is the 100×100 identity matrix, and the stiffness matrix is:

$$\mathbf{K} = K_1 \begin{bmatrix} 2 & -1 & & & 0 \\ & 2 & -1 & & \\ & & \ddots & \ddots & \\ & & & 2 & -1 \\ sym & & & & 1 \end{bmatrix} \tag{1.3}$$

As a result, the fundamental frequency of the system is 0.1759 Hz.

Figure 1.1 Chain-like system

Note that there is no energy dissipating mechanism (zero damping matrix) included in this model. Figure 1.2 shows the free vibration response of the mass M_{50} in the middle of the chain with an zero initial condition except that the mass at the right end (M_{100}) has an initial displacement of 0.01 m. The response appears to be random even though the system is deterministic with a simple initial condition and zero excitation. Figures 1.3 and 1.4 zoom into

www.EngineeringBooksPdf.com

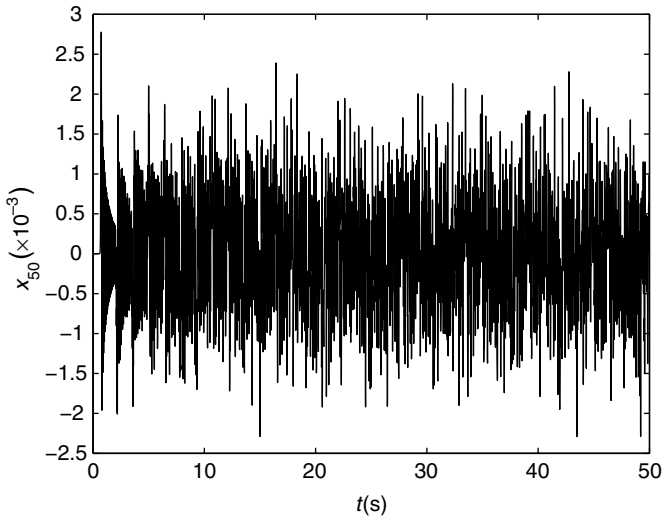

Figure 1.2 Response of the 100-DOF system

the first and last 5 s of the response. In the first 0.6 s, the response of this mass was almost zero since the wave has not yet arrived. After the first wave front arrived, the response reached the maximum and decayed rapidly even though there was no energy dissipating mechanism in the mathematical model. The response mitigation was due to energy transportation by the traveling

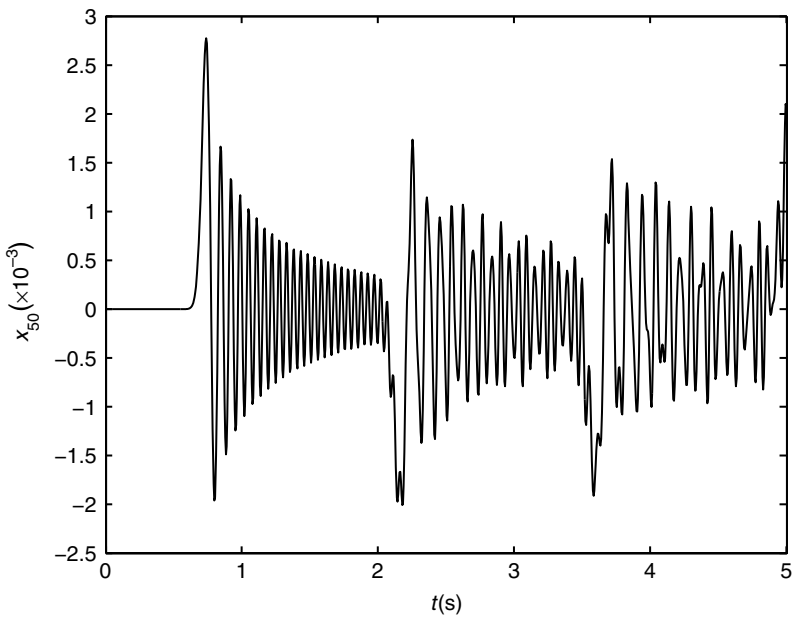

Figure 1.3 Response of the 100-DOF system (first 5 s)

www.EngineeringBooksPdf.com

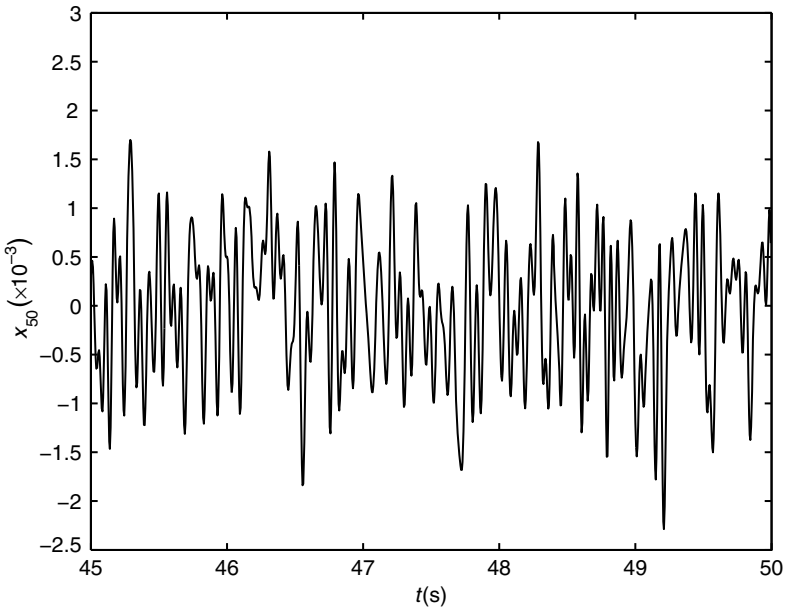

Figure 1.4 Response of the 100-DOF system (last 5 s)

waves. On the other hand, if only the response of the last 5 s was observed, it appeared as a random response even though there were some major frequency contents. This simple example illustrates an important feature of many phenomena in which only part of the complete picture can be observed. If one focuses on the measurement of the motion of M_{50} between 0.75 and 2.0 s, it may be misinterpreted as the presence of an efficient energy dissipation mechanism. If only the last 5 s of the response was measured, it may be interpreted as a random response. However, the existence of uncertainty is only due to the fact that the information of the motion of all the other masses is unknown and the combination of the motion by the waves of different frequencies blurs the underlying behavior of the system. Its meaning is similar to a Chinese proverb: 'A blind touches an elephant.'

This example demonstrates that even for a deterministic system without any external loading, the response *appears to be uncertain*. In the real world, there are many types of unmodeled behavior/dynamics of complex physical phenomena (e.g., chaotic systems) and one possible approach is to treat them as random variables or random processes. Then, statistical moments are used to represent the overall behavior. This type of error is regarded hereafter as a *modeling error*. Another main source of uncertainty is due to the finite amount of information carried by the data. Due to the finite amount of the measurement, and hence the finite amount of information, identification results can be determined up to finite precision so uncertainty gets into the picture. Finally, due to the finite precision of data acquisition, measurement error is induced, including electrical noise and *quantization error*.

George Soros also recognized in financial markets and economics that uncertainty is due to lack of information [251]. In the context of social sciences, it is more appropriate to regard incomplete information as incomplete knowledge. Of course, in financial markets or

www.EngineeringBooksPdf.com

economics, the problems are even more complicated since there are more sources of uncertainties, such as the actions and psychological effects of the participants. They may be modeled as nonlinear time-varying feedback systems.

1.4 Organization of the Book

There are several important questions the author attempts to address in this book:

- Does small posterior uncertainty imply accurate data fitting?
- What are the properties of a suitable model class for parametric identification?
- Does accurate data fitting imply a suitable model class for identification?
- What is the relationship among model sensitivity, posterior uncertainty and robustness of a model class?
- Is it possible to obtain a model class that is powerful in data fitting and also robust to measurement noise and modeling error?

In order to address these questions, this book proceeds with six chapters. Chapter 1 gives the general introduction and literature review of applications in different disciplines of engineering. The purpose for model updating is discussed and the types of uncertainty considered in this book are also exploited.

Chapter 2 introduces the basic concepts of conditional probabilities and the Bayes' Theorem. Simple examples are given for illustration. Then, the Bayesian model updating framework will be formulated for input–output data and the updated probability density function (PDF) is derived. Model identifiability is introduced for different topology of the updated PDF. Monte Carlo simulation and Markov Chain Monte Carlo simulation are introduced. Comments will be made on the deterministic and probabilistic approaches. Linear and nonlinear regression problems will be considered. Then, several applications are presented to demonstrate the fundamental ideas and the efficacy of the Bayesian approach. First, a Bayesian method is derived to investigate the temperature effect on the modal frequencies of buildings. Measurements of a 22-story building for six months are used for the investigation. By dimensional analysis, the relationship between the squared modal frequencies and temperature is parabolic and the uncertain coefficients are obtained by the Bayesian approach. The second application is on the selection of noise parameters for the Kalman filter. The latter is well known for state estimation and parametric identification but the noise parameters are usually assigned by the user's judgement. Here, the Bayesian approach will be applied to select these parameters. A half-or-double algorithm will be presented for the optimization problem. The third application is on an ambient air quality predictive system by use of the extended Kalman filter. A semi-empirical time-varying air quality predictive system is introduced and real ambient air quality and meteorological data will be used for identification of the uncertain parameters. The performance of this system is compared with artificial neural networks.

Chapters 3 and 4 introduce two recently developed Bayesian methods for updating the mathematical models of dynamical systems. Chapter 3 presents the Bayesian spectral density approach. The spectral density estimator is defined to take into account of the aliasing and leakage effect. The statistical properties of the spectral density estimator are examined and

www.EngineeringBooksPdf.com

the posterior PDF can then be obtained. The information entropy method is introduced for optimal placement of sensors. The duffing oscillator is used to demonstrate the methodology for nonlinear model updating. It turns out that the probabilistic method is useful in solving the originally unidentifiable problem. Then, the method is applied for structural health monitoring of the East Asian Hall at the University of Macau during two severe typhoons in 2008. It was observed that the modal frequencies of the building were reduced by a notable percentage but the reduction could be recovered after the typhoons. The last application focuses on hydraulic jump. The Bayesian spectral density approach was applied to identify the frequency of the surface roller in different types of hydraulic jumps.

The Bayesian time-domain approach is introduced in Chapter 4 for random input. An exact formulation is first presented for stationary response and the computational difficulties in applying this formulation will be pointed out. Then, an approximated expansion of the likelihood function is introduced to render the computation of the posterior PDF feasible. This method is applicable for multi-degree-of-freedom linear systems subjected to stationary or non-stationary input. An application on modal updating (i.e., identification of modal quantities such as natural frequencies, modal damping ratios and mode shapes) with unmeasured non-stationary earthquake ground motion is used for demonstration. Comparison in terms of computational efficiency, accuracy and applicability will be given between the frequency-domain and time-domain approaches. A brief introduction is given for extended readings on other Bayesian methods developed recently by the author and his collaborators.

Chapter 5 addresses the problem of model updating with eigenvalue–eigenvector measurements. The generalized eigenvalue problem is considered and the posterior PDF of the model parameters is derived. Solution of the nonlinear optimization problem is obtained by solving a series of linear optimization problems and a tailor-made iterative algorithm is introduced. There is no requirement to match the measured modes with the model modes, which is a difficult task required by most existing methods. The Bayesian framework allows for quantification of the uncertainty of the estimation. Applications on finite-element model updating and structural health monitoring with a building model and a three-dimensional structural frame are used for demonstration.

In the previous chapters, the problem of parametric identification was considered for a given class of models with uncertain parameters. However, selection of a suitable class of models for identification is very important in practice and this is addressed in Chapter 6. The Bayes' theorem is used to derive the relative plausibility of some given model classes. It turns out that the plausibility of a model class depends on its evidence, which quantifies how likely the data can be obtained given a model class. Expressions are given using asymptotic expansion for globally identifiable cases. Comparisons with the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are provided analytically. Then, the physical significance of the evidence will be elaborated in detail. The transitional Markov Chain Monte Carlo simulation is introduced to compute the evidence for general cases regardless of the identifiability. Linear and nonlinear regression problems will be revisited for model class selection consideration. Special treatment is introduced for evaluation of the evidence integral. Finally, the Bayesian model class selection method is applied to two problems in civil engineering. The first application is to determine the number of modes to be included in modal updating for a building. The Bayesian model class selection approach is applied in conjunction with the Bayesian spectral density approach. The second application is on the selection of seismic attenuation

www.EngineeringBooksPdf.com

prediction model class. There are several well-known seismic attenuation prediction models (e.g., Boore–Joyner–Fumal) and the Bayesian method is used to select a proper model class for this application. Candidates of model classes will be generated and they will be evaluated by the Bayesian model class selection method with real strong motion records from China. The physical significance of the most plausible empirical expression is also discussed. Finally, prior distribution will be revisited and an interesting and inspirational example will be given to conclude the book.

www.EngineeringBooksPdf.com

2

Basic Concepts and Bayesian Probabilistic Framework

Keywords: Bayes' theorem; conditional probability; information entropy; Kalman Filter; Markov Chain Monte Carlo simulation; model identifiability; particulate matter; regression problem; reliability; structural health monitoring

In this chapter, the basic concepts of conditional probability are introduced and the Bayes' theorem will be immediately followed. Examples are given for different cases of discrete events and continuous-valued parametric estimation. Then, the Bayesian model updating framework is presented. Model identifiability issues will be exploited. Comments are given on the comparison between deterministic and probabilistic methods. Challenges, difficulties and advantages of probabilistic methods are addressed. Linear and nonlinear regression problems will be illustrated. Since the topology of the updated probability density function can be very complicated, an adaptive Markov Chain Monte Carlo simulation method is introduced to generate samples for the representation of the distribution. Finally, three applications are presented to illustrate the fundamental ideas of Bayesian inference. First, a Bayesian method is derived to construct the relationship between the squared fundamental frequency of buildings and the ambient temperature. By observing the temperature effect on the length and Young's modulus, the relationship turns out to be quadratic if the building can be modeled as a beam. Six-month measurement from a 22-story building is utilized and the Bayesian method estimates the coefficients and their associated uncertainty. The second application is on the well-known Kalman filter. It is also a Bayesian updating tool, that is useful for state estimation and parameter identification but the noise parameters are normally assigned by the user based on prior information or experience. In this study, the Bayesian approach is used to estimate these parameters. The last application is on ambient air quality prediction problems for PM_{10} concentration. An online time-varying prediction system is developed and the Kalman filter is used to update the system. With a properly chosen parametric model, the predictive system over-performs the well-known artificial neural network, especially in predicting the data of high PM_{10} concentration.

Bayesian Methods for Structural Dynamics and Civil Engineering Ka-Veng Yuen
© 2010 John Wiley & Sons (Asia) Pte Ltd

www.EngineeringBooksPdf.com

2.1 Conditional Probability and Basic Concepts

Use A and B to denote two events. The *conditional probability* of event A provided the occurrence of event B is given by:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} \tag{2.1}$$

if $P(B) > 0$. (If $P(B) = 0$, then $P(A \wedge B) = 0$ and $P(A|B)$ is meaningless.) The symbol $\wedge$ denotes the logical operator *and* for two (or more) events so $P(A \wedge B)$ denotes the probability of the occurrence of both events.

The notion of conditional probability does not necessarily imply a *reason–consequence* relationship and *time–order* relationship for two events. For example, it is no wonder that the conditional probability that a person with cancer provided that he/she is under chemotherapy is close to 1.0. (It is not strictly 1.0 due to incorrect diagnosis.) However, the high conditional probability surely does not imply that chemotherapy is the reason for cancer. It is simply due to the fact that only people with positive cancer screening results undergo chemotherapy. If the conditional probability $P(A|B)$ is large, there are three possibilities. (1) Event A is a consequence of Event B, or opposite. In the chemotherapy example, cancer (A) is the reason but not the consequence of undergoing chemotherapy (B). (2) There is a higher level event C that induces both events A and B. For example, in Macao, the probability of a heavily traffic jammed day, provided good ambient air quality, is high. Obviously, there is no direct reason–consequence relationship between these two events. Note that local emission by vehicles is not the main source for the ambient/background pollutants in Macao though it affects seriously the air quality at the street level. However, a good ambient air quality day and a heavily traffic jammed day are consequences of a heavily raining day as the precipitation washes out the pollutants from the air, and a heavily raining day triggers a heavily traffic jammed day. (3) There are more than one higher level or intermediate events in the reason–consequence tree for events A and B. This is a combination/extension of cases 1 and 2.

The *law of total probability* is very useful in the context of conditional probability. If an event A is subdivided into N (either finite or countably infinite) *mutually exclusive events*, $A_1, A_2, \ldots, A_N$, then the probability of another event B is given by [177]:

$$P(B) = \sum_{n=1}^{N} P(B \wedge A_n) = \sum_{n=1}^{N} P(B|A_n)P(A_n) \tag{2.2}$$

for a mutually exclusive partitioning of the event A, i.e., $A_n \wedge A_{n'} = \phi$, if $n \neq n'$ and $A = \bigvee_{n=1}^{N} A_n$. The symbol $\vee$ denotes the logical operator *or* for two (or more) events so $P(A \vee B)$ denotes the probability of the occurrence of at least one of the events.

The continuous analogy of the law of total probability is also commonly used. In this case, the probability of event B can be expressed as:

$$P(B) = E[P(B|X)] \equiv \int_{-\infty}^{\infty} P(B|X)p(X)\mathrm{d}X \tag{2.3}$$

where $E[.]$ denotes the *mathematical expectation* and $p(X)$ is the *probability density function* (PDF) that describes the (real) random variable X. This is valid because the real axis can be

www.EngineeringBooksPdf.com

subdivided into infinitely many intervals which do not have intersection. Then, the events of the random variable X falling into different intervals are mutually exclusive.

Furthermore, it can also be applied to (real) random variables:

$$p(Y) = E[p(Y|X)] = \int_{-\infty}^{\infty} p(Y|X)p(X)\mathrm{d}X \tag{2.4}$$

where $p(Y|X)p(X) = p(X, Y)$ is the *joint PDF* of X and Y. Therefore, $p(Y)$ is simply the *marginal PDF* of Y.

2.1.1 Bayes' Theorem for Discrete Events

By exchanging the role of events A and B in Equation (2.1) and using the fact that $P(A \wedge B) = P(B \wedge A)$, the Bayes' theorem can be obtained:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.5}$$

if $P(B) > 0$. If the event A is partitioned into N mutually exclusive events, $A_1, A_2, \ldots, A_N$, the probability $P(B)$ in the denominator can be replaced by using Equation (2.2):

$$P(A_n|B) = \frac{P(B|A_n)P(A_n)}{\sum_{n=1}^{N} P(A_n)P(B|A_n)} \tag{2.6}$$

for $n = 1, 2, \ldots, N$.

Example. Medical Screening Test
Consider a rare virus and its screening test data. Use C and N to denote a carrier and a non-carrier of this virus. Also, use '+' and '−' to denote a person with positive and negative testing results, respectively. It is known that the number of carriers is 0.5% of the whole population of the city, i.e., $P(C) = 0.005$. Laboratory results show that this screening test has a probability of 0.3% showing *false positive* among all the non-carriers, i.e., $P(\text{'+'}|N) = 0.003$. Furthermore, 0.6% of the testing results are positive, i.e., $P(\text{'+'}) = 0.006$.

The *false positive rate* is defined as the probability of non-carriers among all the positive testing results. By using the Bayes' theorem, it can be estimated as follows:

$$\begin{aligned}
\text{false positive rate} \equiv P(N|\text{'+'}) &= \frac{P(\text{'+'}|N)P(N)}{P(\text{'+'})} \\
&= \frac{P(\text{'+'}|N)[1 - P(C)]}{P(\text{'+'})} \\
&= \frac{0.003(1 - 0.005)}{0.006} \\
&= 0.4975
\end{aligned}$$

In other words, this is the probability of wrong detection among all the positive testing results. Even though the test appears to be accurate in the sense that the conditional probability

www.EngineeringBooksPdf.com

$P(\text{`}+\text{'}|N)$ is small, the false positive rate is close to 50% and it is a well-known indicator for the *Type I error*. Similarly, the conditional probability $P(N|\text{`}-\text{'})$ can be computed by Equation (2.5):

$$\begin{aligned} P(N|\text{`}-\text{'}) &= \frac{P(\text{`}-\text{'}|N)P(N)}{P(\text{`}-\text{'})} \\ &= \frac{[1-P(\text{`}+\text{'}|N)][1-P(C)]}{1-P(\text{`}+\text{'})} \\ &= \frac{(1-0.003)(1-0.005)}{1-0.006} \\ &= 0.9980 \end{aligned}$$

By using the definition of conditional probability in Equation (2.1), the probability of the joint events can be obtained:

$$P(N \wedge \text{`}+\text{'}) = P(\text{`}+\text{'}|N)P(N) = P(\text{`}+\text{'}|N)[1-P(C)] = 0.003(1-0.005) = 0.002\,985$$

$$\begin{aligned} P(N \wedge \text{`}-\text{'}) &= P(\text{`}-\text{'}|N)P(N) = [1-P(\text{`}+\text{'}|N)][1-P(C)] \\ &= (1-0.003)(1-0.005) = 0.992\,015 \end{aligned}$$

Furthermore, the probability of the other two joint events can be obtained by the law of total probability:

$$P(C \wedge \text{`}+\text{'}) = P(\text{`}+\text{'}) - P(N \wedge \text{`}+\text{'}) = 0.006 - 0.002\,985 = 0.003\,015$$

$$P(C \wedge \text{`}-\text{'}) = P(\text{`}-\text{'}) - P(N \wedge \text{`}-\text{'}) = (1-0.006) - 0.992\,015 = 0.001\,985$$

These results are summarized in Table 2.1 and different conditional probabilities and rates can be computed using this table. For example, the *sensitivity* is defined as the probability of correct detection among all the carriers:

$$\text{sensitivity} \equiv P(\text{`}+\text{'}|C) = \frac{P(C \wedge \text{`}+\text{'})}{P(C)} = \frac{0.003\,015}{0.005} = 0.603$$

On the other hand, the *specificity* is defined as the probability of correct diagnosis among all the non-carriers:

$$\text{specificity} \equiv P(\text{`}-\text{'}|N) = \frac{P(N \wedge \text{`}-\text{'})}{P(N)} = \frac{0.994}{0.995} = 0.998\,99$$

Table 2.1 Distribution of the virus screening test

Actual status\Testing result	'+'	'−'	Sum
C	0.003 015	0.001 985	0.005
N	0.002 985	0.992 015	0.995
Sum	0.006	0.994	1.000

www.EngineeringBooksPdf.com

The sensitivity indicates the probability for picking up the true carriers while the specificity denotes the probability of a correct negative testing result for a virus-free person. Reliable results are anticipated for the screening tests with sensitivity and specificity close to unity. However, this is very difficult to achieve for a rare virus. The low value of sensitivity in this case implies a substantial probability of failure in detecting a virus-carrier (38.7%).

Finally, similar to the false positive rate, the *false negative rate* can be computed:

$$\text{false negative rate} \equiv P(C|\text{`}-\text{'}) = \frac{P(C \wedge \text{`}-\text{'})}{P(\text{`}-\text{'})} = \frac{0.001\,985}{0.994} = 0.001\,997$$

and it is commonly used to indicate the level of *Type II error*.

2.1.2 Bayes' Theorem for Continuous-valued Parameters by Discrete Events

For continuous-valued uncertain parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_N]^T$, the concern is to update their probability density function. Consider a neighborhood around $\boldsymbol{\theta}_0 = [\theta_{10}, \theta_{20}, \ldots, \theta_{N0}]^T$, that is a hypercube in the parameter space $\boldsymbol{\Theta}$:

$$C = \left\{\theta_{10} - \frac{\Delta\theta_1}{2} < \theta_1 < \theta_{10} + \frac{\Delta\theta_1}{2}, \theta_{20} - \frac{\Delta\theta_2}{2} < \theta_2 < \theta_{20} + \frac{\Delta\theta_2}{2}, \ldots, \right.$$
$$\left. \theta_{N0} - \frac{\Delta\theta_N}{2} < \theta_N < \theta_{N0} + \frac{\Delta\theta_N}{2}\right\} \tag{2.7}$$

The event B in Equation (2.5) is defined as the occurrence of the parameter vector $\boldsymbol{\theta}$ falling into this hypercube. For small $\Delta\theta_1$, $\Delta\theta_2$,...,$\Delta\theta_N$, the probability of event B is:

$$P(B) = p(\boldsymbol{\theta}) \prod_{l=1}^{N} \Delta\theta_l \tag{2.8}$$

where $p(\boldsymbol{\theta}) = p(\boldsymbol{\theta} = \boldsymbol{\theta}_0)$ is the joint PDF of the uncertain parameter vector $\boldsymbol{\theta}$. Similarly, the conditional probability is given by $P(B|A) = p(\boldsymbol{\theta}|A) \prod_{l=1}^{N} \Delta\theta_l$. Then, the Bayes' theorem in Equation (2.5) is converted into another form:

$$p(\boldsymbol{\theta}|A) = \frac{P(A|\boldsymbol{\theta})p(\boldsymbol{\theta})}{P(A)} \tag{2.9}$$

This form is applicable to the identification of continuous-valued uncertain parameters with observation of discrete events and $p(\boldsymbol{\theta}|A)$ is regarded as the *updated PDF* or *posterior PDF* of the parameter vector $\boldsymbol{\theta}$.

Example. Imperfect dice
An imperfect dice was drawn independently for N times whereas '1' appeared N_1 times. The aim here is to update the probability of the occurrence of '1' in a single draw (denoted as P_1) and this can be achieved by the Bayes' theorem. The conditional PDF of the uncertain parameter P_1 given the value of N_1 is:

$$p(P_1|N_1) = \frac{P(N_1|P_1)p(P_1)}{P(N_1)}$$

www.EngineeringBooksPdf.com

The probability $P(N_1)$ in the denominator serves as a normalizing constant. Given N and P_1, the probability for a particular number of occurrences is:

$$P(N_1|P_1) = \frac{N!}{N_1!(N-N_1)!} P_1^{N_1}(1-P_1)^{N-N_1}$$

In order to establish the updated PDF for P_1, a triangular distribution is used for the prior PDF:

$$p(P_1) = \begin{cases} 12P_1 & \text{if } P_1 \in [0, 1/6] \\ 12/5(1-P_1) & \text{if } P_1 \in (1/6, 1] \end{cases}$$

which is a spread distribution over the support [0, 1]. The peak of the PDF occurs at $P_1 = 1/6$, which is taken from a perfect dice since no measurement is available at the stage of constructing the prior distribution. The updated PDF for P_1 is shown in Figure 2.1 with $N = 200$ and $N_1 = 39$. The most probable value, mean, and standard deviation of the estimate are 0.194, 0.198 and 0.027, respectively. In the view point of frequentists, this probability is simply $39/200 = 0.195$.

Then, consider a larger number of draws ($N = 1000$) and '1' appeared 149 times. By using the Bayes' theorem, the updated PDF for P_1 is shown in Figure 2.2. The most probable value, mean, and standard deviation of the estimate are 0.150, 0.151 and 0.011, respectively. This distribution concentrates in a narrower range due to the extra information gained from the additional samples. With the updated PDF, different confidence intervals can be constructed for the uncertain parameter P_1.

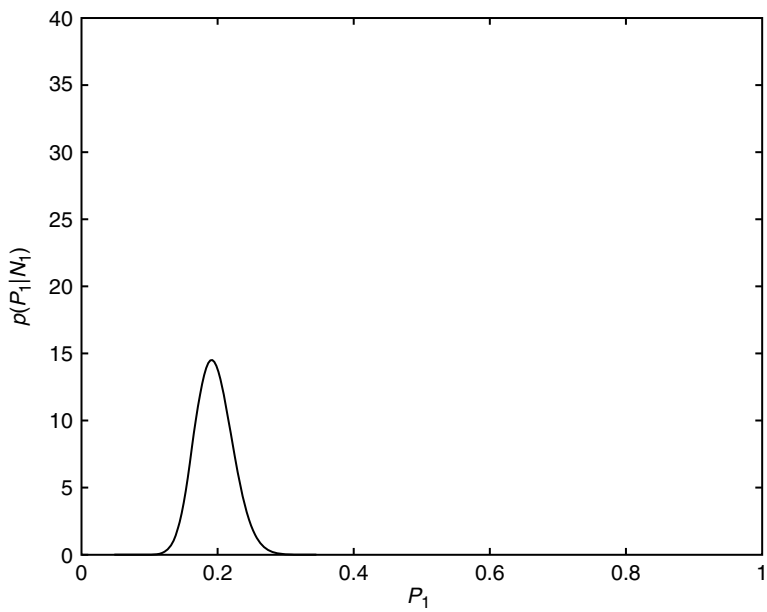

Figure 2.1 Posterior PDF $p(P_1|N_1)$, $N = 200$

www.EngineeringBooksPdf.com

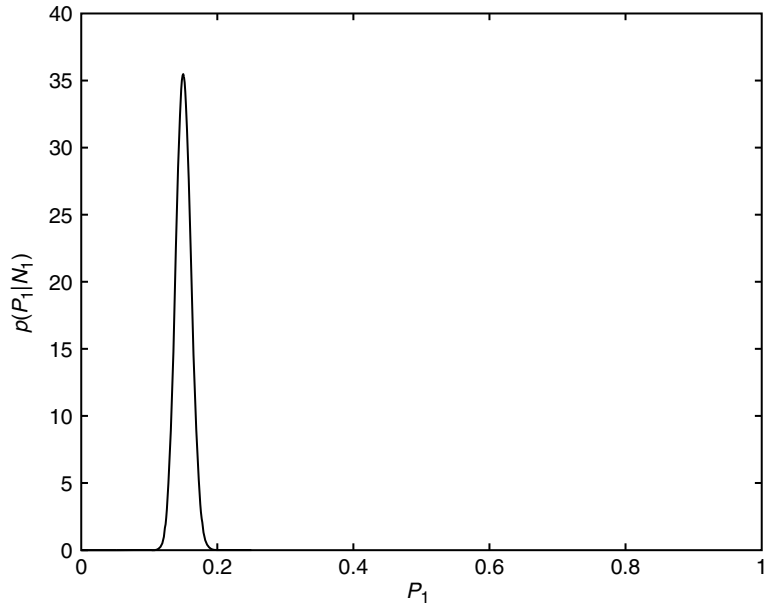

Figure 2.2 Posterior PDF $p(P_1|N_1)$, $N = 1000$

2.1.3 Bayes' Theorem for Discrete Events by Continuous-valued Parameters

The reciprocal form of Equation (2.9) can be obtained easily for the *updated probability* of an event with measurement of continuous-valued variables:

$$P(A|\boldsymbol{\theta}) = \frac{p(\boldsymbol{\theta}|A)P(A)}{p(\boldsymbol{\theta})} \tag{2.10}$$

This form can be used to update the probability of events given the measurement of continuous-valued variables.

Example. Interpretation of Test Results
An aptitude test is designed to assess students' capability for a particular subject and the top 5% students are defined as having excellent command. From previous experience, it is known that the distribution of the scores, θ, of all the excellent students follows a log-normal distribution:

$$p(\theta|A) = \frac{1}{\sqrt{2\pi}\sigma\theta} \exp\left[-\frac{(\ln\theta - \mu)^2}{2\sigma^2}\right] \tag{2.11}$$

with mean 550 and standard deviation 12. In order to determine the values of the parameters μ and σ^2 in the distribution, the expressions for the mean and variance are derived:

$$M = E[\theta|A] = \exp\left(\mu + \sigma^2/2\right) \tag{2.12}$$

www.EngineeringBooksPdf.com

and

$$V = E[(\theta - E[\theta])^2|A] = \left[\exp(\sigma^2) - 1\right] \exp(2\mu + \sigma^2) \tag{2.13}$$

Therefore, the parameters μ and σ^2 can be computed as follows:

$$\mu = \ln\left(\frac{M^2}{\sqrt{V + M^2}}\right) = 6.3097 \tag{2.14}$$

and

$$\sigma^2 = \ln\left(\frac{V + M^2}{M^2}\right) = 4.7592 \times 10^{-4} \tag{2.15}$$

On the other hand, the scores of the rest of the students (group B) also follow the log-normal distribution with mean 480 and standard deviation 15 so the parameters are $\mu_B = 6.1733$ and $\sigma_B^2 = 9.7609 \times 10^{-4}$ for this group. Then, by using the law of total probability, the score distribution of all students is given by:

$$\begin{aligned} p(\theta) &= p(\theta|A)P(A) + p(\theta|B)P(B) \\ &= \frac{1}{20\sqrt{2\pi}\sigma\theta} \exp\left[-\frac{(\ln\theta - \mu)^2}{2\sigma^2}\right] + \frac{19}{20\sqrt{2\pi}\sigma_B\theta} \exp\left[-\frac{(\ln\theta - \mu_B)^2}{2\sigma_B^2}\right] \end{aligned} \tag{2.16}$$

Figure 2.3 shows the probability distribution of the scores for top students (solid line) and ordinary students (dashed line). Figure 2.4 shows the unconditional score distribution for all students.

By the Bayes' theorem in Equation (2.10), one can update the probability of a student with excellent command of the subject given his/her test score:

$$P(A|\theta) = \left\{1 + \frac{19\sigma}{\sigma_B} \exp\left[\frac{(\ln\theta - \mu)^2}{2\sigma^2} - \frac{(\ln\theta - \mu_B)^2}{2\sigma_B^2}\right]\right\}^{-1} \tag{2.17}$$

and this is shown in Figure 2.5. By this figure, different confidence intervals can be constructed for a student to be classified as having excellent command of the subject. For example, a student who scored 533 or above has 90% probability to have excellent command of the subject. This threshold value may be smaller than intuition as it is significantly smaller than the mean of the group. However, the probability for an ordinary student to achieve a score of 533 or above is only 0.038%.

2.1.4 Bayes' Theorem between Continuous-valued Parameters

In most engineering applications, the updating concern is on continuous-valued uncertain parameters with measurements of continuous-valued variables. Even though there is quantization for the measurements, it is more convenient to treat them as continuous-valued variables and probability density functions are used to characterize their statistical behavior. With the same treatment for event B, event A can also represent the occurrence of another parameter vector $\boldsymbol{\psi}$

www.EngineeringBooksPdf.com

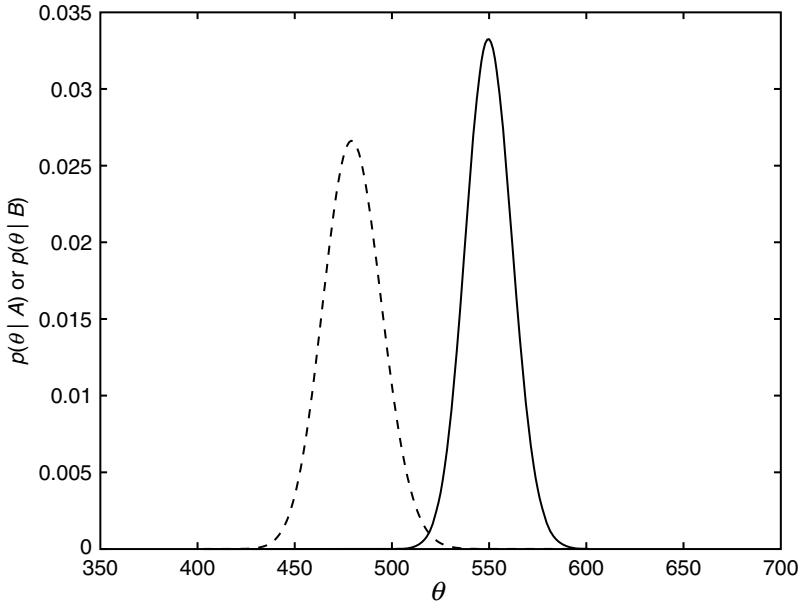

Figure 2.3 Conditional probability density functions of the scores

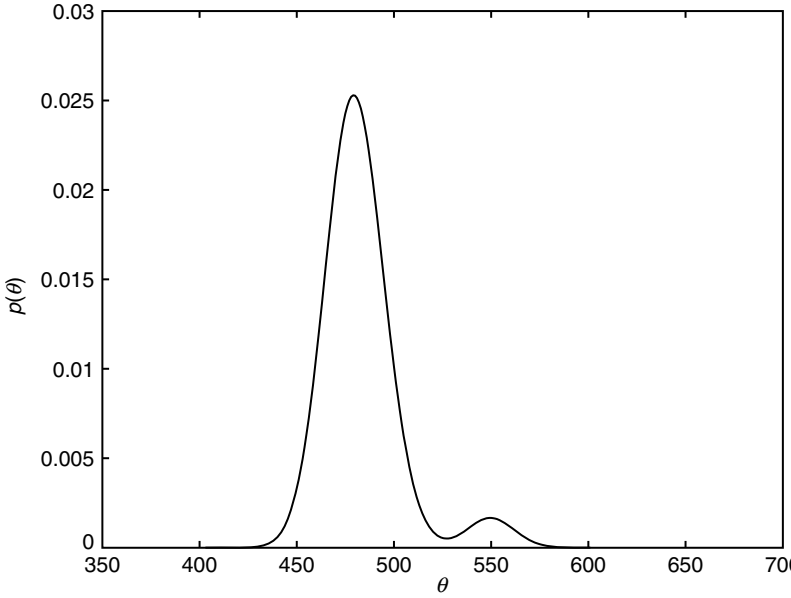

Figure 2.4 Probability density function of the scores

www.EngineeringBooksPdf.com

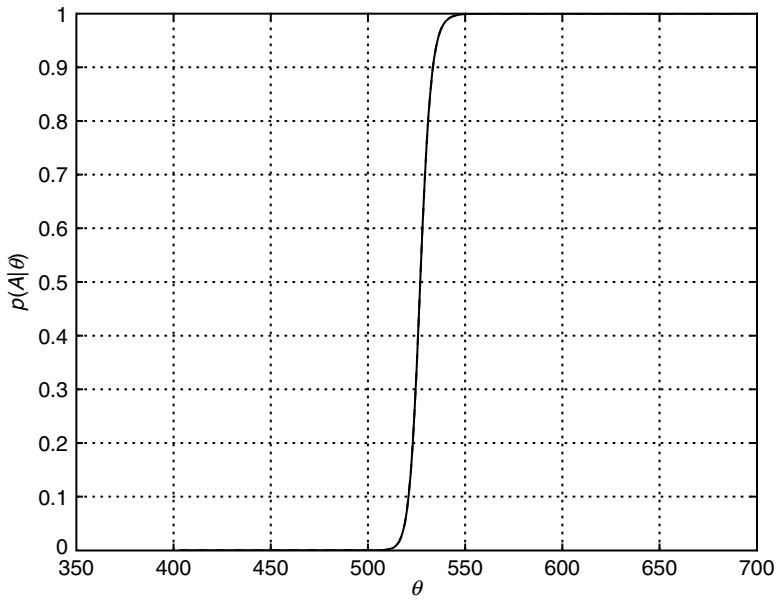

Figure 2.5 Updated probability for different scores

falling into a hypercube. Then, the Bayes' theorem gives the expression for the updated PDF of the parameter vector $\boldsymbol{\theta}$:

$$p(\boldsymbol{\theta}|\boldsymbol{\psi}) = \frac{p(\boldsymbol{\psi}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{\psi})} \tag{2.18}$$

This form with probability densities is useful for most engineering applications and the methodologies presented in this book are built on it. The vector $\boldsymbol{\theta}$ represents the uncertain parameters for identification and the vector $\boldsymbol{\psi}$ represents the measurements of the system. Further discussion will be given in the following sections.

2.1.5 Bayesian Inference

In science and engineering problems, there are various uncertain parameters necessary to be determined for modeling and other purposes. The Bayes' theorem offers the possibility for inferencing uncertain models/systems from their measurements. There are two levels of system identification. The first is *parametric identification*, in which a class of mathematical models for a particular physical phenomenon or system is given with unknown parameters to be identified. The second level deals with the selection of a suitable class of mathematical models for parametric identification. This is significantly more difficult but more important than the first level since parametric identification results will be by no means meaningful if one fails to obtain a *suitable* class of models. However, due to the difficulty of this problem, it is usually determined by user's judgement. Chapters 2–5 focus on parametric identification and Chapter 6 addresses the problem of model class selection.

www.EngineeringBooksPdf.com

Use $\mathcal{D}$ to denote the measured data of a system and consider it as the vector $\boldsymbol{\psi}$ in Equation (2.18). Then, the updated/posterior probability density function (PDF) of the parameters $\boldsymbol{\theta}$ is:

$$p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}) = \kappa_0 p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) p(\boldsymbol{\theta}|\mathcal{C}) \tag{2.19}$$

where $\mathcal{C}$ denotes the class of probabilistic and physical models/rules used for the problem of concern and $\kappa_0 = 1/p(\mathcal{D}|\mathcal{C})$ is a normalizing constant such that integrating the right hand side over the parameter space $\boldsymbol{\Theta}$ yields unity:

$$\int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}) \mathrm{d}\boldsymbol{\theta} = \kappa_0 \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}|\mathcal{C}) p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) \mathrm{d}\boldsymbol{\theta} = 1 \tag{2.20}$$

The likelihood function $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$ represents the contribution of the measured data in establishing the posterior distribution. It reflects how likely the measurements are observed from the model with a particular set of parameters. The likelihood function can be constructed given the class of probabilistic and physical models of the problem and it is the key of Bayesian updating. If a large amount of measurement is available, the likelihood function will be the dominant factor for the Bayesian inference.

2.1.5.1 Prior Distributions

The prior distribution $p(\boldsymbol{\theta}|\mathcal{C})$ denotes the prior information of the parameters and it is based on previous knowledge or user's judgement. In some applications, the prior distribution is treated as a constant and it is absorbed into the normalizing constant but this type of prior distribution does not satisfy the property of the PDF that its integral throughout the parametric space is unity. In general, a prior distribution that does not satisfy this property is referred to as an *improper prior*. Using a constant improper prior distribution yields the maximum likelihood solution.

Another popular choice is the class of *conjugate prior distributions*. A prior distribution is said to be conjugate to a class of likelihood functions $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$ if the resulting posterior distributions $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$ are in the same family as the prior distribution [214]. For example, let's say the likelihood function has the form of *exponential distribution* (of x):

$$p(x|\theta, \mathcal{C}) = \alpha\theta \exp(-\theta x) \tag{2.21}$$

where x is the measurement and θ is the uncertain parameter for identification. Use Gamma distribution as the prior distribution:

$$p(\theta|\mathcal{C}) = \theta^{k-1} \frac{\exp(-\theta/\alpha)}{\alpha^k \Gamma(k)} \tag{2.22}$$

where $k > 0$ and $\alpha > 0$ are the shape and scale parameters that characterize the distribution. The Gamma function is given by:

$$\Gamma(k) = \int_0^\infty x^{k-1} \exp(-x) \mathrm{d}x \tag{2.23}$$

www.EngineeringBooksPdf.com

Using integration by parts, it can be easily shown that $\Gamma(k+1) = k\Gamma(k)$, $\forall k > 0$ and $\Gamma(k) = (k-1)!$ if k is a positive integer.

By the Bayes' theorem, the posterior PDF of θ is given by:

$$p(\theta|x, \mathcal{C}) = \kappa_0 \theta^k \exp\left[-\theta(x + 1/\alpha)\right] \tag{2.24}$$

which remains to be Gamma distributed for θ with shape parameter $k+1$ and scale parameter $(x + 1/\alpha)^{-1}$. Therefore, the Gamma distribution is the conjugate prior distribution for the exponential type of likelihood functions. Conjugate prior distributions are popular because the prior and posterior distributions have the same form so it is easy to quantify the contribution of the updating process by the data.

The maximum entropy method is a theoretically sound approach, especially for the case if the bounds of the parameters are known. The uncertainty of a random variable/vector can be quantified by the information entropy that depends on its probability density function [115, 117, 233]:

$$H(p) \equiv -E[\ln p(\boldsymbol{\theta})] = -\int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}) \ln p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \tag{2.25}$$

where $\boldsymbol{\Theta}$ is the domain of the parameter vector $\boldsymbol{\theta}$. The maximum entropy prior distribution can be found by maximizing the information entropy with the prescribed constraint on the bounds of the parameters. This approach is particularly suitable for the case that only the bounds of the random variable are known since the resultant prior distribution carries the least amount of information among all admissible distributions compatible with the boundary conditions. Further comments on the selection of prior distribution will be given in Chapter 6.

Example. Maximum Entropy Distribution for Given Mean and Variance
In this example, the maximum entropy distribution is to be determined for given values of mean μ and variance σ^2. This distribution can be found by maximizing the information entropy in Equation (2.25) among all the distributions with the target mean and variance. This can be achieved by using the Lagrange multiplier. First, define the Lagrange function L as follows:

$$\begin{aligned} L(p) &= -\int_{-\infty}^{\infty} p(\theta) \ln p(\theta) \mathrm{d}\theta + \lambda_1 \left(\int_{-\infty}^{\infty} p(\theta) \mathrm{d}\theta - 1 \right) + \lambda_2 \left(\int_{-\infty}^{\infty} \theta p(\theta) \mathrm{d}\theta - \mu \right) \\ &\quad + \lambda_3 \left(\int_{-\infty}^{\infty} (\theta - \mu)^2 p(\theta) \mathrm{d}\theta - \sigma^2 \right) \\ &= \int_{-\infty}^{\infty} \Big\{ -p(\theta) \ln p(\theta) + \left[\lambda_1 + \lambda_2 \theta + \lambda_3 (\theta - \mu)^2 \right] \\ &\qquad \times p(\theta) - (\lambda_1 + \lambda_2 \mu + \lambda_3 \sigma^2) f(\theta) \Big\} \mathrm{d}\theta \end{aligned} \tag{2.26}$$

where f is an arbitrary function that satisfies $\int_{-\infty}^{\infty} f(\theta) \mathrm{d}\theta = 1$. Note that maximizing the information entropy is equivalent to maximizing the Lagrange function if the contraints on the mean and variance are satisfied. Then, a function F is defined as the integrand of the last integral of Equation (2.26):

$$F(p) = -p(\theta) \ln p(\theta) + \left[\lambda_1 + \lambda_2 \theta + \lambda_3 (\theta - \mu)^2 \right] p(\theta) - (\lambda_1 + \lambda_2 \mu + \lambda_3 \sigma^2) f(\theta) \tag{2.27}$$

www.EngineeringBooksPdf.com

The necessary condition for an extremum to exist is that the variation of the Lagrange function L is zero, i.e., $\delta L = 0$. This is equivalent to solving $\partial F/\partial p = 0$:

$$-\ln p(\theta) - 1 + \lambda_1 + \lambda_2\theta + \lambda_3(\theta - \mu)^2 = 0 \tag{2.28}$$

and the solution is given by

$$p(\theta) = \exp\left[\lambda_3\theta^2 + (\lambda_2 - 2\lambda_3\mu)\theta + (\lambda_3\mu^2 + \lambda_1 - 1)\right] \tag{2.29}$$

By the three mathematical constraints on the volume of the distribution, the mean and variance of the random variable, it can be easily shown that $\lambda_1 = 1 - \ln(\sqrt{2\pi}\sigma)$, $\lambda_2 = 0$ and $\lambda_3 = -1/2\sigma^2$. Therefore, for given values of mean and variance, the probability distribution that has the extremum information entropy is:

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(\theta-\mu)^2}{2\sigma^2}\right] \tag{2.30}$$

which is a Gaussian distribution. Furthermore, the second derivative of the function F is:

$$\frac{\partial^2 F}{\partial p^2} = -\frac{1}{p(\theta)} < 0 \tag{2.31}$$

so the resultant Gaussian distribution is the maximum entropy solution. In this case, the information entropy for a Gaussian distribution is given by:

$$H = -\frac{1}{2}(1 + \ln 2\pi) + \ln\sigma \tag{2.32}$$

If the random variable is known to lie within the interval $[a, b]$ with given mean μ and variance σ^2, the maximum entropy solution will be a truncated Gaussian distribution:

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma\left[\phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)\right]}\exp\left[-\frac{(\theta-\mu)^2}{2\sigma^2}\right], \quad \theta \in [a, b] \tag{2.33}$$

where ϕ is the cumulative distribution function (CDF) of the standard Gaussian random variable:

$$\phi(\theta) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\theta}\exp\left(-\frac{\alpha^2}{2}\right)\mathrm{d}\alpha \tag{2.34}$$

There is a one-to-one mapping between this function and the error function:

$$\phi(\theta) = \frac{1}{2} + \frac{1}{2}erf\left(\frac{\theta}{\sqrt{2}}\right) \tag{2.35}$$

where the error function is given by:

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp\left(-\alpha^2\right)\mathrm{d}\alpha \tag{2.36}$$

and it can be computed using the function 'erf' in MATLAB® [171].

www.EngineeringBooksPdf.com

Another useful special case is that the random variable lies within a finite interval $[a, b]$ without knowing the mean, variance or other moments. The above method can be applied and the Lagrange function in Equation (2.26) will be modified to exclude the terms with λ_2 and λ_3. It turns out that the maximum entropy distribution is the uniform distribution in $[a, b]$:

$$p(\theta) = \frac{1}{b-a}, \text{ if } \theta \in [a, b] \tag{2.37}$$

and zero otherwise.

2.1.6 Examples of Bayesian Inference

Example. Gaussian Random Variable
Consider a Gaussian random variable X with mean μ and variance σ^2. These two parameters are unknown and they are the uncertain parameters for identification. Independent measurements of the random variable X are available: $\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$ so the likelihood function is given by:

$$\begin{aligned} p(\mathcal{D}|\mu, \sigma^2, \mathcal{C}) &= \prod_{n=1}^{N} p(x_n|\mu, \sigma^2, \mathcal{C}) \\ &= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_n-\mu)^2}{2\sigma^2}\right] \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right] \end{aligned} \tag{2.38}$$

where the model class $\mathcal{C}$ denotes the Gaussian probabilistic model for the random variable. Improper prior is used for the uncertain parameters and it is absorbed into the normalizing constant. In this case, the statistical inference will rely solely on the likelihood of the data and the posterior PDF is given by:

$$\begin{aligned} p(\mu, \sigma^2|\mathcal{D}, \mathcal{C}) &= \frac{\kappa_0}{\sigma^N} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right] \\ &= \frac{\kappa_0}{\sigma^N} \exp\left[-\frac{1}{2\sigma^2}\left(N\mu^2 - 2\mu\sum_{n=1}^{N}x_n + \sum_{n=1}^{N}x_n^2\right)\right] \end{aligned} \tag{2.39}$$

where κ_0 is the normalizing constant. This solution is the maximum likelihood solution. Note that the joint posterior PDF is non-Gaussian although the conditional PDF $p(\mu|\sigma^2, \mathcal{D}, \mathcal{C})$ is Gaussian.

To obtain the joint posterior PDF, it requires only the values of $\sum_{n=1}^{N} x_n$ and $\sum_{n=1}^{N} x_n^2$ or, equivalently, the sample average and sample variance. The most probable values of the

www.EngineeringBooksPdf.com

parameters can be obtained by maximizing the posterior PDF or, equivalently, minimizing the objective function:

$$J(\mu, \sigma^2|\mathcal{D}, \mathcal{C}) = N \ln \sigma + \frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - \mu)^2 \tag{2.40}$$

which is the negative logarithm of the posterior PDF without including the constant. By solving $\partial J(\mu, \sigma^2|\mathcal{D}, \mathcal{C})/\partial\mu = 0$, the most probable value (mode) for the mean parameter μ is:

$$\mu^\star = \frac{1}{N} \sum_{n=1}^{N} x_n \tag{2.41}$$

which is the sample average. Equation (2.39) states that the conditional PDF $p(\mu|\sigma^2, \mathcal{D}, \mathcal{C})$ is Gaussian with mean $\mu^\star$ and variance σ^2/N. The conditional mean does not depend on the value of σ^2 but the conditional variance does.

On the other hand, the marginal PDF for μ is given by:

$$\begin{aligned} p(\mu|\mathcal{D}, \mathcal{C}) &= \int_0^\infty p(\mu, \sigma^2|\mathcal{D}, \mathcal{C}) \mathrm{d}\sigma^2 \\ &= \int_0^\infty \frac{\kappa_0}{\sigma^N} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - \mu)^2\right] \mathrm{d}\sigma^2 \\ &= \kappa_0 \Gamma\left(\frac{N}{2} - 1\right) \left[\frac{1}{2} \sum_{n=1}^{N} (x_n - \mu)^2\right]^{-\frac{N}{2}+1} \end{aligned} \tag{2.42}$$

which is non-Gaussian but it approaches to Gaussian for large N.

Similarly, the most probable value (mode) for the variance parameter σ^2 can be obtained by solving $\partial J(\mu, \sigma^2|\mathcal{D}, \mathcal{C})/\partial\sigma^2 = 0$:

$$\sigma^{2\star} = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu^\star)^2 \tag{2.43}$$

which is the (biased) sample variance.

Again, by observing Equation (2.39), the conditional PDF $p(\sigma^2|\mu, \mathcal{D}, \mathcal{C})$ follows the inverse Gamma distribution, denoted by $IG(\alpha, \beta)$:

$$p(X) = \frac{\beta^\alpha}{\Gamma(\alpha) X^{\alpha+1}} \exp\left(-\frac{\beta}{X}\right) \tag{2.44}$$

with the shape and scale parameters given by

$$\begin{aligned} \alpha &= \frac{N}{2} - 1 \\ \beta &= \frac{1}{2} \sum_{n=1}^{N} (x_n - \mu)^2 \end{aligned} \tag{2.45}$$

www.EngineeringBooksPdf.com

and $\Gamma(.)$ is the Gamma function given in Equation (2.23). If $\alpha > 2$, the mean and variance of this distribution are:

$$E[X] = \frac{\beta}{\alpha - 1} \tag{2.46}$$

and

$$E[(X - E[X])^2] = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)} \tag{2.47}$$

The marginal PDF for σ^2 is:

$$\begin{aligned} p(\sigma^2|\mathcal{D}, \mathcal{C}) &= \int_{-\infty}^{\infty} p(\mu, \sigma^2|\mathcal{D}, \mathcal{C})\mathrm{d}\mu \\ &= \int_{-\infty}^{\infty} \frac{\kappa_0}{\sigma^N} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu)^2\right]\mathrm{d}\mu \\ &= \frac{\kappa_0}{\sigma^{N-1}}\sqrt{\frac{2\pi}{N}} \exp\left(-\frac{N\sigma^{2\star}}{2\sigma^2}\right) \end{aligned} \tag{2.48}$$

which is also inverse Gamma $IG(\alpha, \beta)$ with shape and scale parameters given by:

$$\alpha = \frac{N-3}{2}; \quad \beta = \frac{N}{2}\sigma^{2\star} \tag{2.49}$$

By using the fact that $p(X)$ in Equation (2.44) is a PDF, the following definite integral has a closed-form solution:

$$\int_0^{\infty} X^{-\alpha-1} \exp(-\beta/X)\mathrm{d}X = \beta^{-\alpha}\Gamma(\alpha) \tag{2.50}$$

and this can be used to determine the normalizing constant:

$$\kappa_0 = \left(\frac{N}{2}\right)^{\frac{N}{2}-1} \frac{(\sigma^\star)^{N-3}}{\Gamma\left(\frac{N-3}{2}\right)\sqrt{\pi}} \tag{2.51}$$

Consider 10 samples of the random variable X, i.e., $N = 10$, with sample average $\mu^\star = \frac{1}{N}\sum_{n=1}^{N} x_n = 5$ and sample variance $\sigma^{2\star} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu^\star)^2 = 4$. The posterior PDF can be computed by Equation (2.39) and it is shown in Figure 2.6. It is a skewed distribution and its contours with 75%, 50%, 25%, 10% and 1% of the maximum probability density value are shown in Figure 2.7. The posterior PDF is symmetric about the line $\mu = \mu^\star (= 5)$.

Then, consider 400 samples ($N = 400$) with the same sample average and sample variance. The posterior PDF is shown in Figure 2.8. By taking a larger number of data points, the posterior PDF concentrates in a significantly smaller region and the distribution becomes closer to Gaussian. This can be visualized by the oval-like contours in Figure 2.9.

www.EngineeringBooksPdf.com

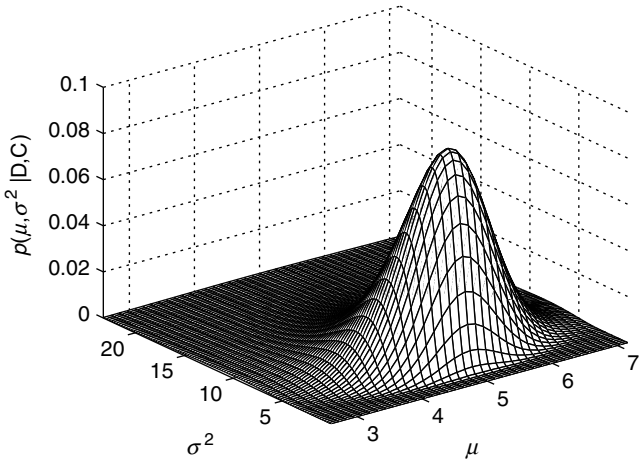

Figure 2.6 Posterior PDF

Example. Exponential Random Variable
The PDF of an exponential random variable X with mean μ is given by:

$$p(X) = \frac{1}{\mu} \exp\left(-\frac{X}{\mu}\right) \tag{2.52}$$

This random variable can also be treated as a scaled Chi-square random variable with two degrees of freedom. Specifically, X can be expressed as:

$$X = \frac{\mu}{2}(Y_1^2 + Y_2^2) \tag{2.53}$$

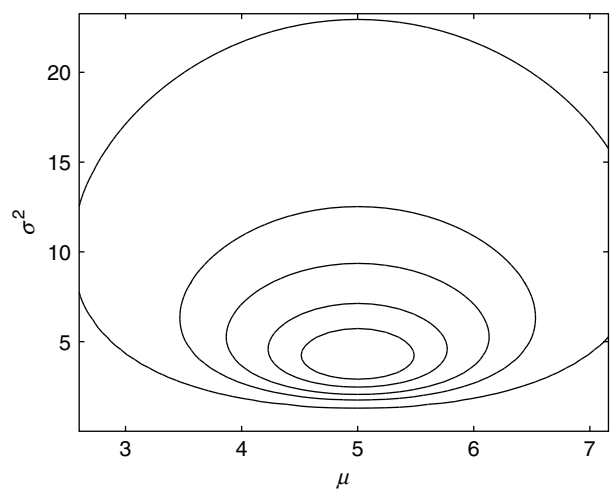

Figure 2.7 Contours of the posterior PDF

www.EngineeringBooksPdf.com

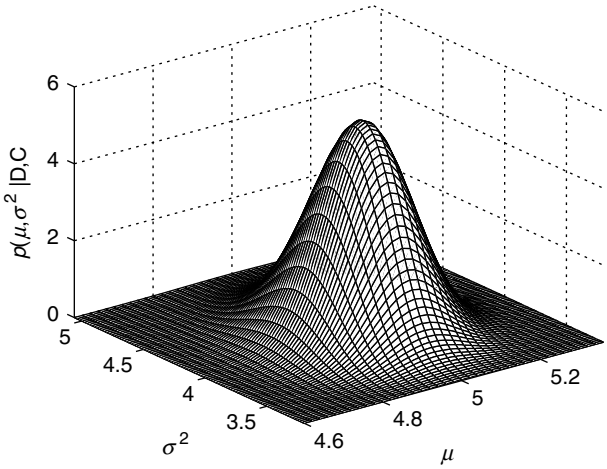

Figure 2.8 Posterior PDF ($N = 400$)

where Y_1 and Y_2 are independent standard Gaussian random variables. In this example, the mean parameter is unknown and it is the uncertain parameter for identification. The variance of this random variable is given by $\sigma^2 = \mu^2$. Independent measurements of X are available: $\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$ (with $N \geq 3$) so the likelihood function is given by:

$$p(\mathcal{D}|\mu, \mathcal{C}) = \prod_{n=1}^{N} p(x_n|\mu, \mathcal{C}) = \prod_{n=1}^{N} \frac{1}{\mu} \exp\left(-\frac{x_n}{\mu}\right) = \frac{1}{\mu^N} \exp\left(-\frac{1}{\mu}\sum_{n=1}^{N} x_n\right) \tag{2.54}$$

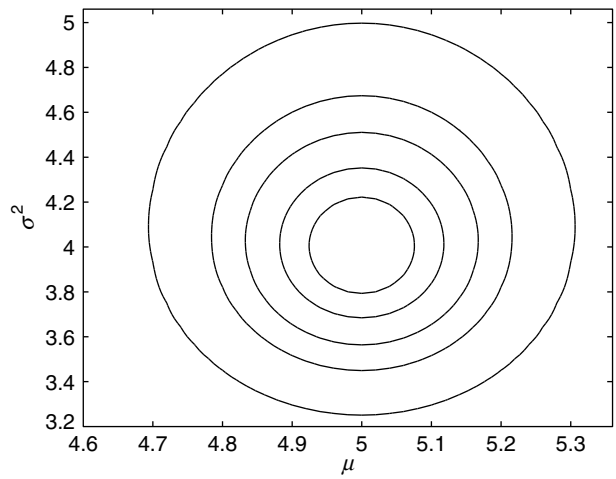

Figure 2.9 Contours of the posterior PDF ($N = 400$)

www.EngineeringBooksPdf.com

Uniform prior in a sufficiently large interval is used for the uncertain parameter so that the Bayesian inference relies solely on the likelihood of the data:

$$p(\mu|\mathcal{C}) = \begin{cases} \dfrac{1}{\mu_u - \mu_l}, & \text{if } \mu \in [\mu_l, \mu_u] \\ 0, & \text{otherwise} \end{cases} \tag{2.55}$$

Then, the posterior PDF of the parameter can be readily obtained:

$$p(\mu|\mathcal{D}, \mathcal{C}) = \begin{cases} \dfrac{\kappa_0}{(\mu_u - \mu_l)\mu^N} \exp\left(-\dfrac{1}{\mu}\displaystyle\sum_{n=1}^{N} x_n\right), & \text{if } \mu \in [\mu_l, \mu_u] \\ 0, & \text{otherwise} \end{cases} \tag{2.56}$$

which is a truncated inverse Gamma distribution with $\alpha = N - 1$ and $\beta = \sum_{n=1}^{N} x_n$. If the interval $[\mu_l, \mu_u]$ covers a sufficiently large range, the mean of the distribution can be approximated by:

$$E[\mu|\mathcal{D}, \mathcal{C}] \approx \frac{1}{N-2} \sum_{n=1}^{N} x_n \tag{2.57}$$

Furthermore, the normalizing constant is given by:

$$\kappa_0 \approx \frac{(\mu_u - \mu_l)}{\Gamma(N-1)} \sum_{n=1}^{N} x_n \tag{2.58}$$

The error in this approximation is due to the truncation of the distribution and the correct value of κ_0 should be larger.

By solving $\partial p(\mu|\mathcal{D}, \mathcal{C})/\partial\mu = 0$, the most probable value (mode) for the mean parameter μ is equal to the sample average:

$$\mu^\star = \frac{1}{N} \sum_{n=1}^{N} x_n \tag{2.59}$$

provided that $\frac{1}{N}\sum_{n=1}^{N} x_n \in [\mu_l, \mu_u]$. For a given value of the sample average, the posterior PDF of the mean parameter μ is uniquely determined regardless of the individual values of the samples. Other statistical moments may be utilized for validation of the selected distribution. Consider $N = 10$ independent samples with sample average equal to 5. The posterior PDF can be computed by Equation (2.56) and it is shown in Figure 2.10. It is skewed with a long tail on the right side. However, if the number of data points is increased to $N = 100$ with the same sample average, the posterior PDF depicted in Figure 2.11 is more concentrated since more information for the uncertain mean parameter is obtained from the additional samples. Meanwhile, the distribution approaches to Gaussian.

www.EngineeringBooksPdf.com

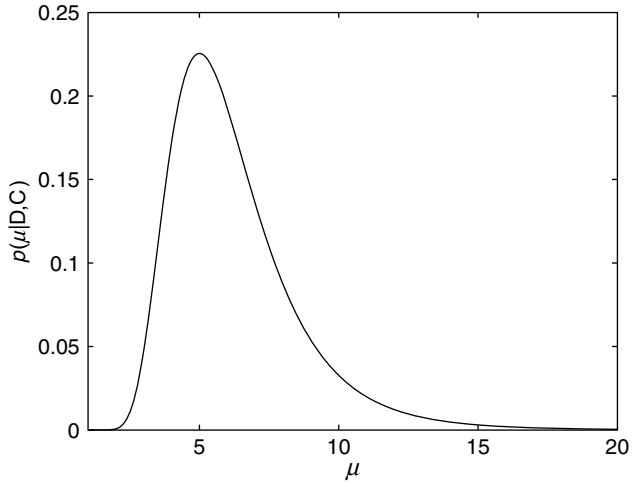

Figure 2.10 Posterior PDF ($N = 10$)

Example. Occurrence Rate of an Event
Consider the problem of updating the occurrence rate of an event, such as a certain level of earthquakes or typhoons in a particular region. The discrete Poisson distribution is a well-known probabilistic model for this purpose and the probability of exactly k (≥ 0) occurrences of an event in a specified time interval is given by:

$$P(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{2.60}$$

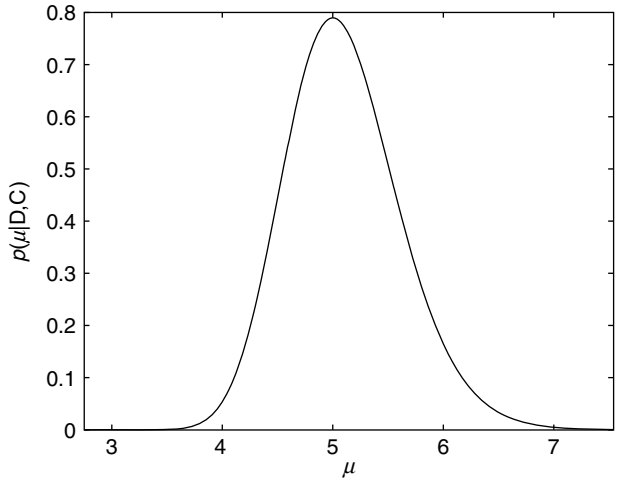

Figure 2.11 Posterior PDF ($N = 100$)

www.EngineeringBooksPdf.com

where λ is the *occurrence rate*. Assume that observations of N years are available and there are $k_1, k_2, \ldots, k_N$ occurrences in each of the years. By assuming that these occurrences in different years are statistically independent, the posterior PDF of the occurrence rate λ is given by:

$$\begin{aligned} p(\lambda|k_1, k_2, \ldots, k_N, \mathcal{C}) &= \frac{\kappa_0 p(\lambda|\mathcal{C})}{\prod_{n=1}^{N} k_n!} \prod_{n=1}^{N} \frac{\lambda^{k_n} \mathrm{e}^{-\lambda}}{k_n!} \\ &= \kappa_0 p(\lambda|\mathcal{C}) \exp\left(-N\lambda + \sum_{n=1}^{N} k_n \ln \lambda\right) \end{aligned} \tag{2.61}$$

Here, a uniform prior distribution for the occurrence rate λ is used in a sufficiently large range so the inference is contributed solely by the likelihood of the measurement. In this case, the posterior PDF in Equation (2.61) is truncated Gamma distributed with shape parameter and scale parameter given by:

$$\begin{aligned} \alpha &= \sum_{n=1}^{N} k_n \\ \beta &= N^{-1} \end{aligned} \tag{2.62}$$

By maximizing the posterior PDF with respect to λ, the most probable value of the occurrence rate can be obtained from solving $\partial p(\lambda|k_1, k_2, \ldots, k_N, \mathcal{C})/\partial\lambda = 0$:

$$\lambda^{\star} = \frac{1}{N} \sum_{n=1}^{N} k_n \tag{2.63}$$

The associated uncertainty of this estimate is completely described by the posterior PDF shown in Figure 2.12 for $N = 10$ and $\frac{1}{N}\sum_{n=1}^{N} k_n = 5$. The mean and standard deviation of the distribution are approximately 5.1 and 0.71, respectively.

Consider a special case of a rare event that did not happen in the previous N years. By using the Poisson model again, the most probable value of the occurrence rate is equal to zero, i.e., $\lambda^{\star} = 0$. However, this does not imply that the event will never happen. By Equation (2.61) with an improper prior, the posterior PDF of λ is:

$$p(\lambda|k_1 = 0, k_2 = 0, \ldots, k_N = 0, \mathcal{C}) = N\mathrm{e}^{-N\lambda} \tag{2.64}$$

which is an exponential distribution. The expected value of λ is given by:

$$E[\lambda|\mathcal{D}, \mathcal{C}] = N^{-1} \tag{2.65}$$

with variance N^{-2}. This example demonstrates an important distinction between Bayesians and frequentists. In the viewpoint of frequentists, the occurrence rate of this event cannot be estimated (if not zero) since the *relative frequency* of the occurrence of this event is zero in the observed duration, which is not sufficiently long. However, the Bayesian approach utilizes the complete information of the data to estimate the plausibility of different values of the occurrence rate.

www.EngineeringBooksPdf.com

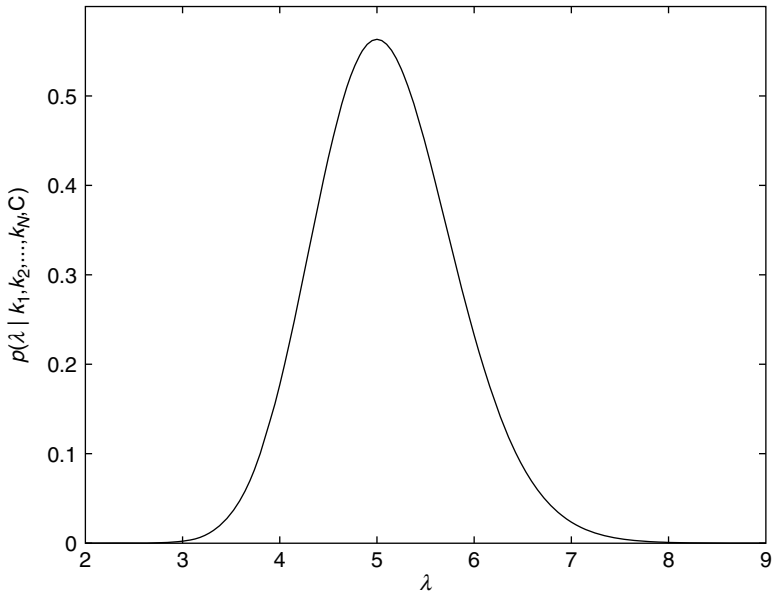

Figure 2.12 Posterior PDF $p(\lambda|k_1, k_2, \ldots, k_N, \mathcal{C})$

Example. Does Small Parametric Uncertainty Imply Good Data Fitting?
Consider a quantity Q and its two measured values: $\hat{Q}_1$ and $\hat{Q}_2$. By the specifications of the instrument, it is known that the measurement noise for both data points is Gaussian with zero mean and variance σ_0^2. Now, the value of Q is updated using the data by the Bayes' theorem:

$$p(Q|\hat{Q}_1, \hat{Q}_2, \mathcal{C}) = \kappa_0(2\pi\sigma_0^2)^{-1} \exp\left[-\frac{(Q-\hat{Q}_1)^2 + (Q-\hat{Q}_2)^2}{2\sigma_0^2}\right]$$

$$= \kappa_0(2\pi\sigma_0^2)^{-1} \exp\left[-\frac{(\hat{Q}_1-\hat{Q}_2)^2}{4\sigma_0^2}\right] \exp\left[-\frac{\left(Q - \frac{\hat{Q}_1+\hat{Q}_2}{2}\right)^2}{\sigma_0^2}\right] \quad (2.66)$$

Here, an improper prior is used and it is absorbed into the normalizing constant. Also, the model class $\mathcal{C}$ describes all the assumptions made for the measurements and the noise model. It is clear that the posterior distribution for Q is Gaussian with mean $(\hat{Q}_1 + \hat{Q}_2)/2$. The variance is $\sigma_0^2/2$ and it does not depend on the values of $\hat{Q}_1$ and $\hat{Q}_2$.

Now, consider the case with $\hat{Q}_1 = 1$, $\hat{Q}_2 = 5$ and $\sigma_0^2 = 0.01$. By Equation (2.66), the mean and variance are 3 and 0.005, respectively. Figure 2.13 shows the posterior PDF for three cases: (1) using $\hat{Q}_1$ only (the peak on the left hand side), (2) using $\hat{Q}_2$ only (the peak on the right hand side), and (3) using both $\hat{Q}_1$ and $\hat{Q}_2$ (the peak in the middle). The posterior variance with

www.EngineeringBooksPdf.com

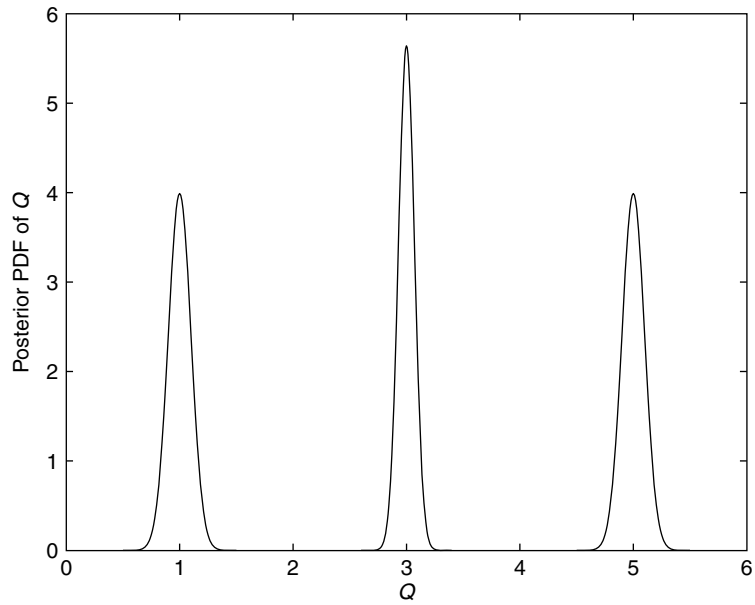

Figure 2.13 Posterior PDF of Q with $\hat{Q}_1$ or $\hat{Q}_2$ or both

both data points is $\sigma_0^2/2 = 0.005$ and the coefficient of variation is equal to $0.1/3\sqrt{2} = 2.36\%$. However, the data fitting is by no means a good one since the likelihood for each data point is small, i.e., $(\sqrt{2\pi}0.1)^{-1}\exp(-2^2/0.02) = 5.52 \times 10^{-87}$. Therefore, the maximum likelihood value for using both $\hat{Q}_1$ and $\hat{Q}_2$ will be 3.05×10^{-173}.

If σ_0^2 is reduced by half to 0.005, the posterior variance of Q will be even smaller, i.e., 0.0025, and the coefficient of variation is $0.05/3\sqrt{2} = 1.18\%$. However, the maximum likelihood value will become 3.18×10^{-799}. Therefore, small posterior uncertainty does not necessarily associate with a large likelihood value or good fitting to the data. In this case, the smaller posterior uncertainty is simply due to the underestimation of the measurement noise.

2.2 Bayesian Model Updating with Input–output Measurements

2.2.1 Input–output Measurements

In this section, the general Bayesian framework is presented. It was originally presented for structural model updating using input–output measurements in Beck and Katafygiotis [19]. Consider a linear or nonlinear dynamical system with input–output relationship:

$$\mathbf{x}(t) = \varphi(t; \mathbf{F}, \mathbf{x}(0), \boldsymbol{\theta}_m) \tag{2.67}$$

where t denotes the time, $\mathbf{x}(t) \in \mathbb{R}^{N_d}$ is the model response vector at time t, $\mathbf{x}(0)$ is the initial condition, and $\mathbf{F}$ is an external input/excitation, which is assumed to be measured. The input–output relationship is governed by the function φ, which is characterized by the model parameters

www.EngineeringBooksPdf.com

$\boldsymbol{\theta}_m$. For static systems, the relationship in Equation (2.67) is reduced to $\mathbf{x}(t) = \varphi(\mathbf{F}(t), \boldsymbol{\theta}_m)$and t denotes an index.

Assume that discrete response data are available at $N_o(\leq N_d)$ observed DOFs, i.e., some selected components of $\mathbf{x}(t)$ and/or their linear combinations. Use Δt to denote the sampling time step. Due to measurement noise and modeling error, referred to hereafter as prediction error, the measured response $\mathbf{y}_n \in \mathbb{R}^{N_o}$ (at time $t = n\Delta t$) will differ from the model response $\mathbf{L}_o\mathbf{x}(n\Delta t)$ corresponding to the measured degrees of freedom where $\mathbf{L}_o$ denotes an $N_o \times N_d$ observation matrix, which is determined by the configuration of the sensing system. Therefore:

$$\mathbf{y}_n = \mathbf{L}_o\mathbf{x}(n\Delta t) + \boldsymbol{\epsilon}_n \tag{2.68}$$

The prediction error is modeled as a discrete zero-mean Gaussian white noise vector process $\boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon}_n \in \mathbb{R}^{N_o}$, and it satisfies the following correlation structure:

$$E\left[\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_{n'}^T\right] = \boldsymbol{\Sigma}_\epsilon\delta_{nn'} \tag{2.69}$$

where $E[.]$ denotes the mathematical expectation, $\boldsymbol{\Sigma}_\epsilon$ denotes the $N_o \times N_o$ covariance matrix of the prediction error process $\boldsymbol{\epsilon}$ and $\delta_{nn'}$ denotes the Kronecker delta:

$$\delta_{nn'} = \begin{cases} 1 & \text{if } n = n' \\ 0 & \text{if } n \neq n' \end{cases} \tag{2.70}$$

2.2.2 Bayesian Parametric Identification

Use $\boldsymbol{\theta}$ to denote the parameter vector for identification. It includes the model parameters $\boldsymbol{\theta}_m$ and the parameters that determine the elements of the upper right triangular part of the prediction-error covariance matrix $\boldsymbol{\Sigma}_\epsilon$ (symmetry defines the lower triangular part of this matrix).

The dynamic data $\mathcal{D}$ consists of the measured time histories at N discrete time steps of the excitation and system response. Assume equal variances and stochastic independence for the prediction errors of different channels of measurements so the covariance matrix for the prediction errors is:

$$\boldsymbol{\Sigma}_\epsilon = \sigma_\epsilon^2\mathbf{I}_{N_o} \tag{2.71}$$

where $\mathbf{I}_{N_o}$ is the $N_o \times N_o$ identity matrix. Then, the updated PDF of the uncertain parameters in $\boldsymbol{\theta} = [\boldsymbol{\theta}_m^T, \sigma_\epsilon^2]^T$ given the data $\mathcal{D}$ and model class $\mathcal{C}$ can be expressed as:

$$p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}) = \kappa_0 p(\boldsymbol{\theta}|\mathcal{C})(2\pi)^{-\frac{NN_o}{2}}\sigma_\epsilon^{-NN_o}\exp\left[-\frac{NN_o}{2\sigma_\epsilon^2}J_g(\boldsymbol{\theta}_m; \mathcal{D}, \mathcal{C})\right] \tag{2.72}$$

where κ_0 is a normalizing constant and $p(\boldsymbol{\theta}|\mathcal{C})$ is the prior PDF of the uncertain parameters in $\boldsymbol{\theta}$, expressing the user's judgement about the relative plausibility of the values of the uncertain parameters before the data are used. The *goodness-of-fit* function is given by:

$$J_g(\boldsymbol{\theta}_m; \mathcal{D}, \mathcal{C}) = \frac{1}{NN_o}\sum_{n=1}^{N}\left|\left|\mathbf{y}_n - \mathbf{L}_o\mathbf{x}(n\Delta t; \boldsymbol{\theta}_m, \mathcal{C})\right|\right|^2 \tag{2.73}$$

www.EngineeringBooksPdf.com

where $\mathbf{x}(n\Delta t; \boldsymbol{\theta}_m, \mathcal{C})$ is the model response based on the assumed class of models and the model parameter vector $\boldsymbol{\theta}_m$ while $\mathbf{y}_n$ is the measured response at time $n\Delta t$. Furthermore, $||.||$ denotes the Euclidean norm (2-norm) of a vector. The most probable model parameter vector $\boldsymbol{\theta}^\star$ is obtained by maximizing the posterior PDF in Equation (2.72). For large N or with improper prior, this is equivalent to minimizing the goodness-of-fit function $J_g(\boldsymbol{\theta}_m; \mathcal{D}, \mathcal{C})$ in Equation (2.73) over all possible values of $\boldsymbol{\theta}_m$:

$$\boldsymbol{\theta}_m^\star = \arg\min_{\boldsymbol{\theta}_m} J_g(\boldsymbol{\theta}_m; \mathcal{D}, \mathcal{C}) \tag{2.74}$$

If $J_g(\boldsymbol{\theta}_m; \mathcal{D}, \mathcal{C})$ is known only implicitly, numerical optimization is needed to search for the optimal model parameters and this can be done by the function 'fminsearch' in MATLAB® [171].

The most probable value of the prediction-error variance in $\boldsymbol{\theta}^\star$ can be obtained also by maximizing the posterior PDF:

$$\frac{\partial p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})}{\partial \sigma_\epsilon^2} = 0 \tag{2.75}$$

and the solution is available in closed form:

$$\sigma_\epsilon^{2\star} = \min_{\boldsymbol{\theta}_m} J_g(\boldsymbol{\theta}_m; \mathcal{D}, \mathcal{C}) = J_g(\boldsymbol{\theta}_m^\star; \mathcal{D}, \mathcal{C}) \tag{2.76}$$

2.2.3 Model Identifiability

The updated PDF in Equation (2.72) provides the complete description of the plausibility of the model parameters but its topology may be very complicated in general, especially if there are a large number of uncertain parameters so the distribution is difficult to be visualized. Use $\mathcal{S}_{mod}(\boldsymbol{\theta}_0; \mathbf{F})$ to denote the set of all model parameters which give the same output at the observed degrees of freedom as the model associated with $\boldsymbol{\theta}_0$ and the input $\mathbf{F}$.

A parameter θ_l of $\boldsymbol{\theta}$ is *model-identifiable* [132] at $\boldsymbol{\theta}^\star$ for model class $\mathcal{C}$ and input $\mathbf{F}$ if there exists a positive value ϵ_l such that:

$$\boldsymbol{\theta} \in \mathcal{S}_{mod}(\boldsymbol{\theta}^\star; \mathbf{F}) \Rightarrow |\theta_l - \theta_l^\star| > \epsilon_l \text{ or } \theta_l = \theta_l^\star \tag{2.77}$$

In other words, $\theta_l^\star$ is uniquely specified within a neighborhood of each of its possible values by $\mathbf{F}$ and $\mathcal{D}$. There are three main categories of identifiability [132]:

1. A parameter θ_l of $\boldsymbol{\theta}$ is *globally model-identifiable* at $\boldsymbol{\theta}^\star$ for model class $\mathcal{C}$ and input $\mathbf{F}$ if:

$$\boldsymbol{\theta} \in \mathcal{S}_{mod}(\boldsymbol{\theta}^\star; \mathbf{F}) \Rightarrow \theta_l = \theta_l^\star \tag{2.78}$$

   In other words, $\theta_l^\star$ is uniquely specified by $\mathbf{F}$ and $\mathcal{D}$. If θ_l is globally model-identifiable at $\boldsymbol{\theta}^\star$, then it is also model-identifiable at $\boldsymbol{\theta}^\star$.
2. A parameter θ_l of $\boldsymbol{\theta}$ is *locally model-identifiable* at $\boldsymbol{\theta}^\star$ for model class $\mathcal{C}$ and input $\mathbf{F}$ if it is model-identifiable but not globally model-identifiable.
3. A parameter θ_l of $\boldsymbol{\theta}$ is *model-unidentifiable* if it is not model-identifiable.

www.EngineeringBooksPdf.com

For dynamic testing, it is the number of observed degrees of freedom, N_o, and their distribution that are the essential factors for identifiability. On the other hand, increasing the number of data points N does not increase the number of effective mathematical constraints, and hence the identifiability. For example, an unidentifiable case is obtained with a 5-min measurement for the identification of a structural model. Then, using a set of measurement of 10 min or longer does not affect the identifiability but it reduces only the variances in some of the principle directions.

For globally model-identifiable cases with large N [19], it turns out that the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$ is approximately Gaussian and this can be explained as follows. The measurement can be subdivided into a number of sub-records with equal length. Posterior PDFs utilizing different sub-records exist and they are similar if different sets of records are associated with similar condition (e.g., level of measurement noise, level of excitation, ...). By using the central limit theorem, the posterior PDF utilizing the complete record is approximately Gaussian even though the sub-records are not independent. Then, the mean is the optimal parameter vector $\boldsymbol{\theta}^\star$ and the covariance matrix is equal to the inverse of the Hessian matrix of the objective function $J(\boldsymbol{\theta}) = -\ln\left[p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})p(\boldsymbol{\theta}|\mathcal{C})\right]$ at $\boldsymbol{\theta}^\star$ (Appendix A):

$$\mathcal{H}^{(l,l')}(\boldsymbol{\theta}^\star) = -\frac{\partial^2}{\partial\theta_l\theta_{l'}} \ln\left[p(\boldsymbol{\theta}|\mathcal{C})p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})\right]\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \tag{2.79}$$

In practice, the prior distribution may be used as a *regularizer* [164, 169] to improve the well-posedness of the inverse problem:

$$p(\boldsymbol{\theta}|\mathcal{C}) = \frac{1}{(2\pi)^{\frac{N_\theta}{2}} \prod_{l=1}^{N_\theta} \sigma_l} \exp\left(-\frac{1}{2}\sum_{l=1}^{N_\theta} \frac{\theta_l^2}{\sigma_l^2}\right) \tag{2.80}$$

It is a Gaussian PDF with zero mean so it decays as the radial distance to the origin increases in any direction. If there exist two or more models that fit the measurement equally well, using this radially decaying prior distribution helps trimming down the set of the optimal parameters to the one with the smallest 2-norm.

Example. Consider a class of models (set of matrices) $\mathcal{C}$ in which the matrices are parameterized as follows:

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix} \theta_1 + \theta_2 & \theta_2 \\ \theta_2 & \theta_2 \end{bmatrix} \tag{2.81}$$

where $\boldsymbol{\theta} = [\theta_1, \theta_2]^T \in \mathbb{R}^2$. The actual values for θ_1 and θ_2 are taken to be $\tilde{\theta}_1 = \tilde{\theta}_2 = 1$ so the actual matrix is:

$$\tilde{\mathbf{A}} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \tag{2.82}$$

The *eigenvalue problem* for this matrix is given by:

$$\tilde{\mathbf{A}}\boldsymbol{\phi} = \lambda\boldsymbol{\phi} \tag{2.83}$$

www.EngineeringBooksPdf.com

where λ and $\boldsymbol{\phi}$ are the *eigenvalue* and *eigenvector* of the matrix $\tilde{\mathbf{A}}$. The eigenvalues can be found by solving the algebraic equation:

$$\left|\tilde{\mathbf{A}} - \lambda \mathbf{I}_2\right| = 0 \tag{2.84}$$

where $\mathbf{I}_2$ is the 2×2 identity matrix. In this case, the actual eigenvalues are $\lambda^{(1)} = 3/2 - \sqrt{5}/2 = 0.382$ and $\lambda^{(2)} = 3/2 + \sqrt{5}/2 = 2.618$ and their associated eigenvectors are $[1, -(\sqrt{5}/2 + 1/2)]^T$ and $[1, \sqrt{5}/2 - 1/2]^T$, respectively. Now, identification of the uncertain parameters in $\boldsymbol{\theta}$ is proceeded with different types of eigenvalue–eigenvector measurements.

Case 1: One measured eigenvalue
Assume that the smaller eigenvalue $\lambda^{(1)}$ is measured with 5% Gaussian measurement noise and five independent measurements are taken: $\hat{\lambda}_1^{(1)} = 0.3860$, $\hat{\lambda}_2^{(1)} = 0.3922$, $\hat{\lambda}_3^{(1)} = 0.4157$, $\hat{\lambda}_4^{(1)} = 0.3592$, and $\hat{\lambda}_5^{(1)} = 0.3615$. To simplify the problem, the standard deviation of the measurement noise is assumed known here and it is $\sigma_1 = 0.0191$. In this case, the likelihood function can be written as follows:

$$p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) = (2\pi)^{-\frac{5}{2}} \sigma_1^{-5} \exp\left[-\frac{1}{2\sigma_1^2} \sum_{n=1}^{5} \left(\lambda^{(1)}(\boldsymbol{\theta}) - \hat{\lambda}_n^{(1)}\right)^2\right] \tag{2.85}$$

where $\lambda^{(1)}(\boldsymbol{\theta})$ is the smaller model eigenvalue of the matrix $\mathbf{A}(\boldsymbol{\theta})$ and $\hat{\lambda}_n^{(1)}$, $n = 1, 2, \ldots, 5$, are its measurements. Figure 2.14 shows the likelihood function $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$ and the optimal parameters are non-unique. This is an unidentifiable case since there exists a ridge on which the likelihood function is maximized. Figure 2.15 shows the trajectory of (θ_1, θ_2) with maximum value of $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$ and the equation of this curve is given by:

$$\overline{\hat{\lambda}^{(1)}}^2 - (\theta_1 + 2\theta_2)\overline{\hat{\lambda}^{(1)}} + \theta_1\theta_2 = 0 \tag{2.86}$$

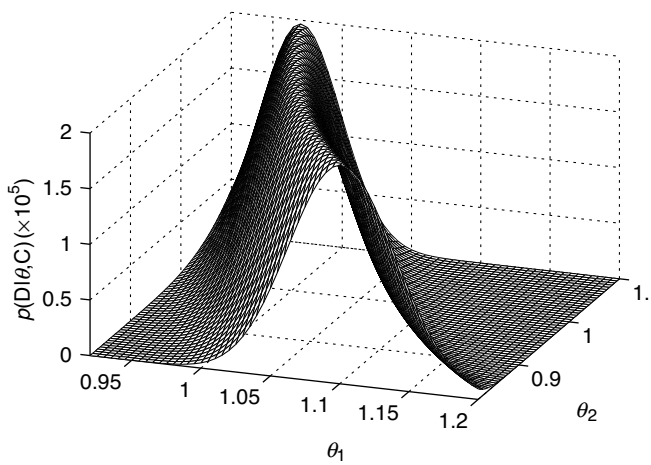

Figure 2.14 Likelihood function with one measured eigenvalue only

www.EngineeringBooksPdf.com

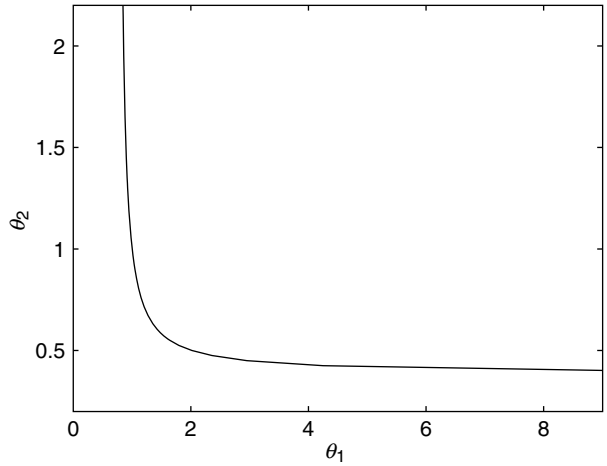

Figure 2.15 Trajectory of (θ_1, θ_2) with maximum likelihood value (unidentifiable case)

where $\overline{\hat{\lambda}^{(1)}} = 0.2\sum_{n=1}^{5}\hat{\lambda}_n^{(1)} = 0.3849$ is the average of the measurements of $\lambda^{(1)}$.

Case 2: Two measured eigenvalues
In addition to the smaller eigenvalue, the larger eigenvalues are also measured with 5% measurement noise and they are $\hat{\lambda}_1^{(2)} = 2.3614$, $\hat{\lambda}_2^{(2)} = 2.5877$, $\hat{\lambda}_3^{(2)} = 2.7070$, $\hat{\lambda}_4^{(2)} = 2.3875$, and $\hat{\lambda}_5^{(2)} = 2.7272$. Again, the standard deviations of the measurement noise are assumed known and they are $\sigma_1 = 0.0191$ and $\sigma_2 = 0.1309$. In this case, the likelihood function is given by:

$$p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) = (2\pi\sigma_1\sigma_2)^{-5}\exp\left[-\frac{1}{2\sigma_1^2}\sum_{n=1}^{5}\left(\lambda^{(1)}(\boldsymbol{\theta}) - \hat{\lambda}_n^{(1)}\right)^2 - \frac{1}{2\sigma_2^2}\sum_{n=1}^{5}\left(\lambda^{(2)}(\boldsymbol{\theta}) - \hat{\lambda}_n^{(2)}\right)^2\right] \tag{2.87}$$

and it is shown in Figure 2.16. There are two local maxima of equal probability density so it is a locally identifiable case. Figure 2.17 shows the contours of $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$ and it can be clearly seen that the probability concentrates in the two regions around the local maxima. These two local maxima lie on the trajectory in Figure 2.15. The additional measured eigenvalue provides extra constraint in reducing the number of most probable values of the parameters from infinity to two.

Case 3: One measured eigenvalue with the corresponding eigenvector
In this case, only the smaller eigenvalue is measured and its value is the same as in the previous cases. In addition, the corresponding eigenvector is also measured. Note that the eigenvector can be determined up to a scaling constant only because $\alpha\boldsymbol{\phi}$ is an eigenvector if $\boldsymbol{\phi}$ is an eigenvector and $\alpha \in \mathbb{R} - \{0\}$. Therefore, instead of using the eigenvector directly, the eigenvector component ratio $r \equiv \phi_2^{(1)}/\phi_1^{(1)}$ is utilized for identification and the five independent observations are:

www.EngineeringBooksPdf.com

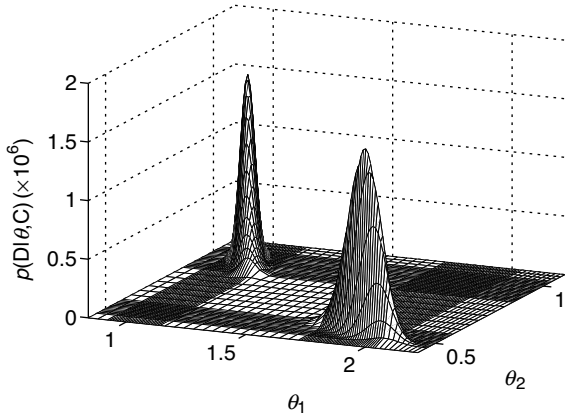

Figure 2.16 Likelihood function with both eigenvalues measured

$\hat{r}_1 = -1.4594$, $\hat{r}_2 = -1.5993$, $\hat{r}_3 = -1.6730$, $\hat{r}_4 = -1.4755$, and $\hat{r}_5 = -1.6855$. Again, the standard deviations of the measurement noise are assumed known and they are $\sigma_1 = 0.0191$ and $\sigma_r = 0.0809$. In this case, the likelihood function is given by:

$$p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) = (2\pi\sigma_1\sigma_r)^{-5} \exp\left[-\frac{1}{2\sigma_1^2}\sum_{n=1}^{5}\left(\lambda^{(1)}(\boldsymbol{\theta}) - \hat{\lambda}_n^{(1)}\right)^2 - \frac{1}{2\sigma_r^2}\sum_{n=1}^{5}(r(\boldsymbol{\theta}) - \hat{r}_n)^2\right] \quad (2.88)$$

and it is shown in Figure 2.18. It is a globally identifiable case since there is a unique global maximum of the likelihood function. Figure 2.19 shows its oval-like contours. Compared with the previous case with two measured eigenvalues, it shows that it is not the number of measured quantities but the effective mathematical constraints induced by the measurements to determine

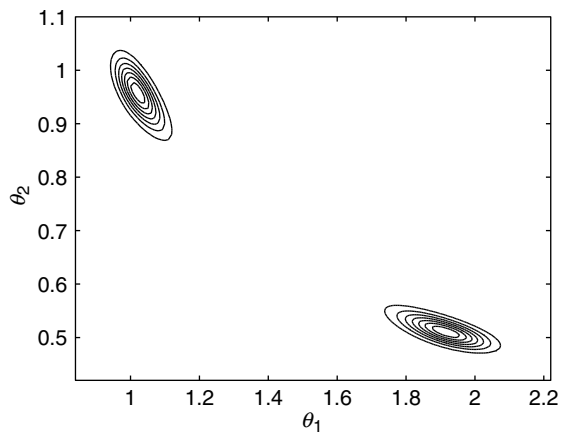

Figure 2.17 Contours of the likelihood function (locally identifiable case)

www.EngineeringBooksPdf.com

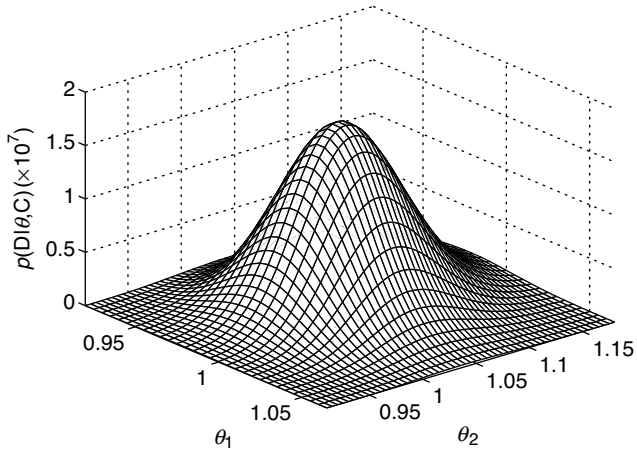

Figure 2.18 Likelihood function with one measured eigenvalue and eigenvector

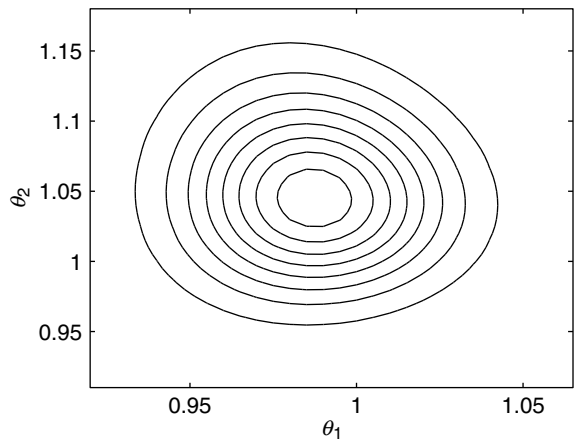

Figure 2.19 Contours of the likelihood function (globally identifiable case)

model-identifiability. In both Case 2 and 3, two quantities were measured but one turned out to be locally identifiable and the other globally identifiable.

2.3 Deterministic versus Probabilistic Methods

In most deterministic identification methods, important observable quantities are measured (e.g., modal frequencies of a building for stiffness identification) and the uncertain model parameters are obtained by minimizing a goodness-of-fit/error function of these measurements

www.EngineeringBooksPdf.com

or their induced quantities:

$$J_g(\boldsymbol{\theta}) = F(\boldsymbol{\theta}; \mathcal{D}, \mathcal{C}) \tag{2.89}$$

which is a measure of the data fitting error. One of the most popular approach is the least-squares type of methods that minimize a weighted sum of squared error:

$$J_g(\boldsymbol{\theta}) = \frac{1}{N} \sum_{n=1}^{N} w_n [x_n(\boldsymbol{\theta}) - y_n]^2 \tag{2.90}$$

where $x_n(\boldsymbol{\theta})$ and y_n, $n = 1, 2, \ldots, N$, are the model predicted quantities and their measurements. The weightings w_n, $n = 1, 2, \ldots, N$, depend on the specific method and uniform weightings are often chosen for simplicity. If the likelihood function happens to have the form of Gaussian for the uncertain parameters, the *optimal parameters* by least-squares and probabilistic methods are equivalent. In general, a deterministic approach *assumes* a particular form of objective function that is compatible with a particular probability distribution function even though it is not specified explicitly.

Probabilistic approaches, particularly the Bayesian approach, utilize the complete information of the data for the statistical inference if the appropriate likelihood function is constructed whereas deterministic methods usually bypass this issue. As a result, probabilistic methods allow for the quantification of the uncertainty of the parametric estimation. Furthermore, using the appropriate probability distribution ensures the correctness of the *optimal parameters*. Although Bayesian inference is attractive for allowing direct quantification of the uncertainty of the parameter estimation, there are main challenges in developing Bayesian methods:

1. In some applications, it is difficult to obtain the likelihood function with an appropriate choice of the type of probability distribution. This is not a trivial task since the probability distribution of the random variables in establishing the likelihood function may be complicated. For example, consider a random process x and its measurements at different time steps with equal spacing: $x_1, x_2, \ldots, x_N$. The auto-correlation function can be estimated by:

$$R_n \equiv \frac{1}{N-n} \sum_{n'=1}^{N-n} x_{n'} x_{n'+n} \tag{2.91}$$

   where $n = 0, 1, 2, \ldots, N-1$ denotes the number of lagging time steps. If R_0, R_1, …, R_{N-1} are used to estimate the model parameters that govern the stochastic process x, it is necessary to construct the likelihood function $p(R_0, R_1, \ldots, R_{N-1}|\boldsymbol{\theta}, \mathcal{C})$. Even though the individual probability distribution of a particular R_n can be obtained, deriving the joint PDF of $R_0, R_1, \ldots, R_{N-1}$ is another story due to the correlation structure of the R_ns. This explains why parametric identification using the correlation function is well-known to be difficult and usually biased. Furthermore, in other cases, even though the proper likelihood function can be written analytically, numerical computation may be very expensive or even prohibited. One example can be referred to in Chapter 4, Section 4.2.
2. In the probabilistic approach, the solution is not simply the *optimal* parameters but also the probability density function that describes the complete picture of the uncertainty. It is a challenging task to demonstrate the representation of the updated PDF since it has

www.EngineeringBooksPdf.com

been shown that its topology may be very complicated (e.g., multiple peaks or even unidentifiable).

Example. Consider a quadratic function:

$$Q = ax + bx^2 \tag{2.92}$$

where the actual values of the uncertain parameters are $\tilde{a} = -3$ and $\tilde{b} = 2$. The function values are observed at $x = 0.25, 0.5, 0.75, 1, \ldots, 2.75, 3$ with independent Gaussian measurement noise ϵ of zero mean and standard deviation σ_0/x:

$$y = Q + \epsilon \tag{2.93}$$

The actual value of σ_0 is taken to be 1.0. In order to have a fair ground for comparison with the least-squares method, an improper prior is taken so the updated PDF for the uncertain parameters is given by:

$$p(a, b, \sigma_0|\mathcal{D}, \mathcal{C}) = \kappa_0 \sigma_0^{-12} \exp\left[-\frac{1}{2\sigma_0^2}\sum_{n=1}^{12} x_n^2 (y_n - ax_n - bx_n^2)^2\right] \tag{2.94}$$

where κ_0 is a constant that does not depend on the uncertain parameters. The most probable values of the parameters can be obtained by maximizing the posterior PDF. In particular, the equations for a and b are given by:

$$\begin{cases} a\sum_{n=1}^{12} x_n^4 + b\sum_{n=1}^{12} x_n^5 = \sum_{n=1}^{12} x_n^3 y_n \\ a\sum_{n=1}^{12} x_n^5 + b\sum_{n=1}^{12} x_n^6 = \sum_{n=1}^{12} x_n^4 y_n \end{cases} \tag{2.95}$$

and these are uncoupled with the one for σ_0. Then, the updated values for a and b are readily obtained:

$$\begin{bmatrix} a^\star \\ b^\star \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^{12} x_n^4 & \sum_{n=1}^{12} x_n^5 \\ \sum_{n=1}^{12} x_n^5 & \sum_{n=1}^{12} x_n^6 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{n=1}^{12} x_n^3 y_n \\ \sum_{n=1}^{12} x_n^4 y_n \end{bmatrix} \tag{2.96}$$

Furthermore, the optimal value for the standard deviation of the measurement noise is obtained by solving $\partial p(a^\star, b^\star, \sigma_0|\mathcal{D}, \mathcal{C})/\partial\sigma_0 = 0$:

$$\sigma_0^\star = \sqrt{\frac{1}{12}\sum_{n=1}^{12} x_n^2 (y_n - a^\star x_n - b^\star x_n^2)^2} \tag{2.97}$$

Figure 2.20 shows a typical result of the problem. The dashed line is the true relationship and the crosses are the measurements. Even though the samples are quite scattered in the low range of x, the Bayesian approach reflects the correct weighting for different measurements. On the other hand, the traditional least-squares method simply minimizes the 2-norm of the difference

www.EngineeringBooksPdf.com

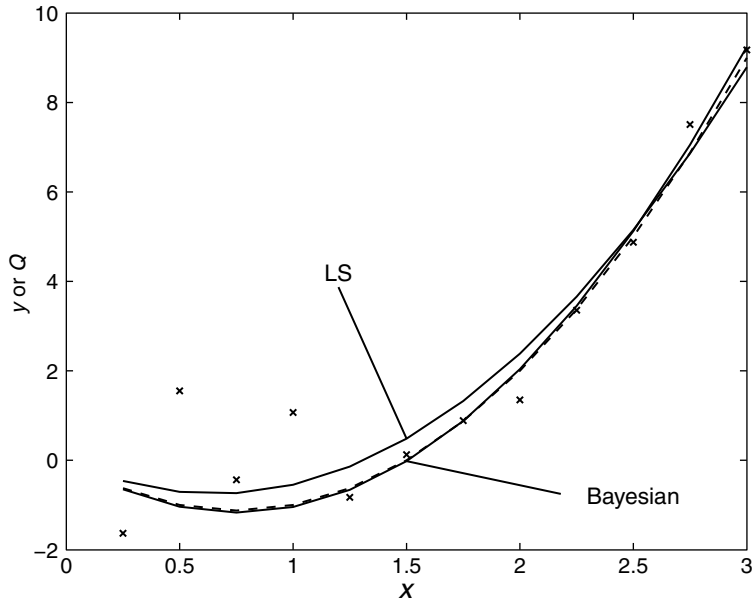

Figure 2.20 Comparison between the Bayesian and least-squares methods

between the measurements and the model-predicted values, and the identified parameters are given by:

$$\begin{bmatrix} a^\star \\ b^\star \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^{12} x_n^2 & \sum_{n=1}^{12} x_n^2 \\ \sum_{n=1}^{12} x_n^3 & \sum_{n=1}^{12} x_n^4 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{n=1}^{12} x_n y_n \\ \sum_{n=1}^{12} x_n^2 y_n \end{bmatrix} \tag{2.98}$$

A significantly different curve is obtained and it is referred to (LS) in the figure. The least-squares method gives unnecessarily too much weighting to the measurements in the region with small values of x and the identified curve is lifted up on the left hand side. Although the least-squares method can be modified to provide correct identification results, appropriate distribution and thus the correct weightings are necessary to achieve this goal. Generally speaking, the likelihood function is not necessary of the Gaussian type so the objective function used in the least-squares method is incorrect. One example is the spectrum and it will be further explained in Chapter 3.

2.4 Regression Problems

2.4.1 Linear Regression Problems

A linear regression relationship for a quantity of concern Q can be written as:

$$Q(\mathbf{x}; \mathbf{b}, \mathcal{C}) = \sum_{l=1}^{N_b} b_l x_l \tag{2.99}$$

www.EngineeringBooksPdf.com

where N_b is the total number of uncertain coefficients, b_ls. The variables $x_1, x_2, \ldots, x_{N_b}$ are the corresponding measured data in establishing the regression formula. The measurement of Q is modeled as:

$$y = Q(\mathbf{x}; \mathbf{b}, \mathcal{C}) + \epsilon \tag{2.100}$$

where ϵ is a Gaussian random variable with zero mean and variance σ_ϵ^2. It is used to represent the measurement noise and modeling error. The uncertain parameters in $\boldsymbol{\theta} = [\mathbf{b}^T, \sigma_\epsilon^2]^T$ include the coefficients b_ls and the prediction-error variance σ_ϵ^2 so the total number of uncertain parameters is $N_b + 1$.

The data $\mathcal{D}$ includes the measurement of $\mathbf{x}$ and the corresponding values for y. By assuming that the prediction errors in different records are statistically independent, the likelihood function is obtained:

$$p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) = (2\pi)^{-\frac{N}{2}} \sigma_\epsilon^{-N} \exp\left[-\frac{N}{2\sigma_\epsilon^2} J_g(\mathbf{b}; \mathcal{D}, \mathcal{C})\right] \tag{2.101}$$

where N is the total number of measured records. The goodness-of-fit function $J_g(\mathbf{b}; \mathcal{D}, \mathcal{C})$ represents the degree of data fitting and it is given by:

$$J_g(\mathbf{b}; \mathcal{D}, \mathcal{C}) = \frac{1}{N} \sum_{n=1}^{N} \left[y(n) - \sum_{l=1}^{N_b} b_l x_l(n) \right]^2 \tag{2.102}$$

A smaller value of this function implies better fitting to the data. Two special but popular choices of prior distributions of the uncertain parameters are considered in the following sections.

2.4.1.1 Independent Uniform Prior

In the case of a uniform prior PDF of the coefficients, the optimal coefficient vector $\mathbf{b}^\star$ can be obtained by minimizing $J_g(\mathbf{b}; \mathcal{D}, \mathcal{C})$. This can be achieved by solving the linear algebraic equation: $\partial J_g(\mathbf{b}; \mathcal{D}, \mathcal{C})/\partial \mathbf{b} = \mathbf{0}$, and the updated coefficient vector is readily obtained:

$$\mathbf{b}^\star = \mathbf{A}^{-1} \begin{bmatrix} \frac{1}{N} \sum_{n=1}^{N} x_1(n) y(n) \\ \frac{1}{N} \sum_{n=1}^{N} x_2(n) y(n) \\ \vdots \\ \frac{1}{N} \sum_{n=1}^{N} x_{N_b}(n) y(n) \end{bmatrix} \tag{2.103}$$

where $\mathbf{A}$ is an $N_b \times N_b$ symmetric matrix and its (l, l') component is:

$$A^{(l,l')} = \frac{1}{N} \sum_{n=1}^{N} x_l(n) x_{l'}(n) \tag{2.104}$$

By using the Cauchy–Schwarz inequality, it can be easily shown that this matrix is invertible if and only if the vectors $[x_l(1), x_l(2), \ldots, x_l(N)]$, $l = 1, 2, \ldots, N_b$, are linearly independent.

www.EngineeringBooksPdf.com

In a practical situation, it is rare that this matrix is perfectly singular but it is possibly ill-conditioned. In this case, it implies that some of the variables x_ls can be expressed in a linear combination of the others. In this case, the order of the regression relationship has to be reduced to give an identifiable case.

Furthermore, the updated fitting-error variance $\sigma_\epsilon^{2\star}$ can be obtained by maximizing the product of the prior PDF and the likelihood function $p(\boldsymbol{\theta}|\mathcal{C})p(\mathcal{D}|\boldsymbol{\theta},\mathcal{C})$, which is proportional to the posterior PDF. It turns out that it is equal to the minimal goodness-of-fit value:

$$\sigma_\epsilon^{2\star} = \min_{\mathbf{b}} J_g(\mathbf{b};\mathcal{D},\mathcal{C}) = J_g(\mathbf{b}^\star;\mathcal{D},\mathcal{C}) \tag{2.105}$$

The posterior PDF $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{C})$ is truncated Gaussian centered at the optimal parameter $\boldsymbol{\theta}^\star = [\mathbf{b}^{\star T}, \sigma_\epsilon^{2\star}]^T$. The Hessian matrix $\mathcal{H}(\boldsymbol{\theta}^\star)$ is the second derivatives of the objective function $J(\boldsymbol{\theta}) = -\ln p(\boldsymbol{\theta}|\mathcal{C})p(\mathcal{D}|\boldsymbol{\theta},\mathcal{C})$ and its elements are given by:

$$\mathcal{H}^{(l,l')}(\boldsymbol{\theta}^\star) = -\frac{\partial^2}{\partial\theta_l\theta_{l'}} \ln\left[p(\boldsymbol{\theta}|\mathcal{C})p(\mathcal{D}|\boldsymbol{\theta},\mathcal{C})\right]\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \tag{2.106}$$

In this case, the Hessian matrix has the following form:

$$\mathcal{H}(\boldsymbol{\theta}^\star) = N\begin{bmatrix} (\sigma_\epsilon^{2\star})^{-1}\mathbf{A} & \mathbf{0}_{N_b\times 1} \\ \mathbf{0}_{1\times N_b} & 0.5(\sigma_\epsilon^{2\star})^{-2} \end{bmatrix} \tag{2.107}$$

For large N, the joint posterior PDF can be well approximated by Gaussian distribution with the covariance matrix:

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \mathcal{H}(\boldsymbol{\theta}^\star)^{-1} = \frac{1}{N}\begin{bmatrix} \sigma_\epsilon^{2\star}\mathbf{A}^{-1} & \mathbf{0}_{N_b\times 1} \\ \mathbf{0}_{1\times N_b} & 2(\sigma_\epsilon^{2\star})^2 \end{bmatrix} \tag{2.108}$$

The diagonal elements of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ are the marginal variance of the corresponding element of $\boldsymbol{\theta}$ and the quantification of the uncertainty of the model parameters allows for the uncertainty analysis of the prediction.

2.4.1.2 Gaussian Prior for the Coefficients and Inverse Gamma Distribution for the Prediction-error Parameter

Another popular choice of the prior PDF is the Gaussian distribution. First, the prior PDF considered in this section is separable:

$$p(\boldsymbol{\theta}|\mathcal{C}) = p(\mathbf{b},\sigma_\epsilon^2|\mathcal{C}) = p(\mathbf{b}|\mathcal{C})p(\sigma_\epsilon^2|\mathcal{C}) \tag{2.109}$$

For the uncertain coefficients in $\mathbf{b}$, the Gaussian prior PDF is taken with mean $\mathbf{b}^\eta$ and covariance matrix $\mathbf{V}$:

$$p(\mathbf{b}|\mathcal{C}) = (2\pi)^{-\frac{N_b}{2}}|\mathbf{V}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\mathbf{b}-\mathbf{b}^\eta)^T\mathbf{V}^{-1}(\mathbf{b}-\mathbf{b}^\eta)\right] \tag{2.110}$$

www.EngineeringBooksPdf.com

However, the prior distribution for prediction-error variance is taken to be the conjugate prior and it is the inverse Gamma distribution in this case:

$$p(\sigma_\epsilon^2|\mathcal{C}) = \frac{\beta_0^{\alpha_0}}{\sigma_\epsilon^{2(\alpha_0+1)}\Gamma(\alpha_0)} \exp\left(-\frac{\beta_0}{\sigma_\epsilon^2}\right) \tag{2.111}$$

Then, the product of the prior PDF and likelihood function is given by:

$$p(\boldsymbol{\theta}|\mathcal{C})p(\mathcal{D}|\boldsymbol{\theta},\mathcal{C}) = (2\pi)^{-\frac{N_b+N}{2}}|\mathbf{V}|^{-\frac{1}{2}}\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\sigma_\epsilon^{-2(\alpha_0+1)-N}\exp\left(-\frac{\beta_0}{\sigma_\epsilon^2}\right)$$
$$\times \exp\left[-\frac{1}{2}(\mathbf{b}-\mathbf{b}^\eta)^T\mathbf{V}^{-1}(\mathbf{b}-\mathbf{b}^\eta) - \frac{1}{2\sigma_\epsilon^2}\sum_{n=1}^{N}\left(y(n)-\sum_{l=1}^{N_b}b_l x_l(n)\right)^2\right] \tag{2.112}$$

For a given value of σ_ϵ^2, the conditional optimal coefficient vector $\mathbf{b}^\star$ can be obtained by maximizing the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{C})$, or equivalently $p(\boldsymbol{\theta}|\mathcal{C})p(\mathcal{D}|\boldsymbol{\theta},\mathcal{C})$. This can be done by solving the linear algebraic equation: $\partial p(\boldsymbol{\theta}|\mathcal{C})p(\mathcal{D}|\boldsymbol{\theta},\mathcal{C})/\partial\mathbf{b} = \mathbf{0}$ and the solution is:

$$\mathbf{b}^\star(\sigma_\epsilon^2) = (N\sigma_\epsilon^{-2}\mathbf{A}+\mathbf{V}^{-1})^{-1}\left(N\sigma_\epsilon^{-2}\begin{bmatrix}\frac{1}{N}\sum_{n=1}^{N}x_1(n)y(n)\\ \frac{1}{N}\sum_{n=1}^{N}x_2(n)y(n)\\ \vdots \\ \frac{1}{N}\sum_{n=1}^{N}x_{N_b}(n)y(n)\end{bmatrix} + \mathbf{V}^{-1}\mathbf{b}^\eta\right) \tag{2.113}$$

where the matrix $\mathbf{A}$ has (l, l') element:

$$A^{(l,l')} = \frac{1}{N}\sum_{n=1}^{N}x_l(n)x_{l'}(n) \tag{2.114}$$

It can be easily shown that the matrix $N\sigma_\epsilon^{-2}\mathbf{A}+\mathbf{V}^{-1}$ is invertible.

In the special case when the matrix $\mathbf{V}$ has large diagonal elements (i.e., the prior distribution contains very little information of the coefficients), the solution in Equation (2.113) is reduced to the solution in Equation (2.103). In general, the solution in Equation (2.113) can be treated as a weighted average of the solution in Equation (2.103) and the most probable value of the prior distribution and the weightings are $N\sigma_\epsilon^{-2}$ and $\mathbf{V}^{-1}$.

For a given coefficient vector $\mathbf{b}$, the conditional optimal prediction-error variance $\sigma_\epsilon^{2\star}$ can be found by solving $\partial p(\boldsymbol{\theta}|\mathcal{C})p(\mathcal{D}|\boldsymbol{\theta},\mathcal{C})/\partial\sigma_\epsilon^2 = 0$ and it is readily obtained:

$$\sigma_\epsilon^{2\star}(\mathbf{b}) = \frac{NJ_g(\mathbf{b};\mathcal{D},\mathcal{C})+2\beta_0}{N+2(\alpha_0+1)} \tag{2.115}$$

For large N, it is approximately equal to $J_g(\mathbf{b};\mathcal{D},\mathcal{C})$.

By using the expressions of the conditional optimal in Equations (2.113) and (2.115), the optimal parameters can be obtained by the following iterative algorithm. First, take $\sigma_\epsilon^2 = 0$ in Equation (2.113) and compute the conditional optimal values for the uncertain coefficients. Then, with these conditional optimal coefficients, the conditional prediction-error variance

www.EngineeringBooksPdf.com

can be computed by Equation (2.115). By using this value of the prediction-error variance, the conditional optimal for the uncertain coefficients can be updated. By repeating the last two steps, the updated uncertain parameters can be obtained until convergence is achieved.

The conditional PDF $p(\mathbf{b}|\mathcal{D}, \mathcal{C}, \sigma_\epsilon^2)$ is Gaussian with mean $\mathbf{b}^\star(\sigma_\epsilon^2)$ and covariance matrix $(N\sigma_\epsilon^{-2}\mathbf{A} + \mathbf{V}^{-1})^{-1}$. Furthermore, the conditional PDF $p(\sigma_\epsilon^2|\mathcal{D}, \mathcal{C}, \mathbf{b})$ follows the inverse Gamma distribution $IG(\alpha, \beta)$ with shape parameter $\alpha = N/2 + \alpha_0$ and scale parameter $\beta = NJ_g(\mathbf{b}; \mathcal{D}, \mathcal{C})/2 + \beta_0$. Moreover, for the case of large N, the matrix $(N\sigma_\epsilon^{-2}\mathbf{A} + \mathbf{V}^{-1})^{-1}$ can be approximated by $\sigma_\epsilon^2/N\mathbf{A}^{-1}$ and the marginal PDF of the prediction-error variance σ_ϵ^2 is approximately an inverse Gamma distribution $IG(\alpha^\star, \beta^\star)$, where the shape and scale parameter are given by:

$$\begin{aligned} \alpha^\star &= \frac{N - N_b}{2} + \alpha_0 \\ \beta^\star &= \frac{N}{2} J_g(\mathbf{b}^\star; \mathcal{D}, \mathcal{C}) + \beta_0 \end{aligned} \tag{2.116}$$

According to Equations (2.46) and (2.47), the mean and variance are:

$$E[\sigma_\epsilon^2|\mathcal{D}, \mathcal{C}] = \frac{NJ_g(\mathbf{b}^\star; \mathcal{D}, \mathcal{C}) + 2\beta_0}{N - N_b + 2\alpha_0 - 2} \tag{2.117}$$

and

$$E[(\sigma_\epsilon^2 - E[\sigma_\epsilon^2|\mathcal{D}, \mathcal{C}])^2|\mathcal{D}, \mathcal{C}] = \frac{2[NJ_g(\mathbf{b}^\star; \mathcal{D}, \mathcal{C}) + 2\beta_0]^2}{(N - N_b + 2\alpha_0 - 2)^2(N - N_b + 2\alpha_0 - 4)} \tag{2.118}$$

For large N, the variance is approximately equal to $2J_g(\mathbf{b}^\star; \mathcal{D}, \mathcal{C})^2/N = 2(\sigma_\epsilon^{2\star})^2/N$. By observing Equation (2.108), it is the same as the previous case with uniform prior distribution.

2.4.2 Nonlinear Regression Problems

In this section, nonlinear predictive model classes are considered and there is in general no closed-form solution for the updated model parameters in contrast to the previous linear case. An efficient algorithm is introduced in this section to search for the updated parameters. With a similar fashion to Equation (2.99), a nonlinear regression formula takes the following general form:

$$Q(\mathbf{x}; \mathbf{b}, \mathbf{n}, \mathcal{C}) = f(\mathbf{x}; \mathbf{n}) + \sum_{l=1}^{N_b} b_l f_l(\mathbf{x}; \mathbf{n}) \tag{2.119}$$

where f and $f_l, l = 1, 2, \ldots, N_b$, are known nonlinear functions with unknown parameters in $\mathbf{n} \in \mathbb{R}^{N_n}$. Again, the measurement is assumed to be different from its model predicted value due to measurement noise and modeling error:

$$y = Q(\mathbf{x}; \mathbf{b}, \mathbf{n}, \mathcal{C}) + \epsilon \tag{2.120}$$

The random variable is again modeled as a zero mean Gaussian random variable with variance σ_ϵ^2. Given some measurements of $\mathbf{x}$ and y, the likelihood function takes the same form as in

www.EngineeringBooksPdf.com

Equation (2.101) but the goodness-of-fit function has this form:

$$J_g(\mathbf{b}, \mathbf{n}; \mathcal{D}, \mathcal{C}) = \frac{1}{N} \sum_{n=1}^{N} \left[y(n) - f(\mathbf{x}(n); \mathbf{n}) - \sum_{l=1}^{N_b} b_l f_l(\mathbf{x}(n); \mathbf{n}) \right]^2 \tag{2.121}$$

Given a particular parameter vector of $\mathbf{n}$, the closed-form solution for the conditional optimal values for $\mathbf{b}$ and σ_ϵ^2 can be obtained. For example, with the prior PDF used in Section 2.4.1.1, the conditional optimal value for $\mathbf{b}$ can be computed in a similar fashion as Equation (2.103):

$$\mathbf{b}^\star(\mathbf{n}) = \mathbf{A}^{-1} \begin{bmatrix} \frac{1}{N} \sum_{n=1}^{N} f_1(\mathbf{x}(n); \mathbf{n})[y(n) - f(\mathbf{x}(n); \mathbf{n})] \\ \frac{1}{N} \sum_{n=1}^{N} f_2(\mathbf{x}(n); \mathbf{n})[y(n) - f(\mathbf{x}(n); \mathbf{n})] \\ \vdots \\ \frac{1}{N} \sum_{n=1}^{N} f_{N_b}(\mathbf{x}(n); \mathbf{n})[y(n) - f(\mathbf{x}(n); \mathbf{n})] \end{bmatrix} \tag{2.122}$$

and the conditional optimal value for σ_ϵ^2 is given by:

$$\sigma_\epsilon^{2\star}(\mathbf{n}) = \min_{\mathbf{b}} J_g(\mathbf{b}, \mathbf{n}; \mathcal{D}, \mathcal{C}) = J_g(\mathbf{b}^\star, \mathbf{n}; \mathcal{D}, \mathcal{C}) \tag{2.123}$$

Once the conditional optimal values for $\mathbf{b}$ and σ_ϵ^2 are obtained, the value of the goodness-of-fit function can be computed by Equation (2.121) whereas the normalizing constant in the posterior PDF does not affect the parametric identification results. By maximizing the goodness-of-fit function with respect to $\mathbf{n}$, the updated model parameters can be obtained. Therefore, the closed-form solution of the conditional optimal parameters reduces the dimension of the original optimization problem from $N_b + N_n + 1$ to N_n only.

It can be easily shown that the conditional PDF $p(\mathbf{b}|\mathcal{D}, \mathcal{C}, \sigma_\epsilon^2, \mathbf{n})$ is Gaussian with mean $\mathbf{b}^\star(\mathbf{n})$ and covariance matrix $\sigma_\epsilon^2/N\mathbf{A}^{-1}$, which depend on $\mathbf{n}$. Furthermore, the conditional PDF $p(\sigma_\epsilon^2|\mathcal{D}, \mathcal{C}, \mathbf{n})$ follows the inverse Gamma distribution with shape parameter $\alpha^\star$ and scale parameter $\beta^\star$ given by:

$$\begin{aligned} \alpha^\star &= \frac{N - N_b}{2} + \alpha_0 \\ \beta^\star &= \frac{N}{2} J_g(\mathbf{b}^\star, \mathbf{n}^\star; \mathcal{D}, \mathcal{C}) + \beta_0 \end{aligned} \tag{2.124}$$

2.5 Numerical Representation of the Updated PDF

2.5.1 General Form of Reliability Integrals

In probabilistic analysis, the following integral is commonly encountered:

$$\mathcal{R} \equiv \int_{\boldsymbol{\Theta}} Q(\boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \tag{2.125}$$

where Q is the concerned quantity that depends on the model parameter vector $\boldsymbol{\theta}$ and $p(\boldsymbol{\theta})$ is a probability distribution, that may be the prior or posterior PDF of the model parameters. In

www.EngineeringBooksPdf.com

the latter case where system measurements are available, this integral can be updated:

$$\mathcal{R}_{\mathcal{D}} \equiv E[Q|\mathcal{D}, \mathcal{C}] = \int_{\boldsymbol{\Theta}} Q(\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}) \mathrm{d}\boldsymbol{\theta} \tag{2.126}$$

For example, in reliability analysis, the quantity Q is considered as the probability that the structure with parameter vector $\boldsymbol{\theta}$ would fail, i.e., $Q(\boldsymbol{\theta}) = P(F|\boldsymbol{\theta}, \mathcal{C})$. Then, the updated integral becomes the updated robust probability of failure for the structure, when it is subjected to some stochastic excitation [198]:

$$\mathcal{R}_{\mathcal{D}} = P(F|\mathcal{D}, \mathcal{C}) = \int_{\boldsymbol{\Theta}} P(F|\boldsymbol{\theta}, \mathcal{C}) p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}) \mathrm{d}\boldsymbol{\theta} \tag{2.127}$$

This updated robust probability of failure incorporates knowledge about $\boldsymbol{\theta}$ from $\mathcal{C}$ and from the updated information from the data. It is robust because the modeling uncertainties are taken into explicit account [198]. The updated structural reliability of the structure is $P(S|\mathcal{D}, \mathcal{C}) = 1 - P(F|\mathcal{D}, \mathcal{C})$, where S denotes a safe status of the structure.

Another example is the probability of damage. Damage may be quantified as the fractional reduction in the stiffness from that in the undamaged state of a substructure. Assume that the measurements of the structure are available for the undamaged and possibly damaged state, denoted by $\mathcal{D}^{ud}$ and $\mathcal{D}^{pd}$, respectively. Then, the probability that damage in the lth substructure exceeds the damage level d is given by:

$$\begin{aligned} P_l^{dam}(d|\mathcal{D}^{ud}, \mathcal{D}^{pd}, \mathcal{C}) &\equiv P\{\theta_l^{pd} < (1-d)\theta_l^{ud}|\mathcal{D}^{ud}, \mathcal{D}^{pd}, \mathcal{C}\} \\ &= \int_{-\infty}^{\infty} P(\theta_l^{pd} < (1-d)\theta_l^{ud}|\theta_l^{ud}, \mathcal{D}^{pd}, \mathcal{C}) p(\theta_l^{ud}|\mathcal{D}^{ud}, \mathcal{C}) \mathrm{d}\theta_l^{ud} \end{aligned} \tag{2.128}$$

This probability of damage in the substructures can be used to indicate the likely location and severity of damage in the structure. The integral in Equation (2.128) can also be classified in the updated integral in Equation (2.126) if $\mathcal{D} = \{\mathcal{D}^{ud}, \mathcal{D}^{pd}\}$, $\boldsymbol{\theta} = \theta_l^{ud}$ and $Q(\boldsymbol{\theta}) = P(\theta_l^{pd} < (1-d)\theta_l^{ud}|\theta_l^{ud}, \mathcal{D}, \mathcal{C})$. Then, $\mathcal{R}_{\mathcal{D}} = P_l^{dam}(d|\mathcal{D}^{ud}, \mathcal{D}^{pd}, \mathcal{C})$, the probability of damage exceeding severity d in the lth substructure.

2.5.2 Monte Carlo Simulation

Monte Carlo simulation (MCS) is a well-known method for evaluating the integral in Equation (2.125) [225]. It is robust since it does not depend on the type of problems or the number of random variables, i.e., the dimension of the integral. The key of MCS is to simulate random samples of $\boldsymbol{\theta}$ according to the distribution $p(\boldsymbol{\theta})$: $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N$. Then, the integral in Equation (2.125) can be evaluated directly by statistical averaging:

$$\mathcal{R} \approx \frac{1}{N} \sum_{n=1}^{N} Q(\boldsymbol{\theta}_n) \tag{2.129}$$

In evaluating the type of updated robust integral in Equation (2.126) by MCS, it requires the parameter samples distributed according to the updated PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$. Therefore, generation

www.EngineeringBooksPdf.com

of samples according to the updated PDF does not only serve as the representation of the posterior distribution but also for the computation of reliability integrals and other marginal distributions or moments. However, this may not be a trivial task, especially for locally identifiable and unidentifiable cases. In the next section, an adaptive Markov Chain Monte Carlo simulation algorithm is introduced to fulfill this goal.

2.5.3 Adaptive Markov Chain Monte Carlo Simulation

2.5.3.1 Metropolis–Hastings Algorithm

From a simulation perspective, Equation (2.126) suggests that $\mathcal{R}_{\mathcal{D}}$ may be estimated as the average of Q over the samples which are simulated according to $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$. Although a method for generating independent samples according to $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$ is generally not available, the Markov Chain Monte Carlo (MCMC) simulation, in particular, the Metropolis–Hastings (MH) algorithm, offers a feasible way to simulate samples according to an arbitrary distribution, at the expense of introducing dependence among the samples. However, direct application of the MH algorithm to simulate Markov chain samples according to $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$ is not feasible due to the small region $\mathcal{N}$ of probability concentration. Nevertheless, as the adaptive method is built on the MH algorithm, its implementation is discussed in this section.

The Metropolis–Hastings algorithm is a simple procedure to simulate samples according to an arbitrary PDF where the target PDF has to be known up to a scaling constant only. It was originally developed by Metropolis and his co-workers [178] and later generalized by Hastings [100]. Its potential use for structural reliability problems and Bayesian model updating has been recently demonstrated [8, 9, 17]. In the MH method, samples are simulated as the states of a special Markov chain whose limiting stationary distribution is equal to the target PDF. In other words, the PDF of the Markov chain sample $\boldsymbol{\theta}_n$ simulated at the nth Markov step, tends to the target PDF as $n \to \infty$. The Markov chain samples, which are dependent in general, can be used for statistical averaging as if they were independent, although with some reduction of efficiency in the estimator.

Let $p^*(\boldsymbol{\xi}|\boldsymbol{\theta})$ be a chosen PDF, called the *proposal PDF*, which is a PDF for $\boldsymbol{\xi}$ that depends on $\boldsymbol{\theta}$. The role of p^* will become clear shortly. For convenience in notation, define the following function:

$$q(\boldsymbol{\theta}) \equiv p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})p(\boldsymbol{\theta}|\mathcal{C}) = p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})/\kappa_0 \tag{2.130}$$

The value of q can be computed readily for a given $\boldsymbol{\theta}$, while the same is not true for $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$ because evaluating the normalizing constant κ_0:

$$\kappa_0 = 1/p(\mathcal{D}|\mathcal{C}) = 1/\int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})p(\boldsymbol{\theta}|\mathcal{C})\mathrm{d}\boldsymbol{\theta} \tag{2.131}$$

requires a high-dimensional integration. In the following, the MH algorithm is described to simulate N Markov chain samples $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N\}$ with a limiting stationary distribution equal to the target PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$. To start the Markov chain, let $\boldsymbol{\theta}_1$ be a point chosen either deterministically or simulated according to some PDF which approximates $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$. In general, for $n = 1, 2, \ldots, N-1$, the next sample $\boldsymbol{\theta}_{n+1}$ is simulated from the current sample $\boldsymbol{\theta}_n$ by first simulating a *candidate state* $\boldsymbol{\xi}$ from the proposal PDF $p^*(\boldsymbol{\xi}|\boldsymbol{\theta}_n)$ and then computing the

www.EngineeringBooksPdf.com

probability density quotient:

$$\mathcal{Q}(\boldsymbol{\xi}, \boldsymbol{\theta}_n) = \frac{q(\boldsymbol{\xi})p^*(\boldsymbol{\theta}_n|\boldsymbol{\xi})}{q(\boldsymbol{\theta}_n)p^*(\boldsymbol{\xi}|\boldsymbol{\theta}_n)} \tag{2.132}$$

The candidate state $\boldsymbol{\xi}$ is then accepted with probability $\min\{1, \mathcal{Q}\}$ and rejected with the remaining probability $1 - \min\{1, \mathcal{Q}\} = \max\{0, 1 - \mathcal{Q}\}$.

If the candidate state is accepted, it will be taken as the next state of the Markov chain, i.e., $\boldsymbol{\theta}_{n+1} = \boldsymbol{\xi}$. Otherwise, the current state is remained as the next state, i.e., $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n$. In other words, the next state is given by:

$$\boldsymbol{\theta}_{n+1} = \begin{cases} \boldsymbol{\xi} & \text{with probability } \min\{1, \mathcal{Q}\} \\ \boldsymbol{\theta}_n & \text{with probability } \max\{0, 1 - \mathcal{Q}\} \end{cases} \tag{2.133}$$

This process is repeated until N Markov chain samples have been generated.

It can be shown that the next sample $\boldsymbol{\theta}_{n+1}$ will be distributed according to $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$ if the current sample $\boldsymbol{\theta}_n$ is also distributed in the same way. In other words, $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$ is the stationary PDF of the Markov chain. If the Markov chain is started with the initial state $\boldsymbol{\theta}_1$ simulated from a PDF different from $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$, then the Markov chain will be in a transient state and its samples will not be distributed exactly as $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$. Under the assumption of ergodicity, however, the Markov chain will converge to the stationary state, and so the PDF of $\boldsymbol{\theta}_n$ will tend to $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$ as $n \to \infty$. With a finite sample size used in an actual implementation, ergodicity often becomes an issue of whether the Markov chain samples can populate sufficiently well the region of significant probability of the target PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$. See Au and Beck [8, 9] for a more detailed discussion of ergodicity in applying the MH method to reliability problems.

Using the Markov chain samples $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N\}$, $\mathcal{R}_\mathcal{D}$ in Equation (2.126) is estimated as the average $\hat{\mathcal{R}}_\mathcal{D}$ of Q over the samples, which is the same as the usual MCS estimator, except that the samples are simulated from a Markov chain instead of being independent and identically distributed. Nevertheless, the estimator $\hat{\mathcal{R}}_\mathcal{D}$ has similar statistical properties as those of the MCS estimators. In order to reduce the initial transient effect of the Markov chain on the estimate, the first few samples (say 10) may be discarded when computing $\hat{\mathcal{R}}_\mathcal{D}$. Thus, the Markov chain samples $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N\}$ used for computing $\hat{\mathcal{R}}_\mathcal{D}$ are those simulated after the initial transient stage.

The proposal PDF p^* determines the distribution of the candidate state $\boldsymbol{\xi}$ given the current state, and consequently the convergence rate of the estimator $\hat{\mathcal{R}}_\mathcal{D}$ to $\mathcal{R}_\mathcal{D}$. If the candidate state is rejected too often because small values of the quotient $\mathcal{Q}$ are encountered during simulation, the Markov chain will consist of many repeated samples. As a result, the correlation among samples will be large and it slows down the convergence of $\hat{\mathcal{R}}_\mathcal{D}$. Unfortunately, due to the small scale φ of the neighborhood $\mathcal{N}$ of the manifold, it is very difficult to choose an appropriate p^* in the absence of prior knowledge about the manifold so that the acceptance rate of the candidate state is not too small while at the same time the Markov chain samples effectively explore $\mathcal{N}$. Consequently, the region visited by the Markov chain samples will be small compared to $\mathcal{N}$, leading to substantial bias in the estimate for $\mathcal{R}_\mathcal{D}$. As a result, direct application of the MH algorithm to simulate samples according to the target PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$ is not feasible.

www.EngineeringBooksPdf.com

2.5.3.2 Adaptive MH Algorithm

The problems encountered in applying the MH algorithm directly arise from the fact that the updated PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$ is concentrated in a small neighborhood $\mathcal{N}$ of the manifold $\mathcal{S}$ in the parameter space. The process of adapting samples to $\mathcal{N}$ in the MH algorithm is inhibited by the small scale φ of the thickness of $\mathcal{N}$ compared to the size of the proposal PDF required to cover $\mathcal{N}$ with an affordable number of samples. This suggests that direct adaptation using a proposal PDF which varies with a vastly different length scale from that of the target PDF will not be effective. In view of this, a sequence of intermediate PDFs is introduced in Beck and Au [17] to bridge the gap in length scale between the prior PDF $p(\boldsymbol{\theta}|\mathcal{C})$ and the target updated PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$. By utilizing successively the information from the previous adapted intermediate PDFs in the MH algorithm, the region $\mathcal{N}$ can be populated by the Markov chain samples in a gradual manner.

Let $\{p^{(1)}, p^{(2)}, \ldots, p^{(s_0)}\}$ be a chosen sequence of PDFs converging to $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$ $(= p^{(s_0)})$ so that their region of significant probability content gradually diminishes to that of $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$. For example, $p^{(s)}$ may be chosen as the updated PDF from Bayes' Theorem based on an increasing amount of data:

$$p^{(s)} = p(\boldsymbol{\theta}|\mathcal{D}^{(s)}, \mathcal{C}) \tag{2.134}$$

where $\mathcal{D}^{(1)} \subset \mathcal{D}^{(2)} \subset \cdots \subset \mathcal{D}^{(s_0)} = \mathcal{D}$. Thus, starting with the prior PDF $p(\boldsymbol{\theta}|\mathcal{C})$ as the proposal PDF, the MH algorithm is carried out to simulate samples $\{\boldsymbol{\theta}_1^{(1)}, \boldsymbol{\theta}_2^{(1)}, \ldots, \boldsymbol{\theta}_N^{(1)}\}$ with the target PDF $p^{(1)}$. A kernel sampling density $\mathcal{P}^{(1)}$ is constructed as a weighted sum of Gaussian PDFs centered among these samples to approximate $p^{(1)}$ [8, 240]:

$$\mathcal{P}^{(1)}(\boldsymbol{\theta}) = \sum_{n=1}^{N} w_n \, \mathcal{G}(\boldsymbol{\theta}; \boldsymbol{\theta}_n^{(1)}, \boldsymbol{\Sigma}_n) \tag{2.135}$$

where $\mathcal{G}(\boldsymbol{\theta}; \boldsymbol{\theta}_n^{(1)}, \boldsymbol{\Sigma}_n)$ is the multi-dimensional Gaussian PDF evaluated at $\boldsymbol{\theta}$ with mean $\boldsymbol{\theta}_n^{(1)}$ and covariance matrix $\boldsymbol{\Sigma}_n$, and the w_n are the probability weights associated with the Gaussian PDFs, which satisfy $w_n \geq 0$ and $\sum_{n=1}^{N} w_n = 1$. The choice of w_n and $\boldsymbol{\Sigma}_n$ is completely defined by the samples $\{\boldsymbol{\theta}_1^{(1)}, \boldsymbol{\theta}_2^{(1)}, \ldots, \boldsymbol{\theta}_N^{(1)}\}$, and so is $\mathcal{P}^{(1)}$ [8]. Since $\mathcal{P}^{(1)}$ is a weighted sum of Gaussian PDFs, the kernel marginal PDF for a particular component θ_l of $\boldsymbol{\theta}$ can be obtained by analytically integrating Equation (2.135). Similarly, the kernel marginal cumulative distribution function (CDF) of θ_l can be readily computed in terms of a weighted sum of Gaussian CDFs. More importantly, independent samples distributed as $\mathcal{P}^{(1)}$ can be simulated readily, which are approximately distributed as $p^{(1)}$ and so lie in the region of significant probability of $p^{(1)}$.

To continue, $\mathcal{P}^{(1)}$ is used as the proposal PDF for simulating the Markov chain samples $\{\boldsymbol{\theta}_1^{(2)}, \boldsymbol{\theta}_2^{(2)}, \ldots, \boldsymbol{\theta}_N^{(2)}\}$ with the target PDF $p^{(2)}$. These samples are then used to construct the kernel sampling density $\mathcal{P}^{(2)}$ as in Equation (2.135), which gives an approximation to $p^{(2)}$. In general, the kernel sampling density $\mathcal{P}^{(s)}$ (which approximates $p^{(s)}$) is constructed using the Markov chain samples at the sth simulation level which is then used as the proposal PDF for simulating Markov chain samples for the next level with the target PDF $p^{(s+1)}$. This is continued until the $s\text{th}_0$ simulation level, where Markov chain samples for the target updated PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}) = p^{(s_0)}$ are simulated.

www.EngineeringBooksPdf.com

Let $\mathcal{R}^{(s)}$ be the expectation of the concerned quantity $Q(\boldsymbol{\theta})$ when $\boldsymbol{\theta}$ is distributed according to $p^{(s)}$. It is estimated using the Markov chain samples $\{\boldsymbol{\theta}_1^{(s)}, \boldsymbol{\theta}_2^{(s)}, \ldots, \boldsymbol{\theta}_N^{(s)}\}$ by:

$$\mathcal{R}^{(s)} \approx \hat{\mathcal{R}}^{(s)} = \frac{1}{N}\sum_{n=1}^{N} Q(\boldsymbol{\theta}_n^{(s)}) \tag{2.136}$$

Since $p^{(s)}$ converges to $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C})$ as the simulation level s increases to s_0, $\mathcal{R}^{(s)}$ converges to $\mathcal{R}_{\mathcal{D}}$ defined by Equation (2.126):

$$\mathcal{R}_{\mathcal{D}} \approx \hat{\mathcal{R}}^{(s_0)} \tag{2.137}$$

The success of the adaptive strategy relies on effective application of the MH algorithm at each simulation level s, which requires that $p^{(s-1)}$ (which is approximated by $\mathcal{P}^{(s-1)}$) varies with a similar length scale to $p^{(s)}$ for $s = 1, 2, \ldots, s_0$. The choice of the sequence $\{p^{(s)}\}$ is thus important to the success of the adaptive method. If the updated PDF with data $\mathcal{D}$ is of the form [18, 268]:

$$p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}) = \kappa \exp\left[-\frac{J_g'(\boldsymbol{\theta})}{2\varphi^2}\right] \tag{2.138}$$

where J_g' is the goodness-of-fit function normalized by the prediction error covariance, and φ is a measure of the size of the prediction error, then a good choice for the sequence $\{p^{(s)}\}$ is:

$$p^{(s)} = \kappa^{(s)} \exp\left[-\frac{J_g'(\boldsymbol{\theta})}{2(\varphi^{(s)})^2}\right] \tag{2.139}$$

where $(\varphi^{(s)})^2 = 2^{s_0-s}\varphi^2$, $s = 1, 2, \ldots, s_0$, with $2^{s_0} \approx \varphi^{-2}$ if the length scale of the prior PDF is of order one. Thus, the exponent of the posterior PDF is doubled when advancing to the next simulation level. Note also that the computational effort does not depend on the dimension of the parameter space.

2.5.3.3 Statistical Properties of the Estimator

The statistical properties of the estimator $\hat{\mathcal{R}}^{(s)}$ in Equation (2.136) are presented next, assuming the Markov chain generated according to the MH algorithm at each simulation level is ergodic. In spite of the fact that $\hat{\mathcal{R}}^{(s)}$ is computed using dependent samples from a Markov chain, it still has the usual properties of MCS estimators using independent and identically distributed samples [70]. For example, $\hat{\mathcal{R}}^{(s)}$ converges to $\mathcal{R}^{(s)}$ with probability 1 as $N \to \infty$ (Strong Law of Large Numbers), and under similar conditions as those for Monte Carlo estimators, $\hat{\mathcal{R}}^{(s)}$ is Gaussian distributed as $N \to \infty$ (Central Limit Theorem). If the Markov chain is started with the initial state $\boldsymbol{\theta}_1^{(s)}$ distributed as the target PDF $p^{(s)}$, then the Markov chain is stationary, and $\hat{\mathcal{R}}^{(s)}$ is unbiased. Otherwise, $\hat{\mathcal{R}}^{(s)}$ is only asymptotically unbiased, although the bias decays with an increasing number of Markov steps.

www.EngineeringBooksPdf.com

For a fixed proposal PDF for the sth simulation level, assuming that the Markov chain has settled into its stationary state, it can be shown that the coefficient of variation of $\hat{\mathcal{R}}^{(s)}$ is:

$$\delta^{(s)} = \Delta^{(s)}\sqrt{\frac{1+\gamma^{(s)}}{N}} \tag{2.140}$$

where $\Delta^{(s)}$ is the coefficient of variation of $Q(\boldsymbol{\theta})$ when $\boldsymbol{\theta}$ is distributed as $p^{(s)}$, and $\gamma^{(s)}$ is a correlation factor:

$$\gamma^{(s)} = 2\sum_{n=1}^{N-1}\left(1-\frac{n}{N}\right)\rho^{(s)}(n) \tag{2.141}$$

and $\rho^{(s)}(n)$ is the correlation coefficient between the concerned quantity Q evaluated by the Markov chain samples separated by n Markov steps:

$$\rho^{(s)}(n) = \frac{E_{p^{(s)}}\left[\left(Q(\boldsymbol{\theta}_1^{(s)}) - \mathcal{R}^{(s)}\right)\left(Q(\boldsymbol{\theta}_{n+1}^{(s)}) - \mathcal{R}^{(s)}\right)\right]}{(\mathcal{R}^{(s)})^2} \tag{2.142}$$

By estimating the correlation sequence $\{\rho^{(s)}(n)\}$ from the Markov chain samples, $\gamma^{(s)}$ in Equation (2.141), and hence $\delta^{(s)}$ in Equation (2.140), can be estimated in a single simulation run.

2.5.4 Illustrative Example

The adaptive MCMC simulation method is applied to updating the robust reliability for a two-story structural frame, depicted in Figure 2.21. The bay width and story height are 5.0 m and 2.5 m, respectively. The Young's modulus and mass density are taken to be 200 GPa and 7800 kg/m^3, respectively. The beams have a cross-sectional area of 0.01 m^2 and a moment of inertia of 6.0×10^{-4} m^4 but they are 0.02 m^2 and 1.5×10^{-4} m^4 for the columns. As a result, the structure has modal frequencies of 5.20 and 15.4 Hz. The structure is assumed to have 1.0% of critical damping for all modes. A simple model with two degrees of freedom is used in the system identification in order to estimate the inter-story stiffnesses and to assess the reliability of the structure. Specifically, the stiffness matrix is given by:

$$\mathbf{K} = K^{\eta}\begin{bmatrix}\theta_1+\theta_2 & -\theta_2\\ -\theta_2 & \theta_2\end{bmatrix} \tag{2.143}$$

where θ_l is the inter-story stiffness parameter of the lth story. The uncertain stiffness parameters $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ to be identified are multipliers of the nominal interstory stiffness which is equal to $K^{\eta} = 46.08$ MN/m. Furthermore, a lumped mass model (i.e., diagonal mass matrix) is used with a floor mass of 11.17×10^3 kg. There is no combination of (θ_1, θ_2) such that the structural mass and stiffness matrix are fully consistent with the actual frame. The nominal and other possible models deviate from the actual frame system and this is imposed deliberately to simulate the reality.

www.EngineeringBooksPdf.com

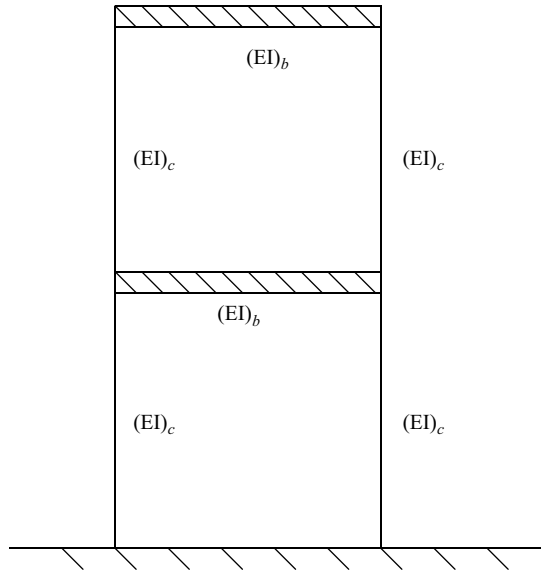

Figure 2.21 Two-story structural frame

Two cases are considered for the available dynamic data that correspond to locally identifiable and unidentifiable cases. For the first case, the modal data $\mathcal{D}$ consists of the identified modal frequencies for both modes of the frame. Rather than performing modal identification on simulated time histories, noisy versions of the modal frequencies are generated and the measurements are $\hat{\Omega}^{(1)} = 5.5$ Hz and $\hat{\Omega}^{(2)} = 14.9$ Hz.

The prior PDF $p(\boldsymbol{\theta}|\mathcal{C})$ is taken to be independent log-normal PDFs with means of 0.9 and 1.2 and unit variance. Using the modal data $\mathcal{D}$, the updated PDF for the stiffness parameter vector $\boldsymbol{\theta}$ is formulated as:

$$p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}) = \kappa_0 \exp\left[-\frac{J_g'(\boldsymbol{\theta})}{2\varphi^2}\right] p(\boldsymbol{\theta}|\mathcal{C}) \tag{2.144}$$

where κ_0 is a normalizing constant, φ is related to the prediction-error coefficient of variation and the modal goodness-of-fit function depends on the type of measurements. In the adaptive procedure, the sequence of intermediate PDFs $\{p^{(s)}\}$ is constructed by successively substituting the sequence of values $\varphi^2 = 1/2^s$ into Equation (2.144) for $s = 1, 2, \ldots, 10$.

2.5.4.1 Locally Identifiable Case

If it is assumed that only one sensor (at either the first or second floor) was used during the modal testing, only the modal frequencies can be identified. In this case, the two stiffness parameters are locally identifiable and the normalized modal goodness-of-fit function is given

www.EngineeringBooksPdf.com

by:

$$J'_g(\boldsymbol{\theta}) = \sum_{m=1}^{2} \left[\frac{\hat{\Omega}^{(m)2} - \Omega^{(m)}(\boldsymbol{\theta})^2}{\hat{\Omega}^{(m)2}} \right]^2 \tag{2.145}$$

where $\hat{\Omega}^{(m)}$ and $\Omega^{(m)}(\boldsymbol{\theta})$ are the measured modal frequency and the model modal frequency of the mth mode, which depend on the structural parameters in $\boldsymbol{\theta}$. It turns out that there exist two local maxima of the updated PDF for the stiffness parameters: [0.73,0.84] and [1.6,0.38]. The structural model with $\theta_1 = 0.73$ and $\theta_2 = 0.84$ is clearly more flexible than the nominal model with $\theta_1 = \theta_2 = 1$. This is expected as the simple structural model for identification with the nominal parameters is obtained by assuming rigid floors.

Figure 2.22 shows the Markov chain samples for the structure at simulation levels $s = 1, 4, 7, 8, 9,$ and 10, corresponding to $\varphi^2 = 1/2, 1/16, 1/128, 1/256, 1/512,$ and$1/1024$ in Equation (2.144). For each simulation level, the first 10 Markov chain samples are ignored in order to avoid the samples in the transient stage. $N = 500$ Markov chain samples are simulated and they are shown with dots in the figure. The Markov chain samples are not all distinct since there is a certain probability to repeat a Markov chain sample in each step. To show the population of samples consistently, the area of the dots are shown proportional to the number of samples at each location. All the samples are located around the two local optimal points of the updated PDF. The marginal kernel PDFs for θ_1 and θ_2, constructed from the Markov chain samples, are shown in Figure 2.23 and the estimated marginal PDFs can be readily obtained from the joint kernel PDF without numerical integration. The dashed lines show typical results for three different runs and the solid lines show the solution obtained by numerical integration for reference. The peak at around (0.73, 0.84) is significantly higher than the other so there are more samples distributed around (0.73, 0.84) in Figure 2.22.

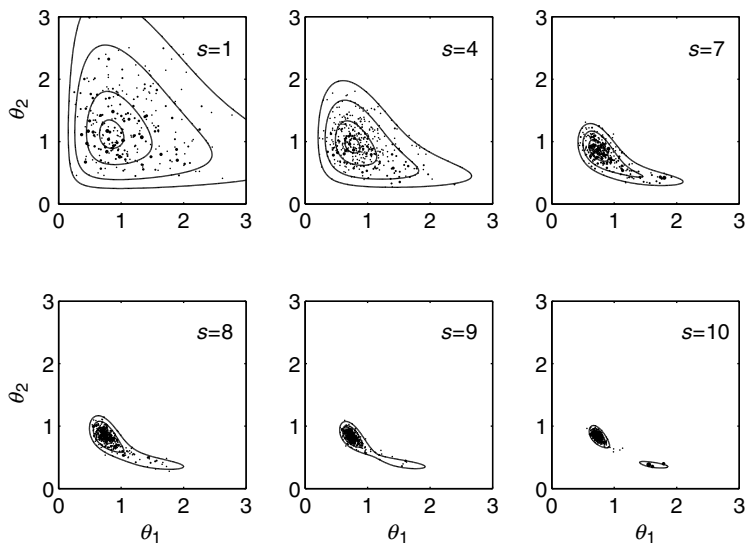

Figure 2.22 Markov chain samples for the structural parameters (locally identifiable case)

www.EngineeringBooksPdf.com

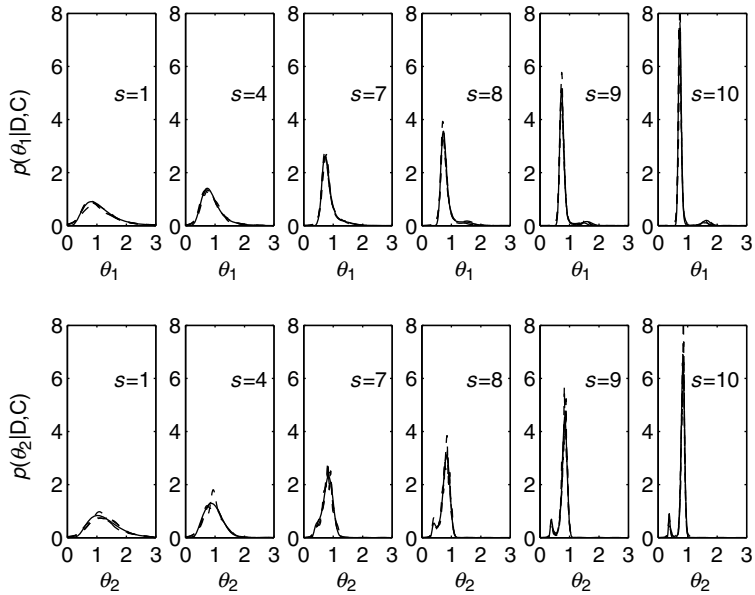

Figure 2.23 Updated marginal PDFs at different simulation levels (locally identifiable case)

The reliability of the structure subjected to stochastic ground motion is updated using the shear building model. For the purpose of this illustrative example, the ground excitation is modeled by stationary Gaussian white noise with spectral intensity $S_{\ddot{x}_g 0} = 6.0 \times 10^{-3}\ \mathrm{m^2/s^3}$. The updated robust failure probability $P(F|\mathcal{D}, \mathcal{C})$ is studied where failure corresponds to the stationary displacement of the top floor exceeding the threshold level $b = 3.0$ cm (0.6% deformation) within a duration of $T = 20$ s. Assuming that the out-crossing events follow a Poisson process, the probability of failure with given parameters can be approximated by [218]:

$$P(F|\boldsymbol{\theta}, \mathcal{C}) \approx 1 - \exp\left[-2T\nu(\boldsymbol{\theta})\right] \tag{2.146}$$

where $\nu(\boldsymbol{\theta})$ is the *mean upcrossing rate* for the displacement of the top floor:

$$\nu(\boldsymbol{\theta}) = \frac{\sigma_{\dot{x}_2}(\boldsymbol{\theta})}{2\pi\, \sigma_{x_2}(\boldsymbol{\theta})} \exp\left[-\frac{b^2}{2\, \sigma_{x_2}(\boldsymbol{\theta})^2}\right] \tag{2.147}$$

In this equation, σ_{x_2} and $\sigma_{\dot{x}_2}$ are the standard deviations of the stationary displacement and velocity response of the top floor, respectively, obtained by the Lyapunov equation [249]:

$$\mathbf{A}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{A}^T + 2\pi S_{\ddot{x}_g 0}\mathbf{1} = \mathbf{0} \tag{2.148}$$

The matrix $\mathbf{A}$ is given by:

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix} \tag{2.149}$$

www.EngineeringBooksPdf.com

where **M**, **C** and **K** are the mass, damping and stiffness matrix, respectively. The matrix **1** is the matrix with all entries being unity. The matrix $\mathbf{\Sigma}$ represents the covariance matrix of the state vector of the displacements and velocities: $[\mathbf{x}^T, \dot{\mathbf{x}}^T]^T$. Equation (2.148) can be solved by the function 'lyap' in MATLAB® [171].

The probability of outcrossing depends not only on the variance of the response but also the frequency content of the signal. It is not surprising that a signal has a higher probability of outcrossing than another with the same variance but a lower frequency content over the same duration of time since there are more cycles in the time duration in the former case. In particular, the ratio $\sigma_{\dot{x}_2}(\boldsymbol{\theta})/\sigma_{x_2}(\boldsymbol{\theta})$ in the formula reflects the averaged frequency of the signal. For example, for a single-degree-of-freedom (SDOF) system subjected to stationary Gaussian white noise, it is equal to the natural frequency of the system. The estimates $\hat{P}(F|\mathcal{D}^{(s)}, \mathcal{C})$, $s = 1, 2, \ldots, 10$, for the updated failure probabilities for each prediction-error level, are shown in Figure 2.24. The circles show typical results from three independent runs. The dashed line and solid line show the solution by numerical integration and the average by 50 simulation runs, respectively. The failure probabilities can be updated without information of the mode shapes.

2.5.4.2 Unidentifiable Case

The second case assumes that only the first modal frequency is observed because, for example, the second mode is not excited significantly during the modal testing. In this case,

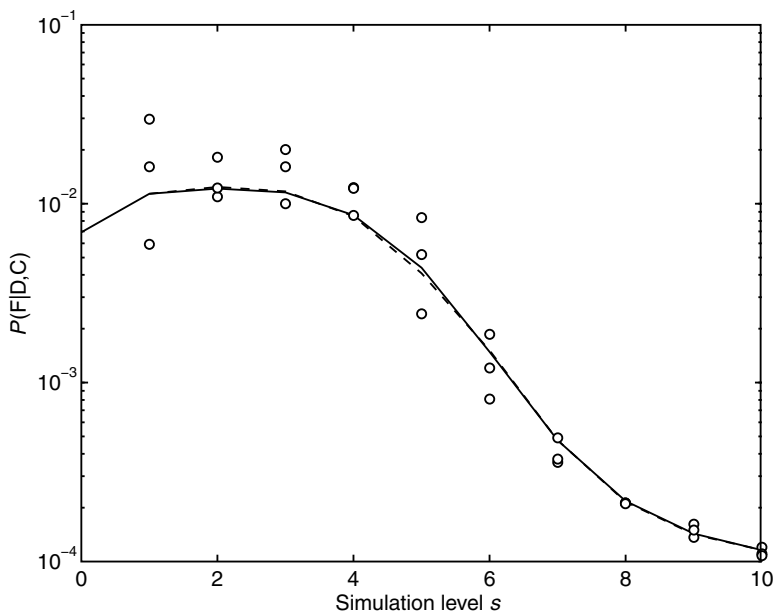

Figure 2.24 Updated failure probabilities (locally identifiable case)

www.EngineeringBooksPdf.com

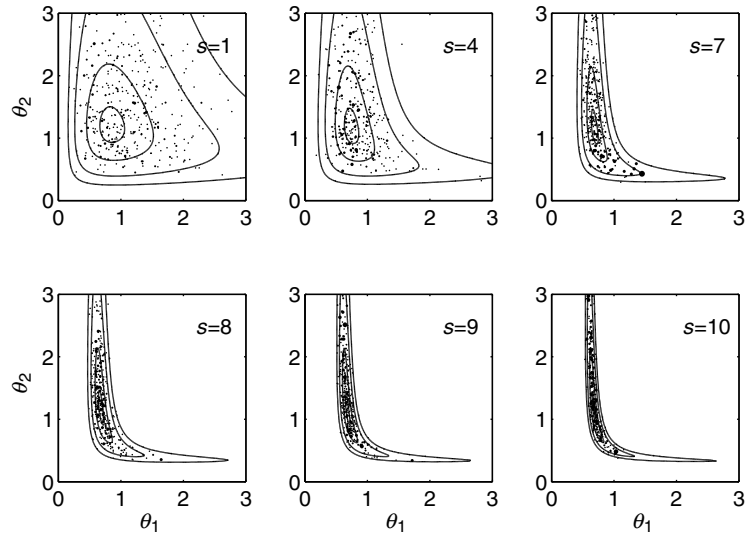

Figure 2.25 Markov chain samples for the structural parameters (unidentifiable case)

the normalized goodness-of-fit function in Equation (2.144) is given by:

$$J_g'(\boldsymbol{\theta}) = \left[\frac{\hat{\Omega}^{(1)^2} - \Omega^{(1)}(\boldsymbol{\theta})^2}{\hat{\Omega}^{(1)^2}} \right]^2 \tag{2.150}$$

which is similar to Equation (2.145). This unidentifiable case produces an extended region in the parameter space that gives high posterior probability values. Figure 2.25 shows the Markov chain samples for the stiffness parameters of the frame. The region of high probability content contains as a subset its counterpart in Figure 2.22 for the locally identifiable case, as expected, since this case is the same as the locally identifiable case except that the data constraint from the second modal frequency is removed. Therefore, the local maxima in the locally identifiable case should be located in the high probability region of this unidentifiable case. The marginal kernel PDFs for θ_1 and θ_2, constructed from the Markov chain samples, are shown in Figure 2.26 and the estimated marginal PDFs can be readily obtained from the joint kernel PDF without numerical integration. Even though it is an unidentifiable case, the region of significant probability for θ_1 concentrates in a very narrow interval and this can be verified by Figure 2.25. By comparing the marginal PDFs of θ_2 at simulation levels $s = 1$ and $s = 10$, it can be concluded that the data do not have much saying on θ_2 in this case.

Figure 2.27 shows the updated failure probability of the frame. Assessment of the impact of damage on the reliability of the structure can be performed even though there are infinitely many most probable parameter estimates in this case and it is also seen from the similarity of Figures 2.24 and 2.27. The loss of information about the second modal frequency does not have much effect on the updated failure probability because the first mode dominates the response of this frame. Even the individual values of θ_1 and θ_2 cannot be identified, different combinations

www.EngineeringBooksPdf.com

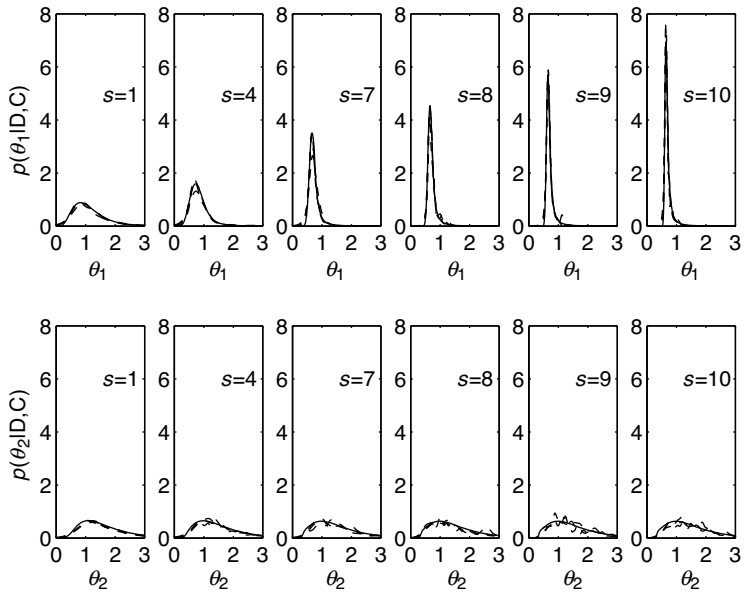

Figure 2.26 Updated marginal PDFs at different simulation levels (unidentifiable case)

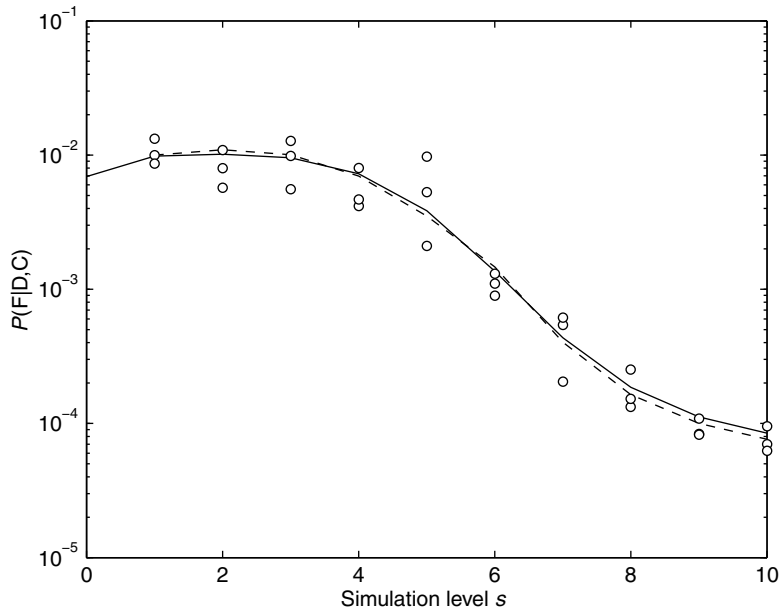

Figure 2.27 Updated failure probabilities (unidentifiable case)

www.EngineeringBooksPdf.com

of these parameters in the high probability density region give similar first modal frequency. Furthermore, the failure criterion considers only a global response assessment (displacement of the top floor). It is worth noting that the failure probability $P(F|\boldsymbol{\theta}, \mathcal{C})$ in this example was obtained by the Poisson approximation. However, the first excursion may be solved more accurately by other simulation methods without using the Poisson approximation, such as shown by Au and Beck [9], Pradlwarter and Schuëller [209] and Pradlwarter *et al.* [210].

2.6 Application to Temperature Effects on Structural Behavior

2.6.1 Problem Description

Structural health monitoring (SHM) is a process for evaluation of the structural safety status from structural response measurements. The ultimate goal is to develop an economical and non-destructive system for the earliest possible detection of damage. It has been attracting much attention in the past two decades. A tremendous amount of efforts have been dedicated in this area [15, 18, 19, 25, 51, 52, 68, 69, 77, 101, 103, 106, 138, 142, 143, 149, 160, 168, 174, 185], [186, 195, 201, 204, 228, 238, 244, 246, 261, 267, 268, 281, 282]. Comprehensive literature reviews on the development of SHM can be found in Doebling *et al.* [69] and Sohn *et al.* [247] and there were also a number of workshops, [1, 44, 187], for example, and special issues of journals, e.g., *Journal of Engineering Mechanics* in July 2000 [90] and January 2004 [26] and *Computer-Aided Civil and Infrastructure Engineering* in January 2001 [278] and May 2006 [22].

In particular, ambient vibration survey (AVS) on field measurements is a popular choice to acquire necessary information for estimating the structural health status. It has attracted much interest because it offers a means of obtaining dynamic data in an economical and efficient manner, without requiring the setup of special dynamic experiments (e.g., actuators) which are usually costly, time consuming, and often obtrusive. A challenging problem arises on field monitoring since the modal parameters, such as the modal frequencies, are affected not only by the health status of a structure but also by various ambient conditions, such as temperature and relative humidity [247]. Since the identified values of the modal parameters may be masked by these substantial effects, it is essential to quantify these ambient effects for the establishment of a reliable SHM system. In this section, the ambient temperature effect on the fundamental frequency of a building is investigated. Six-month daily monitored acceleration time histories of a 22-story reinforced concrete residential building are utilized. The relationship between the fundamental frequency and temperature is derived. Based on this relationship, variation of the fundamental frequency caused by temperature fluctuation can be filtered out and a more reliable structural health monitoring system can be developed.

2.6.2 Thermal Effects on Modal Frequencies of Buildings

Buildings may be modeled as a flexible vertical cantilever beam subjected to gravitational and lateral loadings. A building with relatively uniform properties over the height can be simplified as a continuous beam with its structural properties distributed uniformly along the length [243].

The Timoshenko beam model accounts for both shear deformation as well as the rotational inertia effect. To mimic the behavior of a building, the beam model is taken to be a uniform

www.EngineeringBooksPdf.com

cantilever beam. The boundary condition problem becomes an eigenvalue problem and the squared fundamental frequency is given by [265]:

$$\Omega^2 = \frac{GK_s}{\rho L^2}\lambda \tag{2.151}$$

where G is the shear modulus of the materials, ρ is the mass density, and L is the length of the beam. K_s is the dimensionless shear coefficient of a Timoshenko beam. It depends on the geometry of the cross section and it lies between 1/2 and 1 (e.g., it is 5/6 for a rectangular cross section and 9/10 for a circular cross section). The dimensionless variable λ satisfies the following characteristic equation [265]:

$$\left(\frac{\Omega^2+\lambda_2^2}{\Omega^2-\lambda_1^2}+\frac{\Omega^2-\lambda_1^2}{\Omega^2+\lambda_2^2}\right)\cos\lambda_1\cosh\lambda_2+\left(\frac{\lambda_1}{\lambda_2}-\frac{\lambda_2}{\lambda_1}\right)\sin\lambda_1\sinh\lambda_2=2 \tag{2.152}$$

where λ_1 and λ_2 are given by:

$$\lambda_{1,2}^2=\frac{\Omega^2}{2}(1+\gamma)\left[\frac{1}{(1+\gamma)^2}\sqrt{\frac{4\gamma A}{IL^2}+(1-\gamma)^2}\pm 1\right] \tag{2.153}$$

The variables A and I are the area and second moment of inertia of the cross section of the beam. The coefficient γ is given by:

$$\gamma=\frac{GK_s}{E} \tag{2.154}$$

where E is the modulus of elasticity.

Although there is no closed-form solution for the characteristic equation in Equation (2.152), numerical methods, such as finite element methods, can be used to solve this problem [265, 272]. It was shown that the squared fundamental frequency governed by Equation (2.152) can be approximated as a combination of a bending term and a shear term:

$$\Omega^2=\frac{EI}{\rho AL^4}\left(C_b+C_s\frac{K_sAGL^2}{EI}\right) \tag{2.155}$$

where C_b and C_s are dimensionless coefficients that depend only on the deformed shape of a beam regardless of its material properties. For a slender building, the bending effect dominates the structural behavior, and the shear effect can be neglected. In this case, the beam behaves closely to an Euler–Bernoulli beam and its squared fundamental frequency is given by [59]:

$$\Omega^2=12.36\frac{EI}{\rho AL^4} \tag{2.156}$$

On the other hand, for a short and sturdy building, it behaves more closely to a shear beam, whose squared fundamental frequency can be approximated by:

$$\Omega^2=\frac{\pi^2K_sG}{4\rho L^2} \tag{2.157}$$

Temperature variation induces both geometric and material effects to a structure. As a result, the modal frequencies of a structure are influenced by the temperature and it is more important

www.EngineeringBooksPdf.com

than other ambient factors, such as relative humidity [227]. Thermal expansion and Young's modulus vary linearly in the range of room temperature [188]. Use α_L and α_E to denote the thermal coefficients for expansion and Young's modulus, respectively, so:

$$L = L_0(1 + \alpha_L \Delta\mathcal{T}) \tag{2.158}$$

$$E = E_0(1 + \alpha_E \Delta\mathcal{T}) \tag{2.159}$$

where L_0 and E_0 are the length and Young's modulus at the reference temperature $\mathcal{T}_0$, and $\Delta\mathcal{T} = \mathcal{T} - \mathcal{T}_0$ is the temperature difference from the reference temperature. Similarly, the area, the second moment of inertia and the shear modulus are given by:

$$A = A_0(1 + \alpha_L \Delta\mathcal{T})^2 \tag{2.160}$$

$$I = I_0(1 + \alpha_L \Delta\mathcal{T})^4 \tag{2.161}$$

where A_0 and I_0 are the corresponding beam properties at temperature $\mathcal{T}_0$. Note that under ambient temperature, the variation of the Poission's ratio, ν, is negligible [188] and so the shear modulus, $G = \frac{E}{2(1+\nu)}$, follows the same relationship as the Young's modulus:

$$G = G_0(1 + \alpha_E \Delta\mathcal{T}) \tag{2.162}$$

where G_0 is the shear modulus at temperature $\mathcal{T}_0$.

Since the total mass of the beam/building does not depend on the temperature, the mass density is inversely proportional to the volume of the beam model: $\rho \propto 1/AL$. By considering the thermal effects to all these quantities with Equation (2.155), the thermal dependence of the squared fundamental frequency can be expressed as follows:

$$\Omega^2 = \Omega_0^2(1 + \alpha_E \Delta\mathcal{T})(1 + \alpha_L \Delta\mathcal{T}) \tag{2.163}$$

where Ω_0 is the fundamental frequency at temperature $\mathcal{T}_0$, and b_0, b_1 and b_2 are the uncertain coefficients for identification. Typical values of α_E are $-3.2 \times 10^{-4}/^\circ$C and $-7.2 \times 10^{-3}/^\circ$C for steel and concrete, respectively. On the other hand, typical values of α_L are $1.2 \times 10^{-5}/^\circ$C and $1.3 \times 10^{-5}/^\circ$C for steel and concrete, respectively. As a result, Equation (2.163) shows a decreasing squared fundamental frequency–temperature trend, which is opposite to previous observations [227]. This is because the structural joints become stiffer with an increasing temperature. Therefore, the equivalent Young's modulus–temperature relationship of the Timoshenko beam is modified: $E_{eq} = E_0[1 + (\alpha_E + \alpha_J)\Delta\mathcal{T}]$, where α_J is the coefficient to account for the joint stiffening effect and it is an uncertain coefficient that depends on the details of the structure. It is not surprising that this coefficient is mode-dependent. Finally, Equation (2.163) becomes:

$$\Omega^2 = \Omega_0^2[1 + (\alpha_E + \alpha_J)\Delta\mathcal{T}](1 + \alpha_L \Delta\mathcal{T}) = b_0 + b_1\mathcal{T} + b_2\mathcal{T}^2 \tag{2.164}$$

It remains a quadratic function but the coefficients are uncertain. Equation (2.164) is applicable to all modal frequencies for Timoshenko beam models. Therefore, it is proposed to bridge the squared fundamental frequency and the ambient temperature by a quadratic function, and the coefficients can be estimated by Bayesian analysis.

www.EngineeringBooksPdf.com

2.6.3 Bayesian Regression Analysis

In this section, Bayesian analysis is performed to identify the uncertain coefficients of the quadratic function. The *effective temperature* $\mathcal{T}_n$ is assumed to be different from the measured values since there is measurement noise in the data acquisition process and the temperature in different parts of the building could also be non-uniform. The difference is assumed Gaussian with zero mean and variance $\sigma_{\mathcal{T}}^2$. In this study, this standard deviation is taken to be $\sigma_{\mathcal{T}} = 0.5°\text{C}$ since the difference between the indoor and outdoor temperature measurement was around 1°C and the average value was used as the measured temperature $\hat{\mathcal{T}}_n$ for the nth day. On the other hand, the squared fundamental frequency is identified by the Bayesian spectral density approach to be presented in Chapter 3. Therefore, the uncertain parameters include the coefficients of the quadratic function and the effective temperatures: $\boldsymbol{\theta} = [b_0, b_1, b_2, \mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_N]^T$, where N is the number of data points. The data include the measurements of the temperature and the identified squared fundamental frequencies: $\mathcal{D} = [\hat{\mathcal{T}}_1, \hat{\mathcal{T}}_2, \ldots, \hat{\mathcal{T}}_N, \hat{\Omega}_1^2, \hat{\Omega}_2^2, \ldots, \hat{\Omega}_N^2]$. By assuming that the measurement noises are statistically independent, the likelihood function is given by:

$$p(\mathcal{D}|b_0, b_1, b_2, \mathcal{T}_1, \ \mathcal{T}_2, \ldots, \mathcal{T}_N, \mathcal{C}) = (2\pi\sigma_{\mathcal{T}})^{-N} \prod_{n=1}^{N} \sigma_{\Omega_n^2}^{-1} \times \exp\left[-\frac{1}{2\sigma_{\Omega_n^2}^2}\sum_{n=1}^{N}(b_0 + b_1\mathcal{T}_n + b_2\mathcal{T}_n^2 - \hat{\Omega}_n^2)^2 - \frac{1}{2\sigma_{\mathcal{T}}^2}\sum_{n=1}^{N}(\mathcal{T}_n - \hat{\mathcal{T}}_n)^2\right] \tag{2.165}$$

where $\hat{\Omega}_n^2$ and $\sigma_{\Omega_n^2}$ are the identified squared fundamental frequency and its associated standard derivation of the nth measurement obtained from the Bayesian spectral density approach. The variable $\hat{\mathcal{T}}_n$ is the corresponding measurement of the temperature. Then, by the Bayes' theorem, the updated PDF of the uncertain parameters is given by:

$$p(b_0, b_1, b_2, \mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_N|\mathcal{D}, \mathcal{C}) = \kappa_0 p(b_0, b_1, b_2, \mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_N|\mathcal{C}) \times p(\mathcal{D}|b_0, b_1, b_2, \mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_N, \mathcal{C}) \tag{2.166}$$

In this study, a non-informative prior PDF for the uncertain parameters is used and the prior PDF is absorbed into the normalizing constant. Then, the updated PDF is proportional to the likelihood function:

$$p(b_0, b_1, b_2, \mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_N|\mathcal{D}, \mathcal{C}) \propto p(\mathcal{D}|b_0, b_1, b_2, \mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_N, \mathcal{C}) \tag{2.167}$$

To obtain the optimal parameters, the objective function is defined as the negative logarithm of the likelihood function without taking the terms that do not depend on the uncertain parameters:

$$J(b_0, b_1, b_2, \mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_N) = \sum_{n=1}^{N}\left[\frac{1}{2\sigma_{\Omega_n^2}^2}(b_0 + b_1\mathcal{T}_n + b_2\mathcal{T}_n^2 - \hat{\Omega}_n^2)^2 + \frac{1}{2\sigma_{\mathcal{T}}^2}(\mathcal{T}_n - \hat{\mathcal{T}}_n)^2\right] \tag{2.168}$$

The optimal values can be obtained by minimizing the objective function but this optimization problem is nonlinear. However, the solution can be obtained by solving a series of linear

www.EngineeringBooksPdf.com

optimization problems. First, the conditional optimal values of the uncertain coefficients can be found by solving $\partial J(\boldsymbol{\theta})/\partial \mathbf{b} = \mathbf{0}$ and it gives the following linear algebraic equation:

$$\begin{bmatrix} \sum_{n=1}^{N} \sigma_{\Omega n}^{-2} & \sum_{n=1}^{N} \mathcal{T}_n \sigma_{\Omega n}^{-2} & \sum_{n=1}^{N} \mathcal{T}_n^2 \sigma_{\Omega n}^{-2} \\ \sum_{n=1}^{N} \mathcal{T}_n \sigma_{\Omega n}^{-2} & \sum_{n=1}^{N} \sigma_{\Omega n}^{-2} & \sum_{n=1}^{N} \mathcal{T}_n \sigma_{\Omega n}^{-2} \\ \sum_{n=1}^{N} \mathcal{T}_n^2 \sigma_{\Omega n}^{-2} & \sum_{n=1}^{N} \mathcal{T}_n \sigma_{\Omega n}^{-2} & \sum_{n=1}^{N} \sigma_{\Omega n}^{-2} \end{bmatrix} \begin{bmatrix} b_0^\star \\ b_1^\star \\ b_2^\star \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^{N} \hat{\Omega}_n^2 \sigma_{\Omega n}^{-2} \\ \sum_{n=1}^{N} \hat{\Omega}_n^2 \mathcal{T}_n \sigma_{\Omega n}^{-2} \\ \sum_{n=1}^{N} \hat{\Omega}_n^2 \mathcal{T}_n^2 \sigma_{\Omega n}^{-2} \end{bmatrix} \tag{2.169}$$

On the other hand, the conditional optimal value for $\mathcal{T}_n$ can be found by solving $\partial J(\boldsymbol{\theta})/\partial \mathcal{T}_n = 0$, $n = 1, 2, \ldots, N$, and it gives the following cubic equation:

$$\frac{2b_2^2}{\sigma_{\Omega_n^2}^2} \mathcal{T}_n^{3\star} + \frac{3b_1 b_2}{\sigma_{\Omega_n^2}^2} \mathcal{T}_n^{2\star} + \left[\frac{2b_2(b_0 - \hat{\Omega}_n^2) + b_1}{\sigma_{\Omega_n^2}^2} + \frac{1}{\sigma_{\mathcal{T}}^2} \right] \mathcal{T}_n^\star + \frac{b_1(b_0 - \hat{\Omega}_n^2) + b_1}{\sigma_{\Omega_n^2}^2} - \frac{\hat{\mathcal{T}}_n}{\sigma_{\mathcal{T}}^2} = 0 \tag{2.170}$$

Given the values of b_0, b_1 and b_2, this cubic equation can be solved analytically. If there are three real roots, the one closest to $\hat{\mathcal{T}}_n$ should be taken.

The optimization problem can be solved as follows. First, take $\mathcal{T}_n = \hat{\mathcal{T}}_n$, for $n = 1, 2, \ldots, N$, and then the conditional optimal values of the coefficients b_0, b_1 and b_2 can be obtained by solving Equation (2.169). By plugging these values into Equation (2.170), the conditional optimal values of the effective temperatures can be updated. This procedure is repeated until convergence is achieved. Furthermore, the Hessian matrix of the objective function J in Equation (2.168) at the optimal values is given by:

$$\mathcal{H}(\boldsymbol{\theta}^\star) = \begin{bmatrix} \mathcal{H}_b & \mathcal{H}_{b\mathcal{T}} \\ \mathcal{H}_{b\mathcal{T}}^T & \mathcal{H}_{\mathcal{T}} \end{bmatrix} \tag{2.171}$$

where the submatrix $\mathcal{H}_b$ is given by

$$\mathcal{H}_b = \begin{bmatrix} \sum_{n=1}^{N} \sigma_{\Omega_n^2}^{-2} & \sum_{n=1}^{N} \mathcal{T}_n \sigma_{\Omega_n^2}^{-2} & \sum_{n=1}^{N} \mathcal{T}_n^2 \sigma_{\Omega_n^2}^{-2} \\ \sum_{n=1}^{N} \mathcal{T}_n \sigma_{\Omega_n^2}^{-2} & \sum_{n=1}^{N} \mathcal{T}_n^2 \sigma_{\Omega_n^2}^{-2} & \sum_{n=1}^{N} \mathcal{T}_n^3 \sigma_{\Omega_n^2}^{-2} \\ \sum_{n=1}^{N} \mathcal{T}_n^2 \sigma_{\Omega_n^2}^{-2} & \sum_{n=1}^{N} \mathcal{T}_n^3 \sigma_{\Omega_n^2}^{-2} & \sum_{n=1}^{N} \mathcal{T}_n^4 \sigma_{\Omega_n^2}^{-2} \end{bmatrix} \tag{2.172}$$

The (l, n) components of the submatrix $\mathcal{H}_{b\mathcal{T}}$, $l = 1, 2, 3$; $n = 1, 2, \ldots, N$, are given by:

$$\mathcal{H}_{b\mathcal{T}}^{(1,n)} = \sigma_{\Omega_n^2}^{-2} \tag{2.173}$$

$$\mathcal{H}_{b\mathcal{T}}^{(2,n)} = \sigma_{\Omega_n^2}^{-2} \left(b_0^\star + 2b_1^\star \mathcal{T}_n^\star + 3b_2 \mathcal{T}_n^{\star 2} - \hat{\Omega}_n^2 \right) \tag{2.174}$$

$$\mathcal{H}_{b\mathcal{T}}^{(3,n)} = \sigma_{\Omega_n^2}^{-2} \left(2b_0^\star \mathcal{T}_n^\star + 3b_1^\star \mathcal{T}_n^{\star 2} + 4b_2 \mathcal{T}_n^{\star 3} - 2\hat{\Omega}_n^2 \mathcal{T}_n^\star \right) \tag{2.175}$$

and $\mathcal{H}_{\mathcal{T}}$ is a diagonal matrix with (n, n) component:

$$\mathcal{H}_{\mathcal{T}}^{(n,n)} = \sigma_{\Omega_n^2}^{-2} \left[(b_1^\star + 2b_2^\star \mathcal{T}_n^\star)^2 + 2b_2^\star (b_0^\star + b_1^\star \mathcal{T}_n^\star + b_2 \mathcal{T}_n^{\star 2} - \hat{\Omega}_n^2) \right] + \sigma_{\mathcal{T}}^{-2} \tag{2.176}$$

www.EngineeringBooksPdf.com

For large N, the posterior PDF can be approximated by Gaussian distribution and its covariance matrix is given by the inverse of the Hessian matrix.

2.6.4 Analysis of the Measurements

Influenced by the subtropical climate and seasonal monsoon, the weather of Macao changes from hot and humid to cool and dry from Summer to Winter. A monitoring project was accomplished for a 22-story residential building, namely the East Asia Hall, from June to November in 2008. It was inaugurated for lodging athletes during the 4th East Asian Games held in 2005 and started to serve as a student dormitory of the University of Macau afterwards. The East Asian Hall is a reinforced concrete building with an L-shape floor plan and the two spans have different lengths. The height and the two spans are 64.70 m, 51.90 m and 61.75 m, respectively.

Acceleration time histories of the East Asian Hall were observed daily from 23:00 to 23:30 throughout the monitoring period. The data were taken around midnight to minimize the spatial variation of the temperature. Two state-of-the-art servo-accelerometers, which operate based on an exploration geophone spring–mass system, are mounted in two orthogonal directions at the junction of the two spans on the 18th floor. The sensitivity of the overall acquisition system is 50V/g and the sampling interval is 0.002 s.

The Bayesian spectral density approach for parametric identification and model updating regression analysis are applied. During the monitoring period, four typhoons 'flitted' over Macao. The structural behavior under such violent wind excitation is treated as discordance and the measurements obtained under these events are not taken into account for the analysis. By excluding these fifteen days of measurements, there are 168 pairs of identified squared fundamental frequency and measured temperature in the data set. Figure 2.28(a) shows the variation of the identified squared fundamental frequencies with their associated uncertainties represented by a confidence interval that is bounded by the plus or minus three standard derivations from the estimated values. It is noticed that this confidence interval contains 99.7% of the probability. Since the confidence intervals are narrow compared with the variation

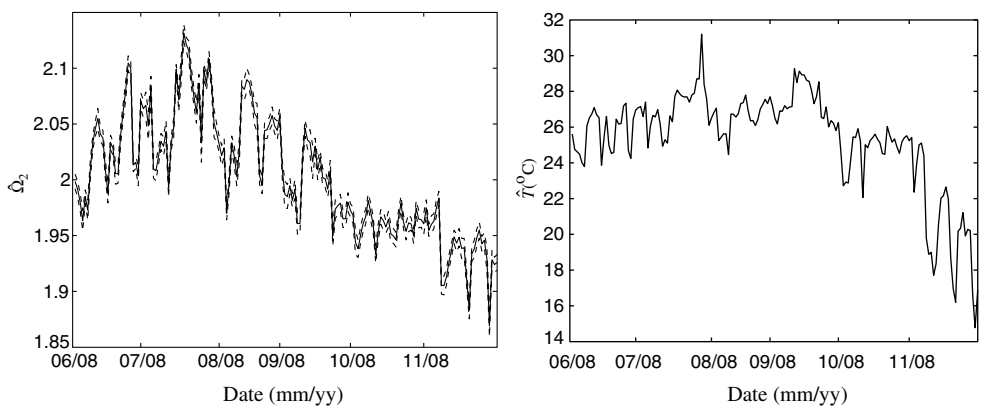

Figure 2.28 Squared fundamental frequency and temperature

www.EngineeringBooksPdf.com

Table 2.2 Identification result of the quadratic function coefficients

Parameter	b_0	b_1	b_2
Estimated value	2.24	−0.000 384	0.001 13
COV	0.0030	0.015	0.011

of the identified values, this indicates that beyond the statistical uncertainty, the natural frequency is influenced by other factors. The difference between the maximum and minimum of the squared fundamental frequency is approximately 13% and this may be considered as an 'alarm of damage'. On the other hand, by comparing the fluctuation of the squared fundamental frequency and the ambient temperature, as shown in Figure 2.28(b), there is strong correlation between the two variables. This confirms that the ambient temperature is an important factor that has to be considered in the development of a reliable dynamic monitoring system.

By using the Bayesian methodology described in Section 2.6.3, the coefficients of the quadratic function are identified and they are shown in Table 2.2. Furthermore, the associated uncertainty can also be quantified and the coefficients of variation (COV) are also shown in the table. Figure 2.29 shows the squared fundamental frequency–temperature quadratic relationship with the optimal values of the coefficients. The identified squared fundamental frequencies and the measured temperatures are represented by the dots. A concave parabolic curve is obtained from the regression analysis, and most of the data points fall in the monotonically increasing region. Note that this curve is valid only in the range from 15°C to 30°C where into most of the data points fall. To extend the applicable region, measurements with lower or higher temperatures will be necessary.

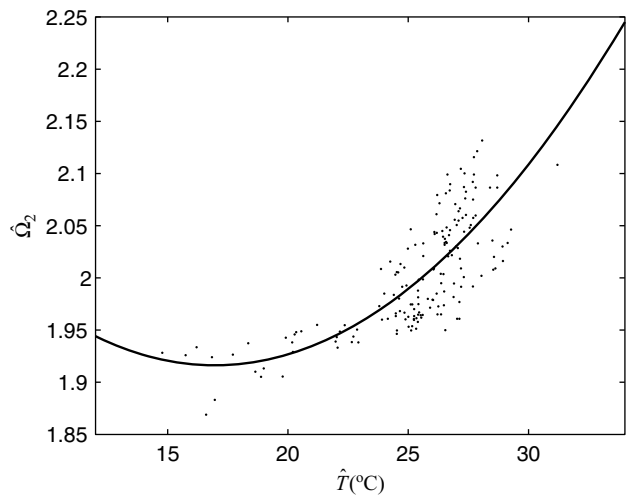

Figure 2.29 Identified squared fundamental frequency versus temperature and the best fitting curve

www.EngineeringBooksPdf.com

2.6.5 Concluding Remarks

In order to develop a reliable structural health monitoring system, it is essential to investigate the ambient effects on the structural behavior. In this study, the Bayesian method was applied to establish the relationship between the squared fundamental frequency and the ambient temperature. It started with the approximation of the Timoshenko beam and it turned out that the relationship is adequately modeled by a quadratic function. By utilizing six-month daily structural acceleration time histories of a 22-story residential building, the uncertain coefficients were obtained. It was observed that the squared fundamental frequency exhibits 13% difference during the monitoring period and the variation follows a similar pattern as that of the ambient temperature. By the regression analysis, a concave relationship was found and most of the data points fall in the monotonically increasing region. The method does not only identify the optimal estimates, but also allows for quantification of the associated uncertainty.

2.7 Application to Noise Parameters Selection for the Kalman Filter

2.7.1 Problem Description

The Kalman filter was developed by Kalman and Bucy for predicting and filtering random signals of linear systems [128, 129]. Thereafter, the extended Kalman filter was applied to system identification for linear or slightly nonlinear dynamic systems, in addition to its capability of state estimation, in many disciplines of science and engineering [57, 179, 190, 250]. The Kalman filter has become a popular tool in structural identification [89, 110, 238, 258] because it provides not only the response prediction and the estimation of the model parameters but also their associated uncertainties. In addition, the algorithm is online, i.e., the model is updated with new data points at every time step. However, the process noise and measurement noise parameters of the Kalman filter are usually assumed to be known [110, 156] and this assumption is difficult to be fulfilled in some of the applications. Different choices of the measurement noise parameters were demonstrated to affect the performance of the Kalman filter [111]. Therefore, recent attention has been devoted to the estimation of these parameters for the Kalman filter. Shi *et al.* (2000) formulated the extended Kalman filter in the frequency domain to estimate the process noise parameter by including the mechanical parameters and the spectral intensity in an augmented state vector for single-degree-of-freedom (SDOF) systems [236]. Ching *et al.* (2004) applied the expectation maximization technique [239] to estimate the measurement noise parameter [53]. However, there is currently no methodology to estimate both noise parameters of the Kalman filter simultaneously for multi-degree-of-freedom systems. In this section, a Bayesian methodology is introduced to update these unknown parameters. First, the fundamentals of the Kalman filter will be presented in the next section.

2.7.2 Kalman Filter

In this section the basic principles of the Kalman filter are presented for linear multi-degree-of-freedom (MDOF) systems. More details can be found elsewhere [36, 84, 128, 129]. Even though it is not often emphasized and it was not shown explicitly in the original formulation [128, 129], the Kalman filter is a Bayesian updating procedure. Consider a second-order

www.EngineeringBooksPdf.com

linear system with N_d degrees of freedom (DOFs) and its equation of motion:

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{C}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) = \mathbf{T}_0\mathbf{F}(t) \tag{2.177}$$

where $\mathbf{x}$ denotes the generalized coordinate vector of the system, and $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$ are the mass, damping and stiffness matrix of the system, respectively; $\mathbf{T}_0$ and $\mathbf{F}$ are the force distributing matrix and the excitation, respectively. The *state vector* $\mathbf{X}$ is defined to include the displacement and velocity vector:

$$\mathbf{X}(t) \equiv [\mathbf{x}(t)^T, \dot{\mathbf{x}}(t)^T]^T \tag{2.178}$$

Then, Equation (2.177) can be transformed to the state-space form:

$$\dot{\mathbf{X}}(t) = \mathbf{A}_c\mathbf{X}(t) + \mathbf{B}_c\mathbf{F}(t) \tag{2.179}$$

where the matrices $\mathbf{A}_c$ and $\mathbf{B}_c$ are given by:

$$\mathbf{A}_c = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix}$$

$$\mathbf{B}_c = \begin{bmatrix} \mathbf{0} \\ \mathbf{M}^{-1}\mathbf{T}_0 \end{bmatrix} \tag{2.180}$$

Equation (2.179) can be discretized to a difference equation by assuming that the excitation is constant within any time interval, i.e., $\mathbf{F}(n\Delta t + \tau) = \mathbf{F}(n\Delta t)$, $^\forall\tau \in [0, \Delta t)$, $n = 0, 1, 2, \ldots$

$$\mathbf{X}_{n+1} = \mathbf{A}\mathbf{X}_n + \mathbf{B}\mathbf{F}_n \tag{2.181}$$

where the matrices $\mathbf{A}$ and $\mathbf{B}$ are given by:

$$\mathbf{A} = \exp(\mathbf{A}_c\Delta t)$$

$$\mathbf{B} = \mathbf{A}_c^{-1}(\mathbf{A} - \mathbf{I})\mathbf{B}_c \tag{2.182}$$

Note that the matrix exponential can be computed by the function 'expm' in MATLAB® [171]; Δt is the sampling time step. In order to simplify the symbols, the following notations are defined:

$$\mathbf{X}_n \equiv \mathbf{X}(n\Delta t)$$

$$\mathbf{F}_n \equiv \mathbf{F}(n\Delta t) \tag{2.183}$$

The excitation $\mathbf{F}$ is modeled as discrete stationary Gaussian white noise with zero mean and covariance matrix $\boldsymbol{\Sigma}_F(\boldsymbol{\theta}_F)$, where $\boldsymbol{\theta}_F$ is a vector which parameterizes the covariance matrix $\boldsymbol{\Sigma}_F$.

To account for the modeling error and measurement noise, the relationship between the output measurements and the state vector is defined by:

$$\mathbf{y}_n = \mathbf{L}_o\mathbf{X}_n + \boldsymbol{\epsilon}_n \tag{2.184}$$

The observation matrix $\mathbf{L}_o$ is $N_o \times 2N_d$, where N_o is the number of observed DOFs. For example, the state vector assembles the displacement vector and the velocity vector of the

www.EngineeringBooksPdf.com

system, whereas the output measurement can be the absolute acceleration of N_o ($< N_d$) DOFs of a structure subjected to earthquake ground motion. The measurement noise $\boldsymbol{\epsilon}$ is assumed to be a Gaussian, independent and identically distributed (i.i.d.) process with zero mean and covariance matrix $\boldsymbol{\Sigma}_\epsilon(\boldsymbol{\theta}_\epsilon)$, where $\boldsymbol{\theta}_\epsilon$ is a vector which parameterizes the covariance matrix $\boldsymbol{\Sigma}_\epsilon$. Furthermore, it is assumed to be statistically independent to $\mathbf{F}$. The essential steps of the Kalman filter algorithm are to predict and filter at each time step with the data set $\mathcal{D}_N = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$. When the measurements up to the nth time step $\mathcal{D}_n = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ are available, the predicting procedure is applied to estimate $\mathbf{y}_{n+1}$ by using the conditional PDF $p(\mathbf{X}_{n+1}|\mathcal{D}_n, \mathcal{C})$, which is multi-variate Gaussian for linear systems. By using Equation (2.181), the predicted state vector at the $(n+1)$th time step can be estimated from the filtered state at the nth time step:

$$\hat{\mathbf{X}}_{n+1|n} \equiv E[\mathbf{X}_{n+1}|\mathcal{D}_n, \mathcal{C}] = \mathbf{A}\hat{\mathbf{X}}_{n|n} \tag{2.185}$$

where $\hat{\mathbf{X}}_{n|n} \equiv E[\mathbf{X}_n|\mathcal{D}_n, \mathcal{C}]$. In addition, the uncertainty of this estimation is represented by its covariance matrix:

$$\mathbf{P}_{n+1|n} \equiv E[(\mathbf{X}_{n+1} - \hat{\mathbf{X}}_{n+1|n})(\mathbf{X}_{n+1} - \hat{\mathbf{X}}_{n+1|n})^T|\mathcal{D}_n, \mathcal{C}] = \mathbf{A}\mathbf{P}_{n|n}\mathbf{A}^T + \mathbf{B}\boldsymbol{\Sigma}_F\mathbf{B}^T \tag{2.186}$$

where $\mathbf{P}_{n|n} \equiv E[(\mathbf{X}_n - \hat{\mathbf{X}}_{n|n})(\mathbf{X}_n - \hat{\mathbf{X}}_{n|n})^T|\mathcal{D}_n, \mathcal{C}]$ represents the filtered state covariance matrix at the nth time step using the measurements up to the nth time step. When the measurement at the $(n+1)$th time step is available, the filtered state vector $\hat{\mathbf{X}}_{n+1|n+1}$ can be found by maximizing the conditional PDF $p(\mathbf{X}_{n+1}|\mathcal{D}_{n+1}, \mathcal{C})$:

$$\hat{\mathbf{X}}_{n+1|n+1} = \mathbf{P}_{n+1|n+1}(\mathbf{P}_{n+1|n}^{-1}\hat{\mathbf{X}}_{n+1|n} + \mathbf{L}_o^T\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{y}_{n+1}) \tag{2.187}$$

and the uncertainty of this estimation is represented by its covariance matrix:

$$\mathbf{P}_{n+1|n+1} = (\mathbf{P}_{n+1|n}^{-1} + \mathbf{L}_o^T\boldsymbol{\Sigma}_\epsilon\mathbf{L}_o)^{-1} \tag{2.188}$$

Obviously, the accuracy of the state estimation and their covariance matrices depend on the process noise and measurement noise parameters but these parameters are unknown in practice. Here, the Bayesian approach is used to select these noise parameters.

Let $\boldsymbol{\theta} = [\boldsymbol{\theta}_F^T, \boldsymbol{\theta}_\epsilon^T]^T$ be the parameter vector that determines the covariance matrices $\boldsymbol{\Sigma}_F$ and $\boldsymbol{\Sigma}_\epsilon$. Since $\mathcal{D} = \mathcal{D}_N$, Equation (2.19) can be rewritten as:

$$p(\boldsymbol{\theta}|\mathcal{D}_N, \mathcal{C}) = \kappa_0 p(\boldsymbol{\theta}|\mathcal{C})p(\mathcal{D}_N|\boldsymbol{\theta}, \mathcal{C}) \tag{2.189}$$

where $\mathcal{C}$ is a prescribed class of models for the system, $p(\boldsymbol{\theta}|\mathcal{C})$ denotes the prior PDF of the parameters and κ_0 is the normalizing constant. Here, a non-informative prior is taken for the parameter vector $\boldsymbol{\theta}$, so it can be absorbed into the normalizing constant κ_0. Therefore, the most probable parameter vector $\boldsymbol{\theta}^\star$ can be found by maximizing the likelihood function $p(\mathcal{D}_N|\boldsymbol{\theta}, \mathcal{C})$. The likelihood function $p(\mathcal{D}_N|\boldsymbol{\theta}, \mathcal{C})$ reflects the contribution of the measured data $\mathcal{D}$ in establishing the posterior PDF of $\boldsymbol{\theta}$. In order to enhance the computational efficiency, the likelihood function can be expressed as the product of the conditional PDFs of the measurements given all the previous time steps [288]:

$$p(\mathcal{D}_N|\boldsymbol{\theta}, \mathcal{C}) = \prod_{n=1}^{N} p(\mathbf{y}_n|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}, \boldsymbol{\theta}, \mathcal{C}) \tag{2.190}$$

www.EngineeringBooksPdf.com

where the conditional PDFs are given by:

$$p(\mathbf{y}_n|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}, \boldsymbol{\theta}, \mathcal{C}) = (2\pi)^{-\frac{N_o}{2}} |\mathbf{Q}_n|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1})^T \mathbf{Q}_n^{-1} (\mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1})\right] \tag{2.191}$$

The one-step-ahead estimator $\hat{\mathbf{y}}_{n|n-1} \equiv E[\mathbf{y}_n|\mathcal{D}_{n-1}, \mathcal{C}]$ can be expressed in terms of the estimator $\hat{\mathbf{X}}_{n|n-1}$ by using Equation (2.184):

$$\hat{\mathbf{y}}_{n|n-1} \equiv E[\mathbf{y}_n|\mathcal{D}_{n-1}, \mathcal{C}] = \mathbf{L}_o \hat{\mathbf{X}}_{n|n-1} \tag{2.192}$$

In addition, the covariance matrix $\mathbf{Q}_n$ appearing in Equation (2.191) is a linear combination of $\mathbf{P}_{n|n-1}$ and $\boldsymbol{\Sigma}_\epsilon$:

$$\mathbf{Q}_n \equiv E[(\mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1})(\mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1})^T|\mathcal{D}_{n-1}, \mathcal{C}] = \mathbf{L}_o \mathbf{P}_{n|n-1} \mathbf{L}_o^T + \boldsymbol{\Sigma}_\epsilon \tag{2.193}$$

Although the matrix $\boldsymbol{\Sigma}_F$ does not explicitly appear in the above expressions, its influence on the likelihood function is reflected in the covariance matrices $\mathbf{P}_{n|n-1}$ and $\mathbf{Q}_n$. Since a non-informative prior distribution is used, the objective function J can be defined as the negative logarithm of the likelihood function without including the constant term:

$$\begin{aligned} J(\boldsymbol{\theta}) = {} & \frac{1}{2}\sum_{n=1}^{N} \ln\left|\mathbf{L}_o \mathbf{P}_{n|n-1} \mathbf{L}_o^T + \boldsymbol{\Sigma}_\epsilon\right| + \\ & \frac{1}{2}\sum_{n=1}^{N} \left[(\mathbf{y}_n - \mathbf{L}_o \hat{\mathbf{X}}_{n|n-1})^T (\mathbf{L}_o \mathbf{P}_{n|n-1} \mathbf{L}_o^T + \boldsymbol{\Sigma}_\epsilon)^{-1} (\mathbf{y}_n - \mathbf{L}_o \hat{\mathbf{X}}_{n|n-1})\right] \end{aligned} \tag{2.194}$$

The optimal parameter vector $\boldsymbol{\theta}^\star$ can be found by minimizing the objective function J. Furthermore, the updated PDF allows one to quantify the uncertainty of the estimation, e.g., to calculate the standard deviation or the contours with equal probability density.

2.7.3 Illustrative Examples

2.7.3.1 SDOF System Subjected to Unknown Input

A single-degree-of-freedom (SDOF) system is subjected to zero-mean stationary white noise excitation with spectral intensity $S_{F0} = 0.0048$ N^2 s. The mass, damping coefficient and stiffness are taken to be $M = 1.0$ kg, $C = 0.4\pi$ N s/m, and $K = (4\pi)^2$ N/m so that the natural frequency and damping ratio of the system are 2.0 Hz and 5.0%, respectively. Its velocity is sampled at a rate of 100 Hz for 10 s. To generate the time history of the velocity measurement, 10% of the rms noise is superimposed onto the velocity time history, i.e., the root-mean-square (rms) of the measurement noise is equal to 10% of the rms value of the noise-free velocity of the oscillator. The actual values of $\tilde{\sigma}_F^2$ and $\tilde{\sigma}_\epsilon^2$ are 3.0 N^2 and 4.8×10^{-5} m^2 s^{-2}, respectively.

To verify the result produced by the Kalman filter, the time histories of the displacement and the velocity estimated by the Kalman filter are compared to their actual values in Figures 2.30 and 2.31, respectively. Here, the actual values of the noise variances are used. The solid lines in both figures represent the actual response, whereas the dotted lines represent the response estimated by the Kalman filter. It is noted that the estimated velocity in Figure 2.31 is close to

www.EngineeringBooksPdf.com

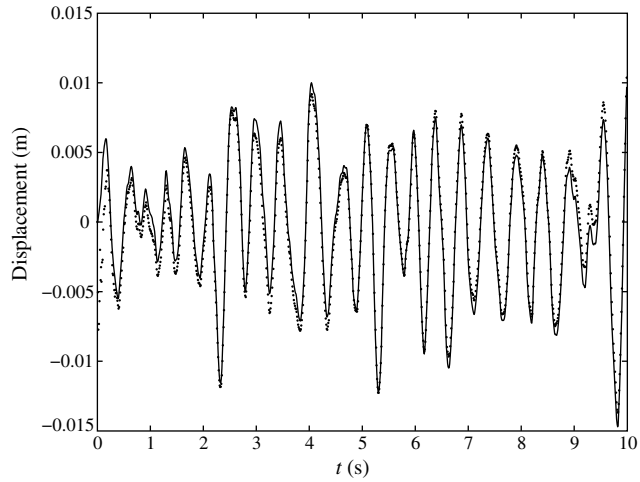

Figure 2.30 Actual and estimated displacement time history

the actual velocity with an rms error of 9.95%, which is virtually the same as the measurement noise level. In addition, the displacement estimated from the Kalman filter in Figure 2.30 is close to the time history of the actual displacement except for the first few iterations that the Kalman filter has not yet obtained sufficient information to update the response. However, the rms error of 0.0011 m, corresponding to a 22.4% error, is large compared to the rms of

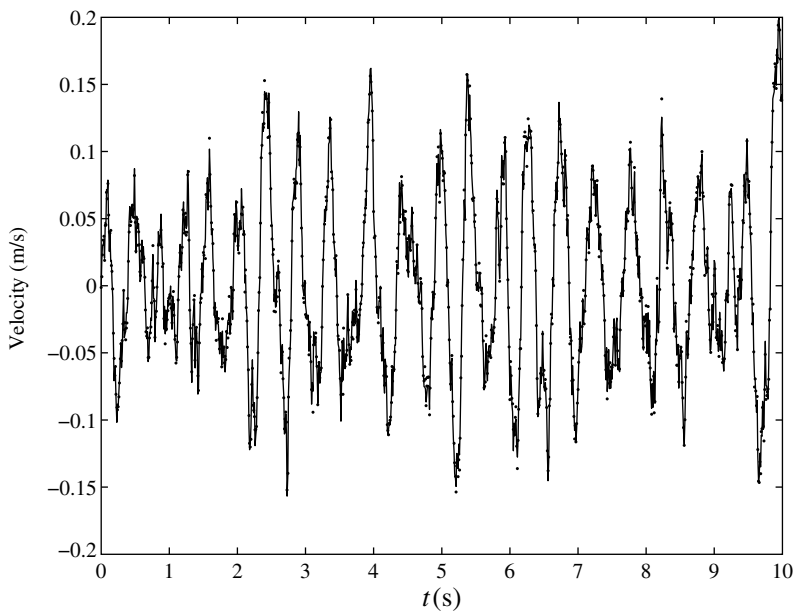

Figure 2.31 Actual and estimated velocity time history

www.EngineeringBooksPdf.com

0.0049 m for the displacement time history. The large rms error is due to the deviation between the two time histories at the first few time steps. To confirm this, the rms error is again computed by ignoring the data in the first second. The rms error of the predicted displacement is 0.0007 m and the rms of the actual displacement is 0.0050 m. The rms error is 14.0% of the rms of the noise-free displacement.

Next, the effect of the choice for the values of σ_F^2 and σ_ϵ^2 on the model performance is studied by comparing the rms errors associated with different combinations of σ_F^2 and σ_ϵ^2. An extra second of data is generated but the state estimation in the first second will be ignored in the calculation of the rms errors to avoid the transient state of the Kalman filter as observed earlier. The variables α_F and α_ϵ denote the following ratios:

$$\alpha_F = \frac{\sigma_F^2}{\tilde{\sigma}_F^2};\ \alpha_\epsilon = \frac{\sigma_\epsilon^2}{\tilde{\sigma}_\epsilon^2} \tag{2.195}$$

where $\tilde{\sigma}_F^2$ and $\tilde{\sigma}_\epsilon^2$ are the actual process noise and measurement noise variance. Table 2.3 shows the rms errors of the estimated displacements and velocities for six combinations of σ_F^2 and σ_ϵ^2. The first row with unity ratios of α_F and α_ϵ represents the case that the actual values of σ_F^2 and σ_ϵ^2 are utilized. When the actual process noise and measurement noise variances are overestimated/underestimated by the same factor, i.e., $\alpha_F = \alpha_\epsilon = 10$, it is found that the rms errors for both the displacement and the velocity remain unchanged. Specifically, by Equations (2.185) – (2.188), it can be shown that the state vector estimation will be identical if the initial covariance matrix, the process noise and the measurement noise variances are multiplied by the same factor.

When $\alpha_F = 10$ and $\alpha_\epsilon = 1$, the rms error of the displacement slightly decreases whereas the rms error of the velocity slightly increases. In addition, when $\alpha_F = 100$ and $\alpha_\epsilon = 1$, the rms error of the displacement further decreases while the rms error of the velocity again slightly increases. Increasing the process noise variance tends to affect the estimation of the velocity more than the estimation of the displacement. On the other hand, it can be shown that the rms error of the estimated displacement/velocity associated with $\alpha_F = 100$ and $\alpha_\epsilon = 1$ is equivalent to that with $\alpha_F = 10$ and $\alpha_\epsilon = 1/10$. The effect of underestimation of the process noise variance is the same as that due to the overestimation of the measurement noise variance. When $\alpha_F = 1$ and $\alpha_\epsilon = 10$, it is found that both the rms errors of the displacement and the velocity increase. In addition, the rms errors further increase when $\alpha_F = 1$ and $\alpha_\epsilon = 100$. Therefore, it is concluded that overestimating the measurement noise parameter or underestimating the process noise variance will increase the rms errors of the displacement and velocity significantly.

Table 2.3 RMS errors of the estimated displacements and velocities for different values of α_F and α_ϵ

α_F	α_ϵ	RMS error of displacement (m)	RMS error of velocity (m/s)
1	1	0.0007	0.0067
10	10	0.0007	0.0067
10	1	0.0006	0.0070
100	1	0.0005	0.0071
1	10	0.0009	0.0103
1	100	0.0016	0.0214

www.EngineeringBooksPdf.com

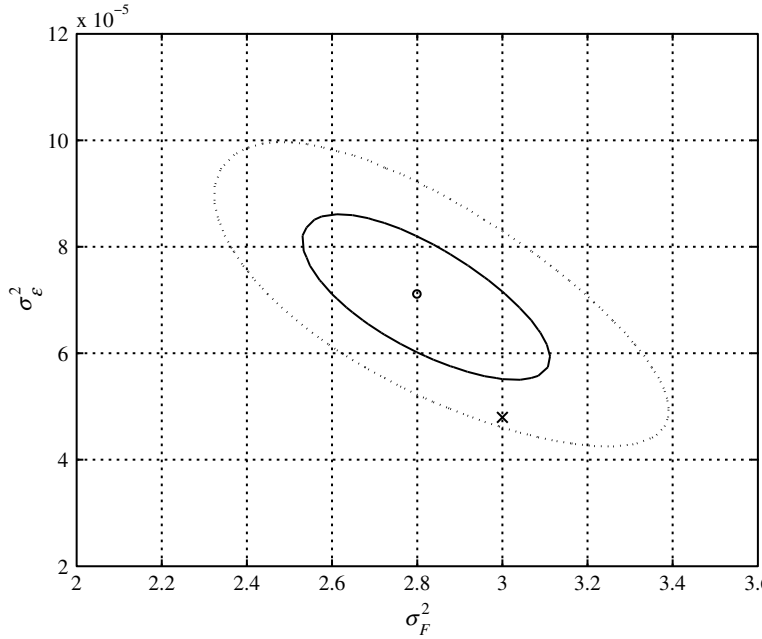

Figure 2.32 Contours of the likelihood function

As demonstrated previously the process noise and the measurement noise parameters directly affect the state vectors estimated by the Kalman filter. Furthermore, the covariance matrix of the state estimation is affected as well. Therefore, accurate estimation of the noise parameters is necessary for good performance of the filter. In this example, the Bayesian approach is applied to select σ_F^2 and σ_ϵ^2. Figure 2.32 shows the contours of the likelihood function $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$ together with the actual noise variances $\tilde{\boldsymbol{\theta}} = [\tilde{\sigma}_F^2, \tilde{\sigma}_\epsilon^2]^T$ and its optimal estimate $\boldsymbol{\theta}^\star$. The two contours correspond to 50% and 10% of the peak value. The optimal values of $\sigma_F^{2\star} = 2.8\text{N}^2$ and $\sigma_\epsilon^{2\star} = 7.1 \times 10^{-5}\text{m}^2/\text{s}^2$ are at reasonable distance to the actual values as the actual noise variances are located within the region with significant probability density. Therefore, the Bayesian approach is validated to give accurate estimation for both noise variances for the linear oscillator.

2.7.3.2 Ten-Story Building Subjected to Ground Acceleration

The second example uses a ten-story shear-building model. It has equal floor mass and interstory stiffness distributed over all stories. The building is subjected to base acceleration adequately modeled by stationary Gaussian white noise excitation with spectral intensity $S_{F0} = 1.6\times10^{-3}\ \text{m}^2/\text{s}^3$. The building has stiffness-to-mass ratio of 4016 s^{-2} so that its fundamental frequency is 1.5 Hz. The Rayleigh damping model is assumed, i.e., the damping matrix is given by $\mathbf{C} = \alpha_1\mathbf{M} + \alpha_2\mathbf{K}$, where $\mathbf{M}$ and $\mathbf{K}$ are the mass and stiffness matrix, respectively. The values of α_1 and α_2 are taken to be 0.0709 s^{-1} and 2.66×10^{-4} s so that the damping ratios

www.EngineeringBooksPdf.com

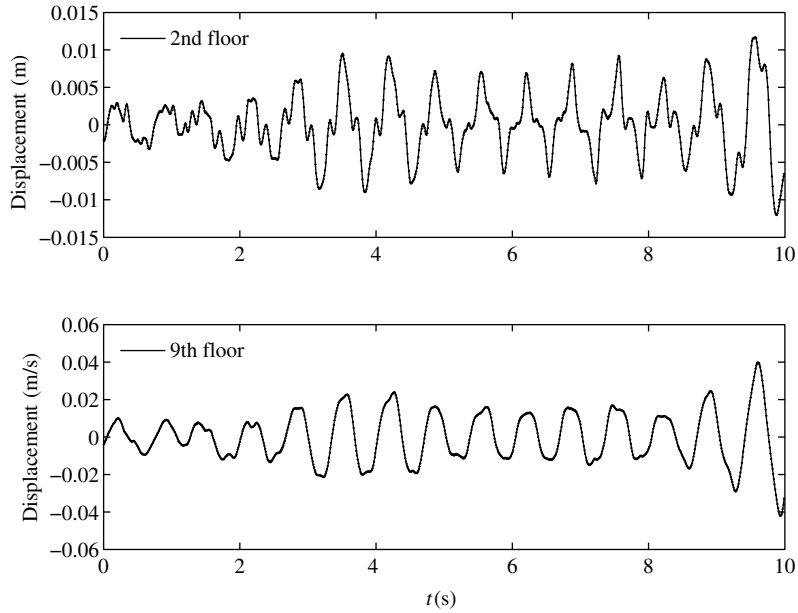

Figure 2.33 Actual and estimated displacement time histories of the 2nd and 9th floors

of the first two modes are 0.5%. By using the absolute acceleration measurements on the 1st, 4th, 7th, and 10th floors, it is attempted to estimate the displacement and the velocity of each floor. The rms of the measurement noise is taken to be 5% rms of the noise-free acceleration of the top floor. The accelerations are measured with a sampling frequency of 100 Hz for 10 s.

Figure 2.33 (Figure 2.34) shows the comparison between the displacement (velocity) estimated with the actual noise variances and their corresponding actual curves of the 2nd and 9th floors. The solid lines represent the actual displacement or velocity, whereas the dotted lines represents its Kalman estimation but the two sets of curves are on top of each other. The observation is further supported by the small rms errors which are only 2.59% and 2.41% of the rms values of the displacements for the 2nd floor and the 9th floor, respectively. In addition, the estimated velocities also agree well with the actual velocities as can be seen in Figure 2.34. The rms errors are 9.44% and 5.47% of the rms values of the velocities for the 2nd floor and the 9th floor, respectively. In conclusion, the Kalman filter is validated to provide an accurate state estimation for the linear MDOF system.

The Bayesian approach is applied to the selection of σ_F^2 and σ_ϵ^2. Figure 2.35 shows the contours, corresponding to 50% and 10% of the peak value, of the likelihood function together with the actual parameter and its optimal estimate. The optimal values of $\sigma_F^{2\star}$ and $\sigma_\epsilon^{2\star}$ are 1.02 m^2/s^4 and 1.22×10^{-2} m^2/s^4, which are close to the actual values (1.00 m^2/s^4 and 1.27×10^{-2} m^2/s^4) in the sense that the actual parameters are located within the region with significant probability density. Therefore, the Bayesian approach is confirmed to give accurate estimation for both noise variances.

www.EngineeringBooksPdf.com

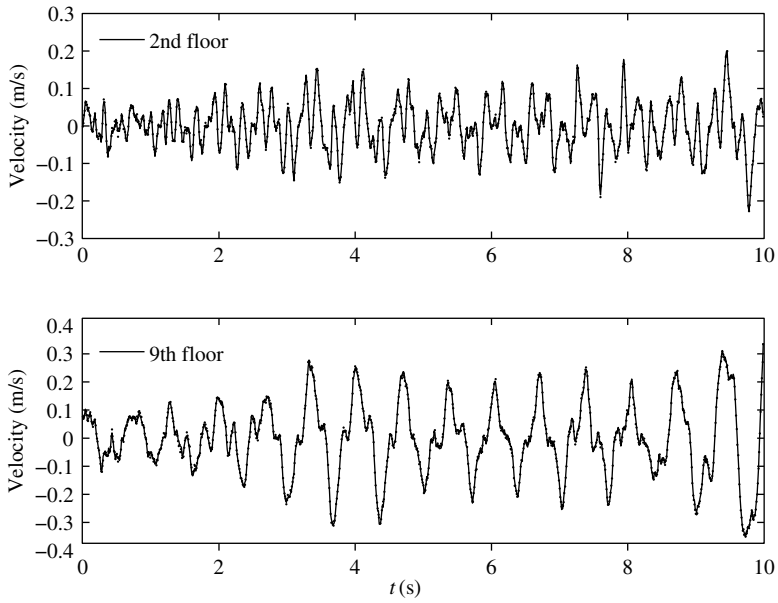

Figure 2.34 Actual and estimated velocity time histories of the 2nd and 9th floors

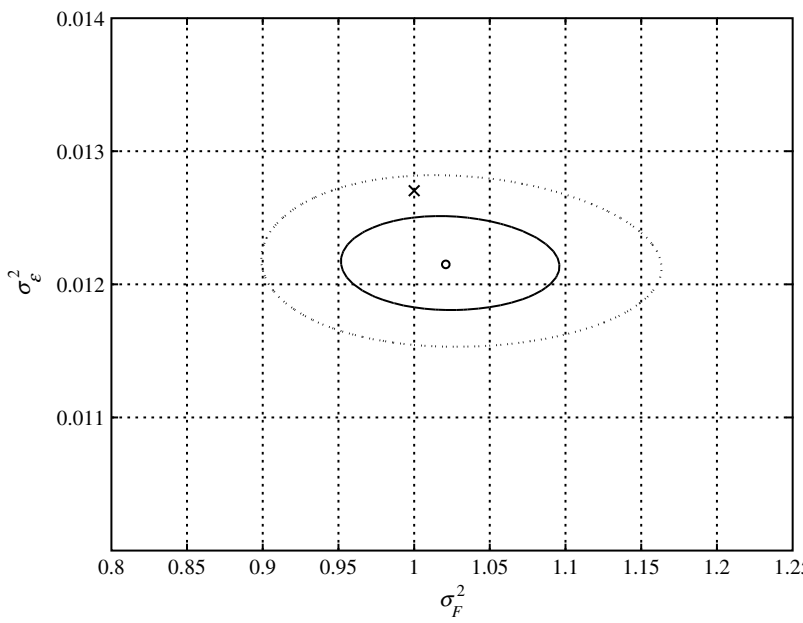

Figure 2.35 Contours of the normalized likelihood function

www.EngineeringBooksPdf.com

2.7.3.3 Half-or-double Optimization Algorithm

An alternative optimization approach, regarded as *half-or-double optimization algorithm* herein, is applied. The initial parameter vector $\boldsymbol{\theta}_0 = (\sigma_{F0}^2, \sigma_{\epsilon 0}^2)$ is chosen arbitrarily and it is taken to be (1, 1) in this example. Then, the objective function in Equation (2.194) is evaluated for the center point and the eight neighboring points:

$$\{(\sigma_F^2, \sigma_\epsilon^2) \in \mathbb{R}^2 | \sigma_F^2 = \sigma_{F0}^2 \alpha_F, \sigma_\epsilon^2 = \sigma_{\epsilon 0}^2 \alpha_\epsilon,\ \alpha_F, \alpha_\epsilon = 1/2, 1, 2\}$$

and the candidate parameter vector is the one that gives the smallest objective function value. If a point $(\sigma_{F0}^2 \alpha_F, \sigma_{\epsilon 0}^2 \alpha_\epsilon)$ gives a smaller objective function value than the center point $(\sigma_{F0}^2, \sigma_{\epsilon 0}^2)$, then the point $(\sigma_{F0}^2/\alpha_F, \sigma_{\epsilon 0}^2/\alpha_\epsilon)$ is no longer required for evaluation due to the convexity of the objective function. After the candidate parameter vector is obtained, it is used as the initial parameter vector in the next iteration. However, it does not require evaluating all the eight neighboring points as some of them were already evaluated in the previous iteration. Specifically, if the candidate vector in the previous iteration was a corner point (i.e., $\alpha_F \neq 1$ and $\alpha_\epsilon \neq 1$), then five neighboring points will have to be evaluated. Otherwise, only three neighboring points will have to be evaluated. The iteration procedure is repeated until the candidate parameter vector remains the same. Use $\boldsymbol{\theta}'$ to denote this sub-optimal parameter vector and they are $\sigma_F^{2\prime} = 1.00\,\mathrm{m}^2/\mathrm{s}^4$ and $\sigma_\epsilon^{2\prime} = 1.56 \times 10^{-2}\,\mathrm{m}^2/\mathrm{s}^4$ in this example. The rms errors of the displacements and the velocities estimated with the optimal and the sub-optimal parameters $\boldsymbol{\theta}^\star$ and $\boldsymbol{\theta}'$ as shown in Table 2.4 are very close to the rms errors of the displacements and the velocities estimated with the actual noise parameter vector. Therefore, the half-or-double optimization algorithm is an alternative for optimization in practice.

2.8 Application to Prediction of Particulate Matter Concentration

2.8.1 Introduction

Nowadays, degradation of air quality is widely observed in most urban areas due to rapid population growth and economic development. As more goods and services are produced, energy is consumed in the fuel combustion processes. At the same time, air pollutants such as particulates, sulfur dioxide, nitrogen dioxide or carbon monoxide are generated and released to the atmosphere. When the air pollutants enter human bodies, they may cause adverse health effects such as bronchitis, asthma, and heart disease, etc. [66, 67]. According to the statistics provided by the World Health Organization (WHO), three million people face terminal health problems worldwide annually by the outdoor air pollution from vehicles and industrial emissions. With such relevance to human health, it necessitates the development of an air quality prediction system since an accurate air quality forecast helps people reduce/cancel outdoor activities on the days of high pollutant concentrations. Meanwhile, more effective and suitable emission control strategies can be implemented if the future air quality is known in advance. Currently, many cities in Canada and the USA have regulations in place to curtail industrial and other activities during times of observed poor air quality. In Canada, the Alberta Environment asked the industry to cut production during morning hours of poor air quality in Calgary and Edmonton. In the United States, an ozone action day is called by the State or Local Air Quality Agency when ozone levels are forecasted to reach unhealthy levels. During

www.EngineeringBooksPdf.com

Table 2.4 RMS errors of the estimated displacements and velocities

RMS error	Using the actual parameters	Using the optimal parameters	Using the sub-optimal by the half-or-double algorithm
x_1	5.99×10^{-5}	5.99×10^{-5}	5.98×10^{-5}
x_2	1.11×10^{-4}	1.11×10^{-4}	1.11×10^{-4}
x_3	1.57×10^{-4}	1.57×10^{-4}	1.57×10^{-4}
x_4	1.96×10^{-4}	1.96×10^{-4}	1.96×10^{-4}
x_5	2.31×10^{-4}	2.31×10^{-4}	2.31×10^{-4}
x_6	2.61×10^{-4}	2.61×10^{-4}	2.61×10^{-4}
x_7	2.86×10^{-4}	2.86×10^{-4}	2.86×10^{-4}
x_8	3.06×10^{-4}	3.06×10^{-4}	3.06×10^{-4}
x_9	3.20×10^{-4}	3.20×10^{-4}	3.20×10^{-4}
x_{10}	3.27×10^{-4}	3.27×10^{-4}	3.27×10^{-4}
$\dot{x}_1$	6.01×10^{-3}	6.02×10^{-3}	6.03×10^{-3}
$\dot{x}_2$	5.97×10^{-3}	5.97×10^{-3}	5.97×10^{-3}
$\dot{x}_3$	6.20×10^{-3}	6.20×10^{-3}	6.20×10^{-3}
$\dot{x}_4$	6.38×10^{-3}	6.39×10^{-3}	6.39×10^{-3}
$\dot{x}_5$	6.52×10^{-3}	6.53×10^{-3}	6.53×10^{-3}
$\dot{x}_6$	6.67×10^{-3}	6.68×10^{-3}	6.68×10^{-3}
$\dot{x}_7$	6.81×10^{-3}	6.81×10^{-3}	6.81×10^{-3}
$\dot{x}_8$	6.93×10^{-3}	6.93×10^{-3}	6.93×10^{-3}
$\dot{x}_9$	7.05×10^{-3}	7.06×10^{-3}	7.05×10^{-3}
$\dot{x}_{10}$	7.17×10^{-3}	7.17×10^{-3}	7.17×10^{-3}

'ozone-action' days, the State or Local Authority encourages the community to take simple voluntary actions to help reducing ground level ozone and hence preventing violation of the National Ambient Air Quality Standards (NAAQSs) [192]. Therefore, the development of an air quality predicting system plays an important role in fulfilling these objectives.

Macao is a coastal city on the West side of the Pearl River Delta Region. It has a geographical area of 28.6 km^2 and a population of 510 000. Five air pollutants (PM_{10}, SO_2, NO_x, O_3 and CO) are monitored by the Government Department entitled the 'Meteorological and Geophysical Bureau' since 1999. PM_{10} is a collective noun for the particles with an aerodynamic diameter of not greater than 10 μm and it is the most dominant air pollutant among them, especially in winter. Figure 2.36 shows the daily averaged PM_{10} concentrations between 2001 and 2005. It is noted that there is a distinct seasonal pattern, i.e., high PM_{10} concentrations are generally observed in winter. Mok and Hoi (2005) found that the seasonal behavior was associated with the swing of the prevailing wind directions caused by the Asian monsoon climates [182]. The monsoon driven winter north-easterly winds bring upon Macao dry and particle enriched air masses leading to higher concentration in that period while the summer south-westerly winds transport humid and clean air from the South China Sea to the region leading to lower values of PM_{10} [182]. Table 2.5 shows the statistics of daily PM_{10} concentrations between 2001 and 2005. It is noted that the annual PM_{10} concentration generally increases during the studied period and the average increment, which is obtained from the slope of the best-fit straight line across the data points, is 2.8 $\mu g\, m^{-3}$/yr. However, there is no obvious trend observed for the maxima and minima of the daily averaged PM_{10} concentrations. The daily averaged PM_{10} concentration does not only depend on the amount of emission in the given area. It is

www.EngineeringBooksPdf.com

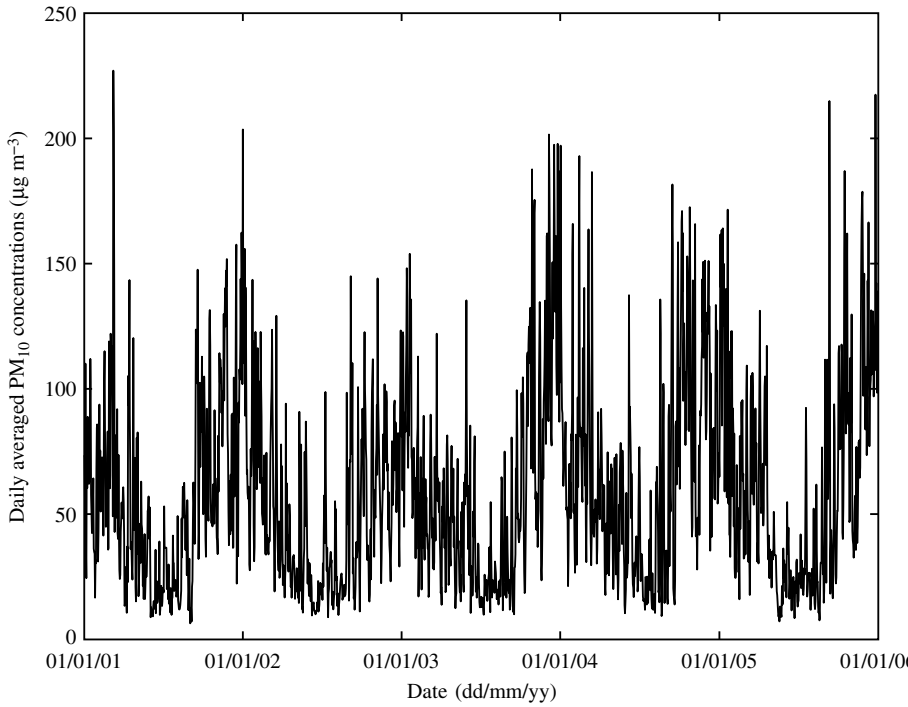

Figure 2.36 Daily averaged PM_{10} concentrations (2001–2005)

also affected by the meteorological conditions which control how much foreign pollutant is brought in and how much local pollutant is transported away. Meanwhile, the meteorological conditions also affect the dilution capacity of the atmosphere. On the contrary, the annual concentration should be less affected by those factors since it is averaged over a longer time period. In other words, the operation of averaging can help smooth the variation in the PM_{10} concentration due to the short-term changes in the meteorological conditions. Therefore, it is a useful indicator to reflect the interannual variability of PM_{10} pollution in Macao. Concluding from the annual PM_{10} concentration between 2001 and 2005, the PM_{10} pollution in Macao is worsening during the studied period.

The last column in Table 2.5 shows the number of days with non-attainment of the air quality standard. Non-attainment is assigned to a particular day when the daily averaged PM_{10}

Table 2.5 Statistics of daily averaged PM_{10} concentrations between 2001 and 2005

Year	Average	Maximum	Minimum	No. of non-attainment days
2001	56.7	227.0	6.4	5
2002	50.3	203.5	8.8	2
2003	59.2	201.5	9.8	16
2004	67.0	197.1	9.3	20
2005	62.2	217.4	7.1	14

www.EngineeringBooksPdf.com

concentration exceeds the threshold of 150 μg m^{-3} specified in the National Ambient Air Quality Standards (NAAQSs) established by the Environmental Protection Agency of the United States (USEPA). The NAAQS standards are adopted here since currently there is not any relevant standard for PM_{10} in Macao. It is noted that the percentage of days with non-attainment of the PM_{10} standard increases at a rate of 0.98% per year. The rise in the number of PM_{10} episode days between 2001 and 2005 coincides with the booming of the Macao economy as there is a 71.0% increase in the GDP per capita of Macao from USD 14 253 to USD 24 369 [163]. The increase in the economic activities leads to extra energy consumption and the subsequent increase in PM_{10} pollution. Therefore, the problem of PM_{10} pollution in Macao is severe and has raised our concern. To tackle the current circumstances, an initial step is to develop an air quality prediction system so that citizens can have chances to minimize the time of exposure to the air pollutants on the PM_{10} episode days.

Successful applications of the Kalman filter in air quality studies can be found in the literature [6, 190, 303]. The objective of the present study is to probe the applicability of the extended Kalman filter to perform state estimation for the time-varying statistical models. In this study, the extended Kalman filter is implemented on two different model classes, namely the time-varying auto-regressive (TVAR) model and the time-varying auto-regressive model with exogenous input (TVAREX). These models are tested with the daily averaged PM_{10} concentrations and the meteorological data provided by the Macao Meteorological and Geophysical Bureau [162]. The PM_{10} data were recorded by using the Tapered Element Oscillating Microbalance (TEOM) at the ambient monitoring station of Macao. It is located at the Taipa Grande Hill which has an altitude of 158.2 m. Therefore, its measurements are representative for indicating the general ambient air quality of Macao. The prediction results of the time-varying models are compared with the results obtained from the artificial neural network algorithm, which has been demonstrated previously as being a successful tool in the area of air quality prediction [83, 97, 107, 122, 184, 193, 206]. In the following section, two statistical model classes are presented.

2.8.2 Extended-Kalman-filter based Time-varying Statistical Models

2.8.2.1 TVAR(p) Model

In this section, the extended Kalman filter is formulated for the time-varying auto-regressive model of order p, which is abbreviated as the TVAR(p) model:

$$x_n = \theta_{1,n-1}x_{n-1} + \theta_{2,n-1}x_{n-2} + \cdots + \theta_{p,n-1}x_{n-p} + f_{n-1} \tag{2.196}$$

where x_n denotes the daily averaged PM_{10} concentration of the nth day. The input f represents the unmodeled dynamics, and it is modeled as a Gaussian independent and identically distributed (i.i.d.) process with zero mean and variance σ_f^2. It represents the neglected factors that influence the PM_{10} concentration. In addition, it is assumed that the measurement of the PM_{10} concentration, denoted as y_n, is contaminated during the measurement process. The relationship between y_n and x_n is given by:

$$y_n = x_n + \epsilon_n \tag{2.197}$$

www.EngineeringBooksPdf.com

where ϵ is the measurement noise and it is also modeled as a Gaussian i.i.d. with zero mean and variance σ_ϵ^2. Furthermore, the stochastic processes f and ϵ are assumed independent. The TVAR(p) model simply implies that the PM_{10} concentration of a day is a weighted sum of its concentrations of the previous p days and the weightings are represented by the time-varying auto-regressive coefficients $\theta_{l,n}$. These unknown coefficients evolve according to the following equation:

$$\theta_{l,n} = \theta_{l,n-1} + w_{l,n-1} \tag{2.198}$$

where $w_{l,n-1}$ denotes the variation of the coefficient $\theta_{l,n-1}$ on the $(n-1)$th day. The stochastic process $\mathbf{w}$ is modeled as a Gaussian i.i.d. with zero mean and covariance matrix $\text{diag}(\sigma_{w1}^2, \sigma_{w2}^2, \ldots, \sigma_{wp}^2)$. Then, the *state vector* $\mathbf{X}_n$ is defined to contain the PM_{10} concentrations of different days and the unknown coefficients to be estimated [104]:

$$\mathbf{X}_n = [x_n, x_{n-1}, \ldots, x_{n-p+1}, \theta_{1,n}, \theta_{2,n}, \ldots, \theta_{p,n}]^T \tag{2.199}$$

The measurement y_n can be expressed in terms of $\mathbf{X}_n$ in the following form:

$$y_n = \mathbf{L}_o \mathbf{X}_n + \epsilon_n \tag{2.200}$$

where the observation matrix $\mathbf{L}_o$ is a row vector in this case:

$$\mathbf{L}_o = [1,\ \mathbf{0}_{1\times(2p-1)}] \tag{2.201}$$

Also, the process noise vector $\mathbf{F}_n$ is defined to contain the process noises f_n and $w_{l,n}$, $l = 1, 2, \ldots, p$:

$$\mathbf{F}_n = [f_n, w_{1,n}, w_{2,n}, \ldots, w_{p,n}]^T \tag{2.202}$$

It is readily followed that $\mathbf{F}_n$ is Gaussian with zero mean and covariance matrix $\mathbf{\Sigma}_F$:

$$\mathbf{\Sigma}_F = \text{diag}(\sigma_f^2, \sigma_{w1}^2, \sigma_{w2}^2, \ldots, \sigma_{wp}^2) \tag{2.203}$$

Then, the TVAR(p) model in Equation (2.196) can be linearized locally to a first-order TVAR vector model:

$$\mathbf{X}_n = \hat{\mathbf{A}}_{n-1|n-1}\mathbf{X}_{n-1} + \mathbf{B}\mathbf{F}_{n-1} + \hat{\mathbf{G}}_{n-1|n-1} \tag{2.204}$$

The matrix $\hat{\mathbf{A}}_{n-1|n-1}$ denotes the estimator of the matrix $\mathbf{A}$ on the $(n-1)$th day, conditional on the measurements $y_1, y_2, \ldots, y_{n-1}$. It is noted that the same notation is used to denote the other estimators conditional on the given measurements. The matrices $\hat{\mathbf{A}}_{n-1|n-1}$ and $\mathbf{B}$ are given by:

$$\hat{\mathbf{A}}_{n-1|n-1} = \begin{bmatrix} \hat{\theta}_{1,n-1|n-1} & \cdots & \hat{\theta}_{p,n-1|n-1} & \hat{x}_{n-1|n-1} & \cdots & \hat{x}_{n-p|n-1} \\ & \mathbf{I}_{p-1} & & & \mathbf{0}_{(p-1)\times(p+1)} & \\ & \mathbf{0}_{p\times p} & & & \mathbf{I}_p & \end{bmatrix} \tag{2.205}$$

$$\mathbf{B} = \begin{bmatrix} 1 & & \mathbf{0}_{1\times p} \\ & \mathbf{0}_{(p-1)\times(p+1)} & \\ \mathbf{0}_{p\times 1} & & \mathbf{I}_p \end{bmatrix} \tag{2.206}$$

www.EngineeringBooksPdf.com

where $\hat{\theta}_{l,n-1|n-1}$ and $\hat{x}_{n-1|n-1}$ denote the estimator of the lth AR coefficient and the PM_{10} concentration on the $(n-1)$th day, conditional on the measurements $y_1, y_2, \ldots, y_{n-1}$. It is noted that the matrix $\mathbf{B}$ is constant. The vector $\hat{\mathbf{G}}_{n-1|n-1}$ is used to compensate the linearization error:

$$\hat{\mathbf{G}}_{n-1|n-1} = \begin{bmatrix} -\hat{\theta}_{1,n-1|n-1}\hat{x}_{n-1|n-1} - \cdots - \hat{\theta}_{p,n-1|n-1}\hat{x}_{n-p|n-1} \\ \mathbf{0}_{(2p-1)\times 1} \end{bmatrix} \tag{2.207}$$

With this TVAR vector model, the predicting and filtering steps for the PM_{10} concentrations can be performed with the Kalman filter. The essential steps of the Kalman filter are to predict and filter the measured PM_{10} concentration alternately. When the measured PM_{10} concentrations up to the $(n-1)$th day are available, the predicting procedure is applied to give the one-step-ahead prediction of the PM_{10} concentration. By using Equations (2.204) and (2.205), the predicted state vector on the nth day can be estimated from the filtered state vector on the $(n-1)$th day:

$$\hat{\mathbf{X}}_{n|n-1} = \begin{bmatrix} \hat{\theta}_{1,n-1|n-1}\hat{x}_{n-1|n-1} + \hat{\theta}_{2,n-1|n-1}\hat{x}_{n-2|n-1} + \cdots + \hat{\theta}_{p,n-1|n-1}\hat{x}_{n-p|n-1} \\ \hat{x}_{n-1|n-1} \\ \vdots \\ \hat{x}_{n-p+1|n-1} \\ \hat{\theta}_{1,n-1|n-1} \\ \vdots \\ \hat{\theta}_{p,n-1|n-1} \end{bmatrix} \tag{2.208}$$

In addition, the uncertainty of the PM_{10} concentration and the estimated model parameters can be quantified by the covariance matrix:

$$\mathbf{P}_{n|n-1} = \hat{\mathbf{A}}_{n-1|n-1}\mathbf{P}_{n-1|n-1}\hat{\mathbf{A}}^T_{n-1|n-1} + \mathbf{B}\boldsymbol{\Sigma}_F\mathbf{B}^T \tag{2.209}$$

When the measurement on the nth day is available, the PM_{10} concentration and the model parameters in the state vector are filtered as follows:

$$\hat{\mathbf{X}}_{n|n} = \mathbf{P}_{n|n}(\mathbf{P}^{-1}_{n|n-1}\hat{\mathbf{X}}_{n|n-1} + \sigma_\epsilon^{-2}\mathbf{L}_o^T y_n) \tag{2.210}$$

The uncertainty of the state estimation is represented by its covariance matrix:

$$\mathbf{P}_{n|n} = \mathbf{P}_{n|n-1} - \frac{1}{\sigma_\epsilon^2 + \sigma^2_{x,n|n-1}}\mathbf{v}_n\mathbf{v}_n^T \tag{2.211}$$

where the vector $\mathbf{v}_n$ and $\sigma^2_{x,n|n-1}$ are the first column vector and the (1, 1) component of the one-step-ahead prediction covariance matrix $\mathbf{P}_{n|n-1}$, respectively.

www.EngineeringBooksPdf.com

2.8.2.2 TVAREX Model

In this section, the Kalman filter is implemented on a time-varying auto-regressive model with exogenous inputs, which is abbreviated as the TVAREX model [105]:

$$x_n = [\theta_{1,n-1}x_{n-1} + \theta_{2,n-1}x'_{n-1} + \theta_{3,n-1}\exp(-\alpha u_n - \beta|\phi_n|)]\exp(-\theta_{4,n-1}r_n) + f_n \tag{2.212}$$

This model is different from the previous one since the TVAR(p) model is an empirical model which does not consider the underlying physical mechanism that governs the variation of the daily averaged PM_{10} concentrations. On the contrary, the TVAREX model also takes into account of the meteorological factors which have significant impacts on the daily averaged PM_{10} concentrations. In this model the symbol x'_{n-1} denotes the hourly averaged PM_{10} concentration before midnight. It is used to reflect the initial condition of PM_{10} concentration on the next day.

The symbols u_n and $|\phi_n|$ denote the magnitude and the absolute angle of the resultant wind velocity vector. The resultant wind velocity vector is obtained by the vector sum of the hourly wind velocity vectors on the day of prediction. The magnitude u_n is associated with the dispersion condition of the nth day. It is small under the conditions of low wind speed, occurrence of reversal in the prevailing wind directions, or a combination of both. When the wind speed is low, the atmosphere becomes stagnant. The particulates generated from the local sources or transported from the upwind areas are trapped in the boundary layer and are hardly removed through advection. When there is a reversal of the prevailing wind directions, the particulate matters which are located downwind of Macao may be transported back due to the directional swing. This condition can also enhance the buildup of PM_{10} concentrations on the day of prediction. The absolute angle of the resultant wind velocity vector $|\phi_n|$ represents the dominant wind direction on the day of prediction. For example, the '0°' refers to the Geographic True North and an absolute angle of 30° includes the angle of +/− 30° from it. Therefore, the entire range of the wind direction is [0°,180°]. The absolute resultant angle indicates the type of the replenishing air masses being transported to Macao. As the angle increases from 0° to 180°, the dominant wind direction changes from northerly to southerly. The air pollution problem in Macao is severe since it is located at the southwest of the Pearl River Delta Region which is composed of several fast developing cities such as Guangzhou and Shenzhen. Particulate matters can be easily transported to Macao from its nearby upwind cities when the prevailing winds are blowing from the northerly directions. On the contrary, Macao is facing the South China Sea located to its South direction. When the winds are blowing from the South China Sea, Macao is less influenced by the transboundary air pollution from the cities located to its North. Therefore, the exponential term containing u_n and $|\phi_n|$ reflects the contribution from the local sources of Macao as well as the cities nearby. The coefficient θ_3 represents the source strength, while the coefficients α and β represent the decaying rate. It should be noted that the values of α and β are fixed and those values are specified through the optimization procedure. Assuming fixed values of α and β ensures that there is a unique most plausible value for each time-varying coefficient on a given day. In this study, the values of α and β are taken to be 2928.8 h/km and 297 deg^{-1}, respectively.

Besides the wind effect, precipitation also affects the PM_{10} concentration as the rainwater washes out part of the particulate matters from the atmosphere. Therefore, the exponential term containing the daily rainfall index r_n, which is defined as the product of the daily rainfall

www.EngineeringBooksPdf.com

amount and the duration of rainfall on the nth day, is used as a discounting factor on the PM_{10} concentration for a rainy day. The coefficient θ_4 controls the magnitude of reduction for a given rainfall index. In this study, the exponential relationship between the PM_{10} concentration and the meteorological variables is assumed since their relationship is considered nonlinear and a similar power law was adopted in previous studies [95, 234]. Finally, the term f represents the modeling error and it is modeled as a Gaussian i.i.d. process with zero mean and variance σ_f^2.

The state vector $\mathbf{X}_n$ is defined to contain the PM_{10} concentration and the unknown coefficients of the nth day:

$$\mathbf{X}_n = [x_n, \theta_{1,n}, \theta_{2,n}, \theta_{3,n}, \theta_{4,n}]^T \tag{2.213}$$

The measurement y_n is also expressed in terms of $\mathbf{X}_n$ with the same form as shown in Equation (2.200), but the observation matrix $\mathbf{L}_o$ becomes a row vector:

$$\mathbf{L}_o = [1, 0, 0, 0, 0] \tag{2.214}$$

Also, the process noise vector $\mathbf{F}_n$ is defined to contain the process noises f_n and $w_{l,n}$, $l = 1, 2, 3, 4$:

$$\mathbf{F}_n = [f_n, w_{1,n}, w_{2,n}, w_{3,n}, w_{4,n}]^T \tag{2.215}$$

It has zero mean and covariance matrix $\mathbf{\Sigma}_F$:

$$\mathbf{\Sigma}_F = diag(\sigma_f^2, \sigma_{w1}^2, \sigma_{w2}^2, \sigma_{w3}^2, \sigma_{w4}^2) \tag{2.216}$$

Again, the state space equation can be linearized to:

$$\mathbf{X}_n = \hat{\mathbf{A}}_{n-1|n-1}\mathbf{X}_{n-1} + \mathbf{B}\mathbf{F}_{n-1} + \hat{\mathbf{G}}_{n-1|n-1} \tag{2.217}$$

The matrices $\hat{\mathbf{A}}_{n-1|n-1}$ and $\mathbf{B}$ of the TVAREX vector model are:

$$\hat{\mathbf{A}}_{n-1|n-1} = \begin{bmatrix} \hat{\theta}_{1,n-1|n-1}e_{1,n} & \hat{x}_{n-1|n-1}e_{1,n} & \hat{x}'_{n-1|n-1}e_{1,n} & e_{1,n}e_{2,n} & -r_n e_{1,n}e_{3,n} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.218}$$

$$\mathbf{B} = \mathbf{I}_5 \tag{2.219}$$

www.EngineeringBooksPdf.com

where the variables $e_{1,n}$, $e_{2,n}$ and $e_{3,n}$ are given by:

$$e_{1,n} = \exp(-\hat{\theta}_{4,n-1|n-1} r_n) \tag{2.220}$$

$$e_{2,n} = \exp(-\alpha u_n - \beta|\phi_n|) \tag{2.221}$$

$$e_{3,n} = \hat{\theta}_{1,n-1|n-1}\hat{x}_{n-1|n-1} + \hat{\theta}_{2,n-1|n-1}\hat{x}'_{n-1|n-1} + \hat{\theta}_{3,n-1|n-1} e_{2,n} \tag{2.222}$$

The variable $e_{1,n}$ is the estimated discounting factor that represents the scavenging effect of rainfall on the predicted PM_{10} concentration for a given value of r_n. The variable $e_{2,n}$ represents the effects of local wind conditions on the predicted PM_{10} concentration per unit value of $\hat{\theta}_{3,n-1|n-1}$. The variable $e_{3,n}$ represents the combined contribution to the predicted PM_{10} concentration from its past history and the local wind conditions on the day of prediction when the effect of rainfall is neglected. The compensation vector $\hat{\mathbf{G}}_{n-1|n-1}$ is:

$$\hat{\mathbf{G}}_{n-1|n-1} = [e_{1,n}(\hat{\theta}_{4,n-1|n-1} r_n e_{3,n} - \hat{\theta}_{1,n-1|n-1}\hat{x}_{n-1|n-1}), 0, 0, 0, 0]^T \tag{2.223}$$

Then, the extended Kalman filter is operated by applying the predicting and filtering procedure for each day. By using Equation (2.217), the one-day-ahead prediction of the state vector on the nth day can be obtained:

$$\hat{\mathbf{X}}_{n|n-1} = [e_{1,n} e_{3,n}, \hat{\theta}_{1,n-1|n-1}, \hat{\theta}_{2,n-1|n-1}, \hat{\theta}_{3,n-1|n-1}, \hat{\theta}_{4,n-1|n-1}]^T \tag{2.224}$$

The covariance matrix of the predicted state vector can be calculated by:

$$\mathbf{P}_{n|n-1} = \hat{\mathbf{A}}_{n-1|n-1}\mathbf{P}_{n-1|n-1}\hat{\mathbf{A}}^T_{n-1|n-1} + \mathbf{B}\boldsymbol{\Sigma}_F\mathbf{B}^T \tag{2.225}$$

Finally, the filtered state vector is given by:

$$\hat{\mathbf{X}}_{n|n} = \mathbf{P}_{n|n}(\mathbf{P}^{-1}_{n|n-1}\hat{\mathbf{X}}_{n|n-1} + \sigma_\epsilon^{-2}\mathbf{L}_o^T y_n) \tag{2.226}$$

and the uncertainty of the filtered state is represented by the covariance matrix:

$$\mathbf{P}_{n|n} = \mathbf{P}_{n|n-1} - \frac{1}{\sigma_\epsilon^2 + \sigma^2_{x,n|n-1}}\mathbf{v}_n\mathbf{v}_n^T \tag{2.227}$$

where the vector $\mathbf{v}_n$ and $\sigma^2_{x,n|n-1}$ are the first column vector and the (1, 1) component of the one-step-ahead prediction covariance matrix $\mathbf{P}_{n|n-1}$, respectively.

2.8.2.3 Artificial Neural Network Based Prediction Model

The artificial neural network (ANN) based prediction model utilized in the present study is the multilayer perceptrons (MLPs). It is adopted as the benchmark to compare with the time-varying statistical models since it has been shown that the MLP architecture could approximate

www.EngineeringBooksPdf.com

any smooth, measurable function between the input and output vectors by a suitable combination of network parameters [108]. In addition, it has been demonstrated previously as a successful tool in the area of air quality prediction [83, 97, 107, 122, 184, 193, 206]. It consists of the input vector $\mathbf{X}$, the hidden layer with the hyperbolic tangent sigmoid transfer function, the output layer with the linear transfer function, and the network output which is the one-day-ahead prediction of the daily averaged PM_{10} concentration. The input vector $\mathbf{X}$ is defined as the column vector which contains the normalized inputs of the TVAREX model as follows:

$$\mathbf{X} = [\overline{x_{n-1}}, \overline{x'_{n-1}}, \overline{u_n}, \overline{|\phi_n|}, \overline{r_n}]^T \tag{2.228}$$

Each component in the input vector is obtained by normalizing the original variable to have a maximum and minimum equal to $+1$ and -1, respectively. These sets of variables are the same as those in the TVAREX model so comparison between the two models will be fair. When a given vector $\mathbf{X}$ is fed to the MLP model, the input vector to the hidden layer is calculated as shown below:

$$\boldsymbol{\eta}^{(1)} = \mathbf{W}^{(1)}\mathbf{X} + \mathbf{b}^{(1)} \tag{2.229}$$

The components of the matrix $\mathbf{W}^{(1)}$ and the vector $\mathbf{b}^{(1)}$ are called the weights and biases of the layer. Here the hidden layer is denoted as layer 1 and the output layer is denoted as layer 2. Then, the output of the hidden layer is calculated from the transfer function $f^{(1)}$, which is chosen to be the hyperbolic tangent sigmoid function in this study:

$$f^{(1)}(\eta_l^{(1)}) = \frac{\exp(\eta_l^{(1)}) - \exp(-\eta_l^{(1)})}{\exp(\eta_l^{(1)}) + \exp(-\eta_l^{(1)})} \tag{2.230}$$

Now the output of the hidden layer (layer 1) becomes the input to the output layer (layer 2). The transfer function used in the output layer is the linear function:

$$f^{(2)}(\boldsymbol{\eta}^{(2)}) = \boldsymbol{\eta}^{(2)} \tag{2.231}$$

where $\boldsymbol{\eta}^{(2)}$ denotes the input to the transfer function of the output layer. Finally, the one-day-ahead predicted daily averaged PM_{10} concentration on the nth day is treated as the output of the network. The data $\mathcal{D}$ includes the N pairs of input vector $\mathbf{X}$ and measured daily averaged PM_{10} concentration y. By tuning the network weights and biases in each layer with the data set $\mathcal{D}$, the network is able to approximate the relationship between the input vector $\mathbf{X}$ and the measurement y if the number of data points N is sufficiently large. This is referred to as the process of *training* or *learning*. In this study, the MLP is trained by using the Levenberg–Marquardt backpropagation algorithm. The parameter vector $\boldsymbol{\theta}$ contains the network weights and biases in the hidden layer and the output layer:

$$\begin{aligned}\boldsymbol{\theta} = [&W_{1,1}^{(1)}, W_{1,2}^{(1)}, W_{1,3}^{(1)}, W_{1,4}^{(1)}, W_{1,5}^{(1)}, b_1^{(1)}, \ldots, W_{s1,1}^{(1)}, W_{s1,2}^{(1)}, W_{s1,3}^{(1)}, W_{s1,4}^{(1)}, W_{s1,5}^{(1)}, b_{s1}^{(1)}, \\ &W_{1,1}^{(2)}, W_{1,2}^{(2)}, W_{1,3}^{(2)}, W_{1,4}^{(2)}, W_{1,s1}^{(2)}, b_1^{(2)}]^T\end{aligned} \tag{2.232}$$

www.EngineeringBooksPdf.com

so the total number of parameters is $7s1 + 1$. At the initial step, the parameter vector $\boldsymbol{\theta}_0$ is initialized with random weights and biases. Then, the error vector $\boldsymbol{\epsilon}$ is given by the difference between the measurements and the network predicted values:

$$\boldsymbol{\epsilon}(\boldsymbol{\theta}) = [y_1 - \hat{x}_{1|0}, y_2 - \hat{x}_{2|1}, \ldots, y_N - \hat{x}_{N|N-1}]^T \tag{2.233}$$

and the goodness-of-fit function is given by:

$$J_g(\boldsymbol{\theta}) = \frac{1}{N}||\boldsymbol{\epsilon}(\boldsymbol{\theta})||^2 \tag{2.234}$$

The parameter vector is updated recursively by using the following equation:

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} - [\mathcal{J}(\boldsymbol{\theta}_{k-1})^T \mathcal{J}(\boldsymbol{\theta}_{k-1}) + \mu_{k-1}\mathbf{I}_{7s1+1}]^{-1} \mathcal{J}(\boldsymbol{\theta}_{k-1})^T \boldsymbol{\epsilon}(\boldsymbol{\theta}_{k-1}) \tag{2.235}$$

where $\mathcal{J} = \partial\boldsymbol{\epsilon}/\partial\boldsymbol{\theta}$ denotes the Jacobian matrix with respect to the parameter vector $\boldsymbol{\theta}$. In the iterative process, the coefficient μ_0 is initialized with some arbitrary positive value at the beginning and it is updated according to the following rule:

$$\mu_{k+1} = \begin{cases} \gamma\mu_k & \text{if } J_g(\boldsymbol{\theta}_k) \geq J_g(\boldsymbol{\theta}_{k-1}) \\ \mu_k/\gamma & \text{if } J_g(\boldsymbol{\theta}_k) < J_g(\boldsymbol{\theta}_{k-1}) \end{cases} \tag{2.236}$$

In this study, the coefficients μ_0 and γ are taken to be 10^{-3} and 10, respectively. The parameter vector is kept updated until $J_g(\boldsymbol{\theta}_{k+1}) - J_g(\boldsymbol{\theta}_k)$ is smaller than a prescribed tolerance or the maximum number of iterations is reached.

2.8.3 Analysis with Monitoring Data

In this section the air quality prediction models are evaluated based on the daily averaged PM_{10} concentrations recorded at the ambient air quality monitoring station in Macao between 2001 and 2005. The station is located at the Taipa Grande Hill, which has an altitude of 159.2 m above sea level. Therefore, the measurements are representative of the general ambient air quality of Macao. To evaluate the performance of each model, some well-known performance indices are utilized and they are the root-mean-square error:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{x}_{n|n-1})^2} \tag{2.237}$$

www.EngineeringBooksPdf.com

the mean absolute percentage error:

$$MAPE = \frac{1}{N}\sum_{n=1}^{N}\frac{|y_n - \hat{x}_{n|n-1}|}{y_n} \times 100\% \tag{2.238}$$

the coefficient of determination:

$$r^2 = \left[\frac{\sum_{n=1}^{N}(y_n - \mu_y)(\hat{x}_{n|n-1} - \mu_x)}{\sqrt{\sum_{n=1}^{N}(y_n - \mu_y)^2}\sqrt{\sum_{n=1}^{N}(\hat{x}_{n|n-1} - \mu_x)^2}}\right]^2 \tag{2.239}$$

and the index of agreement:

$$IA = 1 - \frac{\sum_{n=1}^{N}(y_n - \hat{x}_{n|n-1})^2}{\sum_{n=1}^{N}(|y_n - \mu_y| + |\hat{x}_{n|n-1} - \mu_y|)^2} \tag{2.240}$$

where y_n, $\hat{x}_{n|n-1}$ and N are the measurement of the nth day, the one-day-ahead prediction of the nth day, and the number of total measured days, respectively. The symbols μ_y and μ_x denote the average of the measurements and predictions, respectively. These performance indices are widely used to indicate the performance of the air quality prediction models, for example [97, 107, 122, 184, 191, 193, 234]. In general, good predictive models associate with small values of $RMSE$ and $MAPE$, as well as r^2 and IA close to unity.

Since the data in 2001 and 2002 are used for training the artificial neural network model, the performance indices of the TVAR, TVAREX, and ANN models are calculated by using the data of the following three years. Table 2.6 shows the performance of the TVAR(p) models, $p = 1, 2, \ldots, 10$. It is noted that the RMSE decreases by 5.6% when the order of the TVAR model p increases from 1 to 10. However, the values of $MAPE$ and r^2 increases by 6.9% and 3.4%, respectively. In addition, the IA value is virtually the same for all values of the model order p. Since there is no consensus in the performance indices, it is found that increasing the value of p does not necessarily improve the performance of the TVAR model. Next, the TVAR models are compared to the TVAREX model and the ANN model.

Table 2.6 Performance of the TVAR models

p	$RMSE$ (μg m^{-3})	$MAPE$ (%)	r^2	IA
1	28.12	34.89	0.59	0.87
2	28.08	35.82	0.58	0.87
3	27.55	36.47	0.59	0.87
4	27.38	36.87	0.59	0.87
5	26.95	36.80	0.60	0.87
6	26.86	36.76	0.60	0.87
7	26.73	36.87	0.60	0.87
8	26.61	36.92	0.61	0.87
9	26.57	37.07	0.61	0.87
10	26.56	37.31	0.61	0.87

www.EngineeringBooksPdf.com

For the ANN-based prediction model, different number of neurons from one and ten are considered in the hidden layer. As a too simple model class introduces substantial modeling error and a unnecessarily too complicated model class may lead to over-fitting, the number of neurons in the hidden layer has to be carefully chosen so that a proper trade-off is achieved between the accuracy and the robustness. In addition, 100 sets of different initial parameter vector $\boldsymbol{\theta}_0$ are used in the network training for a given number of neurons. Each initial parameter vector is a column vector containing random weights and biases. Different initial weights and biases are used to train the network since it is noted in Equation (2.235) that different values of $\boldsymbol{\theta}_0$ may lead to different resultant weights and biases in the network optimization process. For some of the starting points $\boldsymbol{\theta}_0$, they may reach the local minimum during the optimization process. The 100 sets of initial parameter vector of $\boldsymbol{\theta}_0$ are generated to increase the chance of reaching the global minimum.

Figure 2.37 shows the statistics of the prediction error by ANNs with different numbers of neurons in the hidden layer. It is noted that the network performance becomes stabilized for four or more neurons. Kermanshahi (1999) proposed a rule of thumb to decide the number of hidden neurons as follows [137]:

$$s_1^\star = \frac{N_I + N_o}{2} + \lambda \tag{2.241}$$

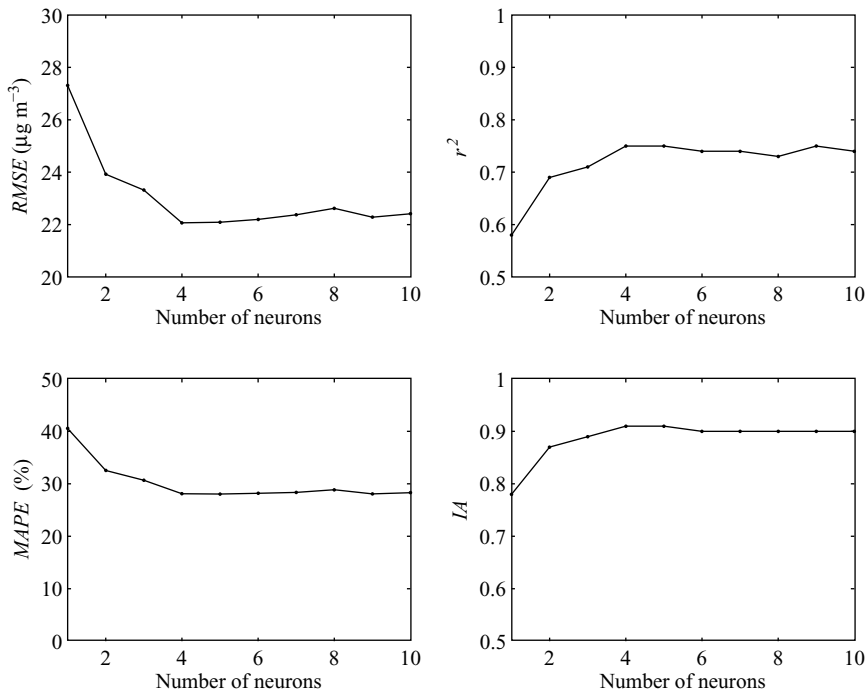

Figure 2.37 Prediction error with different number of neurons

www.EngineeringBooksPdf.com

Table 2.7 Performance of time-varying models and ANN model

Model class	$RMSE$ (μg m^{-3})	$MAPE$ (%)	r^2	IA
TVAR(1)	28.12	34.89	0.59	0.87
TVAREX	19.59	26.90	0.79	0.94
ANN	20.50	25.57	0.78	0.92

where N_I and N_o denote the number of elements in the input layer and the output layer, respectively. The value of λ depends on the user's judgement and it may be taken as one or two. In the present study, $N_I = 5$ and $N_o = 1$ so the optimal number of hidden neurons according to the Kermanshahi's rule of thumb is four or five. This echoes with the previous observation noted in Figure 2.37. Therefore, the network architecture with four neurons in the hidden layer is chosen here to compare with the time-varying models.

Table 2.7 shows the statistics of the prediction error by the time-varying models and the artificial neural network based model. Since the performance of the other TVAR models is similar to the TVAR(1) model, only the error statistics of the TVAR(1) model are shown here for comparison. The error statistics of the ANN prediction model are calculated by using the initial parameter vector which gives the smallest $RMSE$. By Table 2.7, it is noted that the error statistics of the TVAREX model and the ANN model are better than those of the TVAR(1) model. Furthermore, the error statistics of the TVAREX model are slightly better than those of the ANN model. The values of $RMSE$ and $MAPE$ of the TVAREX model are 30.3% and 22.9% less than those of the TVAR(1) model, respectively. In addition, the values of r^2 and IA of the TVAREX model are 33.9% and 8.1% larger than those of the TVAR(1) model, respectively. The $RMSE$ and $MAPE$ of the ANN model are 27.1% and 26.7% less than those of the TVAR(1) model, respectively. The values of r^2 and IA of the ANN model are 32.2% and 5.8% larger than those of the TVAR(1) model, respectively. The variations in the error statistics are consistent and these values confirm that better prediction models can be obtained if the meteorological factors are taken into account.

Figure 2.38 shows the time history of measured daily averaged PM_{10} concentrations and the predictions by the TVAR(1) model. The solid lines represent the measurements whereas the dashed lines represent the one-day-ahead predictions. It is noted that there is a time-delay problem associated with the TVAR(1) model, i.e., the trend of the predictions generally lags behind the trend of the measurements. The problem arises since the prediction is based solely on its own past history so no information can be obtained to predict the sudden development of high concentrations. Furthermore, those influencing factors such as the dispersion condition and the nature of replenishing air masses which can be continental or oceanic are treated as the process noises. Therefore, the process noise is large compared with the rms of the signal and this causes the predicted signal to be delayed.

Figure 2.39 shows the time history of the measured daily averaged PM_{10} concentration and the predictions by the TVAREX model. It is found that the time delay problem is generally resolved. Since the inclusion of the meteorological conditions enhances the effectiveness of the model, the process noise of the system is reduced and it leads to the improvement of the model performance. Finally, Figure 2.40 shows the time history of the measured daily averaged PM_{10} concentrations and the predictions by the ANN model. By comparing Figures 2.39 and 2.40, the TVAREX model is more efficient in capturing the pollution episode days, which correspond to the day with a daily average over 150 μg m^{-3}.

www.EngineeringBooksPdf.com

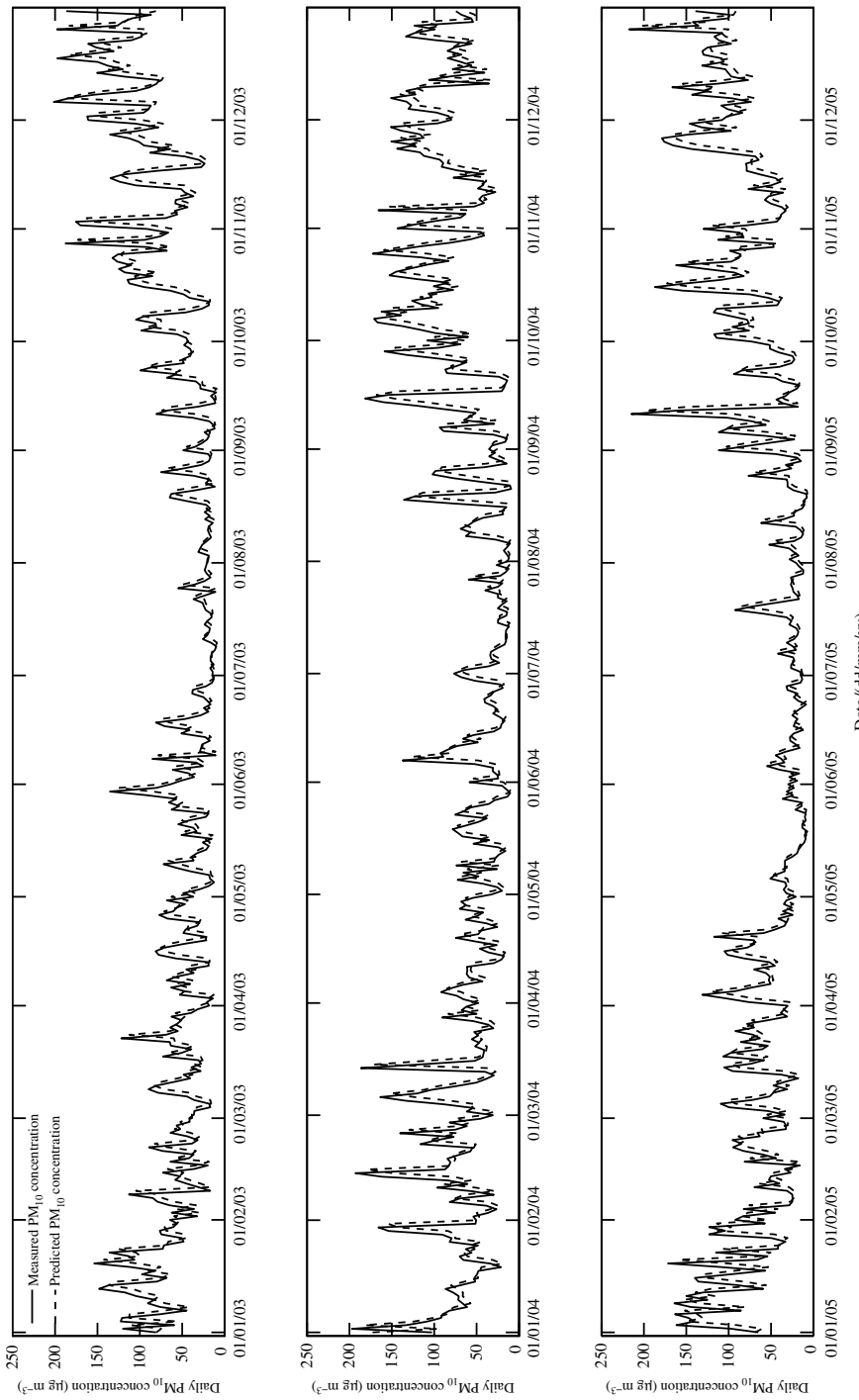

Figure 2.38 Measurement and prediction by the TVAR(1) model (2003–2005)

www.EngineeringBooksPdf.com

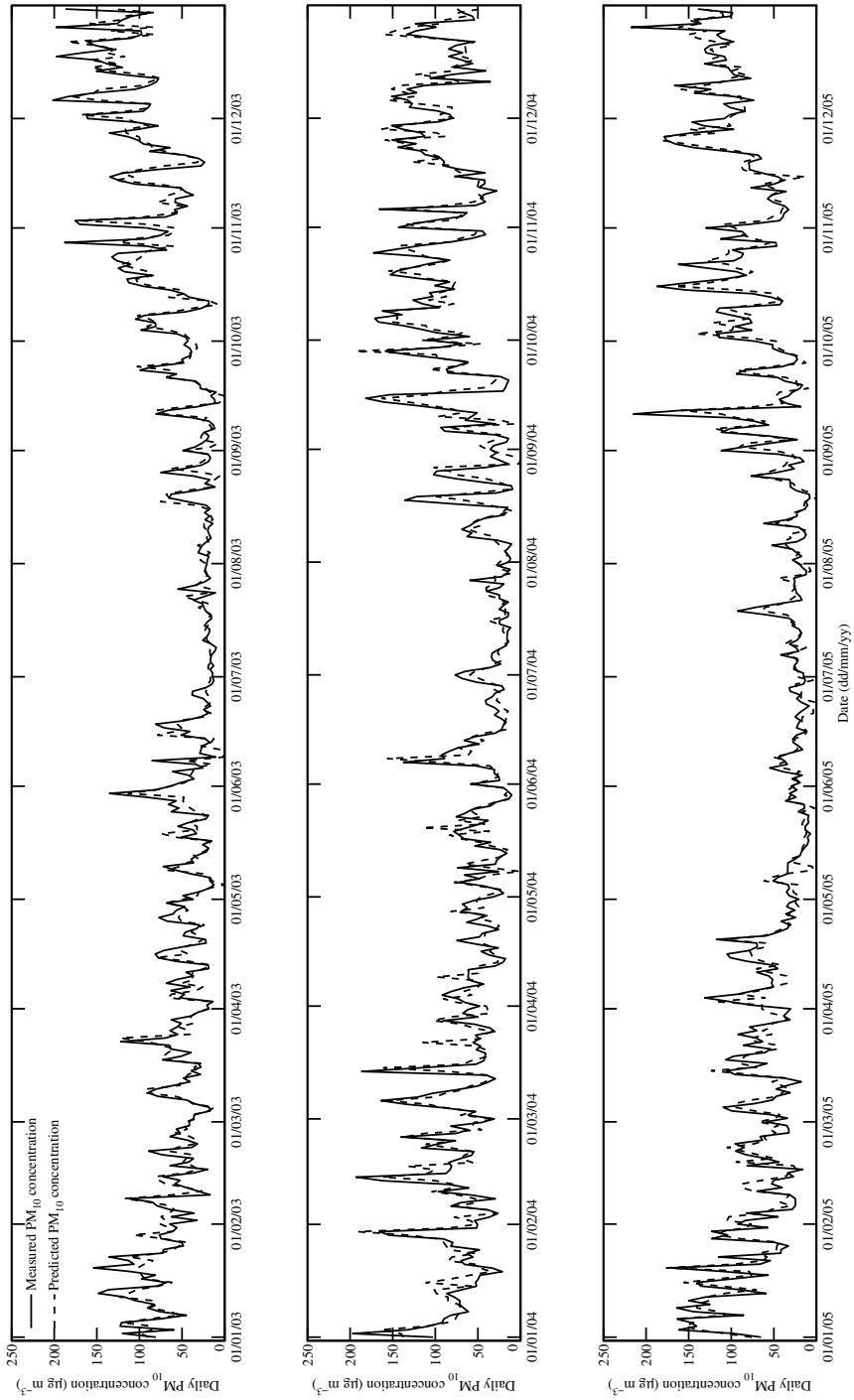

Figure 2.39 Measurement and prediction by the TVAREX model (2003–2005)

www.EngineeringBooksPdf.com

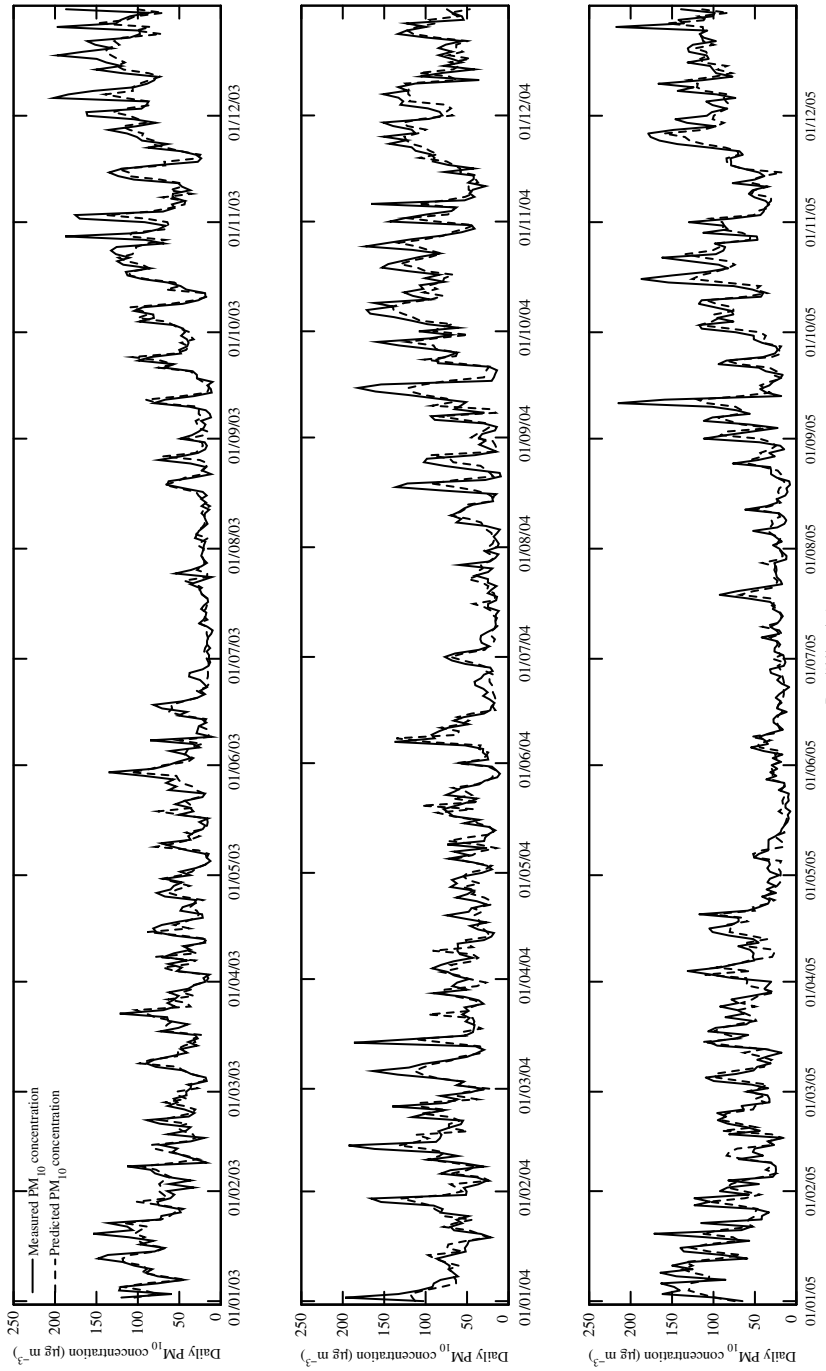

Figure 2.40 Measurement and prediction by the artificial neural network model (2003–2005)

www.EngineeringBooksPdf.com

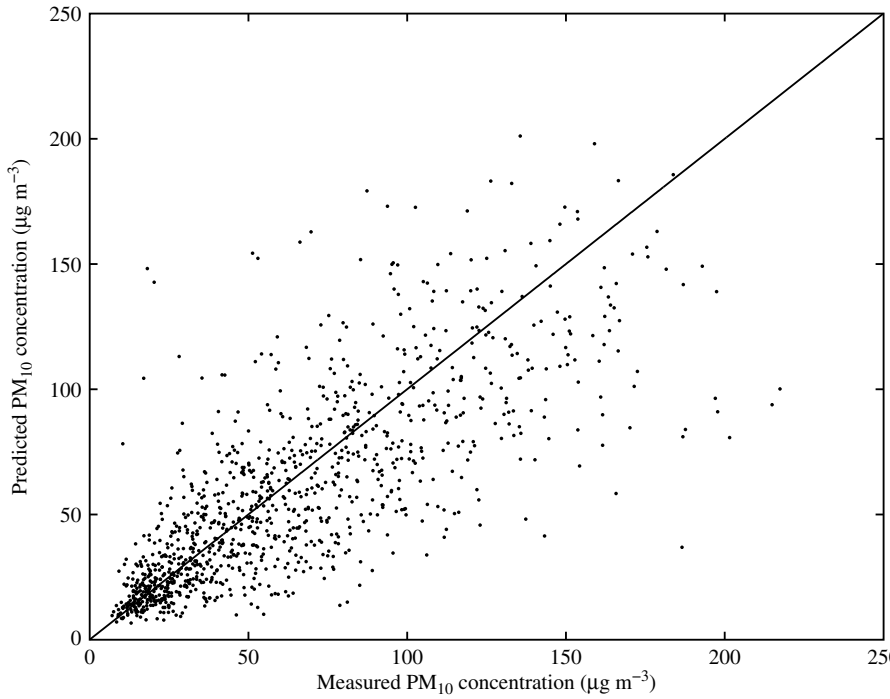

Figure 2.41 Predicted PM_{10} concentrations by the TVAR(1) model against its measurements

Figure 2.41 shows the scatter plot of the TVAR(1) predicted daily averaged PM_{10} concentration against its measurement. A 45° line is also drawn on the figure for reference. When a point falls on the 45° line, the predicted PM_{10} concentration is equal to its measurement. It is noted that a large portion of the points are lying close to the 45° line, indicating that the model results are acceptable. However, there are still a substantial portion of the points far away from the line. As mentioned above, the predictions generally lag behind the measurements. A large prediction error is expected when there is a large fluctuation in the daily averaged PM_{10} concentrations. Therefore, those points lying far above or below the 45° line are associated with the onset and retreat of the PM_{10} episode.

Figure 2.42 shows the scatter plot of the predicted daily averaged PM_{10} concentration by the TVAREX model against its measurements. Compared with the scatter plot of the results of the TVAR(1) model, it is found that the points become more concentrated around the 45° line. It echoes with the observation of the improvement in the time-delay problem as shown in Figure 2.41. Figure 2.43 shows the scatter plot of the predicted daily averaged PM_{10} concentrations by the ANN model against its measurements. It is noted that the measured PM_{10} concentrations are generally higher than the predicted PM_{10} concentrations when the predicted PM_{10} concentrations are above 100 $\mu g\ m^{-3}$. This is consistent with the observation made by comparing Figures 2.39 and 2.40.

www.EngineeringBooksPdf.com

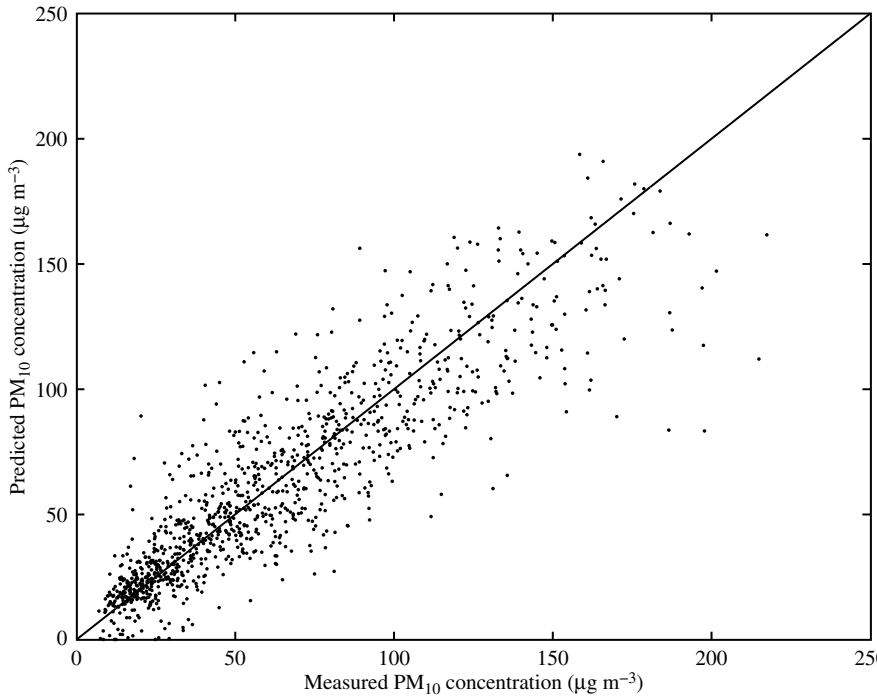

Figure 2.42 Predicted PM_{10} concentrations by the TVAREX model against its measurements

In order to further compare the performance of the TVAREX and ANN models in capturing the PM_{10} pollution episodes, the probability of detection (POD) is utilized:

$$POD \equiv P(\hat{x}_{n|n-1} \geq B | y_n \geq B) \tag{2.242}$$

It is the probability that the model produces correct warning signals when the daily averaged PM_{10} concentration is greater than or equal to a given threshold B. This is shown in Table 2.8 for different thresholds from 0 to 150 μg m^{-3}. The POD values of the two models are similar for thresholds below 100 μg m^{-3}. However, the POD values of the TVAREX model are higher than those of the ANN model when the threshold is over 100 μg m^{-3} so the TVAREX model is more efficient in capturing the PM_{10} episode days. Furthermore, by observing Figure 2.43, the points are all below the 45° line when the measured concentration is higher than 150 μg m^{-3}, implying failure of detection by the ANN model. However, the situation is much better for the TVAREX model as can be seen in Figure 2.42. Note that prediction of the episode days is a difficult task in general since the PM_{10} concentrations can build up rapidly by observing Figures 2.38 – 2.40.

However, the higher values of POD alone do not guarantee better performance of the TVAREX model. For example, a model that consistently over-predicts the PM_{10} concentration

www.EngineeringBooksPdf.com

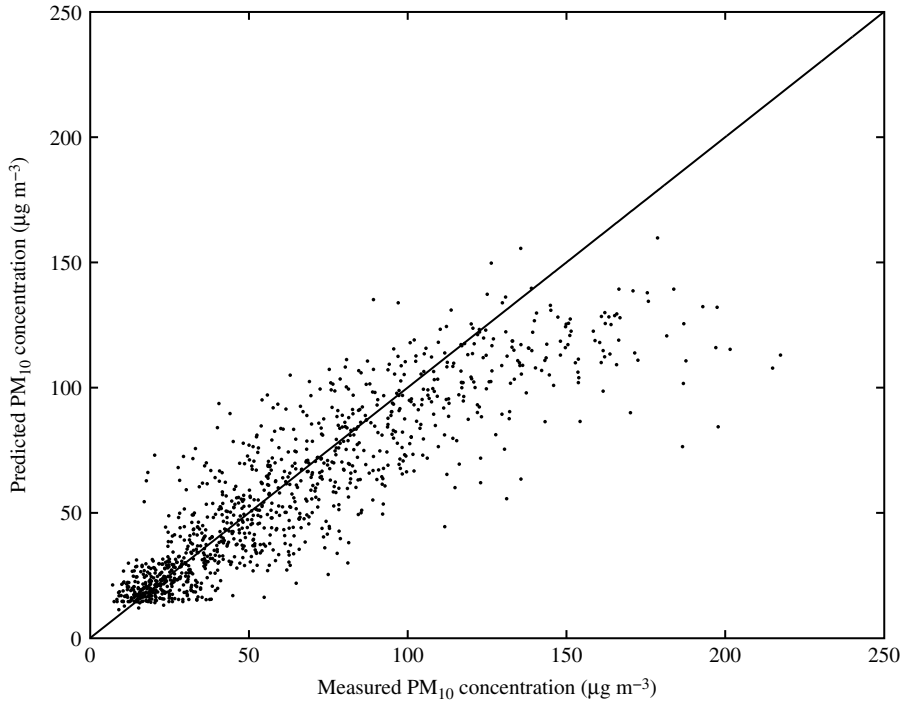

Figure 2.43 Predicted PM_{10} concentrations by the ANN model against its measurements

Table 2.8 Probabilities of detection by the TVAREX and ANN models

B	POD by the TVAREX model	POD by the ANN model
0	1.00	1.00
10	0.98	1.00
20	0.95	0.94
30	0.94	0.92
40	0.92	0.89
50	0.87	0.85
60	0.84	0.85
70	0.84	0.79
80	0.83	0.78
90	0.75	0.73
100	0.71	0.67
110	0.65	0.57
120	0.61	0.34
130	0.54	0.14
140	0.48	0.01
150	0.44	0.00

www.EngineeringBooksPdf.com

Table 2.9 Probabilities of false alarms by the TVAREX and ANN models

B	PFA by the TVAREX model	PFA by the ANN model
0	0.00	0.00
10	0.01	0.01
20	0.06	0.06
30	0.07	0.05
40	0.09	0.08
50	0.11	0.11
60	0.16	0.14
70	0.16	0.13
80	0.17	0.14
90	0.19	0.17
100	0.23	0.16
110	0.25	0.14
120	0.24	0.12
130	0.29	0.32
140	0.37	0.67
150	0.42	0.50

has high values of the POD but it will provide a substantial amount of false alarms. In order to investigate this possibility of Type I error, the probability of false alarm (PFA) is utilized:

$$PFA \equiv P(y_n < B|\hat{x}_{n|n-1} \geq B) \tag{2.243}$$

It is the probability that the measured daily averaged PM_{10} concentration is below the threshold B on a predicted warning day and it is shown in Table 2.9 for different thresholds from 0 to 150 $\mu g\, m^{-3}$. Again, the PFA values of the two models are similar when the threshold is below 100 $\mu g\ m^{-3}$. However, the PFA values of the ANN model are lower those that of the TVAREX model when the threshold is between 100 and 125 $\mu g\ m^{-3}$. By also considering the POD values, it is difficult to judge which model gives better result between 100 and 125 $\mu g\ m^{-3}$. However, when the threshold is over 125 $\mu g\ m^{-3}$, the PFA values of the TVAREX model are lower than those of the ANN model. Therefore, the TVAREX model is more efficient in capturing the PM_{10} pollution episodes without giving too many false alarms. The problem of the ANN prediction model arises here since the network is trained in an offline manner, i.e., the parameter vector is kept fixed in the prediction phase. However, the actual system may be changing from time to time. Therefore, it is suspected that the input–output relationship of the model may have changed in the testing years and this causes the underestimation of the daily averaged PM_{10} concentrations. On the contrary, the TVAREX model is adaptive, i.e., the model parameters are updated whenever a new measurement of the PM_{10} concentration is obtained. The possible changes in the input–output relationship may be adapted by changing the model parameters through Equation (2.208). Therefore, the Kalman filter has the advantage that the model is retuned daily.

www.EngineeringBooksPdf.com

2.8.4 Conclusion

In this study, the time-varying models and the ANN model were applied to forecast the daily averaged PM_{10} concentrations of Macao between 2003 and 2005. The TVAR models and the TVAREX model were implemented with the Kalman filter and it was found that the TVAREX model and the ANN model are superior to the TVAR models based on the judgment of the error statistics and the agreement in the trends between and the predictions and the measurements. Since information of the meteorological conditions on the day of prediction reflects the dispersion conditions and provides a general description of the sources of PM_{10} pollution affecting Macao, the extra inputs of the TVAREX model and the ANN model enhance the efficiency of the models and resolve the time delay problem associated with the TVAR models. It is concluded that the past history of the daily averaged PM_{10} concentration is not sufficient to capture its behavior in the future for the case of Macao. Although the TVAREX model is comparatively better than the TVAR models, it is still preliminary and can be improved by further investigating the mechanism which governs the accumulation and removal of the PM_{10} concentrations [45]. This application demonstrates an important point that an underlying understanding of the physical problem is the most important factor for modeling. The empirical TVAR model is attractive in the sense that it does not require much knowledge on the dynamics of the particulate motion. However, the efficiency of this model is not as satisfactory as the TVAREX model. Moreover, it must be noted that by using a higher order of the TVAR model, the fitting error could be reduced but it does not necessarily reduce the prediction error.

In addition to the model class modification, further improvements can be achieved by adjusting the process noise and measurement noise parameters which are demonstrated to affect the performance of the Kalman filter [288]. By comparing the prediction results of the Kalman filter based TVAREX model with the ANN model, it is found that the error statistics of the TVAREX model are slightly better than the ANN model. In addition, the TVAREX model is more efficient in capturing the PM_{10} pollution episodes. Therefore, it is concluded that the TVAREX model in conjunction with the Kalman filter is applicable to predict the daily averaged PM_{10} concentration and the algorithm has an appealing advantage that the model parameters can be updated when new measurements are acquired.

www.EngineeringBooksPdf.com

3

Bayesian Spectral Density Approach

Keywords: ambient vibration; correlation function; Duffing oscillator; hydraulic jump; information entropy; modal identification; optimal sensor placement; spectral density; structural health monitoring; Wishart distribution

3.1 Modal and Model Updating of Dynamical Systems

The problem of *parametric identification* for mathematical models using input–output or output-only dynamic measurements has received much attention over the years. One important special case is modal identification, in which the parameters for identification are the small-amplitude modal frequencies, damping ratios, mode shapes and modal participation factors of the lower modes of the dynamical system. In other words, the model class in modal identification is the class of linear modal models. Many time-domain and frequency-domain methodologies have been formulated for input excitation and output response measurements [24, 48, 75, 81, 187].

Much attention has also been devoted to modal identification without measuring the input time history. In particular, a lot of effort has been dedicated to the case of free vibration (or impulse response) and to the case of ambient vibration. In the former case, often time-domain methods based on auto-regressive moving average (ARMA) models are employed, using least squares as the core ingredient in their formulations. However, it was found that the least-squares method yields biased estimates [76]. A number of methods have been developed to eliminate this bias, including the instrumental matrix with delayed observations method [76], the correlation fit method [275], the double least-squares method [114, 202] and the total least-squares method [92]. A detailed comparison of these methods can be found in Cooper [61].

Another important practical category is the ambient vibration survey (AVS). It has attracted much interest because it offers a means of obtaining dynamic data in an economical and efficient manner, without requiring the setup of special dynamic experiments (e.g., actuators) which are usually costly, time consuming, and often obtrusive. In AVS, the naturally occurring vibration

Bayesian Methods for Structural Dynamics and Civil Engineering Ka-Veng Yuen
© 2010 John Wiley & Sons (Asia) Pte Ltd

www.EngineeringBooksPdf.com

of the structure is measured under wind, traffic, and micro-tremors, etc. Then, a system identification technique is used to identify the small-amplitude modal frequencies, damping ratios and mode shapes of the lower modes of the structure. The assumption usually made is that the input excitation is a broad-band stochastic process adequately modeled as stationary Gaussian white noise. A number of time-domain methods have been developed to tackle this problem. One example is the random decrement (RD or 'Randomdec') technique [7]. The theoretical RD functions were shown to be a linear combination of the correlation functions and their derivatives [266]. Then, the modal parameters can be obtained by fitting the experimental RD functions with properly selected triggering conditions. Another example is the instrumental variable method [280]. Several other methods are based on fitting the correlation functions using least-squares type of approach [20, 21, 139]. Different ARMA model based least-squares type of methods have also been proposed, e.g., the two-stage least-squares method [60, 85, 86, 207]. Another important type of methods is the prediction error methods [5, 94, 159, 245] that minimize the optimally selected one-step-ahead output prediction error. Beck (1978) recognized the possible usage of the Kalman filter for model identification [15]. Hoshiya and Saito (1984) investigated the applications of the extended Kalman filter to estimate the dynamic properties, such as modal frequencies, modal damping coefficients and participation factors, of linear multi-degree-of-freedom systems [110]. Since then, many methods were proposed as its evolution for linear and nonlinear dynamical systems [94, 109, 127, 145, 212, 236, 245]. Other nonlinear system identification methods based on advanced statistical tools, such as wavelets and higher-order spectra, have also been investigated for the case of known input [141, 299] and for the case of unknown input [222].

Besides the aforementioned time-domain approaches, many frequency-domain methods have also been developed and widely used. Examples are the complex curve fitting method [153], the maximum entropy method [4, 263], the pole/zero assignment technique [271], the simultaneous frequency-domain approach [62], the rational fraction polynomial approach [219], the orthogonal polynomial approach [264], the polyreference frequency-domain approach [73], the multi-reference simultaneous frequency-domain approach [64] and the best-fit reciprocal vectors method [173].

Results of modal/model identification are usually restricted to the *optimal values* of the uncertain parameters. However, there is additional information related to the uncertainty associated with the parameter estimates and it is valuable for further processing. For example, in the case where the identified modal parameters are used to update the theoretical finite-element model of a structure, the updating procedure involves the minimization of a positive definite quadratic function of the differences between the theoretical and the experimental modal parameters. The weighting matrix in this goodness-of-fit function should reflect the uncertainty in the values of the identified modal parameters so it can be chosen as the inverse of the covariance matrix of these parameters. In practice this covariance matrix is usually estimated by computing the statistics of the optimal estimates of the modal parameters from several sets of ambient data. However, this estimation is unreliable unless the number of data sets is large. Recent interest has arisen to determine the uncertainty of the identified parameters using the Bayesian probabilistic approach. The parametric uncertainty can be quantified in the form of probability distribution in Bayesian inference [31]. In Beck [16] and Beck and Katafygiotis [19], the Bayesian probabilistic system identification framework was presented for measured input–output data. In Yuen and Katafygiotis [291], a method based on the state-space model was presented to consider explicitly the measurement noise in both input and output measurements.

www.EngineeringBooksPdf.com

In this chapter, the Bayesian spectral density approach, which is a frequency-domain approach, for modal/model updating using wide-band response data is presented. It utilizes the statistical properties of the spectral density estimator to obtain not only the optimal values of model parameters but also their associated uncertainty by means of the updated probability distribution of the uncertain parameters. Uncertainty quantification is important for many applications, such as damage detection and reliability analysis.

In the next section, standard random vibration analysis for stationary response of linear single-degree-of-freedom systems and multi-degree-of-freedom systems is reviewed. In Section 3.3, the method is presented first for the case of single-channel output measurements. The spectral density estimator follows the Chi-square distribution. Then, the method is presented for multiple-channel output measurements. The spectral density matrix estimator follows the central complex Wishart distribution. Then, by using the Bayesian probabilistic framework, the updated PDF of the model parameters is obtained based on the above statistics of the spectral density estimator. In Section 3.4, the approximations used in this method are verified numerically. The aliasing and leakage effects are also demonstrated. In Section 3.5, an information entropy based method is presented for optimal sensor placement for structural systems. Information entropy is a single measure of the uncertainty of multiple random variables. The optimal sensor configuration is obtained by minimizing this measure with respect to all possible ones. Then, the Bayesian spectral density approach is illustrated by applications in three different areas. In Section 3.6, the nonlinear Duffing oscillator is considered and it turns out that using one set of response measurement gives an unidentifiable case. However, by using two sets of response measurements with different excitation levels, the identification problem becomes globally identifiable. The Bayesian framework is important in this case since it weighs the two sets of data naturally without applying any adhoc weightings. In Section 3.7, applications to an existing 22-story building is demonstrated. Acceleration time histories were obtained for the complete duration of two typhoons with signal no. 8. The modal properties of the building are updated and the structural behavior is discussed. Finally, in Section 3.8, the method is used to analyze hydraulic jump measurements to illustrate the application to hydraulics problems. Three types of hydraulic jumps are investigated, namely the undular jumps, weak jumps and oscillating jumps. It turns out that the weak jumps and oscillating jumps exhibit oscillatory behavior and the apparent frequency depends on the dimensionless Froude number.

3.2 Random Vibration Analysis

3.2.1 Single-degree-of-freedom Systems

Consider a linear single-degree-of-freedom (SDOF) system with the equation of motion:

$$\ddot{x} + 2\zeta\Omega\dot{x} + \Omega^2 x = f(t) \tag{3.1}$$

where Ω and ζ are the natural frequency and damping ratio of the oscillator, respectively. The unmeasured input f is modeled by zero-mean Gaussian white noise with spectral intensity:

$$S_f(\omega) = S_{f0} \tag{3.2}$$

The flat spectrum indicates equal power of the signal at all frequencies. It is well known that for given model parameters the response x is a zero-mean Gaussian random process. The

www.EngineeringBooksPdf.com

auto-correlation function of a stochastic process can be defined as follows:

$$R_x(t, t+\tau) \equiv E[x(t)x(t+\tau)] \tag{3.3}$$

In the case of stationary response, the auto-correlation function of the stochastic process x governed by Equation (3.1) depends only on the time difference τ and it is given by [161, 249]:

$$R_x(\tau) = \frac{\pi S_{f0}}{2\zeta\Omega^3} e^{-\zeta\Omega|\tau|} \left[\cos(\Omega_d \tau) + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin(\Omega_d |\tau|)\right] \tag{3.4}$$

where $\Omega_d = \Omega\sqrt{1-\zeta^2}$ is the damped natural frequency of the oscillator. The power spectral density function is given by the Fourier transform of the auto-correlation function and it is given by [161, 249]:

$$S_x(\omega) = \frac{S_{f0}}{(\Omega^2 - \omega^2)^2 + (2\zeta\Omega\omega)^2} \tag{3.5}$$

It indicates the energy distribution of the stochastic process at different frequencies. For linear dynamical systems, the auto-correlation function provides sufficient information to construct the joint PDF of the response at different time instances since they are Gaussian with zero mean.

3.2.2 Multi-degree-of-freedom Systems

Consider a linear system with N_d degrees of freedom (DOFs) and its equation of motion of the generalized coordinates $\mathbf{x}$:

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{T}_0\mathbf{F}(t) \tag{3.6}$$

where $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$ are the mass, damping and stiffness matrix, respectively, $\mathbf{T}_0 \in \mathbb{R}^{N_d \times N_F}$ is a force distributing matrix and $\mathbf{F}$ is modeled as zero-mean Gaussian white noise vector process with the spectral intensity matrix:

$$\mathbf{S}_F(\omega) = \mathbf{S}_{F0} \tag{3.7}$$

Using modal analysis, the linear transformation between the generalized coordinates and the modal coordinates is given by:

$$\mathbf{x}(t) = \mathbf{\Phi} \cdot \mathbf{q}(t) \tag{3.8}$$

where $\mathbf{q}(t) = [q^{(1)}(t), q^{(2)}(t), \ldots, q^{(N_d)}(t)]^T$ is the modal coordinate vector, and $\mathbf{\Phi} = [\boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}^{(2)}, \ldots, \boldsymbol{\phi}^{(N_d)}]$ is the modal matrix that satisfies the generalized eigenvalue problem:

$$\mathbf{M}\mathbf{\Phi}\mathbf{\Lambda} = \mathbf{K}\mathbf{\Phi} \tag{3.9}$$

The diagonal matrix $\mathbf{\Lambda}$ contains the eigenvalues. The mth column vector $\boldsymbol{\phi}^{(m)}$ in the modal matrix associated with the mth lowest eigenvalue is called the *mode shape* of the mth mode of the dynamical system. Here, they are normalized such that:

$$\phi_{l_m}^{(m)} = 1, \qquad m = 1, 2, \ldots, N_d \tag{3.10}$$

www.EngineeringBooksPdf.com

where l_m is an arbitrarily selected measured DOF provided that it is not a node of the mth mode. The reason for not adopting mass normalization is that the mass matrix may be unknown in the identification process.

Then, the uncoupled modal equations of motion are obtained:

$$\ddot{q}^{(m)}(t) + 2\zeta^{(m)}\Omega^{(m)}\dot{q}^{(m)}(t) + {\Omega^{(m)}}^2 q^{(m)}(t) = f^{(m)}(t), \qquad m = 1, 2, \ldots, N_d \tag{3.11}$$

where the modal forcing vector $\mathbf{f}(t) = [f^{(1)}(t), f^{(2)}(t), \ldots, f^{(N_d)}(t)]^T$ is given by:

$$\mathbf{f}(t) = (\mathbf{M\Phi})^{-1}\mathbf{T}_0\mathbf{F}(t) \tag{3.12}$$

so it is also a zero-mean Gaussian white noise vector process with the spectral intensity matrix:

$$\mathbf{S}_f(\omega) = \mathbf{S}_{f0} = (\mathbf{M\Phi})^{-1}\mathbf{T}_0\mathbf{S}_{F0}\mathbf{T}_0^T(\mathbf{M\Phi})^{-T} \tag{3.13}$$

For a given set of model parameters, the response $\mathbf{x}$ is a zero-mean Gaussian process and the (l, l') element of its power spectral density matrix function $\mathbf{S}_x$ is given by [249]:

$$S_x^{(l,l')}(\omega) = \sum_{m=1}^{N_d}\sum_{m'=1}^{N_d}\phi_l^{(m)}\phi_{l'}^{(m')}\frac{S_{f0}^{(m,m')}}{[({\Omega^{(m)}}^2 - \omega^2) + 2i\omega\Omega^{(m)}\zeta^{(m)}][({\Omega^{(m')}}^2 - \omega^2) - 2i\omega\Omega^{(m')}\zeta^{(m')}]} \tag{3.14}$$

Furthermore, the correlation matrix function $\mathbf{R}_x$ for $\mathbf{x}$ has the (l, l') element [249]:

$$R_x^{(l,l')}(\tau) = \lim_{t\to\infty}\sum_{m=1}^{N_m}\sum_{m'=1}^{N_m}\phi_l^{(m)}\phi_{l'}^{(m')}\int_0^t h^{(m)}(t-u)h^{(m')}(t+\tau-u)\mathrm{d}u \tag{3.15}$$

where $h^{(m)}$ is the unit impulse response function of the mth mode:

$$h^{(m)}(\tau) = \begin{cases} \frac{1}{\Omega_d^{(m)}}\mathrm{e}^{-\zeta\Omega^{(m)}\tau}\sin(\Omega_d^{(m)}\tau) & \text{if } \tau \geq 0 \\ 0 & \text{if } \tau < 0 \end{cases} \tag{3.16}$$

The zero value of $h^{(m)}(\tau)$ for negative τ is due to the causality of physical systems. The underlying meaning is that any input should not generate response in the past. Furthermore, the correlation matrix function $\mathbf{R}_q$ for the modal coordinates $\mathbf{q}$ has (m, m') element between

www.EngineeringBooksPdf.com

the mth and m'th mode:

$$
\begin{aligned}
R_q^{(m,m')}(\tau) &= \lim_{t\to\infty}\int_0^t h^{(m)}(t-u)h^{(m')}(t+\tau-u)\mathrm{d}u \\
&= \frac{\pi S_{f0}^{(m,m')}}{\Omega_d^{(m)}\Omega_d^{(m')}}\exp\left(-\zeta^{(m)}\Omega^{(m)}\tau\right)\Bigg[\frac{\zeta^{(m)}\Omega^{(m)}+\zeta^{(m')}\Omega^{(m')}}{(\zeta^{(m)}\Omega^{(m)}+\zeta^{(m')}\Omega^{(m')})^2+(\Omega_d^{(m')}-\Omega_d^{(m)})^2}\cos\left(\Omega_d^{(m)}\tau\right) \\
&\quad + \frac{\Omega_d^{(m')}-\Omega_d^{(m)}}{(\zeta^{(m)}\Omega^{(m)}+\zeta^{(m')}\Omega^{(m')})^2+(\Omega_d^{(m')}-\Omega_d^{(m)})^2}\sin\left(\Omega_d^{(m)}\tau\right) \\
&\quad - \frac{\zeta^{(m)}\Omega^{(m)}+\zeta^{(m')}\Omega^{(m')}}{(\zeta^{(m)}\Omega^{(m)}+\zeta^{(m')}\Omega^{(m')})^2+(\Omega_d^{(m')}+\Omega_d^{(m)})^2}\cos\left(\Omega_d^{(m)}\tau\right) \\
&\quad + \frac{\Omega_d^{(m')}+\Omega_d^{(m)}}{(\zeta^{(m)}\Omega^{(m)}+\zeta^{(m')}\Omega^{(m')})^2+(\Omega_d^{(m')}+\Omega_d^{(m)})^2}\sin\left(\Omega_d^{(m)}\tau\right)\Bigg], \qquad \text{for } \tau \geq 0
\end{aligned}
\tag{3.17}
$$

For negative τ, it can be computed by the following relationship:

$$
R_q^{(m,m')}(\tau) = R_q^{(m',m)}(-\tau) \tag{3.18}
$$

3.3 Bayesian Spectral Density Approach

Use $\boldsymbol{\theta}$ to denote the uncertain parameter vector that determines the dynamical model within a prescribed class of models $\mathcal{C}$. These parameters include:

1. The structural parameters that determine the model matrices $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$.
   In the special case of modal updating, assume that only the lowest N_m modes contribute significantly to the model response and only the modal parameters of these modes are to be identified. Then, the uncertain structural parameters include the modal frequencies $\Omega^{(m)}$, the modal damping ratios $\zeta^{(m)}$, and the components of the measured DOFs of the mode shapes $\boldsymbol{\phi}^{(m)}$, except the ones which are used for normalization, $m = 1, 2, \ldots, N_m$; thus, there are a total number of $N_m(N_o + 1)$ unknown modal parameters in modal identification.
2. The forcing parameters that define the spectral intensity matrix $\mathbf{S}_{F0}$.
   In the special case of modal identification, they are the elements of the upper right triangle (diagonal inclusive) of the $N_m \times N_m$ submatrix of $\mathbf{S}_{f0}$ corresponding to the lowest N_m modes. Symmetry defines the lower triangle.
3. The elements of the upper right triangle (diagonal inclusive) of $\boldsymbol{\Sigma}_\epsilon$, or equivalently $\mathbf{S}_{\epsilon 0}$. Again, symmetry defines the lower triangle of this matrix.

For example, in the case of an SDOF system, $\boldsymbol{\theta} = [\Omega, \zeta, S_{f0}, S_{\epsilon 0}]^T$ is the parameter vector for identification. Recall that the scaling of each mode shape is chosen such that one component of a measured DOF is equal to unity. However, this scaling is arbitrary, and the mode shapes can be identified only up to a constant scaling factor. A different mode shape normalization

www.EngineeringBooksPdf.com

will cause all identified components of the mth mode shape to be scaled by the same constant α_m, and at the same time the values of the elements $S_{f0}^{(m,m')}$ and $S_{f0}^{(m',m)}$ will be scaled by α_m^{-2} if $m = m'$ and α_m^{-1} if $m \neq m'$.

3.3.1 Formulation for Single-channel Output Measurements

This section considers the response measurement of a single-degree-of-freedom system or a single-channel measurement of a multi-degree-of-freedom system. Discrete data is sampled with a time step Δt and y_n denotes the measured response at time $t = n\Delta t$. The measurement is different from the model response:

$$y_n = x(n\Delta t) + \epsilon_n, \qquad n = 0, 1, \ldots, N-1 \tag{3.19}$$

Here, the stochastic process x represents the response of a single-degree-of-freedom (SDOF) system or the response of a particular degree of freedom of a multi-degree-of-freedom (MDOF) system. The prediction error is modeled as a zero-mean discrete (band-limited) white noise process ϵ with variance σ_ϵ^2 and spectral intensity:

$$S_\epsilon(\omega) = S_{\epsilon 0} = \frac{\Delta t}{2\pi}\sigma_\epsilon^2, \ \omega \in \left[-\frac{\pi}{\Delta t}, \frac{\pi}{\Delta t}\right] \tag{3.20}$$

and it is assumed to be statistically independent to the stochastic response x.

3.3.1.1 Spectral Density Estimator

Consider the scaled discrete Fourier transform of the stochastic process x:

$$\mathcal{X}(\omega_k) \equiv \sqrt{\frac{\Delta t}{2\pi N}} \sum_{n=0}^{N-1} x(n\Delta t) \exp(-in\omega_k \Delta t) \tag{3.21}$$

where $\omega_k = k\Delta\omega$, $k = 0, 1, \ldots, N_{nqy}$ with $N_{nqy} = \mathrm{INT}(\frac{N}{2})$, $\Delta\omega = \frac{2\pi}{T}$, $T = N\Delta t$, and INT takes the integer part of a real number. The Nyquist frequency is defined as:

$$\omega_{nyq} = N_{nyq}\Delta\omega \tag{3.22}$$

which is half of the sampling frequency and is the upper bound of the frequency in the spectrum. Then, the discrete estimator of the spectral density function S_x is introduced:

$$S_{x,N}(\omega_k) \equiv |\mathcal{X}(\omega_k)|^2 = \frac{\Delta t}{2\pi N}\left|\sum_{n=0}^{N-1} x(n\Delta t)\exp(-in\omega_k\Delta t)\right|^2 \tag{3.23}$$

where $|.|$ takes the modulus of a complex variable. In a similar fashion to Equation (3.21), the discrete Fourier transform for the stochastic process y can be defined:

$$\mathcal{Y}(\omega_k) = \sqrt{\frac{\Delta t}{2\pi N}} \sum_{n=0}^{N-1} y_n \exp(-in\omega_k \Delta t) \tag{3.24}$$

www.EngineeringBooksPdf.com

Then, the spectral density estimator of y can be computed in analogy to Equation (3.23) as follows:

$$S_{y,N}(\omega_k) = |\mathcal{Y}(\omega_k)|^2 = \frac{\Delta t}{2\pi N}\left|\sum_{n=0}^{N-1} y_n \exp\left(-in\omega_k \Delta t\right)\right|^2 \tag{3.25}$$

3.3.1.2 Statistical Properties of the Spectral Density Estimator

In this section, the statistical properties of the spectral density estimator are investigated in order to construct the likelihood function of the data. Firstly, it can be shown that the spectral density estimator is asymptotically unbiased:

$$\lim_{\substack{T\to\infty \\ \Delta t\to 0^+}} E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] = S_x(\omega_k) + S_{\epsilon 0} \tag{3.26}$$

where $E[.]$ denotes the mathematical expectation. However, this estimator is biased for finite T or finite Δt so Equation (3.26) does not hold without taking the limit. For a given set of model parameters, taking expectation of Equation (3.25) yields:

$$\begin{aligned} E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] &= E[S_{x,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] + S_{\epsilon 0} \\ &= \frac{\Delta t}{2\pi N}\sum_{n,n'=0}^{N-1} e^{i(n'-n)\omega_k \Delta t} R_x[(n'-n)\Delta t] + S_{\epsilon 0} \end{aligned} \tag{3.27}$$

Grouping the terms with the same value of $|n' - n|$, the expected value of the spectral density estimator can be expressed as follows:

$$E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] = \frac{\Delta t}{2\pi N}\left[NR_x(0) + 2\sum_{n=1}^{N-1}(N-n)R_x(n\Delta t)\cos\left(n\omega_k \Delta t\right)\right] + S_{\epsilon 0} \tag{3.28}$$

This can be calculated using the function 'fft' in MATLAB® [171]:

```
an = [N/2 [N - 1 : -1 : 1]];
Nnyq = fix(N/2);
ESx = dt/pi/N * real(fft(an. * R));
ESx = EStemp(1 : Nnyq);
ESy = ESx + Se0;
% R is an array that includes Rx(0), Rx(Δt), ..., Rx((N - 1)Δt).
```

To investigate the statistical properties of the spectral density estimator, it can be rewritten in the following form by using Equation (3.25):

$$S_{y,N}(\omega_k) = \xi_c(\omega_k)^2 + \xi_s(\omega_k)^2 \tag{3.29}$$

www.EngineeringBooksPdf.com

where ξ_c and ξ_s are the scaled Fourier cosine and sine functions:

$$\begin{aligned}\xi_c(\omega_k) &= \sqrt{\frac{\Delta t}{2\pi N}} \sum_{n=0}^{N-1} y_n \cos(n\omega_k \Delta t) \\ \xi_s(\omega_k) &= \sqrt{\frac{\Delta t}{2\pi N}} \sum_{n=0}^{N-1} y_n \sin(n\omega_k \Delta t)\end{aligned} \tag{3.30}$$

The discrete stochastic process y, being the measurement of the system governed by Equations (3.1) or (3.6), and Equation (3.19), is Gaussian with zero mean and so do the random variables $\xi_c(\omega_k)$ and $\xi_s(\omega_k)$. Furthermore, for a given value of T, the random variables $\xi_c(\omega_k)$ and $\xi_s(\omega_k)$ are independent with equal variance asymptotically as $\Delta t \to 0^+$ [277]. Therefore, it follows that the spectral density estimator $S_{y,N}(\omega_k)$ has the following asymptotic behavior:

$$\lim_{\Delta t \to 0^+} S_{y,N}(\omega_k) = \frac{1}{2}[S_x(\omega_k) + S_{\epsilon 0}]\chi_2 \tag{3.31}$$

where χ_2 is *Chi-square distributed* with two degrees of freedom or *exponentially distributed*. The PDF of this random variable $S_k \equiv \lim_{\Delta t \to 0^+} S_{y,N}(\omega_k)$ is, therefore, given by:

$$p(S_k|\boldsymbol{\theta}, \mathcal{C}) = \frac{1}{S_x(\omega_k) + S_{\epsilon 0}} \exp\left[-\frac{S_k}{S_x(\omega_k) + S_{\epsilon 0}}\right] \tag{3.32}$$

For finite N and finite Δt, the random variables $\xi_c(\omega_k)$ and $\xi_s(\omega_k)$ are correlated and their variances are not equal. However, it can be shown using simulation that for a particular range of frequencies the probability distribution of $S_{y,N}(\omega_k)$ can be accurately approximated by a Chi-square distribution in analogy to Equation (3.32):

$$p(S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}) \approx \frac{1}{E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]} \exp\left(-\frac{S_{y,N}(\omega_k)}{E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]}\right) \tag{3.33}$$

where $E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]$ is given by Equation (3.28). Therefore, in the particular range of frequencies, the probability distribution of $S_{y,N}(\omega_k)$ can also be accurately approximated by a Chi-square distribution with two degrees of freedom. For example, in the case of displacement such a range of frequencies corresponds to the lower frequency range and this range increases for higher levels of the prediction error. This will be discussed in Section 3.3.3 and verified in Section 3.4 by simulation.

Furthermore, simulation shows that the random variables $S_{y,N}(\omega_k)$ and $S_{y,N}(\omega_{k'})$ are uncorrelated in the same range of frequencies for the Chi-square distribution, for $k \neq k'$ and $k, k' \in \mathcal{K}$. According to Yaglom [277], uncorrelated Chi-square random variables are independent. Use $\mathcal{K}$ to denote the *frequency index set* that contains the frequency indices for these approximations to be accurate. Given the observed data $\mathcal{D}$, the *spectral set* can be computed by using Equation (3.25):

$$\mathcal{S}_{\mathcal{K}} \equiv \{S_{y,N}(k\Delta\omega)|k \in \mathcal{K}\} \tag{3.34}$$

www.EngineeringBooksPdf.com

By the two approximations, the probability density function of $\mathcal{S}_{\mathcal{K}}$ can be well approximated by the product of the Chi-square distributions:

$$p(\mathcal{S}_{\mathcal{K}}|\boldsymbol{\theta}, \mathcal{C}) \approx \prod_{k\in\mathcal{K}} \frac{1}{E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]} \exp\left(-\frac{S_{y,N}(\omega_k)}{E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]}\right) \tag{3.35}$$

where $E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]$ is calculated by Equation (3.28) with $R_x(n\Delta t)$ calculated using Equation (3.4) for SDOF systems or Equation (3.17) for MDOF systems.

3.3.1.3 Identification with the Bayesian Framework

Using the Bayes' theorem, the posterior/updated PDF of the model parameter vector $\boldsymbol{\theta}$, given the spectral set, $\mathcal{S}_{\mathcal{K}}$ is:

$$p(\boldsymbol{\theta}|\mathcal{S}_{\mathcal{K}}, \mathcal{C}) = \kappa_1 p(\boldsymbol{\theta}|\mathcal{C}) p(\mathcal{S}_{\mathcal{K}}|\boldsymbol{\theta}, \mathcal{C}) \tag{3.36}$$

where κ_1 is a normalizing constant, and the likelihood function $p(\mathcal{S}_{\mathcal{K}}|\boldsymbol{\theta}, \mathcal{C})$ is given by Equation (3.35). The optimal parameters in $\boldsymbol{\theta}^\star$ are obtained by maximizing the updated PDF or, equivalently, by minimizing the objective function:

$$\begin{aligned} J(\boldsymbol{\theta}) &\equiv -\ln[p(\boldsymbol{\theta}|\mathcal{C}) p(\mathcal{S}_{\mathcal{K}}|\boldsymbol{\theta}, \mathcal{C})] \\ &= -\ln p(\boldsymbol{\theta}|\mathcal{C}) + \sum_{k\in\mathcal{K}} \left(\ln E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] + \frac{S_{y,N}(\omega_k)}{E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]}\right) \end{aligned} \tag{3.37}$$

This optimization problem can be solved by the MATLAB® function 'fminsearch' [171]. It has been shown numerically for the globally identifiable case with a large number of data points that the updated PDF can be well approximated by a Gaussian distribution $\mathcal{G}(\boldsymbol{\theta}; \boldsymbol{\theta}^\star, \mathcal{H}(\boldsymbol{\theta}^\star)^{-1})$ with mean $\boldsymbol{\theta}^\star$ and covariance matrix $\mathcal{H}(\boldsymbol{\theta}^\star)^{-1}$, where $\mathcal{H}(\boldsymbol{\theta}^\star)$ denotes the Hessian of $J(\boldsymbol{\theta})$ calculated at $\boldsymbol{\theta} = \boldsymbol{\theta}^\star$:

$$\begin{aligned} \mathcal{H}^{(l,l')}(\boldsymbol{\theta}^\star) &= \left.\frac{\partial^2 J(\boldsymbol{\theta})}{\partial\theta_l \partial\theta_{l'}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \\ &= -\left.\frac{\partial^2 \ln p(\boldsymbol{\theta}|\mathcal{C})}{\partial\theta_l \partial\theta_{l'}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} + \sum_{k\in\mathcal{K}} \left[\frac{\partial^2}{\partial\theta_l \partial\theta_{l'}}\left(\ln E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] + \frac{S_{y,N}(\omega_k)}{E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \end{aligned} \tag{3.38}$$

3.3.1.4 Identification with Multiple Sets of Measurements

If several sets of independent time histories $\mathcal{D}^{(1)}, \mathcal{D}^{(2)}, \ldots, \mathcal{D}^{(N_s)}$ are available, the identification process can be proceeded by calculating the corresponding spectral sets $\mathcal{S}_{\mathcal{K}}^{(1)}, \mathcal{S}_{\mathcal{K}}^{(2)}, \ldots, \mathcal{S}_{\mathcal{K}}^{(N_s)}$ and then the updated PDF is given by:

$$p(\boldsymbol{\theta}|\mathcal{S}_{\mathcal{K}}^{(1)}, \mathcal{S}_{\mathcal{K}}^{(2)}, \ldots, \mathcal{S}_{\mathcal{K}}^{(N_s)}, \mathcal{C}) = \kappa_2 p(\boldsymbol{\theta}|\mathcal{C}) \prod_{s=1}^{N_s} p\left(\mathcal{S}_{\mathcal{K}}^{(s)}|\boldsymbol{\theta}, \mathcal{C}\right) \tag{3.39}$$

www.EngineeringBooksPdf.com

where each of the likelihood functions $p(\mathcal{S}_{\mathcal{K}}^{(s)}|\boldsymbol{\theta}, \mathcal{C})$ can be obtained by Equation (3.35). Again, the updated PDF can be approximated with a Gaussian distribution centered at the optimal parameter vector $\boldsymbol{\theta}^{\star}$ obtained by maximizing the right hand side of Equation (3.39). Nevertheless, it is equivalent to minimize the objective function:

$$\begin{aligned} J(\boldsymbol{\theta}) &\equiv -\ln p(\boldsymbol{\theta}|\mathcal{C}) - \sum_{s=1}^{N_s} \ln p\left(\mathcal{S}_{\mathcal{K}}^{(s)}|\boldsymbol{\theta}, \mathcal{C}\right) \\ &= -\ln p(\boldsymbol{\theta}|\mathcal{C}) + N_s \sum_{k\in\mathcal{K}} \ln E[S_{y,N}(\omega_k|\boldsymbol{\theta}, \mathcal{C})] + \sum_{k\in\mathcal{K}} \sum_{s=1}^{N_s} \frac{S_{y,N}(\omega_k)}{E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]} \\ &= -\ln p(\boldsymbol{\theta}|\mathcal{C}) + N_s \sum_{k\in\mathcal{K}} \ln E[S_{y,N}(\omega_k|\boldsymbol{\theta}, \mathcal{C})] + N_s \sum_{k\in\mathcal{K}} \frac{S_{y,N}^{av}(\omega_k)}{E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]} \end{aligned} \tag{3.40}$$

where:

$$S_{y,N}^{av}(\omega_k) = \frac{1}{N} \sum_{s=1}^{N_s} S_{y,N}^{(s)}(\omega_k) \tag{3.41}$$

is the averaged spectral value.

An alternative approach, which yields equivalent results but is computationally more efficient, is to calculate the average of the above N_s spectral density estimates. The probability distribution of the averaged spectral density estimator is:

$$p(S_{y,N}^{av}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}) = \frac{N_s^{N_s} \left[S_{y,N}^{av}(\omega_k)\right]^{N_s-1}}{(N_s-1)!\{E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]\}^{N_s}} \exp\left(-\frac{N_s S_{y,N}^{av}(\omega_k)}{E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]}\right) \tag{3.42}$$

since the random variable $2N_s S_{y,N}^{av}(\omega_k)/E[S_{y,N}(\omega_k)]$ follows the standard Chi-square distribution with $2N_s$ DOFs. In this case, the objective function can be defined as the negative logarithm of the updated PDF without taking the constant terms that do not depend on the uncertain parameters:

$$J(\boldsymbol{\theta}) = -\ln p(\boldsymbol{\theta}|\mathcal{C}) + N_s \sum_{k\in\mathcal{K}} \ln E[S_{y,N}(\omega_k|\boldsymbol{\theta}, \mathcal{C})] + N_s \sum_{k\in\mathcal{K}} \frac{S_{y,N}^{av}(\omega_k)}{E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]} \tag{3.43}$$

which is identical to Equation (3.40). Therefore, the optimal parameters and associated uncertainty obtained by these two approaches are identical.

Although the above formulation was presented for displacement time history, it can be easily modified for velocity or acceleration measurements. In this case the right hand side of Equations (3.4) or (3.17) can be modified with the corresponding expressions for velocity or acceleration. Of course, the case of relative acceleration with white noise excitation is not realistic since the response variance is infinity. However, the absolute acceleration measurements can be considered for ground excitation. Another choice is to utilize a band-limited excitation model.

www.EngineeringBooksPdf.com

3.3.1.5 Importance of the Proper Distribution for the Spectral Density Estimator

In a practical situation with multiple sets of data, one of the following two approaches is usually followed: (1) the averaged spectrum is calculated and then curve fitting is proceeded to obtain the optimal parameters or (2) the optimal estimates $\boldsymbol{\theta}^{\star(1)}, \boldsymbol{\theta}^{\star(2)}, \ldots, \boldsymbol{\theta}^{\star(N_s)}$ are obtained by curve fitting for each of the N_s sets of data, and then the sample average and sample variance of the parameters are computed as the overall optimal parameters and the associated variance of the model parameters. In the first case, only the optimal estimates are obtained while in the second case the associated uncertainty can also be estimated. It is worth noting however that traditional least-squares curve fitting of the spectrum $\mathcal{S}_{\mathcal{K}}$ is not appropriate. This follows directly from Equation (3.35) that the spectral density estimator is non-Gaussian. Figure 3.1 shows the PDFs of the random variables which are the average of n exponentially distributed random variables with mean equal to unity, for $n = 1, 5, 20$. The exponential distribution ($n = 1$) is clearly different from the Gaussian distribution. However, the distribution of the averaged random variables ($n = 5$ or 20) are bell-shaped and approach Gaussian for large n. This explains why least-squares fitting requires an averaged spectrum to proceed.

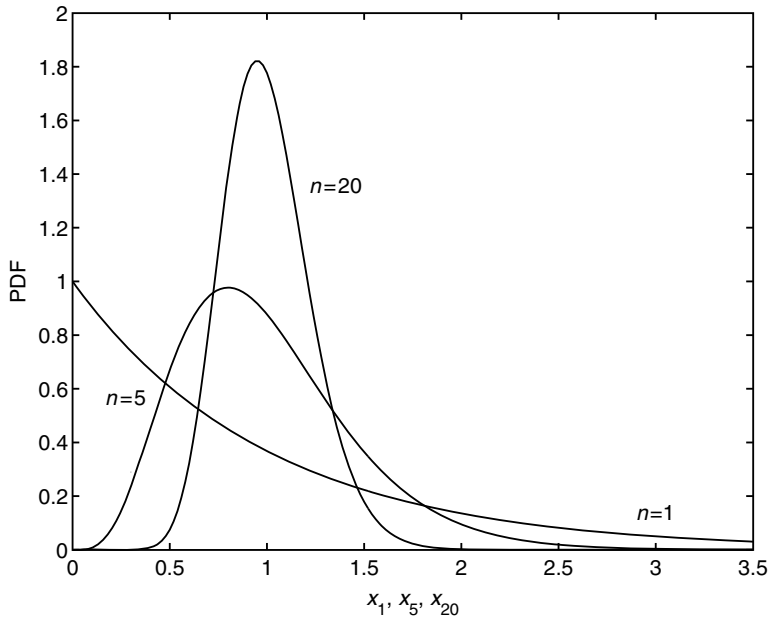

Figure 3.1 Distribution of an exponential random variable and its averages

3.3.2 *Formulation for Multiple-channel Output Measurements*

Assume that discrete response data are available for $N_o (\leq N_d)$ instrumented DOFs and use Δt to denote the sampling time step. Because of measurement noise and modeling error, referred to hereafter jointly as prediction error, the measured response $\mathbf{y}_n \in \mathbb{R}^{N_o}$ at time $t = n\Delta t$ differs

www.EngineeringBooksPdf.com

from the model response $\mathbf{L}_o\mathbf{x}(n\Delta t)$ corresponding to the measured DOFs. Here, $\mathbf{L}_o$ is the $N_o \times N_d$ observation matrix, comprised of zeros and ones. The prediction error is represented by a discrete Gaussian white noise vector process $\boldsymbol{\epsilon}$:

$$\mathbf{y}_n = \mathbf{L}_o\mathbf{x}(n\Delta t) + \boldsymbol{\epsilon}_n \tag{3.44}$$

where $\boldsymbol{\epsilon}_n \in \mathbb{R}^{N_o}$. The discrete process $\boldsymbol{\epsilon}$ has zero mean and it satisfies:

$$E[\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_{n'}^T] = \boldsymbol{\Sigma}_\epsilon\delta_{nn'} \tag{3.45}$$

where $E[.]$ denotes the mathematical expectation, $\delta_{nn'}$ denotes the Kronecker delta, and $\boldsymbol{\Sigma}_\epsilon$ denotes the $N_o \times N_o$ covariance matrix of the prediction-error process $\boldsymbol{\epsilon}$.

3.3.2.1 Spectral Density Matrix Estimator

Consider a discrete stochastic vector process $\mathbf{y}$ and a finite number of discrete data points $\mathcal{D} = \{\mathbf{y}_n, n = 1, 2, \ldots, N\}$. Based on $\mathcal{D}$ a discrete estimator of the spectral density matrix of the stochastic process $\mathbf{y}$ is introduced:

$$\mathbf{S}_{y,N}(\omega_k) = \mathcal{Y}(\omega_k)\mathcal{Y}(\omega_k)^H \tag{3.46}$$

where ^{H} denotes the adjoint or conjugate transpose of a complex vector/matrix. The vector $\mathcal{Y}(\omega_k) \in \mathbb{C}^{N_o}$ denotes the scaled discrete Fourier transform of the vector process $\mathbf{y}$ at frequency ω_k:

$$\mathcal{Y}(\omega_k) = \sqrt{\frac{\Delta t}{2\pi N}} \sum_{n=0}^{N-1} \mathbf{y}_n \exp(-in\omega_k\Delta t) \tag{3.47}$$

where $\omega_k = k\Delta\omega$, $k = 0, 1, \ldots, N_{nqy}$ with $N_{nqy} = \mathrm{INT}(N/2)$, $\Delta\omega = 2\pi/\mathrm{T}$, and $T = N\Delta t$.

Since $\mathbf{x}$ and $\boldsymbol{\epsilon}$ are assumed statistically independent, $\mathbf{q}$ and $\boldsymbol{\epsilon}$ are also independent. By using Equations (3.8) and (3.44), taking expectation of Equation (3.46), yields:

$$\begin{aligned} E[\mathbf{S}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] &= \mathbf{L}_o E[\mathbf{S}_{x,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]\mathbf{L}_o^T + E[\mathbf{S}_{\epsilon,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] \\ &= (\mathbf{L}_o\boldsymbol{\Phi})E[\mathbf{S}_{q,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}](\mathbf{L}_o\boldsymbol{\Phi})^T + E[\mathbf{S}_{\epsilon,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] \end{aligned} \tag{3.48}$$

where $\mathcal{C}$ is a prescribed class of models for the dynamical system. The matrices $\mathbf{S}_{x,N}(\omega_k)$, $\mathbf{S}_{q,N}(\omega_k)$, and $\mathbf{S}_{\epsilon,N}(\omega_k)$ are defined in a similar manner to Equations (3.46) and (3.47) for the vector processes $\mathbf{x}$, $\mathbf{q}$ and $\boldsymbol{\epsilon}$, respectively. It follows easily from Equations (3.45)–(3.47) that $E[\mathbf{S}_{\epsilon,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]$ can be expressed in terms of the covariance matrix $\boldsymbol{\Sigma}_\epsilon$ of the discrete white noise:

$$E[\mathbf{S}_{\epsilon,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] = \frac{\Delta t}{2\pi}\boldsymbol{\Sigma}_\epsilon \equiv \mathbf{S}_{\epsilon 0} \tag{3.49}$$

The term $E[\mathbf{S}_{q,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]$ in Equation (3.48) can be readily obtained by noting that $\mathbf{S}_{q,N}(\omega_k)$ has the elements:

$$S_{q,N}^{(m,m')}(\omega_k) = \frac{\Delta t}{2\pi N} \sum_{n,n'=0}^{N-1} q_n^{(m)} q_{n'}^{(m')} \mathrm{e}^{i(n'-n)\omega_k\Delta t} \tag{3.50}$$

www.EngineeringBooksPdf.com

Grouping together the terms with the same value of $n' - n$, the following expression can be obtained by taking expectation of Equation (3.50):

$$E[S_{q,N}^{(m,m')}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] = \frac{\Delta t}{2\pi N}\left[NR_q^{(m,m')}(0) + \sum_{n=1}^{N-1}(N-n)[R_q^{(m,m')}(n\Delta t)\mathrm{e}^{in\omega_k\Delta t} + R_q^{(m',m)}(n\Delta t)\mathrm{e}^{-in\omega_k\Delta t}]\right] \tag{3.51}$$

where the correlation function of the modal coordinates is given by Equation (3.17). Again, this estimator $E[S_{q,N}^{(m,m')}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]$ can be computed efficiently using the function 'fft' in MATLAB® [171]:

```
an = [N/2 [N - 1 : -1 : 1]];
Nnyq = fix(N/2);
for m1=1:Nm
    ESqtemp = dt/pi/N * real(fft(an. * Rq(m1, m1, :)));
    ESq(m1, m1, :) = ESqtemp(1 : Nnyq);
    % Rq is a 3-array of the correlation function R_q where the first two indices
    % denote the modes and the last denotes the time lag
    for m2 = m1 + 1 : Nm
        ESqtemp = dt/2/pi/N * conj(fft(an. * Rq(m1, m2, :))) + ...
                    fft(an. * Rq(m2, m1, :)))
        ESq(m1, m2, :) = ESqtemp(1 : Nnyq);
        ESq(m2, m1, :) = conj(ESq(m1, m2, :));
    end
end
```

3.3.2.2 Statistical Properties of the Spectral Density Matrix Estimator

Next, the statistical properties of the spectral density matrix estimator $\mathbf{S}_{y,N}(\omega_k)$ are investigated. Denote by $\mathcal{Y}_R(\omega_k)$ and $\mathcal{Y}_I(\omega_k)$ the real and imaginary part, respectively, of $\mathcal{Y}(\omega_k)$ so:

$$\mathcal{Y}(\omega_k) = \mathcal{Y}_R(\omega_k) + i\mathcal{Y}_I(\omega_k) \tag{3.52}$$

Since $\mathbf{y}$ is a zero-mean Gaussian vector process, both $\mathcal{Y}_R(\omega_k)$ and $\mathcal{Y}_I(\omega_k)$, $k = 0, 1, \ldots, N_{nqy}$, are zero-mean Gaussian random vectors. Furthermore, in the limit when $\Delta t \to 0^+$, and for a non-zero and non-Nyquist frequency ω_k, it can be shown that the covariance matrix of the

www.EngineeringBooksPdf.com

vector $[\mathcal{Y}_R(\omega_k)^T, \mathcal{Y}_I(\omega_k)^T]^T$ has the form [295]:

$$\mathbf{G}(\omega_k) = \begin{bmatrix} \mathbf{G}_1(\omega_k) & -\mathbf{G}_2(\omega_k) \\ \mathbf{G}_2(\omega_k) & \mathbf{G}_1(\omega_k) \end{bmatrix} \tag{3.53}$$

This indicates that $\mathcal{Y}_R(\omega_k)$ and $\mathcal{Y}_I(\omega_k)$ have an equal covariance matrix $\mathbf{G}_1(\omega_k)$ but the cross-covariance is anti-symmetric:

$$\mathbf{G}_2(\omega_k)^T = -\mathbf{G}_2(\omega_k) \tag{3.54}$$

or element-wise $E[\mathcal{Y}_{I,l}(\omega_k)\mathcal{Y}_{R,l'}(\omega_k)] = -E[\mathcal{Y}_{R,l}(\omega_k)\mathcal{Y}_{I,l'}(\omega_k)]$. The latter property implies also that the diagonal elements of $\mathbf{G}_2$ are zero, i.e., $E[\mathcal{Y}_{R,l}(\omega_k)\mathcal{Y}_{I,l}(\omega_k)] = 0$, for every l and ω_k. By Equation (3.53) the complex vector $\mathcal{Y}(\omega_k)$ is said to follow the complex multivariate normal distribution [146] as $T \to \infty$ and $\Delta t \to 0^+$ except the zero or Nyquist frequency.

Assume that N_s sets of independent and identically distributed time histories are available: $\mathcal{D}^{(1)}, \mathcal{D}^{(2)}, \ldots, \mathcal{D}^{(N_s)}$. As $T \to \infty$ and $\Delta t \to 0^+$, the corresponding discrete Fourier transforms $\mathcal{Y}^{(s)}(\omega_k)$, $s = 1, 2, \ldots, N_s$, are independent and follow an identical complex N_o-variate normal distribution with zero mean. Then, according to Krishnaiah [146], the averaged spectral density matrix estimator:

$$\mathbf{S}^{av}_{y,N}(\omega_k) = \frac{1}{N_s}\sum_{s=1}^{N_s}\mathbf{S}^{(s)}_{y,N}(\omega_k) = \frac{1}{N_s}\sum_{s=1}^{N_s}\mathcal{Y}^{(s)}(\omega_k)\mathcal{Y}^{(s)}(\omega_k)^\star \tag{3.55}$$

follows the central complex Wishart distribution [93] of dimension N_o with N_s degrees of freedom and mean $E[\mathbf{S}^{av}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] = E[\mathbf{S}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] = 2[\mathbf{G}_1(\omega_k) + i\mathbf{G}_2(\omega_k)]$. The PDF of this distribution is given by:

$$\begin{aligned} p(\mathbf{S}^{av}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}) = & \frac{\pi^{-\frac{N_o(N_o-1)}{2}} N_s^{N_o(N_s-N_o)} |\mathbf{S}^{av}_{y,N}(\omega_k)|^{N_s-N_o}}{(\prod_{s=1}^{N_o}(N_s-s)!)|E[\mathbf{S}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]|^{N_s}} \\ & \times \exp\left(-N_s \operatorname{tr}\{E[\mathbf{S}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]^{-1}\mathbf{S}^{av}_{y,N}(\omega_k)\}\right) \end{aligned} \tag{3.56}$$

where $|\mathbf{A}|$ and $\operatorname{tr}\{\mathbf{A}\}$ denote the determinant and trace of the matrix $\mathbf{A}$, respectively. Note that this PDF exists if and only if $N_s \geq N_o$.

Also, in the special case of a single channel of measurements ($N_o = 1$), the distribution in Equation (3.56) reduces to:

$$p(S^{av}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}) = \frac{N_s^{N_s}\left[S^{av}_{y,N}(\omega_k)\right]^{N_s-1}}{(N_s-1)!\{E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]\}^{N_s}} \exp\left(-\frac{N_s S^{av}_{y,N}(\omega_k)}{E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]}\right) \tag{3.57}$$

which is identical to Equation (3.42). Note that the real random variable $2N_s S^{av}_{y,N}(\omega_k)/E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]$ follows the standard Chi-square distribution with $2N_s$ degrees of freedom.

It can be shown that when $T \to \infty$ and $\Delta t \to 0^+$ the vectors $[\mathcal{Y}_R(\omega_k)^T, \mathcal{Y}_I(\omega_k)^T]^T$ and $[\mathcal{Y}_R(\omega_{k'})^T, \mathcal{Y}_I(\omega_{k'})^T]^T$ are independent if $k \neq k'$ [295] so the complex vectors $\mathcal{Y}(\omega_k)$ and $\mathcal{Y}(\omega_{k'})$ are independent. As a result the random matrices $\mathbf{S}^{av}_{y,N}(\omega_k)$ and $\mathbf{S}^{av}_{y,N}(\omega_{k'})$ are independently

www.EngineeringBooksPdf.com

complex Wishart distributed for $k \neq k'$:

$$p[\mathbf{S}^{av}_{y,N}(\omega_k), \mathbf{S}^{av}_{y,N}(\omega_{k'})] = p[\mathbf{S}^{av}_{y,N}(\omega_k)]p[\mathbf{S}^{av}_{y,N}(\omega_{k'})] \tag{3.58}$$

Although Equations (3.56) and (3.58) are exact only as $\Delta t \to 0^+$, it can be verified by simulation that they are good approximations in a particular frequency range. It will be shown later using simulation that such approximations are indeed accurate if an appropriately chosen bandwidth is considered. The reasons for the violation of these approximations are aliasing and leakage. Therefore, the range of frequency for accurate approximations is the region with large spectral values since the aliasing and leakage effects have relatively minor contribution in such a frequency range. As a result, in the case of displacement measurement such a range of frequencies corresponds to the lower frequency range. This range increases for higher levels of prediction error. This will be shown in the numerical verification in Section 3.4.

3.3.2.3 Identification Based on Spectral Density Estimates

Based on the above discussion regarding the statistical properties of the averaged spectral estimator, the Bayesian spectral density approach for updating the uncertain parameter vector $\boldsymbol{\theta}$ is given as follows. With N_s ($\geq N_o$) independent sets of observed data $\mathcal{D}^{(s)}$, $s = 1, 2, \ldots, N_s$, the corresponding observed spectral density matrix estimators $\mathbf{S}^{(s)}_{y,N}(\omega_k)$, $s = 1, 2, \ldots, N_s$, $k \in \mathcal{K}$, can be obtained using Equations (3.46) and (3.47). Then, the averaged spectral density matrix estimators $\mathbf{S}^{av}_{y,N}(\omega_k)$ is readily obtained by Equation (3.55) to form the *averaged spectral set*:

$$\mathcal{S}^{av}_{\mathcal{K}} \equiv \{\mathbf{S}^{av}_{y,N}(k\Delta\omega) | k \in \mathcal{K}\} \tag{3.59}$$

Here the frequency index set $\mathcal{K}$ represents a range over which Equations (3.56) and (3.58) give a satisfactory approximation and this will be further discussed in Section 3.3.3. Using the Bayes' theorem, the updated PDF of the model parameter vector $\boldsymbol{\theta}$, given the averaged spectral set $\mathcal{S}^{av}_{\mathcal{K}}$, is given by:

$$p(\boldsymbol{\theta}|\mathcal{S}^{av}_{\mathcal{K}}, \mathcal{C}) = \kappa_3 \; p(\boldsymbol{\theta}|\mathcal{C})p(\mathcal{S}^{av}_{\mathcal{K}}|\boldsymbol{\theta}, \mathcal{C}) \tag{3.60}$$

where κ_3 is a normalizing constant. The prior distribution $p(\boldsymbol{\theta}|\mathcal{C})$ expresses the relative plausibility of different values of $\boldsymbol{\theta}$ based on the prior information and engineering judgement. The likelihood function $p(\mathcal{S}^{av}_{\mathcal{K}}|\boldsymbol{\theta}, \mathcal{C})$ expresses the contribution of the measured data. Based on Equations (3.56) and (3.58), it can be calculated as follows:

$$p(\mathcal{S}^{av}_{\mathcal{K}}|\boldsymbol{\theta}, \mathcal{C}) \approx \kappa_4 \prod_{k\in\mathcal{K}} \frac{1}{|E[\mathbf{S}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]|^{N_s}} \exp\left(-N_s \operatorname{tr}\{E[\mathbf{S}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]^{-1}\mathbf{S}^{av}_{y,N}(\omega_k)\}\right) \tag{3.61}$$

where $\kappa_4 = \pi^{-\frac{N_o(N_o-1)N_\omega}{2}} (\prod_{s=1}^{N_o} (N_s - s)!)^{-N_\omega} N_s^{N_o(N_s-N_o)N_\omega} \prod_{k\in\mathcal{K}} |\mathbf{S}^{av}_{y,N}(\omega_k)|^{N_s-N_o}$; N_ω is the number of frequency points to be considered and is equal to the number of the distinct elements in the set $\mathcal{K}$. For a given set of data the constant κ_4 does not depend on the model parameters so it does not affect the optimal parameters and their associate uncertainty. Finally, the most

www.EngineeringBooksPdf.com

probable parameter vector $\boldsymbol{\theta}^\star$ is obtained by minimizing the objective function:

$$J(\boldsymbol{\theta}) = -\ln p(\boldsymbol{\theta}|\mathcal{C}) + N_s \sum_{k\in\mathcal{K}} \left[\ln \left| E[\mathbf{S}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] \right| + \mathrm{tr}\{E[\mathbf{S}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]^{-1} \mathbf{S}_{y,N}^{av}(\omega_k)\} \right] \tag{3.62}$$

In the case where a non-informative prior is used, the first term can be simply neglected. Furthermore, the updated PDF of the parameter vector $\boldsymbol{\theta}$ can be well approximated by a Gaussian distribution $\mathcal{G}(\boldsymbol{\theta}; \boldsymbol{\theta}^\star, \mathcal{H}(\boldsymbol{\theta}^\star)^{-1})$ with mean $\boldsymbol{\theta}^\star$ and covariance matrix $\mathcal{H}(\boldsymbol{\theta}^\star)^{-1}$, where $\mathcal{H}(\boldsymbol{\theta}^\star)$ denotes the Hessian matrix of the objective function J calculated at $\boldsymbol{\theta} = \boldsymbol{\theta}^\star$:

$$\begin{aligned} \mathcal{H}^{(l,l')}(\boldsymbol{\theta}^\star) &= \left. \frac{\partial^2 J(\boldsymbol{\theta})}{\partial\theta_l \partial\theta_{l'}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} = -\left. \frac{\partial^2 \ln p(\boldsymbol{\theta}|\mathcal{C})}{\partial\theta_l \partial\theta_{l'}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \\ &+ N_s \sum_{k\in\mathcal{K}} \left[\frac{\partial^2}{\partial\theta_l \partial\theta_{l'}} \left(\ln \left| E[\mathbf{S}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] \right| + \mathrm{tr}\{(E[\mathbf{S}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}])^{-1} \mathbf{S}_{y,N}^{av}(\omega_k)\} \right) \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \end{aligned} \tag{3.63}$$

Standard optimization algorithms (e.g., the function 'fminsearch' in MATLAB® [171]) can be employed to minimize the objective function J in Equation (3.62) and obtain the optimal parameters and then a central difference can be used to calculate the Hessian matrix $\mathcal{H}(\boldsymbol{\theta}^\star)$ whose inverse is the covariance matrix (Appendix A). An alternative efficient and robust approach, yielding simultaneously both the mean and the covariance matrix of the posterior distribution, is based on adaptive importance sampling [133] or simulated annealing.

3.3.3 Selection of the Frequency Index Set

The two approximations used in the Bayesian spectral density approach are accurate in a particular frequency range. It is recommended to select the frequency index set to include only the range around the peaks in the spectrum even though the Chi-square/Wishart distributions and independence approximations are accurate over a wider range. This selection enhances the computational efficiency without sacrificing substantially the information for the frequency structure of the dynamical system (though it induces loss of information for the prediction-error variance). Another advantage is that the results by this choice will rely less on the 'whiteness assumption' since it requires a flat spectral density function for each mode only around the corresponding peak instead of over the whole frequency range. Moreover, the aliasing and leakage effects generally have less influence on this range since their spectral values are large. Therefore, for practical purpose, the theoretical power spectral density function can be used instead of its estimator in Equation (3.28) or Equation (3.48) with Equation (3.51) since expressions of the correlation functions may be tedious to derive. Such an approximation may affect slightly the damping ratios and the spectral intensity of the excitation but it has a very small effect on the estimation of the modal frequencies of the system.

Another important advantage of this cutoff frequency range is as follows. In most existing probabilistic methods, the uncertainty of the model parameters will tend to zero if the sampling time interval tends to zero with a fixed observed duration (even if it is very short) as long as it is globally identifiable. This is the consequence of the discrete white noise assumption. Note that this phenomenon occurs even for filtered white noise, such as moving average or auto-regressive output of white noise. However, for the Bayesian spectral density method with this

www.EngineeringBooksPdf.com

proposed cutoff frequency range, the sampling time interval does not affect the frequency index set so the same number of frequencies is considered regardless of the sampling time interval (if it is sufficiently small). Therefore, the uncertainty of the model parameters provided by the Bayesian spectral density method will stabilize as the sampling time step tends to zero. This feature of the methodology is appealing for practical purposes.

3.3.4 Nonlinear Systems

The modal decomposition described in Section 3.2.2 is inapplicable for nonlinear systems. Furthermore, analytic solution of the correlation function or the spectral density function of the response of nonlinear systems, especially those with two or more degrees of freedom, is in general unavailable. However, the Chi-square/Wishart and independence approximations of the spectrum are still valid in a particular frequency range [295]. In this case, for a given model parameter vector $\boldsymbol{\theta}$, realizations of the response time histories are simulated and then their spectral density estimators can be computed in a similar manner to that described in Equations (3.46) and (3.47). Finally, the expected values of the spectral density estimators can be approximated by the sample average of these spectral density estimators. It is not recommended to estimate the correlation functions and then use Equations (3.28) or (3.51) to compute the mean spectral density estimator because the correlation function estimated by simulation has a large uncertainty and bias for the components with a large time lag. Therefore, if more efficient tools are available for the computation of the mean spectrum of nonlinear systems, the Bayesian spectral density approach is ready for the identification of their parameters. In Section 3.6, an application for updating the nonlinear Duffing oscillator is presented. It will be demonstrated that uncertainty quantification is very important in this case as the problem becomes ill-posed for one set of measurements.

3.4 Numerical Verifications

In this section, a single-degree-of-freedom linear system is considered to verify the approximations used in the method. The actual parameters are: $\tilde{\Omega} = 4$ rad/s, $\tilde{\zeta} = 1\%$, $\tilde{S}_{f0} = 1$ cm^2 s^{-3} and response time histories are generated using the MATLAB® function 'lsim' [171] for $T = 1000$ s with the time step $\Delta t = 0.005$ s. However, the measurements are assumed to be observed with a sampling time interval $\Delta t = 0.05$ s so $N = 20000$. This is done on purpose to simulate the reality that there is inevitably a frequency content in the signal higher than the Nyquist frequency. Furthermore, measurements are assumed to be noise-free in this example, i.e., $\tilde{S}_{\epsilon 0} = 0$.

Figure 3.2 refers to displacement measurements and shows a comparison among: (i) the theoretical spectral density function S_y corresponding to the continuous process $y = x$ (solid line). According to Equation (3.5), this is given by:

$$S_y(\omega) = \frac{\tilde{S}_{f0}}{(\tilde{\Omega}^2 - \omega^2)^2 + (2\tilde{\zeta}\tilde{\Omega}\omega)^2} \tag{3.64}$$

and (ii) $E[S_{y,N}(\omega_k)|\tilde{\boldsymbol{\theta}}, \mathcal{C}]$, the expected value of the discrete spectral density estimator $S_{y,N}(\omega_k)$, according to Equation (3.28) (dashed line) and (iii) $E[S_{y,N}(\omega_k)|\tilde{\boldsymbol{\theta}}, \mathcal{C}]$ calculated as the average

www.EngineeringBooksPdf.com

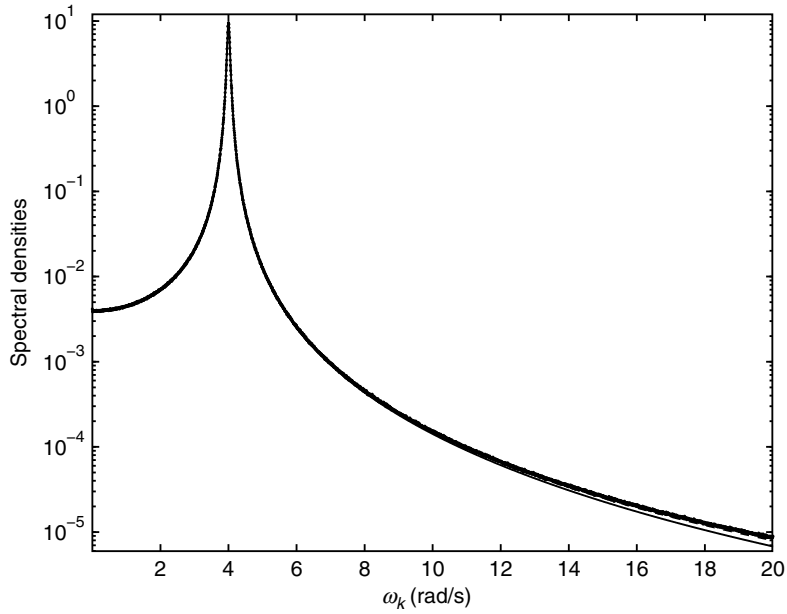

Figure 3.2 $S_y(\omega)$, $E[S_{y,N}(\omega_k)|\tilde{\boldsymbol{\theta}}, \mathcal{C}]$ and $E[S_{y,N}(\omega_k)|\tilde{\boldsymbol{\theta}}, \mathcal{C}]$ from simulation

of $S_{y,N}(\omega_k)$ using four thousand simulation runs (dotted line). It is observed that the latter two agree well so the expression in Equation (3.28) yields an unbiased estimation of the spectrum estimated from discrete measurements.

Figure 3.3 shows the difference between the first two curves in Figure 3.2 relative to the first one, i.e., $E[S_{y,N}(\omega_k)|\tilde{\boldsymbol{\theta}}, \mathcal{C}]/S_y(\omega_k) - 1$. The relative error is large for higher frequencies. One of the reasons for the observed differences, especially in the far right tail, is aliasing. The aliasing effect comes from the fact that the original signal inevitably contains some contents in the frequencies higher than the Nyquist frequency. These contents will appear in the band of the Fourier amplitude spectrum, i.e., $[0, \omega_{nyq}]$ and this phenomenon of overestimating the spectral values in the Fourier amplitude spectrum is referred to as *aliasing*.

Another reason for the difference from the theoretical spectral density is *leakage*. The effect of leakage comes from the fact that the measurement does not contain an exact integer-multiple of the periods of the Fourier components. Instead of showing a single spike in the spectrum, the spectral value is being leaked out to the nearby frequencies. Therefore, it can be observed primarily at frequencies around the natural frequency, causing the expected values of the spectral density estimator to be smaller than the theoretical spectral density.

3.4.1 Aliasing and Leakage

The aliasing and leakage effects come from the manipulation of finite number of data points measured at a finite sampling rate. First, the aliasing effect is demonstrated. Consider a sinusoidal signal with frequency $\Omega = 1.0$ Hz: $x(t) = \sin(2\pi t)$, which is shown by the solid line

www.EngineeringBooksPdf.com

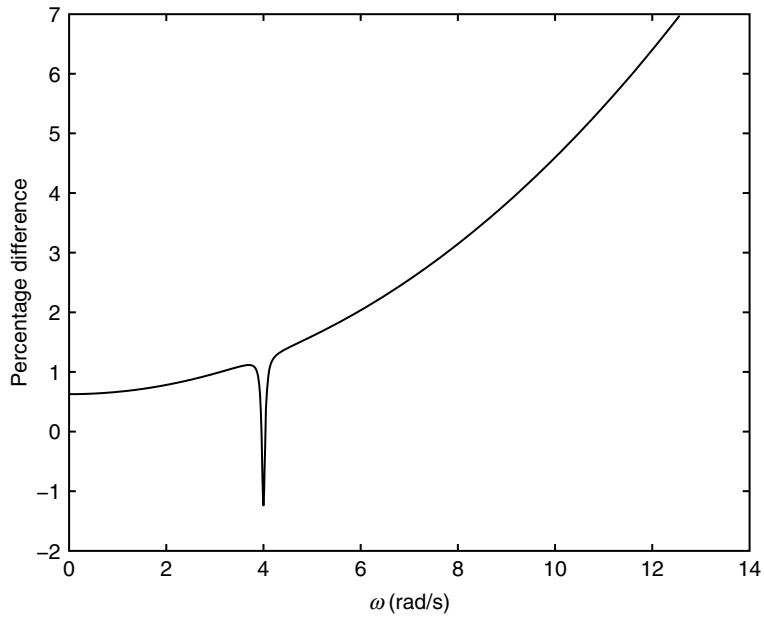

Figure 3.3 Percentage difference between $S_y(\omega_k)$ and $E[S_{y,N}(\omega_k)|\tilde{\boldsymbol{\theta}}, \mathcal{C}]$

in Figure 3.4. The measurements are taken at a sampling rate of 4/3 Hz and these samples are shown by the circles. In this case, the frequency of the original signal is higher than the Nyquist frequency, which is 2/3 Hz. By observing these data points only, misleading information will be obtained and it may be interpreted as a sinusoidal curve with frequency 1/3 Hz, which is shown by the dashed line. Moreover, the discrete Fourier transform of the measured signal shows a spike at $\Omega = 1/3$ Hz. The implication is that if the signal has contents at frequencies higher than the Nyquist frequency (half sampling frequency), these contents will appear in the frequencies lower than the Nyquist frequency. Therefore, the Fourier amplitudes of the low frequencies will be higher than they should be. This phenomenon occurs for general non-sinusoidal signals as there are inevitably signal contents in the frequency range higher than the Nyquist frequency. For large N, the aliasing effect can be quantified by [295]:

$$\lim_{N\to\infty} E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] = S_y(\omega_k) + \sum_{k'=1}^{\infty}\left[S_y\left(\frac{2\pi k'}{\Delta t} + \omega_k\right) + S_y\left(\frac{2\pi k' - \omega_k}{\Delta t}\right)\right] \tag{3.65}$$

Figure 3.5 demonstrates the leakage effect. Consider the same signal $x(t) = \sin(2\pi t)$. Measurements are taken for $T = 20$ s with $\Delta t = 0.1$ s. Figure 3.5(a) shows its Fourier amplitude spectrum $S_{y,N}(\omega)$ and a perfect spike at $\omega = 2\pi$ rad/s $= 1.0$ Hz is observed. Note that exactly 20 cycles are measured in this case. However, if the measurement is taken up to $T = 19.9$ s only, its peak in the Fourier spectrum will leak out and this is shown in Figure 3.5(b). The amplitude value at $\omega = 2\pi$ rad/s is also smaller than that in the case with $T = 20$ s. An even worse case occurs when $T = 19.5$ s, in which 19.5 cycles were measured. The amplitude in the Fourier spectrum is even lower and the leakage range is more spread.

www.EngineeringBooksPdf.com

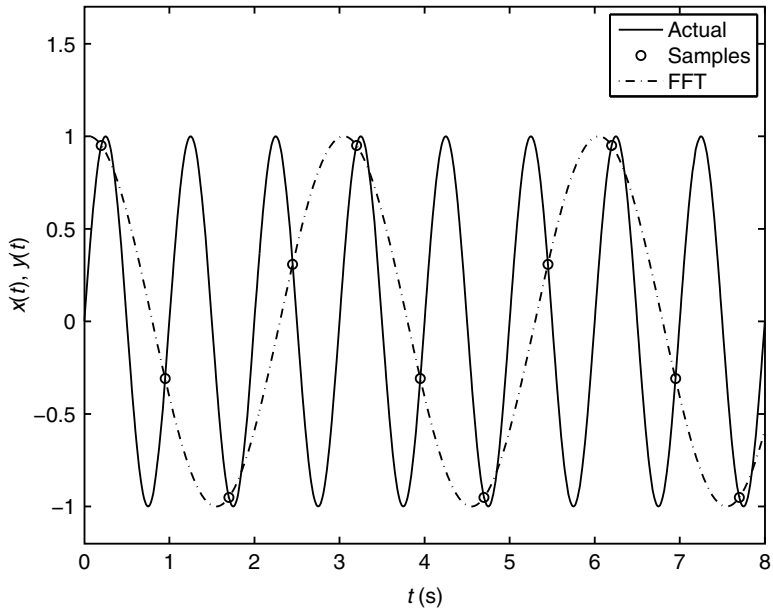

Figure 3.4 Demonstration of the aliasing effect

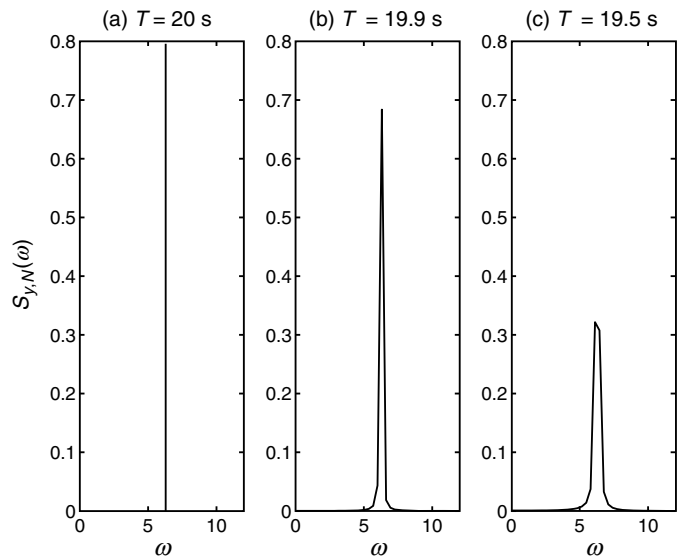

Figure 3.5 Demonstration of the leakage effect

www.EngineeringBooksPdf.com

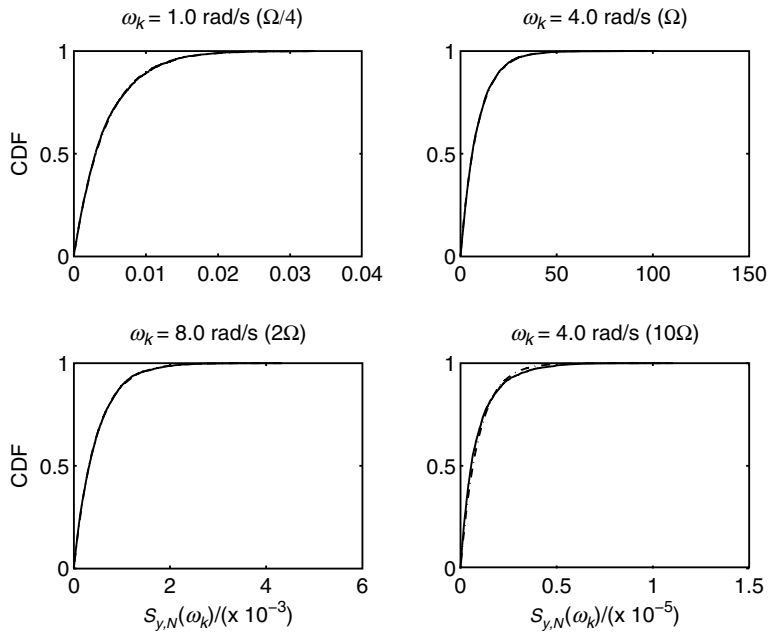

Figure 3.6 CDFs of $S_{y,N}(\omega_k)$ at different frequencies by simulation and the Chi-square approximation (displacement)

Figure 3.6 refers to displacement measurements and shows the comparison between the cumulative distribution function (CDF) of $S_{y,N}(\omega_k)$ obtained using: (i) simulation by the same 4000 simulation runs used for Figure 3.2 (solid line) and (ii) the approximation of a Chi-square distribution with two degrees of freedom according to Equation (3.33) (dashed line). Four different values of ω_k are considered and they belong to different ranges of the spectrum. The Chi-square distribution provides satisfactory approximation in the lower frequency range but the accuracy deteriorates in the higher frequencies (e.g., $\omega_k > 5\tilde{\Omega}$). Figure 3.7 is similar to Figure 3.6 but it considers the absolute acceleration response subjected to ground motion:

$$\ddot{x}_{abs} = \ddot{x} + \ddot{x}_g = -\frac{c}{m}\dot{x} - \frac{k}{m}x \tag{3.66}$$

where $\ddot{x}$ and $\ddot{x}_g$ are the relative acceleration of the mass and the ground acceleration, respectively. In this case the assumption of a Chi-square distribution with two degrees-of-freedom is a good approximation in the higher frequency range while this accuracy deteriorates as one moves to lower frequencies (e.g., $\omega_k < 0.25\tilde{\Omega}$). It is worth noting that for the case of nonzero prediction error (i.e., $\tilde{S}_{\epsilon 0} > 0$), the assumption of a Chi-square distribution provides satisfactory approximation over a larger range of frequencies than the noise-free case because the measurement noise pushes up the spectral density in an expected sense and this dilutes the effects of aliasing.

Figure 3.8 refers to displacement measurements and shows the correlation coefficients between $S_{y,N}(\omega_k)$ and $S_{y,N}(\omega_{k'})$, $k' = 0, 1, \ldots, N_{nqy}$, for four selected values of ω_k. The spectral density estimators in the lower frequency range (e.g., $\omega_k < 2\tilde{\Omega}$) can be considered uncorrelated, thus validating the use of Equation (3.58) with $0 < \omega_k, \omega_{k'} \leq 2\tilde{\Omega}$. However, when one moves to

www.EngineeringBooksPdf.com

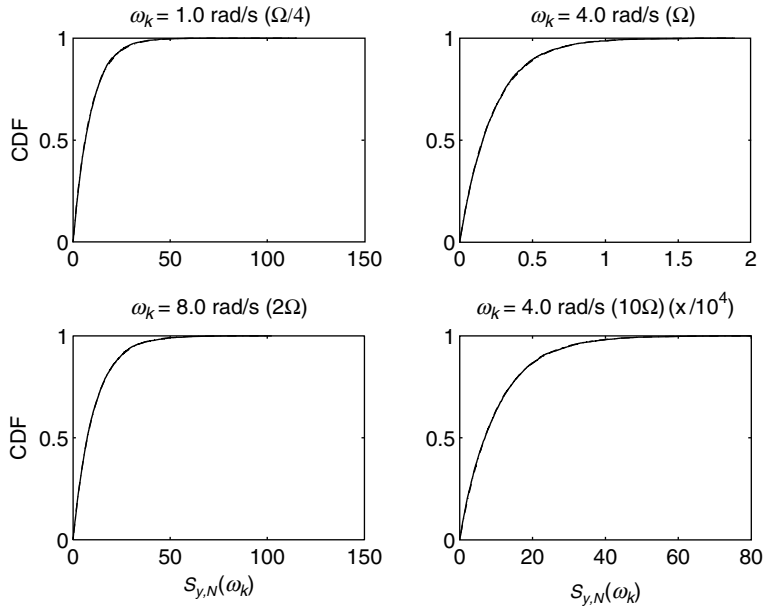

Figure 3.7 CDFs of $S_{y,N}(\omega_k)$ at different frequencies by simulation and the Chi-square approximation (acceleration)

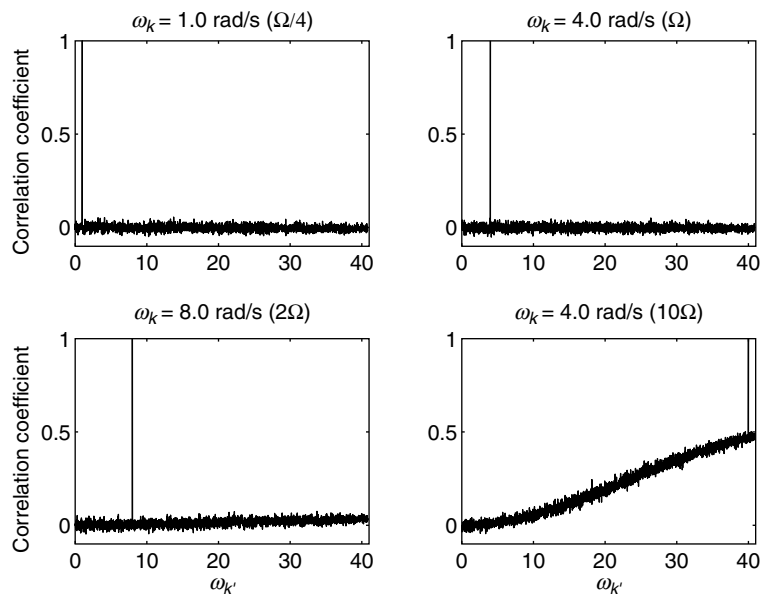

Figure 3.8 Correlation coefficients between $S_{y,N}(\omega_k)$ and $S_{y,N}(\omega_{k'})$ (displacement)

www.EngineeringBooksPdf.com

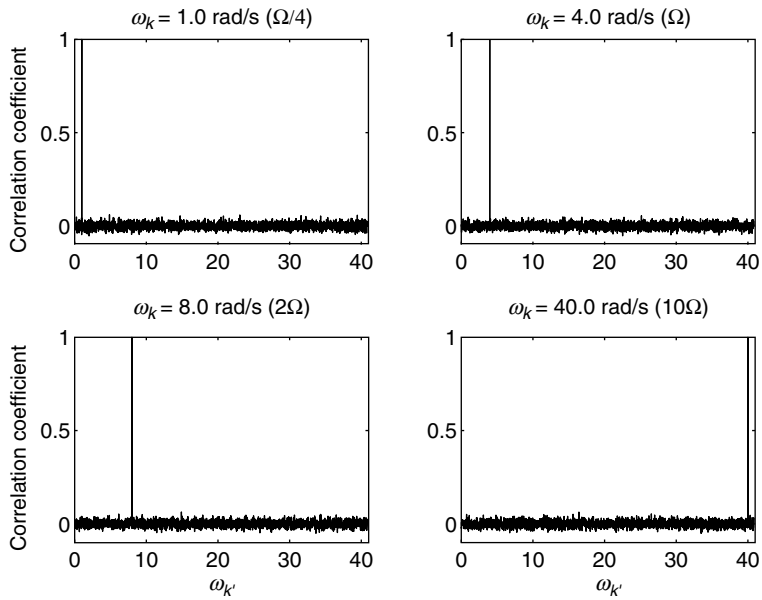

Figure 3.9 Correlation coefficients between $S_{y,N}(\omega_k)$ and $S_{y,N}(\omega_{k'})$ (acceleration)

the high frequencies, the correlation coefficients become large. Figure 3.9 is similar to Figure 3.8 but it considers the absolute acceleration measurements. In this case the spectral density estimators can be considered uncorrelated for the whole frequency range. Again, the presence of a prediction error will in general expand the range of frequencies over which the spectral estimates can be considered to be uncorrelated and, therefore, independent. The aliasing and leakage effects are the reasons for the dependence between the spectral values at different frequencies and they have more effect on the frequency range with smaller spectral values. Therefore, the region around the peaks in the spectrum is utilized for identification.

It will be shown later in this example that utilizing the whole range of frequencies of displacement measurement in Equation (3.61) (i.e., $\mathcal{K} = \{1, 2, \ldots, N_{nqy} - 1\}$) leads to bias and unreasonably small variances, especially for the damping ratio and spectral intensity. In order to obtain reasonable results, an appropriate frequency range is necessary and the recommendation in Section 3.3.3 can be utilized.

3.4.2 Identification with the Spectral Density Approach

To let the identification results solely depend on the contribution from the data, a non-informative prior is taken, i.e., $p(\boldsymbol{\theta}|\mathcal{C})$ is constant and it is absorbed into the normalizing constant. Table 3.1 refers to the identification results using a single set of displacement measurements $\mathcal{D}$. It shows the optimal values $\boldsymbol{\theta}^\star = [\Omega^\star, \zeta^\star, S^\star_{f0}]^T$, the calculated standard deviations σ_Ω, σ_ζ, and $\sigma_{S_{f0}}$, the coefficient of variation (COV) for each parameter and the *normalized distance* for each parameter. This distance represents the absolute value of the difference between

www.EngineeringBooksPdf.com

Table 3.1 Identification results for one set of data and frequency range $(0, 1.2\,\tilde{\Omega}]$

Parameter	Actual $\tilde{\theta}$	Optimal $\theta^\star$	S.D. σ	COV	$\frac{\lvert\tilde{\theta}-\theta^\star\rvert}{\sigma}$
Ω	4.0000	4.0046	0.0063	0.002	0.73
ζ	0.0100	0.0091	0.0015	0.155	0.59
S_{f0}	1.0000	1.0214	0.0393	0.039	0.54

the optimal and target value, normalized by the corresponding posterior standard deviation. Thus, it expresses how many standard deviations the identified value is away from its target value. Here, updating of the modal parameters is performed using only spectral estimates up to the frequency $1.2\tilde{\Omega}$ (i.e., $\mathcal{K} = \{1, 2, \ldots, 764\}$) in Equation (3.61) to ensure that the spectral estimates at different frequencies follow approximately the uncorrelated Chi-square distributions. The marginal updated PDF of the natural frequency Ω and the damping ratio ζ can be obtained by integrating out the spectral intensity. Its contours can be found by using the method in Appendix B and they are shown in Figure 3.10. The estimates of these two parameters can be considered independent. Similarly, Figure 3.11 shows the contours of the marginal updated PDF of the damping ratio ζ and the spectral intensity of the excitation S_{f0}. In contrast to the previous case, the estimates of these two parameters are correlated so the elliptical contours are rotated. Figure 3.12 shows the conditional PDFs of the natural frequency Ω and the damping ratio ζ with the spectral intensity fixed at its optimal value. The conditional PDFs by the Bayesian approach and the Gaussian approximation are plotted with solid lines and dashed

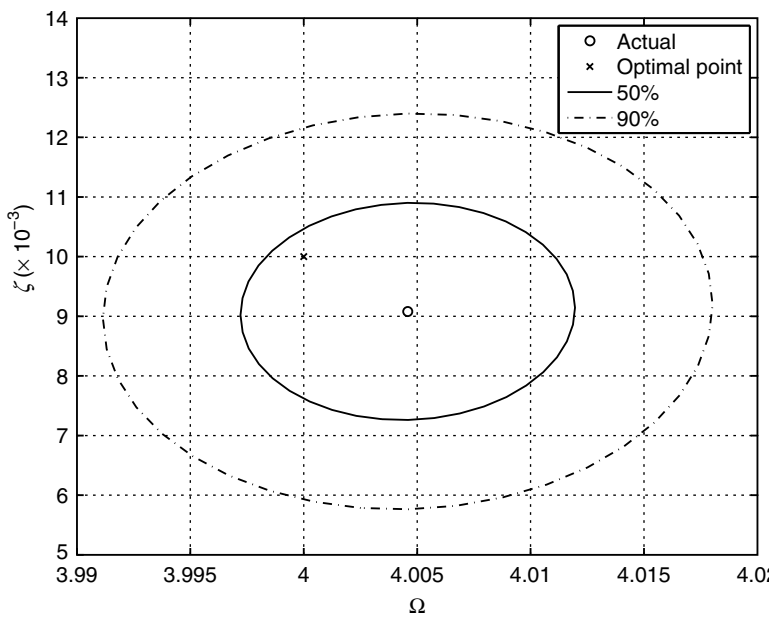

Figure 3.10 Contours of the marginal PDF of Ω and ζ

www.EngineeringBooksPdf.com

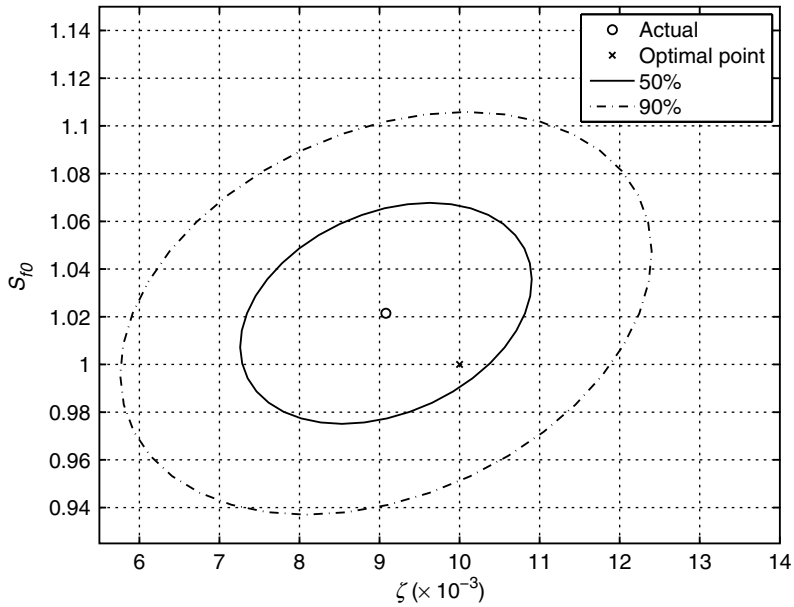

Figure 3.11 Contours of the marginal PDF of ζ and S_{f0}

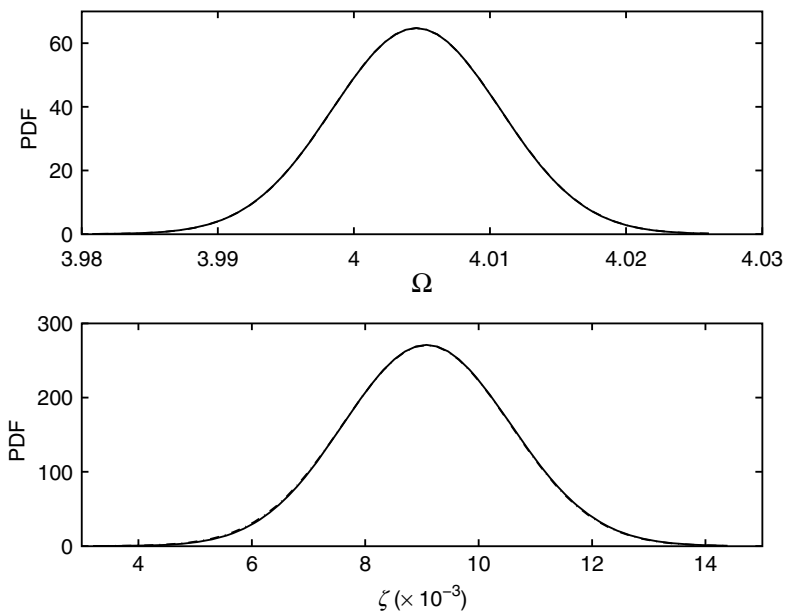

Figure 3.12 Conditional PDFs of Ω and ζ

www.EngineeringBooksPdf.com

lines, respectively, but the two sets of curves are on top on each other, indicating that the Gaussian approximation is accurate and can be used to represent the updated PDF, e.g., for statistical moments computation.

Table 3.2(a) shows the statistics of the optimal parameters obtained from five hundred independent displacement time histories generated using the same actual parameters in $\tilde{\boldsymbol{\theta}}$. Again, only the lower frequency range $(0, 1.2\tilde{\Omega}]$ was used in calculating the updated PDFs of the parameters. The optimal parameter vectors $\boldsymbol{\theta}^{\star(s)}$, $s = 1, 2, \ldots, 500$, were calculated for each time history. The sample average and standard deviation of the optimal parameters calculated from the set $\{\boldsymbol{\theta}^{\star(s)}, s = 1, 2, \ldots, 500\}$ are shown in the third and fourth columns, respectively. The fifth column shows the square root of the sample average of the five hundred different variances obtained by the Bayesian spectral density approach considering each sample response time history separately. It can be seen that the fourth and fifth columns are similar. This implies that the uncertainty of the modal parameters quantified from a single sample is representative for the uncertainty of the optimal parameters obtained from a number of independent data sets with equal length. Also, the uncertainty of the sample average can be quantified by the standard deviation, which is given by the standard deviation in the fourth column divided by $\sqrt{500}$.

Table 3.2(b) is the same as Table 3.2(a) except that the subset frequency range $[0.8\tilde{\Omega}, 1.2\tilde{\Omega}]$ was used in computing the updated PDFs of the parameters. The identification results are again reasonable. As in Table 3.2(a), the values of the fourth and fifth column are similar. It is worth noting that the calculated standard deviation of the parameter S_{f0} is much larger than the corresponding value in Table 3.1. Such increase of the uncertainty is expected in this case since only part of the available information is utilized. Furthermore, the loss of information contributes mainly for the identification of the excitation level but not the time-frequency structure of the signal. On the other hand, the uncertainty of the modal frequency and damping ratio remain almost unchanged, implying that most of the information regarding these two parameters is contained within the selected frequency range.

Table 3.2 Statistics of identification results using 500 sets of data and different frequency ranges

	Parameter	Actual $\tilde{\theta}$	Average of $\theta^\star$	S.D. of $\theta^\star$	$\sqrt{\overline{\sigma^2}}$
(a) $(0, 1.2\tilde{\Omega}]$	Ω	4.0000	4.0004	0.0063	0.0067
	ζ	0.0100	0.0102	0.0018	0.0017
	S_{f0}	1.0000	0.9967	0.0392	0.0385
(b) $[0.8\tilde{\Omega}, 1.2\tilde{\Omega}]$	Ω	4.0000	4.0003	0.0060	0.0068
	ζ	0.0100	0.0103	0.0018	0.0018
	S_{f0}	1.0000	0.9986	0.0745	0.0699
(c) Entire frequency range	Ω	4.0000	4.0031	0.0068	0.0062
	ζ	0.0100	0.0221	0.0117	0.0008
	S_{f0}	1.0000	0.9119	0.0782	0.0156

www.EngineeringBooksPdf.com

Table 3.2(c) shows the results by including all spectral points (except the zero and Nyquist frequency) in the updating process, i.e., $\mathcal{K} = \{1, 2, \ldots, N_{nqy} - 1\}$ in Equation (3.61). It is noted that in this case the identification results do not seem reasonable. In particular, the values in the fourth column are significantly larger than the ones in the fifth column, especially for the damping ratio and spectral intensity. This implies that the uncertainty of the parameters calculated from each time history underestimates the uncertainty of the parameter estimation and this can be explained as follows. When the entire frequency range is used in computing the likelihood function in Equation (3.61), it assumes that the information extracted from each spectral point $S_{y,N}(\omega_k)$ is new information and independent from that extracted from all other points. This is because Equation (3.61) assumes independence of the spectral estimates corresponding to all frequency points. However, as discussed earlier, this approximation of independence deteriorates in the higher frequency range. The effective information that can be extracted from the data should be much less, and the corresponding uncertainty much larger, than that implied by Equation (3.61) when the entire frequency range is utilized.

3.4.3 Identification with Small Amount of Data

Another case is investigated with a very short duration of measurement, namely $T = 60$ s, so it contains roughly 38 fundamental periods of the oscillator. The Bayesian spectral density approach is used for its identification with the frequency index set $\mathcal{K} = \{1, 2, \ldots, 45\}$. Figure 3.13 shows the conditional updated PDFs of Ω and ζ, with all other parameters fixed at their optimal values. It is obvious that the conditional PDFs are non-Gaussian so the Gaussian

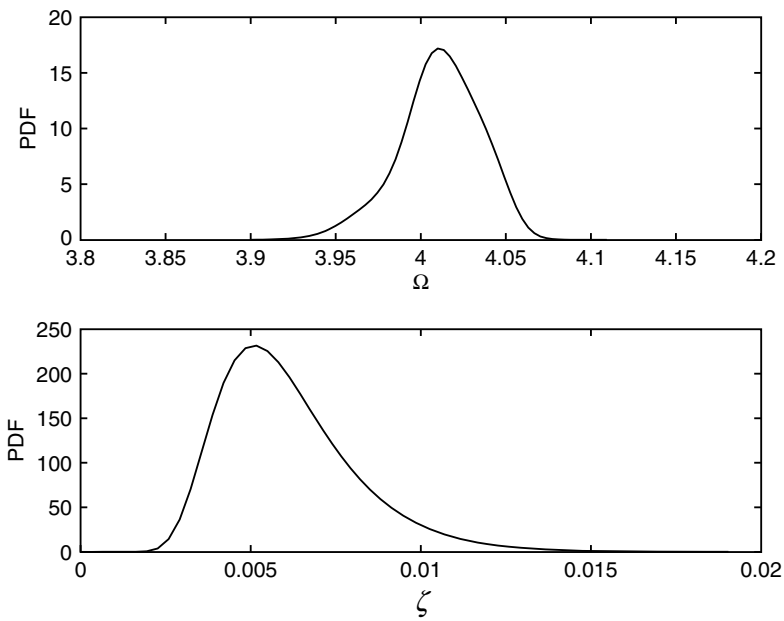

Figure 3.13 Conditional PDFs of Ω and ζ

www.EngineeringBooksPdf.com

approximation is accurate only when the number of data points is sufficiently large. On the other hand, the Bayesian spectral density approach is capable of offering the correct inference without assuming the type of the updated PDF. In the case of a non-Gaussian updated PDF, the statistical moments, such as the variances, can be computed by Monte Carlo simulation. Table 3.3 shows the most probable value (mode), mean and standard deviation for each parameter. Note that the most probable values and the mean values are not equal since the distribution is asymmetric.

Table 3.3 Identification results for one set of data with $T = 60$ s

Parameter	Actual $\tilde{\theta}$	Optimal $\theta^{\star}$	Mean	S.D. σ
Ω	4.0000	4.0108	4.0112	0.0247
ζ	0.0100	0.0051	0.0061	0.0020
S_{f0}	1.0000	0.8658	0.9055	0.1385

3.4.4 Concluding Remarks

The Bayesian spectral density approach for updating the probability density function of the model parameters for general multi-degree-of-freedom systems using wide-band response measurement was presented. The posterior PDF for the model parameters can be accurately approximated by a multi-variate Gaussian distribution when the number of data points is sufficiently large. This property provides an efficient way for quantifying the uncertainty. It should be noted that the Gaussianity approximation of the posterior PDF is not an assumption of the method but the consequence of a large number of data points. The calculated mean and covariance matrix of the posterior distribution offer an estimate of the optimal values of the model parameters and the associated uncertainty. Quantification of these uncertainties is very important if the identification results are utilized for further processing, e.g., for damage detection and updating of reliability. Estimation of the uncertainty does not require calculation of a number of optimal values, obtained from several sets of data, and then calculation of the statistics of these optimal estimates. The Bayesian spectral density approach can be applied also for non-Gaussian input/response [96] since discrete Fourier transformed signals are approximately Gaussian regardless of the distribution of the time-domain signal [230].

3.5 Optimal Sensor Placement

The Bayesian spectral density approach addresses the problem of parametric identification and quantification of the associated uncertainty using response measurement for a given configuration of sensor arrangement. The quality and accuracy of the experimental parameter estimates depend on the number and location of sensors in a structure. In this section, the problem of enhancing the quality of the model parameter estimation in relation to the location and number of sensors is investigated. There are two issues to be addressed. The first one deals with the problem of determining the optimal degrees of freedom to place a prescribed number of sensors in a structure in order to provide the least possible uncertainty of the identification

www.EngineeringBooksPdf.com

results with ambient structural response measurement. The second issue is to quantify the quality of parameter estimates as a function of the number of sensors placed in a structure. The answer to the latter issue is useful for making a cost-effective decision regarding the structural instrumentation and the choice of number of sensors to be placed in the structure.

Previous work regarding the first issue of optimally locating a given number of sensors in a structure has been carried out by Udwadia [262]. In this work, a rational statistical approach was developed based on the Fisher information matrix for the model parameters. Heredia and Esteva (1998) extended this work for the case of large model uncertainties expected in model updating [102]. In both methods, the optimal sensor configuration is the one that minimizes the expected Bayesian loss function, which is defined as the trace of the inverse of the Fisher information matrix, with respect to the sensor configuration [82]. However, only the first issue was considered for optimally locating a given number of sensors in a structure.

Information entropy is a natural single measure of the degree of uncertainty of random variables and random vectors [117, 233]. Papadimitriou *et al.* (2000) proposed to use it for optimal sensor placement since it best corresponds to the objective of structural testing, which is to minimize the uncertainty in the model parameters [196]. It immediately follows that the optimal sensor configuration is selected as the one which minimizes the information entropy. This entropy-based measure resolved the issue related to the arbitrariness in selecting an appropriate norm for the Fisher information matrix in the previous approaches. Moreover, the more important advantage of the information entropy measure is that it provides the basis for comparison between sensor configurations involving different numbers of sensors. This is particularly useful for trading off the cost of instrumentation with the information gained from additional sensors about the state of a structure, thus making cost-effective decisions regarding the optimal instrumentation.

The information entropy of the uncertain parameters $\boldsymbol{\theta}$ is defined as [148, 233]:

$$\begin{aligned} H_{\boldsymbol{\theta}}(\boldsymbol{\Delta}|\mathcal{S}_{\mathcal{K}}^{av}, \mathcal{C}) &= -E[\ln p(\boldsymbol{\theta}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\Delta}, \mathcal{C})] \\ &= -\int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\Delta}, \mathcal{C}) \ln p(\boldsymbol{\theta}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\Delta}, \mathcal{C}) \mathrm{d}\boldsymbol{\theta} \end{aligned} \tag{3.67}$$

where $E[.]$ denotes the mathematical expectation with respect to $\boldsymbol{\theta}$ according to the posterior distribution $p(\boldsymbol{\theta}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\Delta}, \mathcal{C})$, and $\boldsymbol{\Delta}$ is the sensor configuration vector of dimension N_d, comprised of zeros and ones with the position of the non-zero elements specifying the N_o measured DOFs. Therefore, the observation matrix $\mathbf{L}_o$ in Equation (3.44) depends on $\boldsymbol{\Delta}$. The information entropy will be used to assess the possible sensor configurations.

3.5.1 Information Entropy with Globally Identifiable Case

The updated PDF $p(\boldsymbol{\theta}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\Delta}, \mathcal{C})$ in Equation (3.67) is given by Equation (3.60) with Equation (3.61) for the general case of uncertain excitation. The formulation presented here is based on the spectral density estimators obtained from the measured data $\mathcal{D}$ and it depends on the class of structural, excitation and prediction-error models chosen to describe the system. The updated parameter vector $\boldsymbol{\theta}^{\star}$ is obtained by minimizing the objective function $J(\boldsymbol{\theta}) = -\ln p(\boldsymbol{\theta}|\mathcal{C}) p(\mathcal{S}_{\mathcal{K}}^{av}|\boldsymbol{\theta}, \boldsymbol{\Delta}, \mathcal{C})$ with the likelihood function $p(\mathcal{S}_{\mathcal{K}}^{av}|\boldsymbol{\theta}, \boldsymbol{\Delta}, \mathcal{C})$ given by Equation (3.61). Furthermore, the updated PDF of the model parameter vector $\boldsymbol{\theta}$ can be

www.EngineeringBooksPdf.com

approximated [136] by a Gaussian distribution $\mathcal{G}(\boldsymbol{\theta}; \boldsymbol{\theta}^\star, \mathcal{H}(\boldsymbol{\theta}^\star|\boldsymbol{\Delta})^{-1})$ with mean $\boldsymbol{\theta}^\star \equiv \boldsymbol{\theta}(\boldsymbol{\Delta})^\star$ and covariance matrix $\mathcal{H}(\boldsymbol{\theta}^\star|\boldsymbol{\Delta})^{-1}$, where $\mathcal{H}(\boldsymbol{\theta}^\star|\boldsymbol{\Delta})$ denotes the Hessian of the objective function, evaluated at $\boldsymbol{\theta}^\star$. The (l, l') element of the Hessian matrix $\mathcal{H}(\boldsymbol{\theta}^\star|\boldsymbol{\Delta})$ is given by:

$$\mathcal{H}^{(l,l')}(\boldsymbol{\theta}^\star|\boldsymbol{\Delta}) = -\frac{\partial^2}{\partial\theta_l \partial\theta_{l'}} \ln p(\boldsymbol{\theta}|\mathcal{C})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} + N_s \sum_{k\in\mathcal{K}} \left[\frac{\partial^2}{\partial\theta_l \partial\theta_{l'}} \left(\ln \left| E[\mathbf{S}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}] \right| + \mathrm{tr}\{(E[\mathbf{S}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}])^{-1} \mathbf{S}^{av}_{y,N}(\omega_k)\} \right) \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \tag{3.68}$$

The first term on the right hand side vanishes for a non-informative prior distribution that is absorbed into the normalizing constant. In this case, the information entropy can be simplified as follows:

$$H_{\boldsymbol{\theta}}(\boldsymbol{\Delta}|\mathcal{S}^{av}_{\mathcal{K}}, \mathcal{C}) = \frac{1}{2} N_{\boldsymbol{\theta}} [\ln(2\pi) + 1] - \frac{1}{2} \ln \left| \mathcal{H}(\boldsymbol{\theta}^\star|\boldsymbol{\Delta}) \right| \tag{3.69}$$

where $N_{\boldsymbol{\theta}}$ is the number of uncertain parameters in $\boldsymbol{\theta}$. A larger value of the information entropy indicates a higher level of uncertainty of the random variables. In the extreme case where the Hessian matrix is closed to singular (close to an unidentifiable case), the information entropy tends to infinity.

3.5.2 Optimal Sensor Configuration

In a model identification methodology, the information about the parameters of the model is provided by the measured data. The amount of the information depends on the sensor configuration which clearly affects the updated PDF $p(\boldsymbol{\theta}|\mathcal{S}^{av}_{\mathcal{K}}, \boldsymbol{\Delta}, \mathcal{C})$ of the model parameters and, consequently, the uncertainty in the parameter estimates. The sensor configuration should be selected such that the resulting measured data are most informative about the condition of the structure or, equivalently, induces the least possible uncertainty for the parameter estimates. Therefore, the optimal sensor configuration is selected to be the one with minimal information entropy. Although this is a well-posed discrete optimization problem, there is a major complication arising from the explicit dependence of the information entropy on the structural response measurement which is not available at the initial stage of designing the experiment. However, this difficulty can be overcome by considering the limiting case of large number of data sets N_s instead of using Monte Carlo simulation, which is computationally demanding.

In globally identifiable case with large number of data points, the updated PDF $p(\boldsymbol{\theta}|\mathcal{S}^{av}_{\mathcal{K}}, \boldsymbol{\Delta}, \mathcal{C})$ can be well approximated by a Gaussian distribution with a mean equal to the most probable value $\boldsymbol{\theta}^\star$ and covariance matrix equal to the inverse of the Hessian $\mathcal{H}(\boldsymbol{\theta}^\star|\boldsymbol{\Delta})^{-1}$. Moreover, for a large number of data sets, the spectral density estimators $\mathbf{S}^{av}_{y,N}(\omega_k)$ can be replaced, using the law of large numbers, by its expected value $E[\mathbf{S}_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]$ given in Equation (3.48). Substituting the asymptotic values of the spectral density estimators into Equation (3.68), the resulting Hessian matrix $\mathcal{H}(\boldsymbol{\theta}^\star|\boldsymbol{\Delta}, \mathcal{S}^{av}_{\mathcal{K}}) = \mathcal{H}(\boldsymbol{\theta}^\star|\boldsymbol{\Delta})$ can be obtained for large N and the dependence on the details of the measurement is removed. The resulting Hessian matrix $\mathcal{H}(\boldsymbol{\theta}^\star|\boldsymbol{\Delta})$ and the posterior PDF of the parameter vector $\boldsymbol{\theta}$ do no longer depend

www.EngineeringBooksPdf.com

on the data explicitly. The only dependence of $\mathcal{H}(\boldsymbol{\theta}^\star|\boldsymbol{\Delta})$ on the data comes implicitly through $\boldsymbol{\theta}^\star = \boldsymbol{\theta}(\boldsymbol{\Delta}, \mathcal{S}_{\mathcal{K}}^{av})^\star$. Consequently, the information entropy for a given set of data is completely defined by the optimal value $\boldsymbol{\theta}(\boldsymbol{\Delta}, \mathcal{S}_{\mathcal{K}}^{av})^\star$ of the model parameters computed for the given data while the spectral density details of the measured data do not enter in the estimation.

However, since the data is not available in the design process, estimation of the optimal parameters $\boldsymbol{\theta}^\star$ cannot be obtained at this stage. In order to proceed with the design of the optimal sensor configuration, this estimate has to be assumed. In practice, useful designs can be obtained by taking the optimal parameters $\boldsymbol{\theta}^\star$ to be the nominal parameter vector $\boldsymbol{\theta}^\eta$ chosen by the designer. In this case of a Gaussian distribution, the information entropy takes the following form for large N:

$$H_{\boldsymbol{\theta}}(\boldsymbol{\Delta}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\theta}^\eta, \mathcal{C}) = \frac{1}{2} N_{\boldsymbol{\theta}}[\ln(2\pi) + 1] - \frac{1}{2} \ln \left| \mathcal{H}(\boldsymbol{\theta}^\eta|\boldsymbol{\Delta}) \right| \tag{3.70}$$

This depends on the sensor configuration vector in $\boldsymbol{\Delta}$ and the chosen parameters $\boldsymbol{\theta}^\eta$ of the nominal model. The optimal sensor configuration given the nominal model is chosen as the one that minimizes this information entropy measure over all possible configurations.

3.5.3 Robust Information Entropy

The optimal sensor configuration, obtained by minimizing Equation (3.70), depends on the designer's choice of the nominal model determined by the nominal parameter vector $\boldsymbol{\theta}^\eta$. One way to account for the uncertainty in the nominal model is to use a prescribed PDF $p(\boldsymbol{\theta}^\eta|\mathcal{C})$ for $\boldsymbol{\theta}^\eta$. In this case, the optimal sensor configuration becomes the one that minimizes the robust information entropy $E_{\boldsymbol{\theta}'}[H_{\boldsymbol{\theta}}(\boldsymbol{\Delta}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\theta}', \mathcal{C})]$ which is a measure of the overall uncertainty in both $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^\eta$:

$$\begin{aligned} E_{\boldsymbol{\theta}'}[H_{\boldsymbol{\theta}}(\boldsymbol{\Delta}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\theta}', \mathcal{C})] &= \int_{\boldsymbol{\Theta}} H_{\boldsymbol{\theta}}(\boldsymbol{\Delta}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\theta}', \mathcal{C}) p(\boldsymbol{\theta}'|\mathcal{C}) \mathrm{d}\boldsymbol{\theta}' \\ &= \frac{1}{2} N_{\boldsymbol{\theta}}[\ln(2\pi) + 1] - \frac{1}{2} \int_{\boldsymbol{\Theta}} \ln \left| \mathcal{H}(\boldsymbol{\theta}'|\boldsymbol{\Delta}) \right| \, p(\boldsymbol{\theta}'|\mathcal{C}) \, \mathrm{d}\boldsymbol{\theta}' \end{aligned} \tag{3.71}$$

The information entropy $H_{\boldsymbol{\theta}}(\boldsymbol{\Delta}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\theta}^\eta, \mathcal{C})$ given by Equation (3.70) is a special case of the robust information entropy $E_{\boldsymbol{\theta}'}[H_{\boldsymbol{\theta}}(\boldsymbol{\Delta}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\theta}', \mathcal{C})]$ in Equation (3.71) and it corresponds to the choice $p(\boldsymbol{\theta}'|\mathcal{C}) = \delta(\boldsymbol{\theta}' - \boldsymbol{\theta}^\eta)$, where $\delta(.)$ denotes the Dirac delta function. The multi-dimensional integration over $\boldsymbol{\theta}'$ involved in Equation (3.71) can be carried out by Monte Carlo simulation or by an asymptotic expansion developed for these types of integrals [197].

The present formulation for optimal sensor placement in terms of the information entropy provides a rational procedure for comparing the uncertainty of the estimates of the parameter values for different sensor configurations. Specifically, a direct measure of the uncertainty reduction is provided by the change of the information entropy:

$$\Delta H(\boldsymbol{\Delta}; \boldsymbol{\Delta}_{ref}) = E_{\boldsymbol{\theta}'}[H_{\boldsymbol{\theta}}(\boldsymbol{\Delta}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\theta}', \mathcal{C})] - E_{\boldsymbol{\theta}'}[H_{\boldsymbol{\theta}}(\boldsymbol{\Delta}_{ref}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\theta}', \mathcal{C})] \tag{3.72}$$

where $E_{\boldsymbol{\theta}'}[H_{\boldsymbol{\theta}}(\boldsymbol{\Delta}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\theta}', \mathcal{C})]$ and $E_{\boldsymbol{\theta}'}[H_{\boldsymbol{\theta}}(\boldsymbol{\Delta}_{ref}|\mathcal{S}_{\mathcal{K}}^{av}, \boldsymbol{\theta}', \mathcal{C})]$ represent, respectively, the robust information entropy for a sensor configuration $\boldsymbol{\Delta}$ and the reference sensor configuration $\boldsymbol{\Delta}_{ref}$. The reference sensor configuration $\boldsymbol{\Delta}_{ref}$ may correspond to different number of sensors from

www.EngineeringBooksPdf.com

that in the configuration vector $\mathbf{\Delta}$. Equivalently, the reduction in uncertainty can also be quantified by the parameter-uncertainty ratio defined by [196, 296]:

$$\frac{\sigma}{\sigma_{ref}} = \exp\left[\frac{\Delta H(\mathbf{\Delta}; \mathbf{\Delta}_{ref})}{N_{\boldsymbol{\theta}}}\right] \tag{3.73}$$

where σ and σ_{ref} represent the measure of the spread of the posterior PDF of the model parameters for the sensor configurations $\mathbf{\Delta}$ and $\mathbf{\Delta}_{ref}$, respectively. The smaller the value of the parameter uncertainty ratio, the smaller the uncertainty in the parameter estimates, and hence the better identification quality, with $\mathbf{\Delta}$. For more details and examples of the information entropy-based method, please refer to Yuen *et al.* [296].

3.5.4 Discrete Optimization Algorithm for Suboptimal Solution

For a dynamical system with N_d degrees of freedom and N_o sensors, the total number of possible sensor configurations is:

$${}_{N_d}C_{N_o} = \frac{N_d!}{N_o!(N_d - N_o)!} \tag{3.74}$$

which is huge for large N_d and $N_o >> 1$. As a result, an exhaustive search for the optimal configuration may be computationally prohibitive and genetic algorithms [91] can be used to solve this discrete optimization problem. It is particularly useful for this application since it is desired to obtain the measurements from different DOFs being complementary in terms of the information for different modes or different parameters.

Even though genetic algorithms provide a feasible solution for this optimization problem, the following sub-optimal algorithm is an alternative. In order to place N_o sensors to a structure, the problem of placing only one sensor is first considered. In this case, exhaustive search for the N_d possible DOFs can be proceeded and use d_1 to denote the optimal one. Then, optimization problem is continued by adding one more sensor in additional to the one at the $d\text{th}_1$ DOF. In other words, an exhaustive search is carried out for the following $N_d - 1$ possible configurations: $(d_1, 1), (d_1, 2), \ldots, (d_1, d_1 - 1), (d_1, d_1 + 1), \ldots, (d_1, N_d)$ and use d_2 to denote the optimal DOF for the second sensor. This procedure is continued until all the N_o DOFs are found. The total number of configurations to be examined is:

$$N_d + (N_d - 1) + \cdots + (N_d - N_o + 1) = \frac{N_o(2N_d - N_o + 1)}{2} \tag{3.75}$$

which is significantly smaller than $N_d!/N_o!(N_d - N_o)!$. For example, if $N_d = 1000$ and $N_o = 10$, the proposed algorithm requires evaluation of $10(2000 - 10 + 1)/2 = 9955$ configurations but exhaustive search requires $1000!/10!990! \approx 2.6 \times 10^{23}$.

Previous studies showed that the optimal DOFs for N_o sensors do not necessarily include the optimal DOFs for $N_o - 1$ sensors. Therefore, it is emphasized that the sensor configuration obtained by this algorithm is in general not optimal but it is expected to be a suboptimal. More importantly, this suboptimal sensor configuration is *more robust* than the optimal one. In actual practice of modal testing or health monitoring, malfunctioning of some of the sensors is commonly encountered. In this case, the suboptimal configuration may perform better than the optimal configuration when there are a few missing observed DOFs.

www.EngineeringBooksPdf.com

3.6 Updating of a Nonlinear Oscillator

The first application is concerned with the nonlinear *Duffing oscillator* of known mass $M = 1.0$ kg subjected to zero-mean stationary Gaussian white noise f with spectral intensity S_{f0}:

$$M\ddot{x}(t) + C\dot{x}(t) + K_1 x(t) + K_3 x(t)^3 = f(t) \tag{3.76}$$

This oscillator has a nonlinear restoring force: $-K_1 x(t) - K_3 x(t)^3$. A stationary displacement response time history $\mathcal{D}$ was generated with parameter vector $\tilde{\boldsymbol{\theta}} = [\tilde{C}, \tilde{K}_1, \tilde{K}_3, \tilde{S}_{f0}^{(1)}, \tilde{\sigma}_\epsilon^{(1)}]^T$ where $\tilde{C} = 0.1$ kg/s, $\tilde{K}_1 = 5.0$ N/m, $\tilde{K}_3 = 1.0$ N/m^3, $\tilde{S}_{f0}^{(1)} = 0.02$ N^2 s and $(\tilde{\sigma}_\epsilon^2)^{(1)} = 0.0026$ m^2 (15% noise). The sampling time interval is $\Delta t = 0.05$ s, with total measurement duration $T = 400$ s, so the total number of data points is $N = 8000$.

There is no closed-form solution for the auto-correlation function or the power spectral density function of this nonlinear system. Therefore, an equivalent linear system is utilized to obtain the approximated mean of the spectral density estimator. Multiplying Equation (3.76) with $x(t-\tau)$ and taking expectation yields:

$$M\ddot{R}_x(\tau) + C\dot{R}_x(\tau) + K_1 R_x(\tau) + K_3 E[x(t-\tau)x(t)^3] = 0 \tag{3.77}$$

where $R_x(\tau) \equiv E[x(t-\tau)x(t)]$, $\forall t \in \mathbb{R}$, is the auto-correlation function. The term $E[x(t-\tau)x(t)^3]$ can be approximated by neglecting the fourth cumulant term [161]:

$$E[x(t-\tau)x(t)^3] \approx 3\sigma_x^2 R_x(\tau) \tag{3.78}$$

where $\sigma_x^2 = R_x(0)$ is the response variance. Therefore, the differential equation for the approximated response auto-correlation function is readily obtained:

$$M\ddot{R}_x(\tau) + C\dot{R}_x(\tau) + (K_1 + 3\sigma_x^2 K_3)R_x(\tau) = 0 \text{ with } R_x(0) = \sigma_x^2 \text{ and } \dot{R}_x(0) = 0 \tag{3.79}$$

It is a second-order linear ordinary differential equation with constant coefficients, and its analytical solution is given by:

$$R_x(\tau) = \sigma_x^2 \mathrm{e}^{-\zeta_{eq}\Omega_{eq}|\tau|}\left[\cos\left(\Omega_{eq}\sqrt{1-\zeta_{eq}^2}\tau\right) + \frac{\zeta_{eq}}{\sqrt{1-\zeta_{eq}^2}}\sin\left(\Omega_{eq}\sqrt{1-\zeta_{eq}^2}|\tau|\right)\right] \tag{3.80}$$

where Ω_{eq} and ζ_{eq} are the natural frequency and damping ratio of the equivalent linear system and they are given by:

$$\Omega_{eq} = \sqrt{\frac{K_1 + 3\sigma_x^2 K_3}{M}} \tag{3.81}$$

and

$$\zeta_{eq} = \frac{C}{2M\Omega_{eq}} \tag{3.82}$$

www.EngineeringBooksPdf.com

which depend on the response variance σ_x^2. Furthermore, for linear systems, it is well known that the response variance is [157]:

$$\sigma_x^2 = \frac{\pi S_{f0}}{2M\zeta_{eq}\Omega_{eq}^3} \tag{3.83}$$

By solving Equations (3.81)–(3.83), the response variance of the Duffing oscillator can be approximated by:

$$\sigma_x^2 = \frac{K_1}{6K_3}\left(\sqrt{1 + \frac{12\pi K_3 S_{f0}}{CK_1^2}} - 1\right) \tag{3.84}$$

Then, for a given parameter vector $\boldsymbol{\theta}$, the mean spectral density estimator $E[S_{y,N}(\omega_k)|\boldsymbol{\theta}, \mathcal{C}]$ can be obtained by using Equation (3.28). Finally, the updated PDF $p(\boldsymbol{\theta}|\mathcal{S}_{\mathcal{K}}^{(1)}, \mathcal{C})$ is readily obtained using Equation (3.36) with Equation (3.35), where the prior PDF $p(\boldsymbol{\theta}|\mathcal{C})$ is taken as constant over the region where $p(\mathcal{S}_{\mathcal{K}}^{(1)}|\boldsymbol{\theta}, \mathcal{C})$ is large, i.e., a locally non-informative prior PDF [31]. Figure 3.14 shows the zigzag spectrum estimated by the measured time history. The peak in this spectrum is around 2.2 to 2.3 rad/s, which is consistent with the small-amplitude natural frequency of the oscillator, which is $\sqrt{5}/2\pi$ Hz $= 0.356$ Hz $= 2.24$ rad/s. The frequency index set is taken to be $\mathcal{K} = \{1, 2, \ldots, 300\}$, i.e., up to 0.75 Hz or 4.71 rad/s. The smooth curve shows the theoretical mean spectrum of the equivalent linear system. This confirms that the equivalent linear system approximates the Duffing oscillator well in this level of excitation.

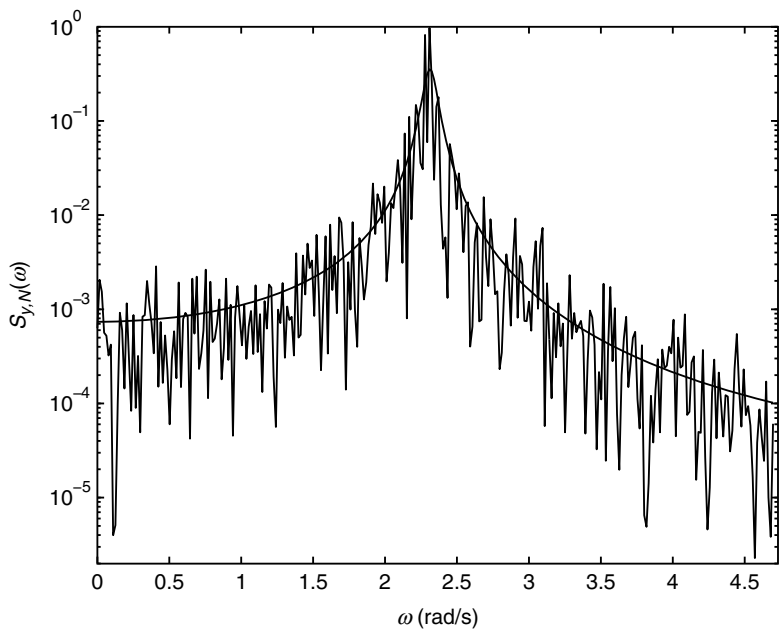

Figure 3.14 Response spectrum of the Duffing oscillator

www.EngineeringBooksPdf.com

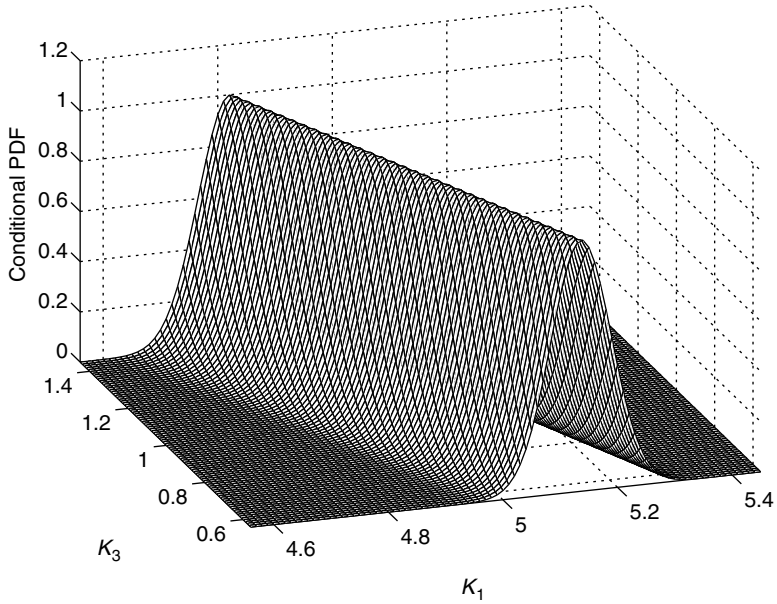

Figure 3.15 Conditional updated PDF

Figure 3.15 shows the conditional updated PDF $p(K_1, K_3|\mathcal{S}_{\mathcal{K}}^{(1)}, \tilde{C}, \tilde{S}_{f0}^{(1)}, (\tilde{\sigma}_\epsilon^2)^{(1)}, \mathcal{C})$ normalized in such a way that the peak value is unity. It is clearly seen that this case is unidentifiable, i.e., given one set of dynamic data, the estimates of K_1 and K_3 suffer from large uncertainty as there are infinitely many combinations of K_1 and K_3 which give similar values for the updated PDF [284].

Another response time history $\mathcal{D}^{(2)}$ was generated for the same oscillator (the same $\tilde{C}$, $\tilde{K}_1$ and $\tilde{K}_3$) for a higher level of excitation: $\tilde{S}_{f0}^{(2)} = 0.04\,\mathrm{N}^2\,\mathrm{s}$ and $(\tilde{\sigma}_\epsilon^2)^{(2)} = 0.0069\,\mathrm{m}^2$ (15% noise). This case is, again, unidentifiable. The updated PDF is plotted together with the previous one in Figure 3.16 and the trajectories of the peaks in the (K_1, K_3) plane have different slopes. By Equation (3.79), the equivalent linear system has a stiffness $K_1 + 3\sigma_x^2 K_3$ so different Duffing oscillators with $K_1 + 3\sigma_x^2 K_3 = K$ (a constant) are associated with the same equivalent linear stiffness. The coefficient $3\sigma_x^2$ depends on the level of excitation S_{f0}, showing that different levels of excitation lead to different slopes of the peak trajectories in the (K_1, K_3) plane. Since the coefficient $3\sigma_x^2$ is always positive, the slope of the peak trajectories in the (K_1, K_3) plane is always negative. This is expected because a smaller value of K_1 can be compensated by a larger value of K_3, and vice versa.

Figure 3.16 suggests that if two dynamic data sets $\mathcal{D}^{(1)}$ and $\mathcal{D}^{(2)}$ are utilized simultaneously, the information from $\mathcal{D}^{(2)}$ is complimentary to $\mathcal{D}^{(1)}$ to provide an extra mathematical constraint for the uncertain parameters, especially for K_1 and K_3. The updated PDF using both sets of data is given by the product of the individuals. As a result, the identification problem will become globally identifiable. Table 3.4 shows the updated values $\boldsymbol{\theta}^\star = [C^\star, K_1^\star, K_3^\star, S_{f0}^{(1)\star}, S_{f0}^{(2)\star}, \sigma_\epsilon^{2(1)\star}, \sigma_\epsilon^{2(2)\star}]^T$ and the calculated standard deviations σ_c, σ_{K_1},

www.EngineeringBooksPdf.com

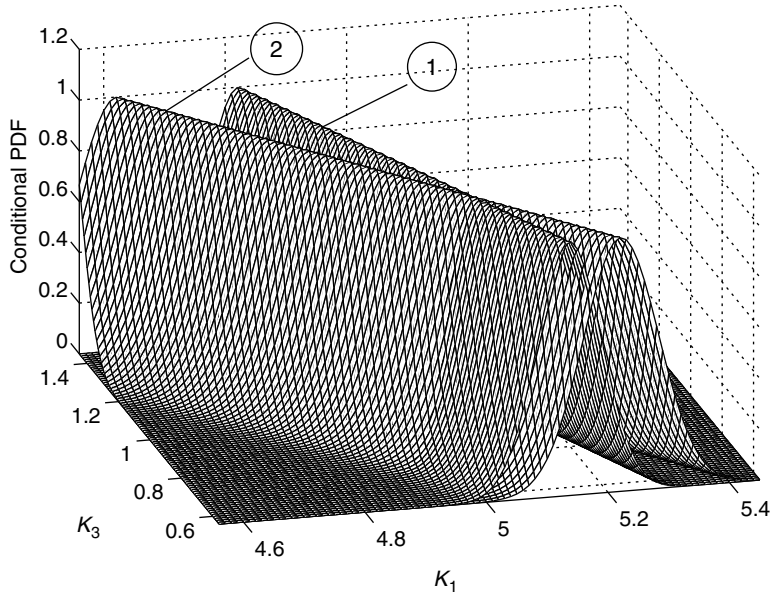

Figure 3.16 Conditional updated PDFs

σ_{K_3}, $\sigma_{S_{f0}^{(1)}}$, $\sigma_{S_{f0}^{(2)}}$, $\sigma_{\sigma_\epsilon^{2(1)}}$ and $\sigma_{\sigma_\epsilon^{2(2)}}$ obtained using both data sets. It also gives the coefficient of variation (COV) for the parameter estimates and the associated normalized distance. This normalized distance represents the absolute value of the difference between the identified optimal value and exact value, normalized with respect to the corresponding calculated standard deviation. The normalized distances in Table 3.4 are between 0.2 to 2.1, suggesting that the procedure is not producing biased estimates. This verifies that the errors are not unusually large nor unreasonably small, compared to the calculated standard deviations.

The optimal estimates of the model parameter vector $\boldsymbol{\theta}$ are not sensitive to the choice of the cutoff frequency, i.e., 0.75 Hz in this case, as long as it is larger than the frequency at which the peak of the response spectral density estimates occurs. Identification using the same sets of

Table 3.4 Comparison of the actual parameters versus the optimal estimates and their statistics for the Duffing oscillator

Parameter	Actual $\tilde{\theta}$	Optimal $\theta^\star$	S.D. σ	COV	$\frac{\lvert\tilde{\theta}-\theta^\star\rvert}{\sigma}$
K_1	5.0000	4.9380	0.1238	0.025	0.50
K_3	1.0000	0.9267	0.2863	0.309	0.26
C	0.1000	0.0866	0.0185	0.214	0.72
$S_{f0}^{(1)}$	0.0200	0.0190	0.0016	0.086	0.64
$S_{f0}^{(2)}$	0.0400	0.0476	0.0037	0.077	2.08
$(\sigma_\epsilon^2)^{(1)}$	0.0026	0.0019	0.0030	1.632	0.26
$(\sigma_\epsilon^2)^{(2)}$	0.0069	0.0055	0.0066	1.205	0.22

www.EngineeringBooksPdf.com

data was also carried out with 10.0 Hz (the Nyquist frequency in this case). The results were virtually the same as those using 0.75 Hz except that there were significant reductions in the uncertainty of the noise levels. In other words, utilizing a larger frequency range gives better estimates for the noise level only. In this case, the frequency range is chosen to include all nonzero frequencies up to $1.5\omega_\rho$ where ω_ρ is the frequency at which the peak of the spectral estimates $S_{y,N}(\omega_k)$ occurs. It is computationally efficient to use such a range without sacrificing the quality of the identification for the model parameter vector $\boldsymbol{\theta}$.

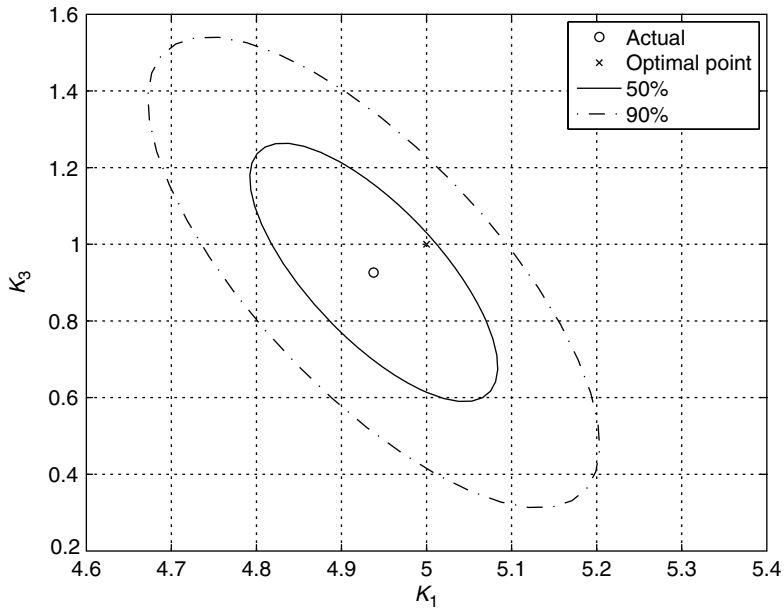

Figure 3.17 Contours of the marginal PDF for K_1 and K_3

Figure 3.17 shows the contours of the marginal PDF of K_1 and K_3. It is not surprising that there is negative correlation on the estimates of these two parameters since the equivalent linear stiffness of the system is contributed by the combination of K_1 and K_3. In this case, the uncertainty for K_3 is significantly larger than that of K_1 since the level of the excitation is mild and the nonlinear cubic term is difficult to be identified.

In order to have better identification results for the nonlinear term, the spectral density of the excitation of the second set of data is increased to $S_{f0}^{(2)} = 0.06$ N^2 s so that the response is of higher level of nonlinearity. Figure 3.18 shows the contours of the marginal PDFs using two sets of data with $S_{f0}^{(1)} = 0.02$ N^2 s and $S_{f0}^{(2)} = 0.06$ N^2 s. It can be clearly seen that the uncertainty of K_3 is significantly reduced since the data contain more information about the nonlinear behavior. However, the optimal value of K_3 deviates from the actual value.

Then, the excitation level is further increased and identification is proceeded again with two sets of measurements with $S_{f0}^{(1)} = 0.02$ N^2 s and $S_{f0}^{(2)} = 0.08$ N^2 s. Figure 3.19 shows the contours of the marginal PDFs. The posterior uncertainty of K_3 is further reduced but its updated value departs further from the actual value and this can be considered as biased. The

www.EngineeringBooksPdf.com

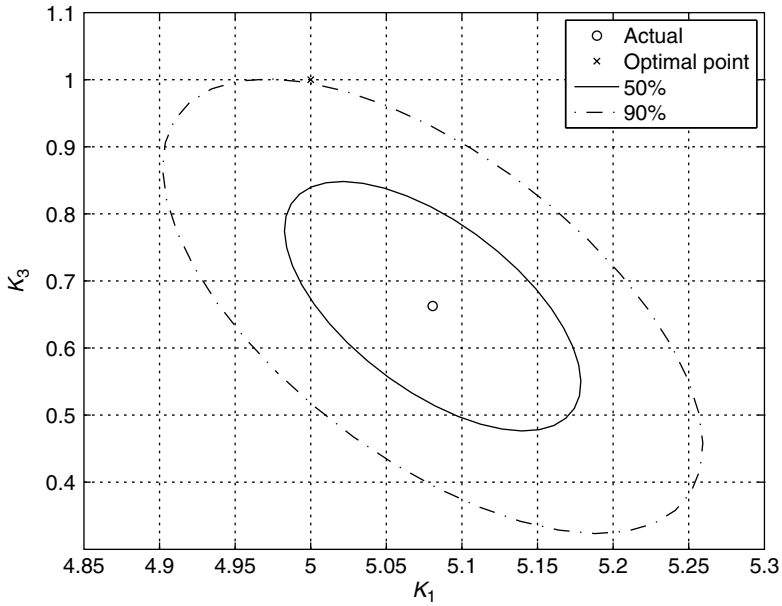

Figure 3.18 Contours of the marginal PDF for K_1 and K_3 with $S_{f0}^{(1)} = 0.02$ N^2 s and $S_{f0}^{(2)} = 0.06$ N^2 s

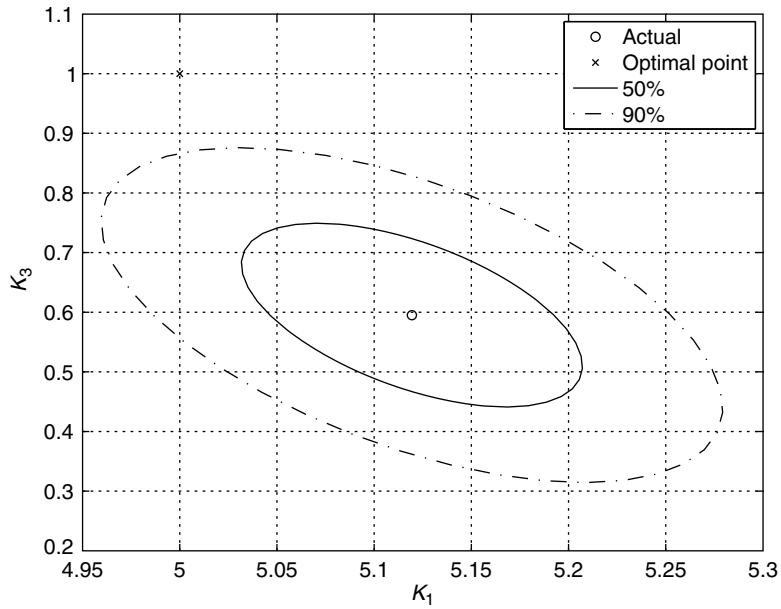

Figure 3.19 Contours of the marginal PDF for K_1 and K_3 with $S_{f0}^{(1)} = 0.02$ N^2 s and $S_{f0}^{(2)} = 0.08$ N^2 s

www.EngineeringBooksPdf.com

reason is that the equivalent linear system governed by Equation (3.80) does not satisfactorily reflect the original nonlinear system in terms of the time–frequency structure. The error of the cumulant approximation in Equation (3.78) induces a modeling error in the spectral density estimator and this contributes to the bias of the identification results.

Figure 3.20 shows the response spectrum for $S_{f0} = 0.08\ \mathrm{N}^2\ \mathrm{s}$. Even though the theoretical mean spectrum of the equivalent linear system agrees well with the spectral density estimator computed using the measurement, the positions of their peaks are slightly different. Therefore, for nonlinear systems subjected to moderate to strong excitation, a more accurate computational method for the response spectrum is necessary. An alternative is to obtain the theoretical spectrum by simulation but this is a computationally demanding task.

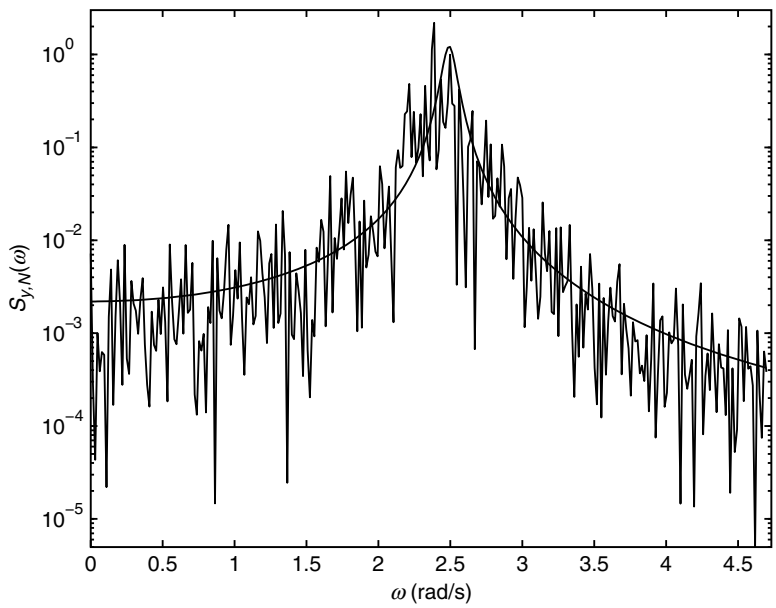

Figure 3.20 Response spectrum of the Duffing oscillator with $S_{f0} = 0.08\ \mathrm{N}^2\ \mathrm{s}$

3.7 Application to Structural Behavior under Typhoons

3.7.1 Problem Description

A tropical cyclone is a common natural phenomenon formed by an unstable overlying atmosphere. In the classification by the Regional Specialized Meteorological Centers, a tropical cyclone with maximum 10-min averaged wind speed over 118 km/h is defined as a typhoon, which is equivalent to category one or above in the Saffir–Simpson hurricane scale. According to the Joint Typhoon Warning Center, there were on average seventeen tropical cyclones per year reaching the typhoon intensity level in the northwestern Pacific Ocean and most of them occurred in the summer. Incidentally, there are many big cities with high

www.EngineeringBooksPdf.com

population and numerous high-rise buildings in the typhoon profolic region so these buildings have to be frequently exposed to this severe aerodynamic condition, especially in the summer. Furthermore, due to the unsteady characteristics of the wind speed and wind pressure, buildings have to withstand strong wind pressure in various directions, which induce many challenging problems for analysis, design and construction [241]. Therefore, it is valuable to investigate the behavior of infrastructures during typhoons in order to have a better understanding for improving future designs and developing more reliable structural health monitoring systems.

Parallel to the remarkable revolution of data acquisition hardwares in recent decades [247], there has also been enormous development on the methodologies in model updating and structural health monitoring. As a result, structural health monitoring for large infrastructures by full-scale field measurement becomes feasible. Abnormal reduction of the modal frequencies of a structure infers loss of structural stiffness so modal frequencies are commonly utilized as the diagnostic indicators. Previous work by Salawu (1997) summarized that a 5% reduction in the modal frequencies indicates probable structural damage [227]. On the other hand, the damping ratio is another indirect indicator that provides information about the status of energy dissipation of a structure. The physical mechanism of energy dissipation is complex and involves a number of factors, such as structural configuration, interfacial mechanism, materials composition, and aerodynamic effects, etc. Therefore, a large uncertainty is encountered in the modeling and quantification. Kareem and Sun (1990) compared several classical methodologies and concluded that the coefficient of variation in the estimation of damping ratios can be up to 80% [131]. Substantial changes of the modal frequencies and damping ratios were observed in recent studies on several landmarked infrastructures during typhoons. For example, Li *et al.* (2000) compared the structural response of a hollow truss tall building system under four typhoons [154], while Xu *et al.* (2003) analyzed the structural behavior of a high-rise reinforced concrete building under typhoon York, which had the strongest strength and longest duration among all the typhoons which have affected Hong Kong since 1983 [276]. These studies also demonstrated that typhoons can produce temporary or permanent effects to structures and more investigations are necessary to understand the typhoon effects to structural systems.

In this study, the acceleration response time histories of the East Asian Hall in Macao were recorded under the passage of two strong typhoons, named Kammuri and Kuri, in August 2008. Globally speaking, the two typhoons had similar tracks in the northwestern Pacific Ocean. They passed through the north of the Philippines, entered the south China sea and landed at the southern coast of China. However, Kammuri passed through the south side of Macao but Nuri the north side. Since the typhoons in the northern hemisphere always spin counterclockwise due to the Coriolis effect, completely different aerodynamic conditions were generated onto the building due to the small variations in their paths. The influences of the structural behavior have also been investigated. The Bayesian spectral density approach is applied to analyze the full-scale field measurement of the East Asia Hall. The modal parameters of the structure and the excitation are identified for different time interval during the typhoons in order to investigate the wind impact on the building. It is observed that the modal frequencies of the instrumented building were reduced substantially during the two typhoons, but the temporary reduction recovered almost instantly after the typhoons dissipated. Moreover, the correlation between the damping ratios and the intensity of the excitation is discussed.

www.EngineeringBooksPdf.com

3.7.2 Meteorological Information of the Two Typhoons

The meteorological information including the wind speed and wind direction is regularly monitored by the Meteorological and Geophysical Bureau of Macao, namely Direcção dos Serviços Meteorológicos e Geofísicos (SMG) in Portuguese. The station is only 1.25 km away from the East Asian Hall. It is important to emphasize that the Bayesian spectral density approach requires output-only measurements. The meteorological information provided by the SMG is the 1-min averaged wind speed and this sampling time step is too large for computing the response time histories of the building. This meteorological information is used for reference only but does not participate in the identification process.

3.7.2.1 Typhoon Kammuri

Kammuri was the 9th tropical cyclone in the western Pacific Ocean and the 3rd typhoon affecting Macao in 2008. The tropical depression was formed on 3 August 2008 at longitude 21.0 N and latitude 105.2 E. Its highest 10-min averaged wind speed and lowest pressure were 95 km/h and 975 hPa (or 975 mbar = 97.5 kPa), respectively. Departing from the northern Philippines, the intensity of the Kammuri kept on increasing and it was classified as a typhoon by the SMG at 13:00 on 4 August (GMT + 08:00). Ten hours later, it reached the severe tropical storm level with gale signal no. 8 and lasted for 13 h. The 'eye' of this typhoon 'flitted' approximately 90 km south from Macao and it moved slowly at 9 km/h only. Finally, the intensity decreased rapidly after it landed at the Guangxi province on 7 August. The official typhoon signals from the SMG, the corresponding wind speed interval, the gust speed and the passage time are listed in Table 3.5. The wind speeds shown in this table are the observed values at the SMG station but not the maximum wind speed of the typhoon. Its track and other details are available on the website of the SMG (http://www.smg.gov.mo/e_index.php). The 10-min averaged wind speed (dots) and wind direction (crosses) are shown in Figure 3.21 starting from 19:00 on 5 August since the wind speed in the period of signal no. 1 was low in this case. Here, the wind direction counts from the north clockwise. From example, 90° implies an easterly wind. When the Kammuri was approaching Macao, the wind direction was around 20° from the north and it remained so until the highest wind speed was reached. Then, it changed rapidly from 20° to 130° within 7 h and the wind speed dropped to 20 km/h only before the typhoon signal was cancelled.

Table 3.5 Typhoon signals announced by SMG

Typhoon signal	10-min averaged wind speed (km/h)	Gust speed (km/h)	Hoisted time (yy/mm/dd GMT + 08:00)	
			Kammuri	Nuri
1	< 41	—	13:00 04/08	20:00 20/08
3	(41, 62)	>110	20:15 05/08	00:30 22/08
8	(63, 117)	>180	07:00 06/08	12:00 22/08
3	(41, 62)	>110	20:00 06/08	00:30 23/08
0	< 41	—	04:00 07/08	11:00 23/08

www.EngineeringBooksPdf.com

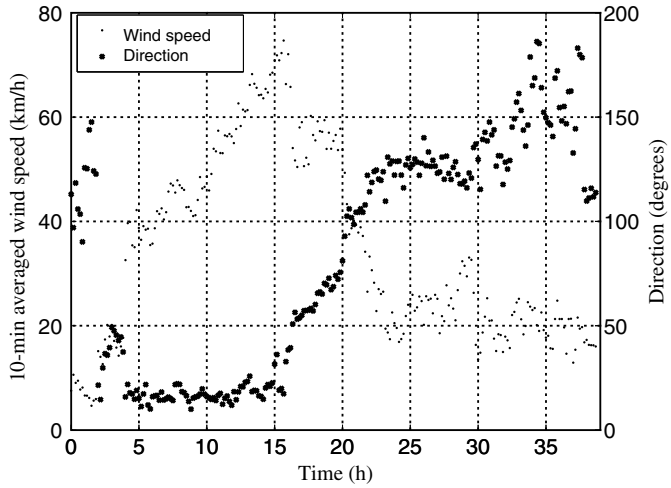

Figure 3.21 Wind characteristics of the Kammuri

3.7.2.2 Typhoon Nuri

After the Kammuri, there were two weeks of calm weather before another typhoon, Nuri, affected Macao from 20 to 23 August 2008. Its track and other information are also available on the website of the SMG. The official typhoon signals from the SMG with the corresponding wind and gust speed interval and the duration are also listed in Table 3.5. Nuri was formed as a tropical depression on August 17. After making the first landing in the Philippines, it was rapidly intensified to a tropical storm on the next day and gradually approached the southern China coast. Then, it 'lashed' straight to Hong Kong. After landing in the Pearl River delta region, the Nuri dissipated rapidly and its wind speed dropped back to 'calm level.' The track was similar to the Kammuri, except that it passed through the northeast side of Macao. The maximum 10-min averaged wind speed and the lowest pressure reached 140 km/h and 955 hPa, respectively, so the strength of the Nuri was stronger than the Kammuri. With a four-day duration, the typhoon signal of the Nuri was 'cancelled' after it landed in Hong Kong. The wind characteristics including the 10-min averaged wind speed and the wind direction of the Nuri are shown in Figure 3.22 starting from 19:00 on 20 August. Even though the strength of the Nuri was higher than the Kammuri, the observed wind speed generated in Macao by the Nuri was lower than that by the Kammuri. Moreover, the main wind direction of the Nuri covered all 360° because its track was closer to Macao. The 10-min averaged wind speed increased from less than 10 km/h to approximately 60 km/hr and the wind direction remained 'on north' around the peak of the wind speed. Afterwards, the wind speed dropped rapidly and the wind direction changed drastically to an almost opposite direction because the Nuri landed in Hong Kong, which is very close to Macao. After the 52nd hour, the typhoon signal of the Nuri was downgraded to signal no. 3 and the wind speed dropped to 20 km/h.

www.EngineeringBooksPdf.com

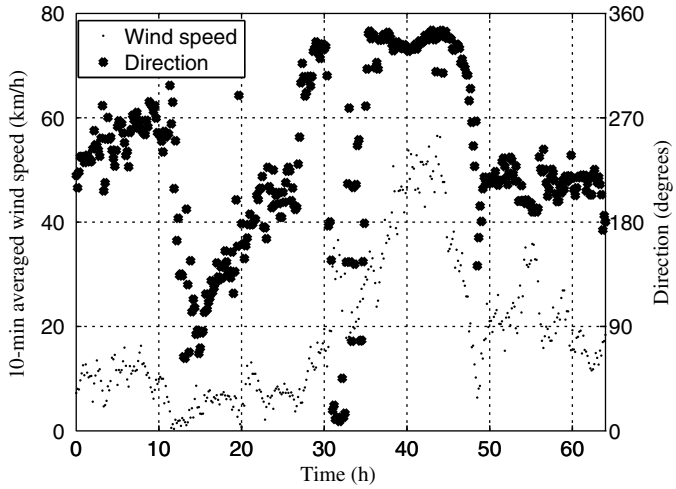

Figure 3.22 Wind characteristics of the Nuri

3.7.3 Analysis of Monitoring Data

3.7.3.1 Instrumentation Arrangement and Measurements

The East Asian Hall, depicted in Figure 3.23, is a 22-story 64.70 m reinforced concrete building. It has an L-shape floor layer with asymmetric spans of 51.90 m and 61.75 m. In contrast to

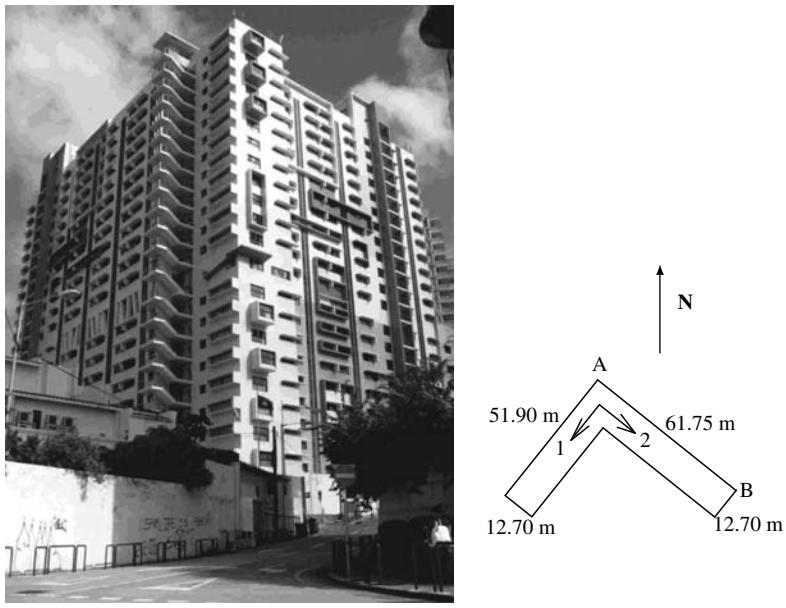

Figure 3.23 Picture and plan view of the East Asia Hall

www.EngineeringBooksPdf.com

most of the previously monitored buildings with rectangular or circular floor plans and large height-to-width aspect ratios [154, 155, 276], the East Asian Hall has an L-shape floor plan and a small aspect ratio. Therefore, the structural behavior cannot be approximated by a flexural beam. Because of the particular geometry, the structural dynamic behavior induced by the wind pressure fluctuation is more complex. Field measurements were collected with two state-of-the-art servo-accelerometers based on a standard exploration geophone spring–mass system, and the sensitivity was selected to be 50 V/g. The accelerometers were installed at the junction of the two spans in two orthogonal directions on the 18th floor, which is 53.5 m from the ground level.

Acceleration time histories were obtained for the whole duration of these two typhoons with a sampling frequency of 500 Hz. In order to demonstrate the change of vibration level, the response rms of each 10-min interval is computed for the Kammuri and is depicted in Figure 3.24. The 10-min averaged wind speed is also shown in the lower part of the figure for comparison. When the signal no. 8 was 'hoisted', the acceleration response was approximately ten times of that under calm situation. One may wonder why the peaks of the response rms and the wind speed were several hours apart. When a building is 'attacked' by a typhoon, both the magnitude of the wind speed and the wind direction affect the structural response. According to the Task Committee on Wind Forces, Committee on Loads and Stresses, Structural Division, ASCE [260], the drag and lift forces on an L-shape prism vary significantly with different attacking angles. By Figure 3.21, the wind direction changed rapidly from 20° to 130° within 7 h after the maximum wind speed was reached. By Figures 3.21 and 3.24, the maximum structural response occurred between the 17th and 19th hour and the attacking angle was in the range from 60° to 75° from the north with the 'eye' of the typhoon at (21.4 N, 113.3 E). Meanwhile, the wind speed was around 55–60 km/h, which was significantly lower

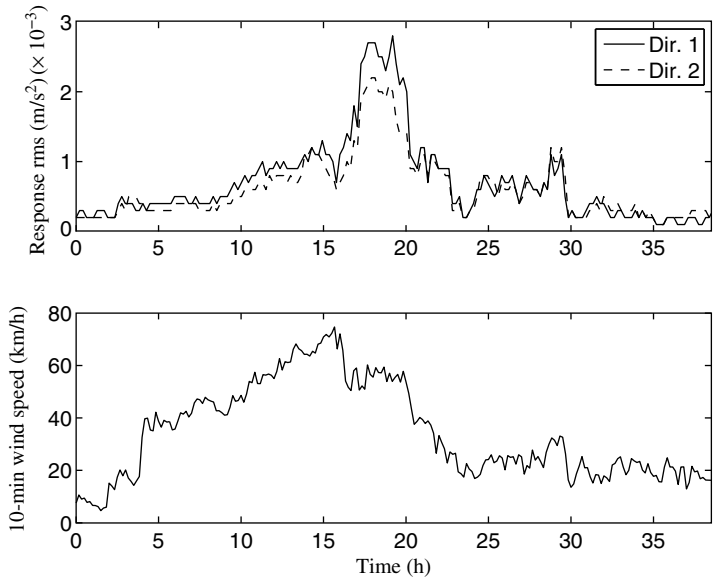

Figure 3.24 Variation of the response rms (Kammuri)

www.EngineeringBooksPdf.com

the maximum wind speed (75 km/h) that occurred at the 16th hour. This will be discussed further in the following section. The same plot is also shown in Figure 3.25 for the Nuri. The time difference between the peak wind speed and the peak response rms is not obvious in this case since the wind field is completely different as the Nuri was about to land after generating the highest wind speed to Macao.

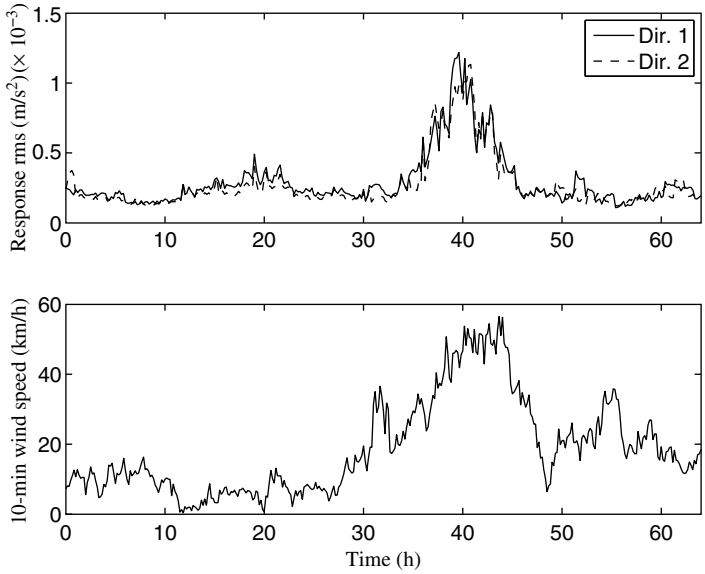

Figure 3.25 Variation of the response rms (Nuri)

The acceleration spectra can be obtained by Equation (3.25) and these are shown in Figure 3.26 for different phases of the Kammuri. Except for the case of signal no. 8, the spectra have high values around zero frequency and decrease until about 0.5 Hz. This is consistent with some well-known wind spectra, such as the Davenport spectrum or Harris spectrum. However, this cannot be observed in the spectrum under signal no. 8 and it suggests that the time–frequency content of the wind excitation in the storm level is significantly different from the others. Furthermore, it is observed that the low-frequency spectral values are all in the order of 10^{-6} m^2 s^{-3}, regardless of the level of the typhoon. Another noteworthy point is that the first peak in the first measured channel does not appear in the second channel. This implies that the mode shape component associated with the second measured channel is roughly zero for the first mode.

3.7.3.2 Identification of Modal Parameters

Since the wind pressure is nonstationary if the whole duration time window is considered, the structural response is also nonstationary [47]. Therefore, special treatments have to be taken in the identification process. In this study the acceleration measurement of the first 10 min of each half-hour interval is used for identification in order to investigate any possible change

www.EngineeringBooksPdf.com

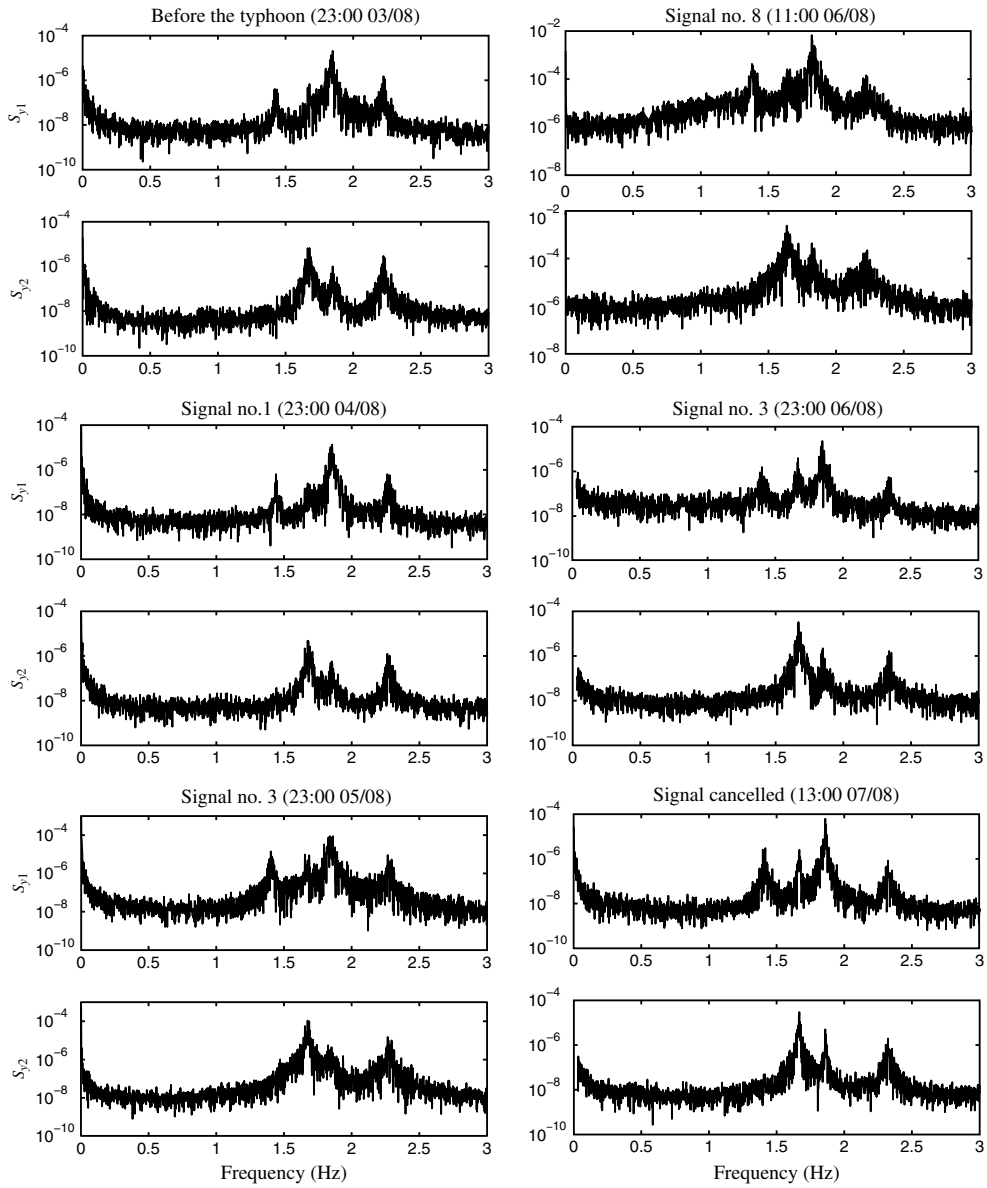

Figure 3.26 Spectra of the measurements for different typhoon signal levels (Kammuri)

of the structural properties in different phases of the typhoons. For each 10-min record, it is partitioned into four sets with an equal time duration, i.e., $N_s = 4$, and then the averaged spectrum can be obtained for identification. The reason for this arrangement is that even though the response is nonstationary in the whole time window, it can be considered approximately stationary in each of the 2.5-min interval. According to the acceleration spectra in Figure 3.26,

www.EngineeringBooksPdf.com

the first four modes have the most significant contribution to the structural response and they are considered for identification in this study, i.e., $N_m = 4$.

The diagonal elements of the excitation spectral intensity matrix describes the energy of the corresponding modes. The variations of their identified values are shown in Figures 3.27 and 3.28 for the Kammuri and Nuri, respectively. Note that comparing the spectral intensities of the modal forces for the different modes is meaningless since their values depend on the normalization of the mode shapes as discussed in Section 3.3.1.2. However, the time variation of the spectral intensity of a mode indicates the variation of the excitation energy with respect

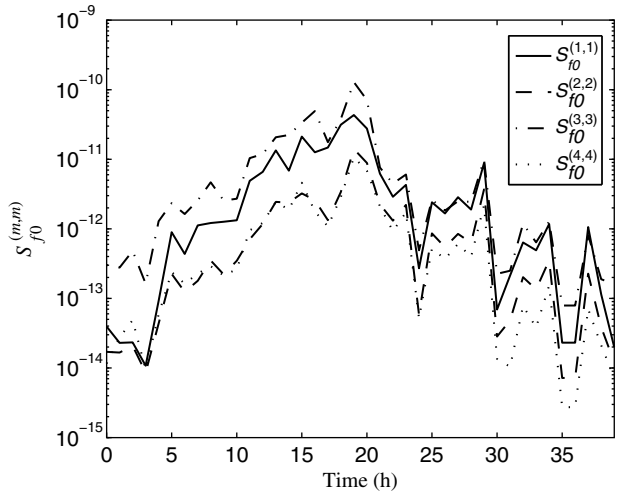

Figure 3.27 Variation of the identified modal spectral intensity of the excitation by the Kammuri

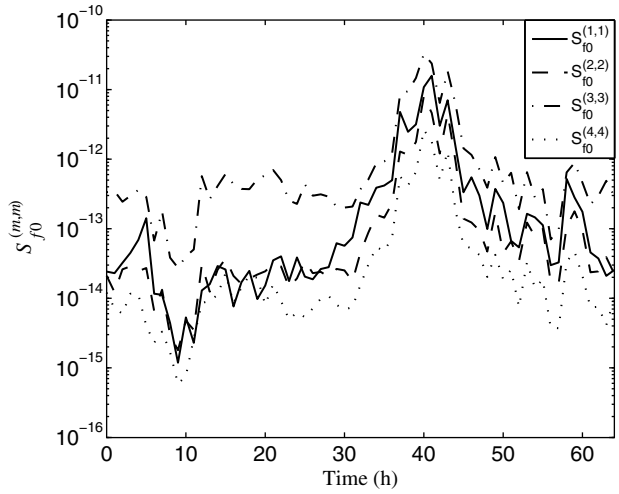

Figure 3.28 Variation of the identified modal spectral intensity of the excitation by the Nuri

www.EngineeringBooksPdf.com

to time. For the Nuri, the peak values of the wind speed and the spectral intensity of the modal forces occurred simultaneously at around the 40th hour, which shows that the structure experienced the maximum excitation during the highest wind speed. However, in the case of the Kammuri, the peaks of the spectral intensities of the modal forces (19th hour) do not occur at the same time with the maximum wind speed (16th hour). This echoes with the discrepancies between the peaks of the acceleration rms and the wind speed.

The modal frequencies of the building are also identified. They correspond to the equivalent linear system of the building since this structure may exhibit nonlinear behavior under this level of excitation. Figure 3.29 shows the variation of the modal frequencies with the associated plus and minus three standard derivations ($\pm 3\sigma$) confidence intervals during the typhoon Kammuri. This interval includes a probability of 99.7% for the equivalent modal frequency falling in this range since the posterior PDF is approximately Gaussian. When the intensity of the excitation

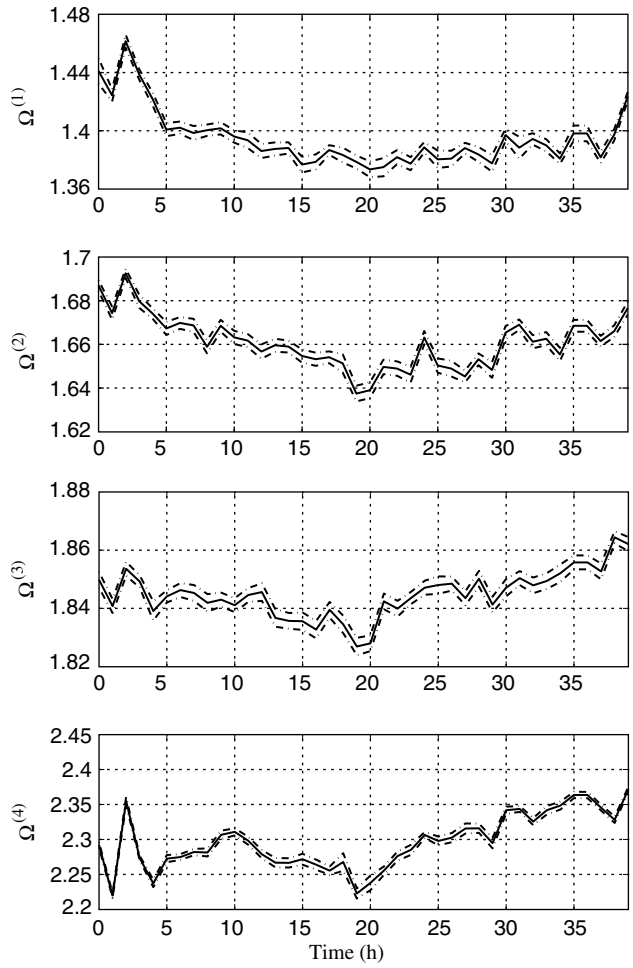

Figure 3.29 Variation of the identified modal frequencies (Kammuri)

www.EngineeringBooksPdf.com

increased, the modal frequencies had an opposite trend. As the typhoon gradually dispersed, the modal frequencies of the building recovered to their initial levels. A similar behavior was also observed during the Nuri. The largest reduction occurred in the first mode by approximately 5% for the Kammuri, and 4% for the Nuri. The confidence interval confirms that the change of the modal frequency was due to change of the structural behavior but not statistical uncertainty. This implies a 10% reduction of the stiffness of the building during the Kammuri but this was recovered immediately after the typhoon signal was cancelled. Furthermore, the restoring force–displacement relationship of the building is softening as the structural stiffness decreases for increasing excitation levels.

Similarly, Figures 3.30 and 3.31 show the variation of the damping ratios and mode shape components, respectively, of the first four modes. The damping ratios and the mode shape components did not show an obvious trend by observing these figures. For the first mode, the

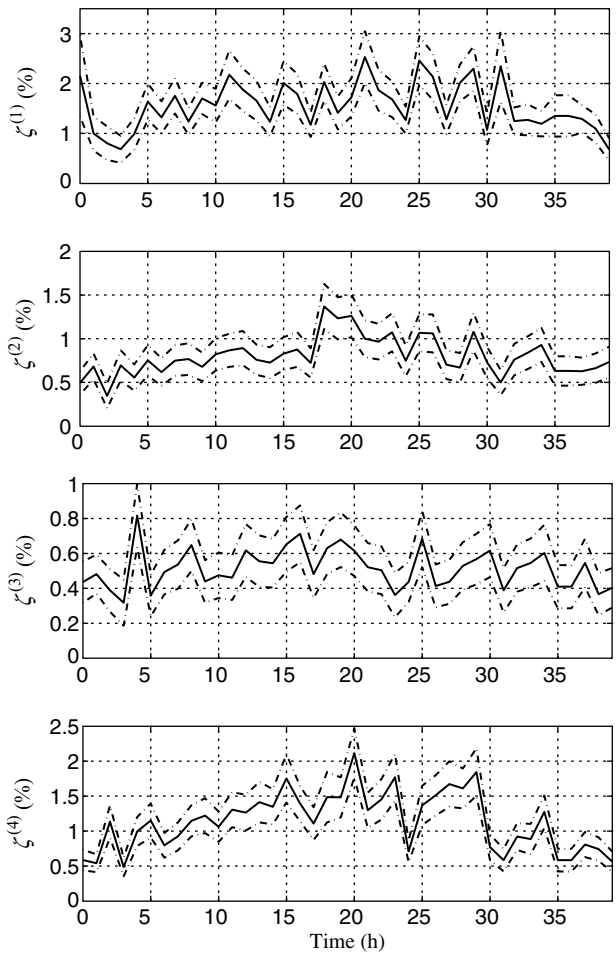

Figure 3.30 Variation of the identified damping ratios (Kammuri)

www.EngineeringBooksPdf.com

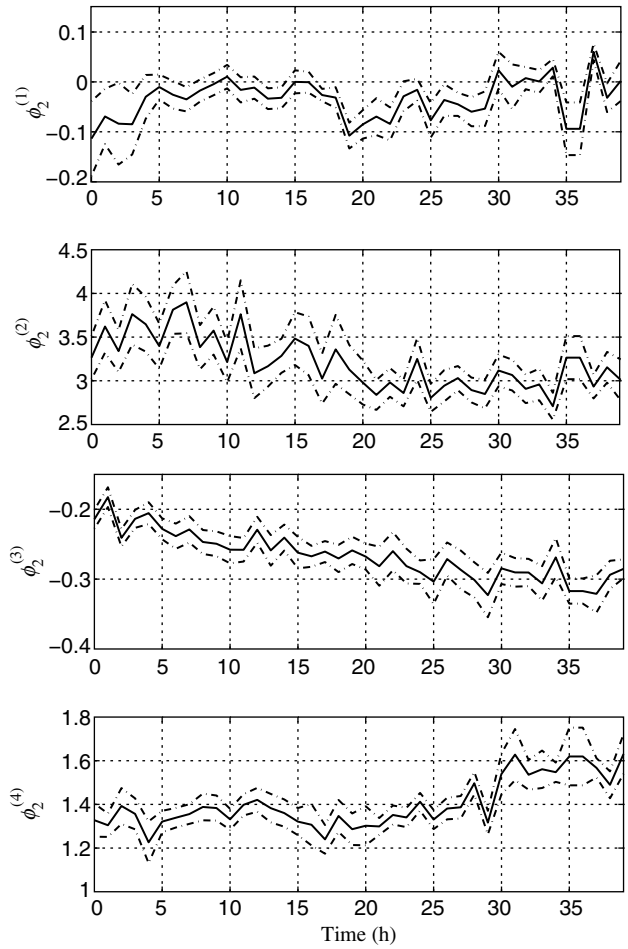

Figure 3.31 Variation of the identified mode shape components (Kammuri)

mode shape component in the second direction is roughly zero. This is consistent with the observation from the acceleration spectra that the first mode did not appear. The average of the mode shape components in the second direction were 3.2, −0.25 and 1.4 for the second to fourth mode. The aspect ratio of the building is roughly 1.0 so its behavior is different from a beam.

Figure 3.32 shows the semi-log scatter plot of the modal frequencies and the corresponding spectral intensities of the modal forces. Due to the possible nonlinear characteristics of the structure, the equivalent modal frequencies decreased with an increasing value of the corresponding modal force. There were obvious negative linear correlations for the first three modes. However, for the fourth mode, the identified values with different typhoons followed different trend lines with different slopes. This indicates that the fourth modal frequency is sensitive to some other factors, such as structural interaction properties, or other ambient conditions (temperature, rainfall and humidity, etc.). Figure 3.33 shows a similar plot for the relationship

www.EngineeringBooksPdf.com

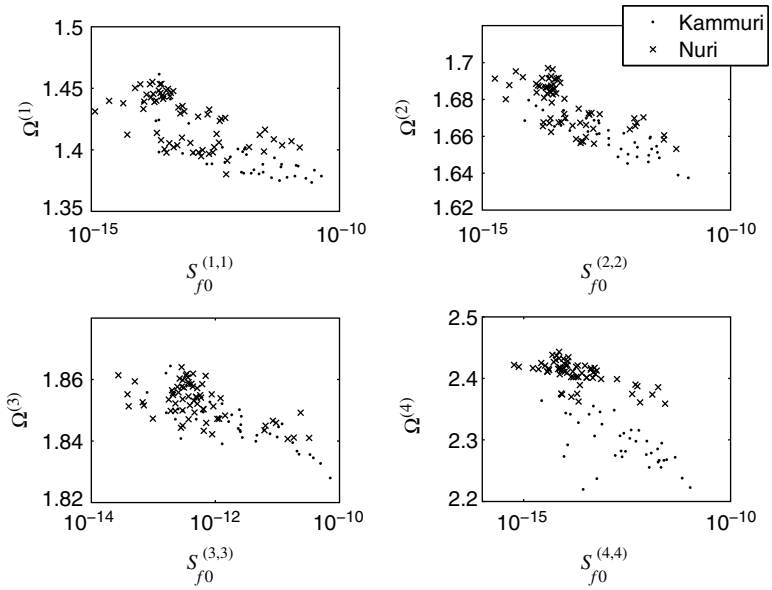

Figure 3.32 Identified modal frequencies versus the corresponding spectral intensities of the modal forces

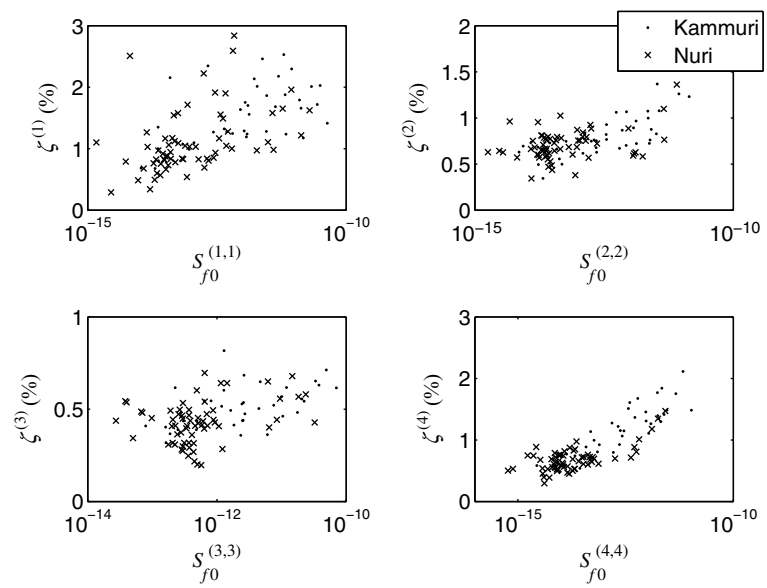

Figure 3.33 Identified modal damping ratios versus the corresponding spectral intensities of the modal forces

www.EngineeringBooksPdf.com

between the damping ratios and the spectral intensities of the corresponding modal forces. Although the points are more scattered than those in Figure 3.32, it is still clear to recognize the positive correlation between the damping ratio and the spectral intensity. Averages of the identified damping ratios are listed in Table 3.6. The damping ratios observed under the Kammuri were higher than those under the Nuri for all modes since the structural response due to the Kammuri was larger. The smallest damping ratio was associated with the third mode with an average of about 0.5%, and the other modes were in the range from 0.68% to 1.56%. The response amplitude-dependent damping ratios indicate that the building may have probably gone through an hysteretic behavior under the typhoons. However, since the modal frequencies and damping ratios were recovered after the dissipation of the typhoons, this behavior was not deteriorating.

Table 3.6 Average of the identified damping ratios for different modes

	Average of damping ratio (%)	
Mode	Kammuri	Nuri
1	1.559	1.196
2	0.803	0.706
3	0.515	0.428
4	1.142	0.681

3.7.3.3 Vortex Shedding

Due to the particular geometry of the East Asian Hall, which is categorized as an irregular prism, the effect of *vortex-shedding* has to be considered [181]. When the flow passes through a body, vortices are generated in the downstream of the body and detach from either side of the body alternately. Since the air flow generates periodically low-pressure vortices on the downstream, the body is pushed to move towards the low-pressure zone. Because the low-pressure zone occupies periodically changing region, vibration of the body is induced in the perpendicular direction to the flow. This behavior is referred to as vortex shedding.

The *Strouhal number* S_r is commonly used to study vortex shedding:

$$S_r = \frac{\Omega D}{V} \tag{3.85}$$

where Ω is the frequency of vortex shedding, D is the characteristic length scale of the body and V is the wind speed. It was found that the Strouhal number is roughly 0.2 regardless of the Reynolds number [140]. Therefore, the frequency of vortex shedding can be estimated by:

$$\Omega = \frac{S_r V}{D} \tag{3.86}$$

In this case, the frequency of vortex shedding is roughly in the range of $\Omega = 0.05$–0.4 Hz, and the value depends on the attacking angle. However, this range is significantly lower than the modal frequencies of the structure so this effect does not induce resonant type of motion for this building.

www.EngineeringBooksPdf.com

The vortex-induced oscillation contributes mainly to the lower frequency range of the spectrum. This phenomenon becomes more notable as the wind speed increases since the vortex shedding frequency is proportional to the wind speed and it will be closer to the fundamental frequency of the East Asian Hall. However, self-excited motion is not expected since the vortex-shedding oscillation frequency is far from the resonant frequencies of the building [203].

3.7.4 Concluding Remarks

The Bayesian spectral density approach was applied to identify the modal parameters of the East Asia Hall by using the field measurement during the passage of two typhoons. The spectral density approach was demonstrated to be flexible for applications and suitable for the current usage. The magnitude of the acceleration was highly correlated to the wind speed but the wind direction was also an important factor. The aerodynamic phenomenon including the vortex-shedding and possible self-excited effect were carefully considered in this type of study but they did not have critical effects on the building in the present study. By observing the change of the identified values of the modal frequencies in the whole process of the two typhoons, it was concluded that the modal frequencies had an obvious decreasing trend with the spectral intensities of the corresponding modal forces while it was opposite for the damping ratios. Furthermore, substantial reductions of the modal frequencies were observed under signal no. 8 but they were recovered instantly after the typhoons were dissipated.

3.8 Application to Hydraulic Jump

3.8.1 Problem Description

Hydraulic jump is an open channel flow phenomenon. It is the transition from supercritical flow in the upstream to a subcritical flow starting at the jump end in the downstream. Since hydraulic jump is similar to those of waves breaking in deep water and of coastal bores, it is commonly used to study the flow characteristics of bores. In the literature, many researchers illustrated the existence of oscillating characteristics, especially the repetitive mechanism of the formation of surface rollers. Rouse *et al.* (1959) performed wind tunnel tests to study the structure of hydraulic jumps, and this work concluded that the surface roller as a flow separation is caused by the divergence of streamlines due to an abrupt increase in the flow area [224]. Peregrine and Svenden (1978) suggested that turbulence is initiated at the front toe [205] and later Hoyt and Sellin (1989) proposed a similar suggestion that the initial stage of turbulent flow resembles that of a mixing layer [113]. Ghafar *et al.* (1995) showed experimentally that the local scour is due to a hydraulic jump formed on the sand bed and the existence of the oscillating characteristics [88]. Yeh and Mok (1990) proposed that the surface roller formation in bores is unsteady but periodical [279]. They suggested that once the surface roller is formed, a large eddy of the roller is advected towards the downstream and the next surface roller is then created. By quantifying the appearing frequencies of the systematically spaced turbulence patches behind the bores, the surface roller generation–advection cycle in bores is confirmed.

This study investigates the frequency content and energy dissipation efficiency of surface fluctuation for three types of hydraulic jumps, namely the undular jumps (Froude number

www.EngineeringBooksPdf.com

<1.7), weak jumps (1.7 < Froude number < 2.5) and oscillating jumps (2.5 < Froude number < 4.5). Time histories of the surface fluctuation were recorded by a resistant-type wave gauge at the jump end. In the literature, it was observed that the characteristics of surface fluctuation are highly related to the cyclic mechanism of the surface roller. Since the free surface interacts with the release of bulk eddies by the roller, the roller frequencies are found to be similar to the frequency of surface fluctuations in the range of Froude numbers used in the experiments. The frequencies of the surface roller in the literature are estimated by engineering judgement so there exists a large discrepancy by different studies. In this study, an investigation is continued in this direction for the surface fluctuation in hydraulic jumps using the Bayesian spectral density approach. The major frequency of the surface roller found in this approach is not by judgement but by the statistical inference from the measurements. It turns out that both theoretical and experimental results strongly support the existence of the periodicity of the surface roller. Moreover, uncertainty of the natural frequency can be quantified. This is important information for the comparison between the experimental and theoretical results.

3.8.2 Fundamentals of Hydraulic Jump

For an open channel flow, the dimensionless *Froude number* of the approaching flow can be defined as follows:

$$F_r = \frac{V_1}{\sqrt{g y_1}} \tag{3.87}$$

where V_1 and y_1 are the water flow speed and depth immediately before the jump, and g is the gravitational constant. The flow is *supercritical* when $F_r > 1$ and a hydraulic jump may happen in this case. Figure 3.34 shows the details of the hydraulic jump. By approximating the inflow and outflow with a uniform velocity distribution and hydrostatic pressure, the water depth ratio between the downstream and upstream is readily obtained by the well-known Bélanger equation, which is essentially the consequence of conservation of mass and conservation of momentum:

$$\frac{y_2}{y_1} = \frac{1}{2}\left(\sqrt{1 + 8{F_r}^2} - 1\right) \tag{3.88}$$

where y_2 is the downstream water depth.

3.8.3 Roller Formation-advection Model

A surface roller in a hydraulic jump is formed on the front face and the eddy is elliptical with the major axis parallel to the front surface. As it is released to the downstream, the eddy will become more circular in shape [279]. Mok and Ieong (2002) proposed that the roller generation–advection frequency Ω can be estimated by [183]:

$$\Omega = \frac{V_1}{L} \tag{3.89}$$

www.EngineeringBooksPdf.com

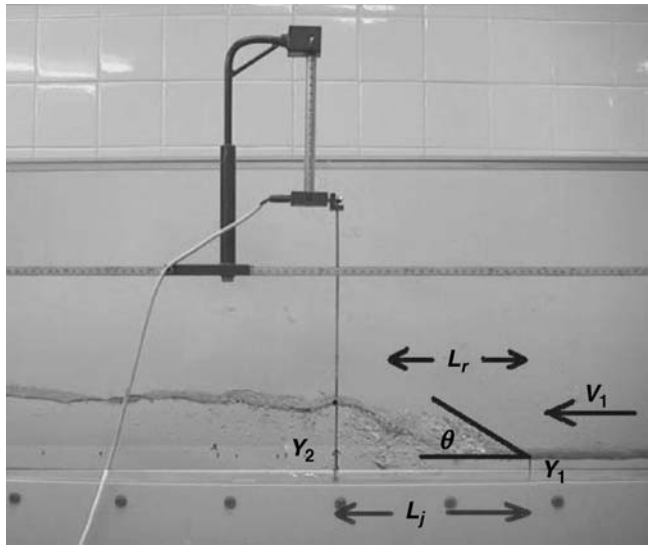

Figure 3.34 Illustration of the hydraulic jump

where V_1 is the upstream supercritical flow velocity and L is the typical roller length and is given by twice the horizontal projected length of the major axis of the elliptical roller:

$$L = \frac{2L_{rj}}{\cos\theta} \tag{3.90}$$

where L_{rj} is the horizontal roller length between the jump toe and the end of the surface roller. It can be found by the empirical relationship in Rouse *et al.* [224]. In this present study that the Froude numbers are between 1 and 3, the ratio L_{rj}/y_2 increases approximately linearly with the Froude number:

$$\frac{L_{rj}}{y_2} \approx F_r - 1 \tag{3.91}$$

The angle θ in Figure 3.34 may be estimated by the slope of the jump toe. Use L_j to denote the horizontal distance between the jump toe and jump end [215, 216] and it can be found in the results of Bradley and Peterka [32] or by using the following empirical relationship [46]:

$$L_j = 220 y_1 \tanh \frac{F_r - 1}{22} \tag{3.92}$$

3.8.4 Statistical Modeling of the Surface Fluctuation

Here, a simple linear single-degree-of-freedom system is used to characterize the surface fluctuation dynamics:

$$\ddot{x} + 2\zeta\Omega\dot{x} + \Omega x = f(t) \tag{3.93}$$

www.EngineeringBooksPdf.com

where x is the vertical displacement of the surface fluctuation, and Ω and ζ are the natural frequency and damping ratio of the oscillator, respectively. The input f on the right hand side is modeled as zero-mean Gaussian white noise with spectral intensity $S_f(\omega) = S_{f0}$. It is well known that for given model parameters the response x is also a Gaussian process with a zero mean, auto-correlation function [161, 249]:

$$R_x(\tau) = \frac{\pi S_{f0}}{2\zeta\omega^3} \mathrm{e}^{-\zeta\Omega|\tau|} \left[\cos(\Omega_d \tau) + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin(\Omega_d |\tau|) \right] \tag{3.94}$$

and spectral density function [161, 249]:

$$S_x(\omega) = \frac{S_{f0}}{(\Omega^2 - \omega^2)^2 + (2\zeta\omega\Omega)^2} \tag{3.95}$$

where $\Omega_d = \Omega\sqrt{1-\zeta^2}$ is the damped natural frequency of the oscillator.

Discrete data with a time step Δt are collected in the experiment. Let y_n be the measured surface fluctuation at time $t = n\Delta t$. There is a difference between the measured response y_n and the actual response $x(n\Delta t)$ due to measurement noise and this is modeled here by a discrete white noise process ϵ with zero mean and variance σ_ϵ^2:

$$y_n = x(n\Delta t) + \epsilon_n, \quad n = 0, 1, \ldots, N-1 \tag{3.96}$$

where ϵ has a spectral intensity

$$S_\epsilon(\omega) = S_{\epsilon 0} = \frac{\Delta t}{2\pi}\sigma_\epsilon^2 \tag{3.97}$$

By using the measurements $y_0, y_1, \ldots, y_{N-1}$, the spectral density estimator can be computed:

$$S_{y,N}(\omega_k) = \frac{\Delta t}{2\pi N} \left| \sum_{n=0}^{N-1} y_n \exp(-in\omega_k \Delta t) \right|^2 \tag{3.98}$$

and can be computed efficiently by the MATLAB® function'fft' as shown in Section 3.3.1.2.

Let $\boldsymbol{\theta} = [\Omega, \zeta, S_{f0}, S_{\epsilon 0}]^T$ denote the vector of the uncertain model parameters to be identified. The most probable values of the parameters and their associated uncertainty are updated by the Bayesian spectral density approach.

3.8.5 Experimental Setup and Results

The laboratory experiment was conducted in a rectangular re-circulating channel tank, which is 5.0 m long, 7.6 cm wide and 25 cm deep. The side walls are made of 'Plexiglas'. Hydraulic jumps on the horizontal floor were generated by regulating an upstream sluice gate and a fixed downstream blockage at the end of the channel. An immersed resistance-type gauge, which consists of two stainless-steel wires 1.5 mm in diameter and 12.5 mm apart, was set at the location of the jump end. The sampling frequency was taken to be 50 Hz for all cases to measure the surface fluctuation time histories. The total number of samples acquired in each case was 327 68 ($= 2^{15}$), corresponding to approximately 655 s.

Experiments were performed for 23 values of the Froude number (from 1.38 to 3.02) and there were 6 cases of undular jumps, 12 cases of weak jumps and 5 cases of oscillating jumps.

www.EngineeringBooksPdf.com

For each experiment, the upstream flow rate was recorded to compute the upstream flow speed and the corresponding Froude number. Then the gauge was placed at the location of the jump end of each Froude number according to the Bradley–Peterka curve [32]. The upstream water depth y_1 was recorded and then the downstream water depth y_2 can be estimated by Equation (3.88).

The surface fluctuation time histories were transformed to the frequency domain according to Equation (3.98). Figure 3.35 shows the spectra for three different Froude numbers and they represent the three types of hydraulic jumps: the undular jump, weak jump and oscillating jump. For the undular jumps with Froude number $F_r = 1.38$, there is no apparent frequency in the spectrum and this happens generally to the other five cases of undular jumps with Froude numbers: 1.43, 1.47, 1.52, 1.58 and 1.63. The surface roller is not apparent for undular jumps. For the weak jump with $F_r = 1.68$, the spectral values are much larger than those in the undular jumps. Although the values in low frequencies are still large, there is an apparent frequency around 10 Hz. For the oscillating jump with $F_r = 2.60$, there is a clear peak at around 5 Hz in the spectrum without strong components in the low-frequency range. Furthermore, the spectral values are larger than the other two cases by an order of magnitude. However, the spectral intensity of the measurement noise is so small that it can be neglected for identification.

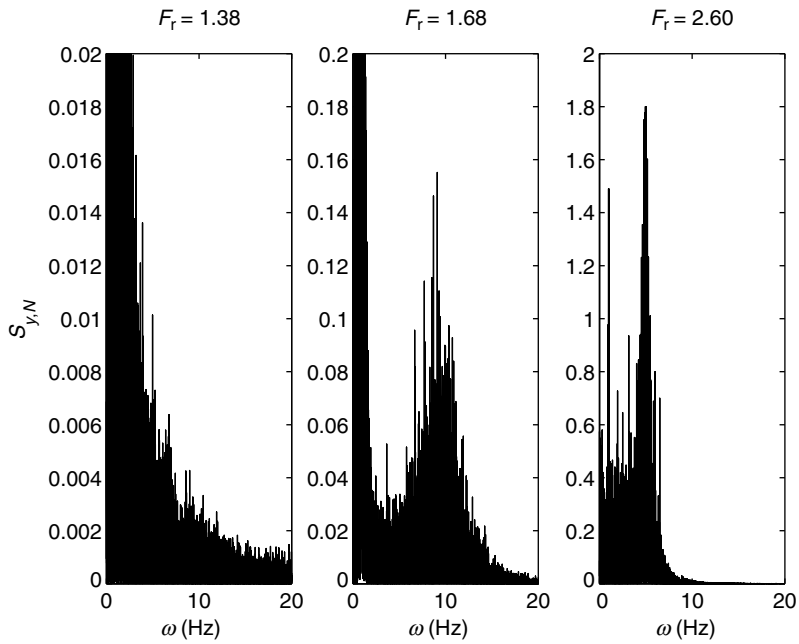

Figure 3.35 Measured spectra for different types of hydraulic jumps

Since there is no apparent frequency in the measurements of undular jumps, identification was proceeded for the weak and oscillating jumps only and the results are summarized in Table 3.7. The first column shows the Froude numbers. The second to the fourth columns show the identified values of the natural frequencies, damping ratios and spectral intensities of the

www.EngineeringBooksPdf.com

Table 3.7 Identification results with different Froude numbers

Froude number	Ω	ζ	S_{f0}
1.68	10.38 (0.048)	0.032 (0.0008)	2.77 (0.104)
1.73	10.73 (0.047)	0.031 (0.0007)	2.24 (0.084)
1.79	11.94 (0.053)	0.030 (0.0007)	2.89 (0.107)
1.82	11.17 (0.102)	0.037 (0.0011)	1.53 (0.059)
1.89	10.05 (0.059)	0.038 (0.0011)	1.11 (0.051)
1.93	8.911 (0.071)	0.043 (0.0013)	1.20 (0.046)
1.98	8.195 (0.074)	0.049 (0.0015)	1.85 (0.067)
2.09	7.990 (0.041)	0.042 (0.0009)	2.17 (0.081)
2.20	6.641 (0.034)	0.051 (0.0012)	3.24 (0.108)
2.31	6.685 (0.031)	0.046 (0.0011)	3.83 (0.129)
2.40	5.357 (0.031)	0.058 (0.0014)	3.36 (0.112)
2.50	5.169 (0.023)	0.055 (0.0013)	4.39 (0.140)
2.60	5.004 (0.025)	0.060 (0.0015)	3.89 (0.124)
2.70	4.984 (0.032)	0.067 (0.0018)	4.02 (0.136)
2.80	4.824 (0.023)	0.061 (0.0015)	3.18 (0.098)
2.91	5.489 (0.063)	0.069 (0.0020)	2.36 (0.113)
3.02	4.557 (0.039)	0.084 (0.0026)	2.71 (0.093)

input. The numbers in parentheses are the standard deviations of the corresponding identified values by the Bayesian spectral density approach. Figure 3.36 shows the identified parameters versus the corresponding Froude numbers. The confidence intervals of plus and minus three standard deviations are also shown in the figure. The dashed lines show the boundaries between

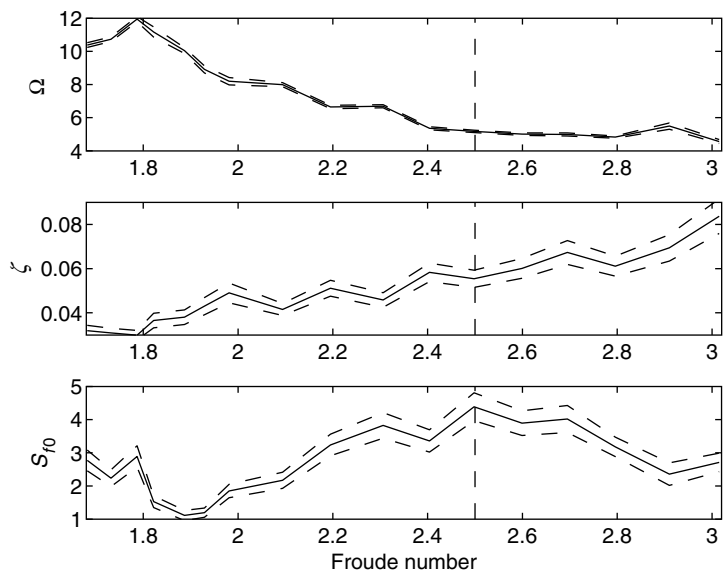

Figure 3.36 Identified parameters with different Froude numbers

www.EngineeringBooksPdf.com

weak jumps and oscillating jumps. It can be clearly seen that the surface roller frequency decreases with an increasing Froude number. The confidence intervals confirm that the changes are not simply due to statistical uncertainty only.

The surface roller frequency can also be estimated by another approach. From the results of Rouse *et al.* (1959), the horizontal length L_{rj} of the surface roller can be estimated by Equation (3.91) [224]. As the measurements of the volume flow rate and the upstream water depth y_1 are available, the upstream velocity and the Froude number can be readily obtained. The apparent frequencies of different Froude numbers can then be estimated by the Mok–Ieong formula in Equation (3.89) [183]. Figure 3.37 shows the comparison bewteen the frequencies estimated by the Mok–Ieong formula and by the identification results with the measured spactra. The crosses represent the values of the weak jumps and the circles represent the oscillating jumps. The apparent frequencies estimated by the measurements are in general higher than the values estimated by the Mok–Ieong formula for the weak jumps. In the case of weak jumps, the values of L_{rj} obtained from Equation (3.91) are relatively small compared to the oscillating jumps. As a result, the apparent frequency estimated by Equation (3.89) is more sensitive to the interpolation error of L_{rj}. This is one of the main reasons for the discrepancies between the theoretical and experimental frequencies for weak jumps. Nevertheless, the experimental results are in general in good agreement with Equation (3.89).

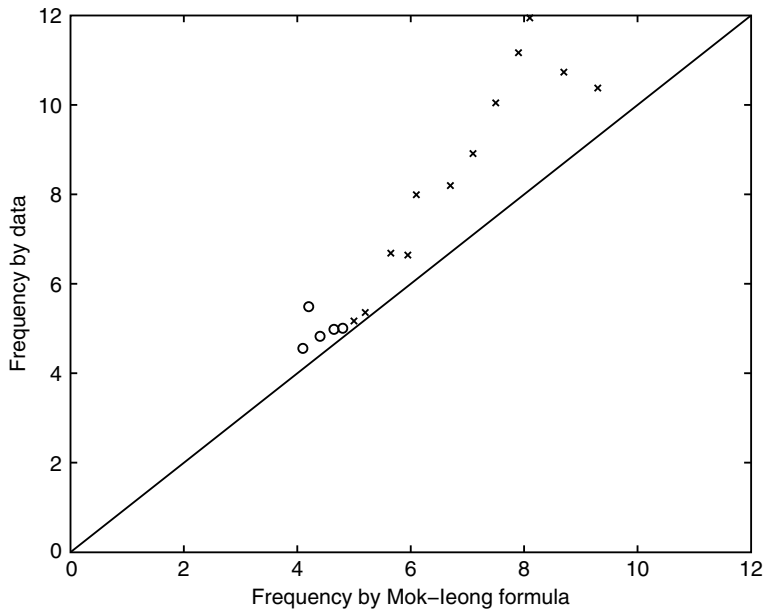

Figure 3.37 Comparison of the frequencies estimated by the Mok–Ieong formula and the data

On the other hand, the damping ratio increases with an increasing Froude number, confirming that hydraulic jumps with larger values of Froude number are more efficient in energy dissipation. However, the spectral intensity of the excitation does not show an obvious trend with the Froude number. According to Streeter and Wylie [256], the fractional head losses of

www.EngineeringBooksPdf.com

hydraulic jump are in the ranges of 0–5%, 5%–15% and 15%–45% for the undular jumps, weak jumps and oscillating jumps, respectively. From Table 3.7, the identified damping ratio increases from 3.2% to 8.4% as the Froude number increases from 1.68 to 3.02 and this agrees with the results in Streeter and Wylie [256] that the fractional head loss increases with a larger value of the Froude number.

3.8.6 Concluding Remarks

The Bayesian spectral density approach was applied to update a simple hydraulic jump surface roller model. The posterior PDF for the model parameters can be accurately approximated by a multi-variate Gaussian distribution. This property provides an efficient way to quantify the parametric uncertainty. The calculated mean and covariance matrix of this distribution offer an estimate of the optimal values of the model parameters and the associated uncertainty. The quantification of these uncertainties is useful to provide the confidence intervals of the model parameters. It turns out that there is significant evidence for the periodic behavior of weak jumps and oscillating jumps, but not the undular jumps. In the former cases, the frequency decreases but the damping ratio increases with an increasing Froude number. This is consistent with the previous study that oscillating jumps dissipate energy more efficiently than weak jumps. However, a more accurate model is necessary for further understanding of the time-frequency structure of hydraulic jumps.

www.EngineeringBooksPdf.com

www.EngineeringBooksPdf.com

4

Bayesian Time-domain Approach

Keywords: ambient vibration; Bayesian inference; best estimaton; conditional probability; correlation function; modal analysis; modal identification; nonstationary response; seismic response; structural health monitoring

4.1 Introduction

Chapter 3 presented the Bayesian spectral density approach for the parametric identification of the multi-degree-of-freedom dynamical model using the measured response time history. The methodology is applicable for linear models and can also be utilized for weakly nonlinear models by obtaining the mean spectrum with equivalent linearization or strongly nonlinear models by obtaining the mean spectrum with simulations. The stationarity assumption in modal/model identification for an ambient vibration survey is common but there are many cases where the response measurements are better modeled as nonstationary, e.g., the structural response due to a series of wind gusts or seismic responses. In the literature, there are very few approaches which consider explicitly nonstationary response data, for example, [226, 229]. Meanwhile, extension of the Bayesian spectral density approach for nonstationary response measurement is difficult since construction of the likelihood function is nontrivial in the frequency domain. Estimation of the time-dependent spectrum requires a number of data sets, which are associated with the same statistical time–frequency properties but this is impossible to achieve in practice.

The Bayesian time-domain approach presented in this chapter addresses this problem of parametric identification of linear dynamical models using a measured nonstationary response time history. This method has an explicit treatment on the nonstationarity of the response measurements and is based on an approximated probability density function (PDF) expansion of the response measurements. It allows for the direct calculation of the updated PDF of the model parameters. Therefore, the method provides not only the most probable values of the model parameters but also their associated uncertainty using one set of response data only. It is found that the updated PDF can be well approximated by an appropriately selected multi-variate Gaussian distribution centered at the most probable values of the parameters if the problem is

Bayesian Methods for Structural Dynamics and Civil Engineering Ka-Veng Yuen
© 2010 John Wiley & Sons (Asia) Pte Ltd

www.EngineeringBooksPdf.com

globally identifiable and the number of data points is large. Then, statistical moments, such as variance, of the random variables can be computed efficiently.

In the next section, an exact formulation is presented with the stationary response measurement of a linear single-degree-of-freedom system. However, it turns out to be computationally prohibited since it requires computation of the matrix determinant and the solution of linear simultaneous algebraic equations of huge dimensions. In Section 4.3, the problem is formulated for the nonstationary response measurement of linear multi-degree-of-freedom systems. Standard random vibration analysis for the nonstationary response of linear systems is also reviewed. In Section 4.4, an approximated expansion of the likelihood function is introduced and the Bayesian time-domain approach is built on this approximation to resolve the computational difficulties. In Section 4.5, simulated data with a linear SDOF system is used to verify the approximation in the methodology. Before concluding the method in Section 4.7, Section 4.6 presents an application with the measurement of a 10-story building subjected to nonstationary earthquake ground motion. The importance of considering explicitly the nonstationarity of the response is discussed. In Section 4.8, the spectral density approach and the time-domain approach are compared in terms of efficiency, accuracy and applicability. Finally, extended readings of some recently developed Bayesian updating methods are briefly introduced in Section 4.9.

4.2 Exact Bayesian Formulation and its Computational Difficulties

Consider a single-degree-of-freedom (SDOF) system with the equation of motion:

$$\ddot{x} + 2\zeta\Omega\dot{x} + \Omega^2 x = f(t) \tag{4.1}$$

where Ω and ζ are the natural frequency and damping ratio of the oscillator, respectively. For demonstration purpose only, stationary input and response are considered in this section. The input f is modeled as a zero-mean stationary Gaussian white noise with spectral intensity:

$$S_f(\omega) = S_{f0} \tag{4.2}$$

It is well known that the stationary response x is a zero-mean Gaussian random process with the auto-correlation function:

$$R_x(\tau) = \frac{\pi S_{f0}}{2\zeta\Omega^3} e^{-\zeta\Omega|\tau|} \left[\cos(\Omega_d \tau) + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin(\Omega_d |\tau|)\right] \tag{4.3}$$

and the power spectral density function:

$$S_x(\omega) = \frac{S_{f0}}{(\Omega^2 - \omega^2)^2 + (2\zeta\Omega\omega)^2} \tag{4.4}$$

where $\Omega_d \equiv \Omega\sqrt{1-\zeta^2}$ is the damped natural frequency of the oscillator [157, 161, 249].

Discrete data are sampled with a time step Δt and y_n are then used to denote the measured response at time $t = n\Delta t$. Due to measurement noise and modeling error, there is a difference between the measured response y_n and the model response $x(n\Delta t)$, referred to hereafter as prediction error. It is assumed that the prediction error can be adequately represented by a

www.EngineeringBooksPdf.com

discrete white noise process ϵ with zero mean and variance σ_ϵ^2:

$$y_n = x(n\Delta t) + \epsilon_n \qquad n = 1, 2, \ldots, N \tag{4.5}$$

so this process ϵ satisfies:

$$E[\epsilon_n \epsilon_{n'}] = \sigma_\epsilon^2 \delta_{nn'} \tag{4.6}$$

where $\delta_{nn'}$ denotes the Kronecker delta, which is given by:

$$\delta_{nn'} = \begin{cases} 1 & \text{if } n = n' \\ 0 & \text{if } n \neq n' \end{cases} \tag{4.7}$$

Furthermore, the stochastic response x and the prediction error ϵ are assumed statistically independent.

Therefore, the set of measurements $\mathcal{D}$ include the data points $\{y_1, y_2, \ldots, y_N\}$. Define a column vector $\mathbf{Y}$:

$$\mathbf{Y} = [y_1, y_2, \ldots, y_N]^T \tag{4.8}$$

Since x and ϵ are Gaussian, the measured response y is also Gaussian. It follows that the likelihood function for a given set of data $\mathcal{D}$ is given by:

$$p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) = (2\pi)^{-\frac{N}{2}} |\boldsymbol{\Gamma}(\boldsymbol{\theta})|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\mathbf{Y}^T \boldsymbol{\Gamma}(\boldsymbol{\theta})^{-1}\mathbf{Y}\right] \tag{4.9}$$

where $\mathcal{C}$ is the prescribed class of models governed by Equation (4.1) with the parameterization $\boldsymbol{\theta} = [\Omega, \zeta, S_{f0}, \sigma_\epsilon]^T$ for the dynamical model to be identified. The notation $|\mathbf{A}|$ is used to denote the determinant of a matrix $\mathbf{A}$. The likelihood function $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$ is an N-variate Gaussian distribution of the measurement vector $\mathbf{Y}$ with zero mean and covariance matrix $\boldsymbol{\Gamma}(\boldsymbol{\theta})$ where the (n, n') element $\Gamma^{(n,n')}(\boldsymbol{\theta}) \equiv E[y_n y_{n'}]$ is given by:

$$\Gamma^{(n,n')}(\boldsymbol{\theta}) = R_x[(n' - n)\Delta t|\boldsymbol{\theta}] + \sigma_\epsilon^2 \delta_{nn'} \tag{4.10}$$

where R_x is the auto-correlation function given by Equation (4.3) for a given set of the model parameter vector $\boldsymbol{\theta}$.

Herein, the identification of the model parameter vector $\boldsymbol{\theta}$, given some measured data $\mathcal{D}$, is concerned. Using the Bayes' theorem, the updated PDF of the model parameter vector $\boldsymbol{\theta}$ is given by:

$$p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}) = \kappa_0 p(\boldsymbol{\theta}|\mathcal{C}) p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) \tag{4.11}$$

where κ_0 is the normalizing constant such that integrating the right hand side of Equation (4.11) over the parameter space yields unity. The prior PDF $p(\boldsymbol{\theta}|\mathcal{C})$ in Equation (4.11) reflects the prior information of the parameters without using the data. The likelihood function $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$ is the dominant factor when the number of data points is large. It reflects the contribution of the measured data $\mathcal{D}$ in establishing the updated (posterior) PDF of the parameters. The relative plausibility between two values of $\boldsymbol{\theta}$ does not depend on the normalizing constant κ_0. It depends only on the relative values of the prior PDF $p(\boldsymbol{\theta}|\mathcal{C})$ and the relative values of $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$. In

www.EngineeringBooksPdf.com

order to search for the most probable model parameter vector $\boldsymbol{\theta}$, denoted by $\boldsymbol{\theta}^\star$, the product of the prior PDF and the likelihood function $p(\boldsymbol{\theta}|\mathcal{C})p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$ is maximized.

However, for a large number of observed data points, repeated evaluations of the factor $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$ for different values of $\boldsymbol{\theta}$ becomes computationally prohibitive. It is obvious from Equation (4.9) that it requires the computation of the solution $\mathbf{X}$ of the algebraic equation $\boldsymbol{\Gamma}(\boldsymbol{\theta})\mathbf{X} = \mathbf{Y}$ and the determinant of the $N \times N$ matrix $\boldsymbol{\Gamma}(\boldsymbol{\theta})$. This task is computationally very expensive for large N even though the former can be done efficiently by pre-conditioners [43, 49, 124]. Repeated evaluations of the likelihood function for thousands of times in the optimization process is computationally prohibitive for large N. Therefore, the exact Bayesian approach described above, based on direct use of the measured data $\mathcal{D}$, becomes practically infeasible. In the next section, the model updating problem will be formulated with a nonstationary response measurement. Standard random vibration analysis will be reviewed. Then, an approximated approach is introduced and it overcomes the computational obstacles and renders the problem practically feasible.

4.3 Random Vibration Analysis of Nonstationary Response

Consider a linear dynamical system with N_d degrees of freedom (DOFs) and its equation of motion is given by:

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{T}_0\mathbf{F}(t) \tag{4.12}$$

where $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$ are the mass, damping and stiffness matrices, respectively, $\mathbf{T}_0 \in \mathbb{R}^{N_d \times N_F}$ is a force distributing matrix, and the input $\mathbf{F}$ is a zero-mean Gaussian nonstationary stochastic process which is modeled by:

$$\mathbf{F}(t) = A(t)\mathbf{g}(t) \tag{4.13}$$

where $\mathbf{F}(t) \in \mathbb{R}^{N_F}$, A is a scalar modulating function, and $\mathbf{g}$ is a Gaussian stationary stochastic process with zero mean and spectral density matrix function $\mathbf{S}_g(\omega) \in \mathbb{C}^{N_F \times N_F}$. Then, the auto-correlation function of the forcing process $\mathbf{F}$ is given by:

$$\mathbf{R}_F(t, t+\tau) \equiv E\left[\mathbf{F}(t)\mathbf{F}(t+\tau)^T\right] = A(t)A(t+\tau)\mathbf{R}_g(\tau) \tag{4.14}$$

where $\mathbf{R}_g$ is the auto-correlation matrix function for the stationary process $\mathbf{g}$.

If the matrix $\mathbf{M}^{-1}\mathbf{C}$ can be diagonalized by the same set of eigenvectors as the matrix $\mathbf{M}^{-1}\mathbf{K}$, the damping model is said to be *classical*. Caughey and O'Kelly (1965) showed the following necessary and sufficient condition [41]:

$$\mathbf{C}\mathbf{M}^{-1}\mathbf{K} = \mathbf{K}\mathbf{M}^{-1}\mathbf{C} \tag{4.15}$$

In this case, it is obvious that $(\mathbf{M}^{-1}\mathbf{C})(\mathbf{M}^{-1}\mathbf{K}) = (\mathbf{M}^{-1}\mathbf{K})(\mathbf{M}^{-1}\mathbf{C})$ so the matrices $\mathbf{M}^{-1}\mathbf{C}$ and $\mathbf{M}^{-1}\mathbf{K}$ share the same set of eigenvectors [189]. Since $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$ are symmetric matrices, $\mathbf{M}^{-1}\mathbf{C}$ and $\mathbf{M}^{-1}\mathbf{K}$ are diagonalizable with the same eigenvector matrix. As a result, Equation (4.12) can be decoupled by introducing the modal coordinates $\mathbf{q} = [q^{(1)}, q^{(2)}, \ldots, q^{(N_d)}]^T$:

$$\mathbf{x}(t) = \boldsymbol{\Phi} \cdot \mathbf{q}(t) \tag{4.16}$$

www.EngineeringBooksPdf.com

where $\mathbf{\Phi} \in \mathbb{R}^{N_d \times N_d}$ is the modal matrix comprised of N_d linearly independent eigenvectors of $\mathbf{M}^{-1}\mathbf{C}$ or $\mathbf{M}^{-1}\mathbf{K}$. These eigenvectors are referred to as the mode shapes $\boldsymbol{\phi}^{(m)}$, $m = 1, 2, \ldots, N_d$, of the dynamical model:

$$\mathbf{\Phi} = \left[\boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}^{(2)}, \ldots, \boldsymbol{\phi}^{(N_d)}\right] \tag{4.17}$$

and they are normalized in the following manner:

$$\boldsymbol{\phi}_{l_m}^{(m)} = 1, \qquad m = 1, 2, \ldots, N_d \tag{4.18}$$

where l_m is an arbitrary measured DOF which is not a node of the mth mode. The reason for not adopting mass normalization is that the mass matrix may be unknown in the identification process.

Then, by pre-multiplying $\mathbf{\Phi}^{-1}\mathbf{M}^{-1}$ to Equation (4.12), the following equation can be obtained:

$$\ddot{\mathbf{q}} + \mathbf{\Phi}^{-1}\mathbf{M}^{-1}\mathbf{C}\mathbf{\Phi}\dot{\mathbf{q}} + \mathbf{\Phi}^{-1}\mathbf{M}^{-1}\mathbf{K}\mathbf{\Phi}\mathbf{q} = \mathbf{\Phi}^{-1}\mathbf{M}^{-1}\mathbf{T}_0\mathbf{F}(t) \tag{4.19}$$

Since $\mathbf{\Phi}$ is the matrix comprised of the eigenvectors of $\mathbf{M}^{-1}\mathbf{C}$ and $\mathbf{M}^{-1}\mathbf{K}$, $\mathbf{\Phi}^{-1}\mathbf{M}^{-1}\mathbf{C}\mathbf{\Phi}$ and $\mathbf{\Phi}^{-1}\mathbf{M}^{-1}\mathbf{K}\mathbf{\Phi}$ are diagonal with the eigenvalues of $\mathbf{M}^{-1}\mathbf{C}$ and $\mathbf{M}^{-1}\mathbf{K}$ on their diagonal entries. Therefore, Equation (4.19) can be rewritten as N_d independent ordinary differential equations of the modal coordinates:

$$\ddot{q}^{(m)}(t) + 2\zeta^{(m)}\Omega^{(m)}\dot{q}^{(m)}(t) + {\Omega^{(m)}}^2 q^{(m)}(t) = A(t) f^{(m)}(t), \qquad m = 1, 2, \ldots, N_d \tag{4.20}$$

where $\mathbf{f}(t) = [f^{(1)}(t), f^{(2)}(t), \ldots, f^{(N_d)}(t)]^T$ is the modal forcing vector given by:

$$\mathbf{f}(t) = (\mathbf{M}\mathbf{\Phi})^{-1}\mathbf{T}_0\mathbf{g}(t) \tag{4.21}$$

The modal forcing $\mathbf{f}$ is a linear combination of the components of $\mathbf{g}$ so it is also a zero-mean Gaussian vector process. It is stationary with the spectral density matrix function:

$$\mathbf{S}_f(\omega) = (\mathbf{M}\mathbf{\Phi})^{-1}\mathbf{T}_0\mathbf{S}_g(\omega)\mathbf{T}_0^T(\mathbf{M}\mathbf{\Phi})^{-T} \tag{4.22}$$

and the auto-correlation matrix function:

$$\mathbf{R}_f(\tau) = \int_{-\infty}^{\infty} \mathbf{S}_f(\omega)\mathrm{e}^{i\omega\tau}\mathrm{d}\omega \tag{4.23}$$

It is well known that the model response $\mathbf{x}$ is a zero-mean Gaussian random process with a correlation function between x_l and $x_{l'}$ [161]:

$$\begin{aligned} R_x^{(l,l')}(t, t+\tau) = &\sum_{m=1}^{N_m}\sum_{m'=1}^{N_m} \phi_l^{(m)}\phi_{l'}^{(m')} \\ &\times \int_0^{t+\tau}\int_0^t A(u)A(v)h^{(m)}(t-u)h^{(m')}(t+\tau-v)R_f^{(m,m')}(u-v)\mathrm{d}u\mathrm{d}v \end{aligned} \tag{4.24}$$

www.EngineeringBooksPdf.com

where $h^{(m)}$ denotes the modal unit impulse response function for the mth mode. It is the modal response of Equation (4.20) due to unit impulse input $A(t)f^{(m)}(t) = \delta(t)$:

$$h^{(m)}(t) = \begin{cases} \frac{1}{\Omega_d^{(m)}} \mathrm{e}^{-\zeta\Omega^{(m)}t} \sin(\Omega_d^{(m)} t) & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases} \tag{4.25}$$

where $\Omega_d^{(m)} = \Omega^{(m)}\sqrt{1 - \zeta^{(m)2}}$ is the damped natural frequency of the mth mode. Causality is the reason for $h^{(m)}(t) = 0$ when $t < 0$. Here, it is assumed that only N_m lower modes contribute significantly to the model response.

The time-dependent power spectral density is given by the Fourier transform of the correlation function [161]:

$$S_x^{(l,l')}(t, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_x^{(l,l')}(t, t+\tau) e^{-i\omega\tau} \mathrm{d}\tau \tag{4.26}$$

Assume that discrete data at times $t = n\Delta t, n = 1, 2, \ldots, N$, are available at $N_o (\leq N_d)$ measured DOFs. Also, due to measurement noise and modeling error, there is prediction error, i.e., a difference between the measured response $\mathbf{y}_n \in \mathbb{R}^{N_o}$ and the model response at time $t = n\Delta t$ corresponding to the measured degrees of freedom. The latter is given by $\mathbf{L}_o\mathbf{x}(n\Delta t)$, where $\mathbf{L}_o \in \mathbb{R}^{N_o \times N_d}$ is an observation matrix. In the case of measuring the displacement time histories, this observation will be comprised of 'zeros' and 'ones.' The row and column numbers of the 'ones' are the number of the measured channels and the corresponding measured degree of freedom, respectively. Therefore, the measured response $\mathbf{y}_n$ at time $n\Delta t$ can be expressed as follows:

$$\mathbf{y}_n = \mathbf{L}_o\mathbf{x}(n\Delta t) + \boldsymbol{\epsilon}_n \tag{4.27}$$

It is assumed that the prediction error can be adequately represented by a discrete zero-mean Gaussian white noise $\boldsymbol{\epsilon}$ with the following $N_o \times N_o$ covariance matrix:

$$E[\boldsymbol{\epsilon}_n \boldsymbol{\epsilon}_{n'}^T] = \boldsymbol{\Sigma}_\epsilon \delta_{nn'} \tag{4.28}$$

where $\delta_{nn'}$ is the Kronecker delta given by Equation (4.7). Furthermore, the prediction error $\boldsymbol{\epsilon}$ and the stochastic response $\mathbf{x}$ are assumed statistically independent.

The measurement $\mathbf{y}$ is a zero-mean discrete Gaussian random process with correlation matrix function $\mathbf{R}_y$:

$$\begin{aligned} \mathbf{R}_y(n, n') &= E[\mathbf{y}_n \mathbf{y}_{n'}^T] \\ &= \mathbf{L}_o \mathbf{R}_x(n\Delta t, n'\Delta t)\mathbf{L}_0^T + \boldsymbol{\Sigma}_\epsilon \delta_{nn'} \end{aligned} \tag{4.29}$$

where $\mathbf{R}_x$ denotes the correlation matrix function of the model response $\mathbf{x}$ and is given by Equation (4.24).

www.EngineeringBooksPdf.com

4.4 Bayesian Updating with Approximated PDF Expansion

Let $\boldsymbol{\theta}$ denote the uncertain parameter vector for identification and it includes the following parameters:

1. The structural parameters that determine the model matrices $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$.
   In the special case of modal updating, assume that only the lowest N_m modes contribute significantly to the response and only the modal parameters of these modes are to be identified. Then, the structural parameters are the modal frequencies $\Omega^{(m)}$, modal damping ratios $\zeta^{(m)}$, and the elements of the first N_m mode shapes $\phi^{(m)}$, except those elements which are equal to unity for normalization purposes, $m = 1, 2, \ldots, N_m$. Thus, there are a total number of $N_m(N_o + 1)$ unknown structural modal parameters.
2. The forcing parameters defining the spectral density matrix function $\mathbf{S}_g$ and the modulating function A.
   In the special case of modal identification, the first part of the forcing parameters are the elements of the upper right triangular part (diagonal inclusive) of the $N_m \times N_m$ submatrix of $\mathbf{S}_{f0}$ of the modal forces (symmetry defines the lower triangular part).
3. The elements of the upper right triangular part (diagonal inclusive) of $\boldsymbol{\Sigma}_\epsilon$ (symmetry defines the lower triangular part of this matrix) or the parameters to be used to determine this matrix.

Recall here that the scaling of each mode shape is chosen such that one of its components corresponding to a measured DOF is equal to unity. However, such scaling is arbitrary and therefore the mode shape vectors can be identified only up to a scaling constant. A different mode shape normalization will cause all identified components of that mode shape to be scaled by the same constant α_m and at the same time the elements $S_{f0}^{(m,m')}$ and $S_{f0}^{(m',m)}$ will be scaled by α_m^{-2} if $m = m'$ and α_m^{-1} if $m \neq m'$.

Using the Bayes' theorem, the updated PDF of the model parameters given some measured response $\mathcal{D}$ is:

$$p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}) = \kappa_1 p(\boldsymbol{\theta}|\mathcal{C}) p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) \tag{4.30}$$

where $\mathcal{C}$ is a prescribed class of models for the dynamical system with the prescribed parameterization. κ_1 is a normalizing constant such that integrating the right hand side of Equation (4.30) over the parameter space yields unity. The factor $p(\boldsymbol{\theta}|\mathcal{C})$ denotes the prior PDF of the parameters and is based on previous knowledge or engineering judgement. In the case where no prior information is available, it can be treated as a constant and absorbed into the normalizing constant. The likelihood function $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$ is the dominant factor on the right hand side to reflect the contribution of the measured data in establishing the posterior distribution. The difficulty here is to construct the likelihood $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$ as the random components of the data points are correlated, which is in contrast to the case in Chapter 2 with input–output data. By using the Bayes' theorem, the likelihood function can be expanded to a product of a conditional PDF and a reduced-order joint PDF:

$$\begin{aligned} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) &= p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N|\boldsymbol{\theta}, \mathcal{C}) \\ &= p(\mathbf{y}_N|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N-1}, \boldsymbol{\theta}, \mathcal{C}) p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N-1}|\boldsymbol{\theta}, \mathcal{C}) \end{aligned} \tag{4.31}$$

www.EngineeringBooksPdf.com

Then, in the same manner, the reduced-order joint PDF $p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N-1}|\boldsymbol{\theta}, \mathcal{C})$ can be further expanded. By continuing this process, the likelihood function can be factorized as a product of a reduced-order likelihood function and a number of conditional PDFs:

$$p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) = p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_p}|\boldsymbol{\theta}, \mathcal{C}) \prod_{n=N_p+1}^{N} p(\mathbf{y}_n|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}, \boldsymbol{\theta}, \mathcal{C}) \tag{4.32}$$

This expansion is exact but it does not resolve the computational difficulties encountered in the exact formulation. This is because the computation of each of the conditional PDFs $p(\mathbf{y}_n|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}, \boldsymbol{\theta}, \mathcal{C})$ for large n requires similar computational effort as in the computation of the likelihood function $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$ in the direct exact formulation and the right hand side of the expansion involves many such conditional PDFs. In order to overcome this computational obstacle, the following approximation is introduced ($n > N_p$) [290]:

$$p(\mathbf{y}_n|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}, \boldsymbol{\theta}, \mathcal{C}) \approx p(\mathbf{y}_n|\mathbf{y}_{n-N_p}, \mathbf{y}_{n-N_p+1}, \ldots, \mathbf{y}_{n-1}, \boldsymbol{\theta}, \mathcal{C}) \tag{4.33}$$

so the likelihood function expansion in Equation (4.32) can be approximated by:

$$p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) \approx p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_p}|\boldsymbol{\theta}, \mathcal{C}) \prod_{n=N_p+1}^{N} p(\mathbf{y}_n|\mathbf{y}_{n-N_p}, \mathbf{y}_{n-N_p+1}, \ldots, \mathbf{y}_{n-1}, \boldsymbol{\theta}, \mathcal{C}) \tag{4.34}$$

In other words, the conditional PDFs depending on more than N_p previous data points are approximated by conditional PDFs depending on only the last N_p data points. The sense of this approximation is that data points belonging too far in the past do not have significant information on the system response of the present point. Of course, one expects this to be legitimate, if N_p is so large that the correlation functions have decayed to negligible values. However, it is found that a value for N_p of the order of $2\pi/\Omega\Delta t$ is sufficient, where Ω is the fundamental frequency (in rad/s) of the system and Δt is the sampling time step. In order words, the value of N_p is chosen to cover roughly one fundamental period of the system. For example, if the sampling time step is $1/25$ of the fundamental period of the system, i.e., $\Delta t = 2\pi/25\Omega$, a value of $N_p = 25$ will be sufficient for the expansion to be accurate. In the case of multi-degree-of-freedom systems (i.e., multi-mode systems), such a selected value of N_p covers more than one period of the higher modes so the approximation is even more accurate for the higher modes. Although the fundamental frequency Ω is an unknown parameter, it can be roughly estimated from the first peak of the response spectrum to obtain the value of N_p prior to the identification. The identification result is not sensitive to the selected value of N_p if it is sufficiently large.

This behavior can be explained as follows. Consider three random variables x, y and z and assume that one is interested in the conditional probability density $p(x|y, z)$ with $p(z|y) > 0$ for the prediction of x by y and z. By using the Bayes' theorem, the following relationship can be obtained:

$$p(x, z|y) = p(x|y, z)p(z|y) \tag{4.35}$$

If x is independent of z given y, the PDF $p(x, z|y)$ can be expanded as follows:

$$p(x, z|y) = p(x|y)p(z|y) \tag{4.36}$$

www.EngineeringBooksPdf.com

In this case, the condition PDF can be simplified as:

$$p(x|y, z) = p(x|y) \tag{4.37}$$

Now, let x be dependent on z. If y and z are fully dependent (or negatively fully dependent), the variables y and z follow a linear relationship: $z = ay + b$. Then, the above equation $p(x|y, z) = p(x|y)$ is still valid. If y and z are almost fully dependent (or almost negatively fully dependent), this equation still holds approximately. The point here is that when considering conditional probability densities, some of the conditioning information may be redundant/repeating and can be omitted without significantly sacrificing the accuracy. This argument can be applied to the case of oscillatory response measurements since measurements a half period apart are negatively highly correlated. Furthermore, response measurements one period apart are highly correlated. Using the recommended value of N_p leads to significant reduction of the computational demand with minor effects on the accuracy of the likelihood function expansion only.

Example. Consider four Gaussian random variables x_1, x_2, x_3 and x_4 with zero mean and a covariance matrix:

$$\begin{bmatrix} 1 & -0.95 & 0.95^2 & -0.95^3 \\ -0.95 & 1 & -0.95 & 0.95^2 \\ 0.95^2 & -0.95 & 1 & -0.95 \\ -0.95^3 & 0.95^2 & -0.95 & 1 \end{bmatrix} \tag{4.38}$$

This example demonstrates the situation that x_1, x_2, x_3 and x_4 are the stochastic response of a dynamical system and any two consecutive random variables x_n and x_{n+1}, $n = 1, 2, 3$, are a half period apart. In order to predict x_1 conditional on x_2, the conditional mean and variance can be computed by using Equations (C.23) and (C.24) in Appendix C:

$$E[x_1|x_2] = -0.95x_2$$
$$\sigma^2_{x_1|x_2} = 1 - 0.95^2 = 0.0975$$

Similarly, x_1 can be estimated with the measurement of x_3:

$$E[x_1|x_3] = 0.95^2 x_3$$
$$\sigma^2_{x_1|x_3} = 1 - 0.95^4 \approx 0.1426$$

or with the measurement of x_4:

$$E[x_1|x_4] = -0.95^3 x_4$$
$$\sigma^2_{x_1|x_4} = 1 - 0.95^6 \approx 0.2262$$

The prediction-error variance in all cases is significantly reduced from the unconditional variance of x_1 (i.e., 1.0) because even the lowest correlation coefficient (between x_1 and x_4) is $-0.95^3 \approx -0.8574$, which is considered negatively highly correlated. Of course, it is expected that using x_2 yields a prediction with the smallest uncertainty.

www.EngineeringBooksPdf.com

Now, consider to predict x_1 with the measurements of x_2 and x_3. Again, by using Equations (C.23) and (C.24), the optimal predictor and its variance can be obtained:

$$E[x_1|x_2, x_3] = \begin{bmatrix} -0.95 & 0.95^2 \end{bmatrix} \begin{bmatrix} 1 & -0.95 \\ -0.95 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} = -0.95x_2$$

$$\sigma^2_{x_1|x_2,x_3} = 1 - \begin{bmatrix} -0.95 & 0.95^2 \end{bmatrix} \begin{bmatrix} 1 & -0.95 \\ -0.95 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -0.95 \\ 0.95^2 \end{bmatrix} = 0.0975$$

Finally, the variable x_1 can be predicted with all the other three variables and the optimal predictor is:

$$\begin{aligned} E[x_1|x_2, x_3, x_4] &= \begin{bmatrix} -0.95 & 0.95^2 & -0.95^3 \end{bmatrix} \begin{bmatrix} 1 & -0.95 & 0.95^2 \\ -0.95 & 1 & -0.95 \\ 0.95^2 & -0.95 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_2 \\ x_3 \\ x_4 \end{bmatrix} \\ &= -0.95x_2 \end{aligned}$$

Furthermore, the associated variance is given by:

$$\begin{aligned} \sigma^2_{x_1|x_2,x_3,x_4} &= 1 - \begin{bmatrix} -0.95 & 0.95^2 & -0.95^3 \end{bmatrix} \begin{bmatrix} 1 & -0.95 & 0.95^2 \\ -0.95 & 1 & -0.95 \\ 0.95^2 & -0.95 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -0.95 \\ 0.95^2 \\ -0.95^3 \end{bmatrix} \\ &= 0.0975 \end{aligned}$$

It is quite surprising that the inclusion of x_3 and x_4 in the conditioning information does not reduce the prediction-error variance even though they are highly correlated with the target variable x_1. In this case, $p(x_1|x_2, x_3, x_4) = p(x_1|x_2, x_3) = p(x_1|x_2)$.

To account for the measurement noise, consider four Gaussian random variables y_1, y_2, y_3 and y_4 with zero mean and the covariance matrix:

$$\begin{bmatrix} 1.01 & -0.95 & 0.95^2 & -0.95^3 \\ -0.95 & 1.01 & -0.95 & 0.95^2 \\ 0.95^2 & -0.95 & 1.01 & -0.95 \\ -0.95^3 & 0.95^2 & -0.95 & 1.01 \end{bmatrix}$$

Here, a value of 0.01 is added to the diagonal elements to represent a 10% rms measurement noise (or 1% of variance). In this case, the optimal predictor and its associated variance are:

$$E[y_1|y_2] = -\frac{0.95}{1.01} y_2 \approx -0.9406 y_2$$

$$\sigma^2_{y_1|y_2} = 1.01 - \frac{0.95^2}{1.01} \approx 0.1164$$

www.EngineeringBooksPdf.com

Similarly, y_1 can be predicted by the measurement of y_3:

$$E[y_1|y_3] = \frac{0.95^2}{1.01} y_3 \approx 0.8936 y_3$$

$$\sigma^2_{y_1|y_3} = 1.01 - \frac{0.95^4}{1.01} \approx 0.2036$$

or by the measurement of y_4:

$$E[y_1|y_4] = -\frac{0.95^3}{1.01} y_4 \approx -0.8489 y_4$$

$$\sigma^2_{y_1|y_4} = 1.01 - \frac{0.95^6}{1.01} \approx 0.2822$$

The prediction-error variances are larger than the previous cases without measurement noise but the conditioning information is still capable to reduce significantly the prediction-error variance since the correlation among these random variables is strong. Now, consider to predict y_1 with the measurements of y_2 and y_3:

$$E[y_1|y_2, y_3] = \begin{bmatrix} -0.95 & 0.95^2 \end{bmatrix} \begin{bmatrix} 1.01 & -0.95 \\ -0.95 & 1.01 \end{bmatrix}^{-1} \begin{bmatrix} y_2 \\ y_3 \end{bmatrix} = -0.8684 y_2 + 0.0767 y_3$$

$$\sigma^2_{y_1|y_2,y_3} = 1 - \begin{bmatrix} -0.95 & 0.95^2 \end{bmatrix} \begin{bmatrix} 1.01 & -0.95 \\ -0.95 & 1.01 \end{bmatrix}^{-1} \begin{bmatrix} -0.95 \\ 0.95^2 \end{bmatrix} = 0.1057$$

Finally, the variable y_1 can be predicted with all the other three variables:

$$E[y_1|y_2, y_3, y_4] = \begin{bmatrix} -0.95 & 0.95^2 & -0.95^3 \end{bmatrix} \begin{bmatrix} 1.01 & -0.95 & 0.95^2 \\ -0.95 & 1.01 & -0.95 \\ 0.95^2 & -0.95 & 1.01 \end{bmatrix}^{-1} \begin{bmatrix} y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

$$= -0.8679 y_2 + 0.0713 y_3 - 0.0063 y_4$$

and the prediction-error variance is given by:

$$\sigma^2_{y_1|y_2,y_3,y_4} = 1 - \begin{bmatrix} -0.950.95^2 & -0.95^3 \end{bmatrix} \begin{bmatrix} 1.01 & -0.95 & 0.95^2 \\ -0.95 & 1.01 & -0.95 \\ 0.95^2 & -0.95 & 1.01 \end{bmatrix}^{-1} \begin{bmatrix} -0.95 \\ 0.95^2 \\ -0.95^3 \end{bmatrix}$$

$$= 0.1057$$

The additional measurement of y_3 improves only slightly the prediction for y_1 but the additional measurement of y_4 has virtually no effect on the prediction. This example is useful to demonstrate the approximation used in the Bayesian time-domain approach. The random variable y_1 is predicted by the measurements of y_2, y_3 and y_4, which are 0.5, 1 and 1.5 periods apart from y_1, respectively. It turns out that including the data points within one period is sufficient. Furthermore, in a usual situation, the sampling time step is much less than half of

www.EngineeringBooksPdf.com

a fundamental period so there are a number of data points in one period. The points more than one period apart will become even more redundant conditioning information. This will be further discussed in Section 4.5.

4.4.1 Reduced-order Likelihood Function

Use the vector $\mathbf{Y}_{n,n'}$ to denote the zero-mean random vector comprised of the response measurements from time $n\Delta t$ to $n'\Delta t$ ($n \leq n'$) in a time-descending order:

$$\mathbf{Y}_{n,n'} = [\mathbf{y}_{n'}^T, \mathbf{y}_{n'-1}^T, \ldots, \mathbf{y}_n^T]^T, \qquad n \leq n' \tag{4.39}$$

Then, the joint PDF $p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_p}|\boldsymbol{\theta}, \mathcal{C})$ follows an $N_o N_p$-variate Gaussian distribution with zero mean and covariance matrix $\boldsymbol{\Sigma}_{\mathrm{Y}_{1,N_p}}$:

$$\boldsymbol{\Sigma}_{\mathrm{Y}_{1,N_p}} = E[\mathbf{Y}_{1,N_p}\mathbf{Y}_{1,N_p}^T] = \begin{bmatrix} \boldsymbol{\Gamma}_{N_p,N_p} & \text{sym} \\ \vdots & \ddots \\ \boldsymbol{\Gamma}_{1,N_p} & \cdots \boldsymbol{\Gamma}_{1,1} \end{bmatrix} \tag{4.40}$$

where the submatrix $\boldsymbol{\Gamma}_{n,n'}$ has the dimension $N_o \times N_o$. Based on Equation (4.29), the matrix $\boldsymbol{\Gamma}_{n,n'}$ is given by:

$$\boldsymbol{\Gamma}_{n,n'} = E\left[\mathbf{y}_n \mathbf{y}_{n'}^T\right] = \mathbf{L}_o \mathbf{R}_x(n\Delta t, n'\Delta t)\mathbf{L}_o^T + \boldsymbol{\Sigma}_\epsilon \delta_{nn'} \tag{4.41}$$

where $\delta_{nn'}$ is the Kronecker Delta, $\mathbf{R}_x$ denotes the auto-correlation matrix function of the model response $\mathbf{x}$, given by Equation (4.24), and $\boldsymbol{\Sigma}_\epsilon$ is the noise covariance matrix defined in Equation (4.28). Since the random vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_p}$ are jointly Gaussian with zero mean, the reduced-order likelihood function $p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_p}|\boldsymbol{\theta}, \mathcal{C})$ is given by:

$$p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_p}|\boldsymbol{\theta}, \mathcal{C}) = (2\pi)^{-\frac{N_o N_p}{2}} |\boldsymbol{\Sigma}_{Y_{1,N_p}}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{Y}_{1,N_p}^T \boldsymbol{\Sigma}_{Y_{1,N_p}}^{-1} \mathbf{Y}_{1,N_p}\right) \tag{4.42}$$

To compute this likelihood function, it involves only the solution of the linear algebraic equation $\boldsymbol{\Sigma}_{Y_{1,N_p}}\mathbf{X} = \mathbf{Y}_{1,Np}$ and the determinant of the matrix $\boldsymbol{\Sigma}_{Y_{1,N_p}}$ which is $N_o N_p \times N_o N_p$ only.

4.4.2 Conditional PDFs

In this section, the general expression is given for the conditional probability densities involving N_p previous points $p(\mathbf{y}_n|\boldsymbol{\theta}; \mathbf{y}_{n-N_p}, \mathbf{y}_{n-N_p+1}, \ldots, \mathbf{y}_{n-1}, \mathcal{C})$ in Equation (4.34), with $n > N_p \geq 1$. First, the random vector $\mathbf{Y}_{n-N_p,n}$ has the zero mean and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Y}_{n-N_p,n}}$:

$$\boldsymbol{\Sigma}_{\mathbf{Y}_{n-N_p,n}} = E[\mathbf{Y}_{n-N_p,n}\mathbf{Y}_{n-N_p,n}^T] = \begin{bmatrix} \boldsymbol{\Gamma}_{n,n} & & \text{sym} \\ \vdots & \ddots & \\ \boldsymbol{\Gamma}_{n-N_p,n} & \cdots & \boldsymbol{\Gamma}_{n-N_p,n-N_p} \end{bmatrix} \tag{4.43}$$

www.EngineeringBooksPdf.com

where the submatrix $\mathbf{\Gamma}_{n,n'}$ is given by Equation (4.41). Then, this matrix is partitioned as follows:

$$\mathbf{\Sigma}_{\mathbf{Y}_{n-N_p,n}} = \begin{bmatrix} \mathbf{\Sigma}_{11,n} & \mathbf{\Sigma}_{12,n} \\ \mathbf{\Sigma}_{12,n}^T & \mathbf{\Sigma}_{22,n} \end{bmatrix} \tag{4.44}$$

where the matrices $\mathbf{\Sigma}_{11,n}$, $\mathbf{\Sigma}_{12,n}$ and $\mathbf{\Sigma}_{22,n}$ have dimensions $N_o \times N_o$, $N_o \times N_o N_p$ and $N_o N_p \times N_o N_p$, respectively. $\mathbf{\Sigma}_{11,n}$ and $\mathbf{\Sigma}_{22,n}$ are the unconditional covariance matrices of the prediction target variables and the conditioning variables, and the matrix $\mathbf{\Sigma}_{12,n}$ quantifies their correlation.

Since the measured response has zero mean, the optimal estimator $\hat{\mathbf{y}}_n$ of $\mathbf{y}_n$ conditional on $\mathbf{Y}_{n-N_p,n-1}$ is given by (refer to Appendix C or Brockwell and Davis [35] for the proof):

$$\hat{\mathbf{y}}_n \equiv E[\mathbf{y}_n | \mathbf{y}_{n-N_p}, \mathbf{y}_{n-N_p+1}, \ldots, \mathbf{y}_{n-1}, \boldsymbol{\theta}, \mathcal{C}] = \mathbf{\Sigma}_{12,n} \mathbf{\Sigma}_{22,n}^{-1} \mathbf{Y}_{n-N_p,n-1} \tag{4.45}$$

and the covariance matrix $\mathbf{\Sigma}_{e,n}$ of the prediction error $\mathbf{e}_n = \mathbf{y}_n - \hat{\mathbf{y}}_n$ is given by:

$$\begin{aligned} \mathbf{\Sigma}_{e,n} &\equiv E\left[\mathbf{e}_n \mathbf{e}_n^T \middle| \mathbf{y}_{n-N_p}, \mathbf{y}_{n-N_p+1}, \ldots, \mathbf{y}_{n-1}, \boldsymbol{\theta}, \mathcal{C} \right] \\ &= \mathbf{\Sigma}_{11,n} - \mathbf{\Sigma}_{12,n} \mathbf{\Sigma}_{22,n}^{-1} \mathbf{\Sigma}_{12,n}^T \end{aligned} \tag{4.46}$$

Therefore, the conditional probability density $p(\mathbf{y}_n | \mathbf{y}_{n-N_p}, \mathbf{y}_{n-N_p+1}, \ldots, \mathbf{y}_{n-1}, \boldsymbol{\theta}, \mathcal{C})$ follows an N_o-variate Gaussian distribution with mean $\hat{\mathbf{y}}_n$ and covariance matrix $\mathbf{\Sigma}_{e,n}$:

$$\begin{aligned} &p(\mathbf{y}_n | \mathbf{y}_{n-N_p}, \mathbf{y}_{n-N_p+1}, \ldots, \mathbf{y}_{n-1}, \boldsymbol{\theta}, \mathcal{C}) \\ &\qquad = (2\pi)^{-\frac{N_o}{2}} |\mathbf{\Sigma}_{e,n}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (\mathbf{y}_n - \hat{\mathbf{y}}_n)^T \mathbf{\Sigma}_{e,n}^{-1} (\mathbf{y}_n - \hat{\mathbf{y}}_n) \right] \end{aligned} \tag{4.47}$$

Therefore, the reduced-order likelihood function and the conditional PDFs in the approximated expansion are available and the procedure can be summarized as follows:

1. Use Equation (4.30) with the likelihood function $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})$ being computed through the approximation in Equation (4.34).
2. The reduced-order likelihood function $p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_p} | \boldsymbol{\theta}, \mathcal{C})$ can be calculated using Equation (4.42) along with Equation (4.40) and (4.41).
3. Each of the conditional probability densities in Equation (4.34) can be calculated by Equation (4.47) along with Equation (4.43)–(4.46).

The most probable parameter vector $\boldsymbol{\theta}^\star$ is obtained by minimizing the objective function which is the negative logarithm of the posterior PDF (without including the terms that do not depend on $\boldsymbol{\theta}$):

$$\begin{aligned} J(\boldsymbol{\theta}) = &- \ln p(\boldsymbol{\theta}|\mathcal{C}) + \frac{1}{2} \ln \left| \mathbf{\Sigma}_{Y_{1,N_p}} \right| + \frac{1}{2} \mathbf{Y}_{1,N_p}^T \mathbf{\Sigma}_{Y_{1,N_p}}^{-1} \mathbf{Y}_{1,N_p} \\ &+ \frac{1}{2} \sum_{n=N_p+1}^{N} \left[\ln |\mathbf{\Sigma}_{e,n}| + (\mathbf{y}_n - \hat{\mathbf{y}}_n)^T \mathbf{\Sigma}_{e,n}^{-1} (\mathbf{y}_n - \hat{\mathbf{y}}_n) \right] \end{aligned} \tag{4.48}$$

www.EngineeringBooksPdf.com

The reason for minimizing the objective function $J(\boldsymbol{\theta})$ instead of maximizing the posterior PDF directly is that the former has a better computational condition. It is found that the updated PDF of the parameter vector $\boldsymbol{\theta}$ can be well approximated by a Gaussian distribution $\mathcal{G}(\boldsymbol{\theta}; \boldsymbol{\theta}^\star, \mathcal{H}(\boldsymbol{\theta}^\star)^{-1})$ with mean $\boldsymbol{\theta}^\star$ and covariance matrix $\mathcal{H}(\boldsymbol{\theta}^\star)^{-1}$, where $\mathcal{H}(\boldsymbol{\theta}^\star)$ denotes the Hessian of the objective function $J(\boldsymbol{\theta})$ calculated at $\boldsymbol{\theta} = \boldsymbol{\theta}^\star$.

In the special case of stationary response, the matrices $\boldsymbol{\Sigma}_{e,n} = \boldsymbol{\Sigma}_e$ and $\boldsymbol{\Sigma}_{12,n}\boldsymbol{\Sigma}_{22,n}^{-1} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$ required in the computation of the optimal estimator $\hat{\mathbf{y}}_n$ do not depend on the time step n. Therefore, the objective function in Equation (4.48) can be simplified as follows:

$$J(\boldsymbol{\theta}) = -\ln p(\boldsymbol{\theta}|\mathcal{C}) + \frac{1}{2}\ln\left|\boldsymbol{\Sigma}_{Y_{1,N_p}}\right| + \frac{1}{2}\mathbf{Y}_{1,N_p}^T\boldsymbol{\Sigma}_{Y_{1,N_p}}^{-1}\mathbf{Y}_{1,N_p} + \frac{N-N_p}{2}\ln|\boldsymbol{\Sigma}_e| + \\ + \frac{1}{2}\sum_{n=N_p+1}^{N}(\mathbf{y}_n - \hat{\mathbf{y}}_n)^T\boldsymbol{\Sigma}_e^{-1}(\mathbf{y}_n - \hat{\mathbf{y}}_n) \tag{4.49}$$

In this case, it requires only the computation of the inverse and determinant of the matrices $\boldsymbol{\Sigma}_{Y_{1,N_p}}$ $(= \boldsymbol{\Sigma}_{22})$ and $\boldsymbol{\Sigma}_e$, which are $N_oN_p \times N_oN_p$, and $N_o \times N_o$, respectively. These sizes are much smaller than the $NN_o \times NN_o$ matrix $\boldsymbol{\Sigma}_{Y_{1,N}}$ in the direct formulation.

Although the above formulation is presented for displacement time histories, it can be modified easily for velocity or acceleration measurements by using the corresponding modal impulse response functions for velocity or acceleration in Equation (4.24). Of course, the case of relative acceleration with white noise excitation is not realistic since the response variance is infinity. However, absolute acceleration measurements can be considered for ground excitation or relative acceleration measurements with non-white excitation.

If the right hand side of Equation (4.13) is replaced by $\sum_{r=1}^{N_r} A_r(t|\boldsymbol{\theta}_A)\mathbf{g}_r(t)$, the Bayesian time-domain methodology is applicable for excitations having different modulating functions, e.g., ambient vibrations with a series of wind gusts.

4.5 Numerical Verification

This example considers the identification of a single-degree-of-freedom system with a simulated noisy stationary displacement measurement. The actual parameters in $\tilde{\boldsymbol{\theta}} = [\tilde{\Omega}, \tilde{\zeta}, \tilde{S}_{f0}, \tilde{\sigma}_\epsilon]^T$ for data generation are $\tilde{\Omega} = 5$ rad/s, $\tilde{\zeta} = 0.01$, $\tilde{S}_{f0} = 10$ cm^2 s^{-3} and $\tilde{\sigma}_\epsilon = 0.3545$ cm. The chosen value of $\tilde{\sigma}_\epsilon$ corresponds to a 10% prediction-error level, i.e., the rms of the noise is 10% of the rms of the noise-free response. The sampling time step is $\Delta t = 0.05$ s, and the total time interval is $T = 200$ s, i.e., $N = 4000$. However, the system response was generated with a finer time step of 0.005 s to simulate the realistic situation that the actual signal has a larger range of frequency content than that the measurement can detect. Furthermore, the simulation starts with a zero initial condition for a total simulation interval of 210 s and the measurement is taken by ignoring the first 10 s in order to obtain a stationary response. The non-informative prior distribution $p(\boldsymbol{\theta}|\mathcal{C})$ is utilized so the parameters are inferred solely by the measurements.

Figure 4.1 validates the approximation in Equation (4.34). The top subplot shows the auto-correlation function $R_x(n\Delta t|\tilde{\boldsymbol{\theta}})$ with the actual parameters. The optimal predictor $\hat{y}_{2001}$, at the 2001st time step, and the corresponding standard deviation σ_ϵ are calculated from

www.EngineeringBooksPdf.com

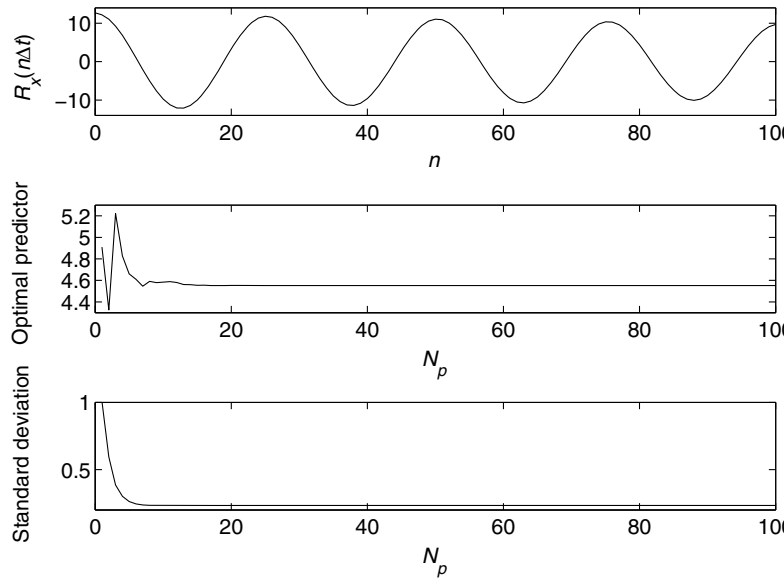

Figure 4.1 Auto-correlation function, predicted response and standard deviation of the prediction error

Equations (4.45) and (4.46). They are shown in the middle and bottom subplots, respectively, as a function of the number of previous points N_p used for the prediction. For values of N_p beyond approximately 15 (corresponding to slightly over half of a period of the oscillator), the optimal predictor of the response and its associated uncertainty stabilizes. This implies that increasing the value of N_p beyond $N_p = 15$ in Equation (4.34) will not further improve the quality of the identification results. Thus, it is conservative to suggest that N_p is chosen to contain approximately one period of the oscillator. It is worth noting that the value of N_p at which stabilization occurs is rather insensitive to the prediction error level (10% in this case). If a larger value of N_p is utilized, the incremental improvement in the identification results will be insignificant but the computational effort will increase rapidly. The low value of N_p at which the values of $\hat{y}_{2001}$ and σ_ϵ stabilize is rather surprising, given that the auto-correlation function is far from being decayed to negligible values but this was explained in Section 4.4. The computational efficiency of the Bayesian time-domain is built on the appreciation of this low value of N_p.

Table 4.1 refers to the identification results using a single set of displacement measurements $\mathcal{D}$. It shows the optimal values in $\boldsymbol{\theta}^\star = [\Omega^\star, \zeta^\star, S_{f0}^\star, \sigma_\epsilon{}^\star]^T$, the calculated standard deviations $\sigma_\Omega, \sigma_\zeta, \sigma_{S_{f0}}$ and σ_{σ_ϵ}, the coefficient of variation (COV) for each parameter and the normalized distance for each parameter. This distance represents the absolute value of the difference between the optimal and actual value of the parameter, normalized by the corresponding calculated standard deviation. Here, the value $N_p = 25$ (corresponding to roughly one period of the oscillator) was used in Equation (4.34). Repeating the identification with $N_p = 50$ yields identical results.

Figure 4.2 shows the contours of the marginal PDF of the natural frequency Ω and the damping ratio ζ. The two elliptical contours enclose the region with probability 0.5 and 0.9,

www.EngineeringBooksPdf.com

Table 4.1 Identification results using one set of response data

Parameter	Actual $\tilde{\theta}$	Optimal $\theta^\star$	S.D. σ	COV	$\lvert\tilde{\theta} - \theta^\star\rvert/\sigma$
Ω	5.0000	4.9840	0.0168	0.003	0.95
ζ	0.0100	0.0113	0.0033	0.327	0.41
S_{f0}	10.000	10.641	0.6256	0.063	1.03
σ_ϵ	0.3545	0.3433	0.0045	0.013	2.51

respectively. The estimates of these two parameters can be treated as independent since the ellipses have horizontal major axes. Figure 4.3 shows the contours of the marginal PDF of the damping ratio ζ and the spectral intensity of the excitation S_{f0}. In contrast to the previous case, the estimates of these two parameters are highly correlated so the elliptical contours are rotated to have inclined major axes. Increasing the damping ratio and the spectral intensity with the same factor yields similar auto-correlation functions (only the damped natural frequency will be affected slightly) and hence a similar probability density given the data.

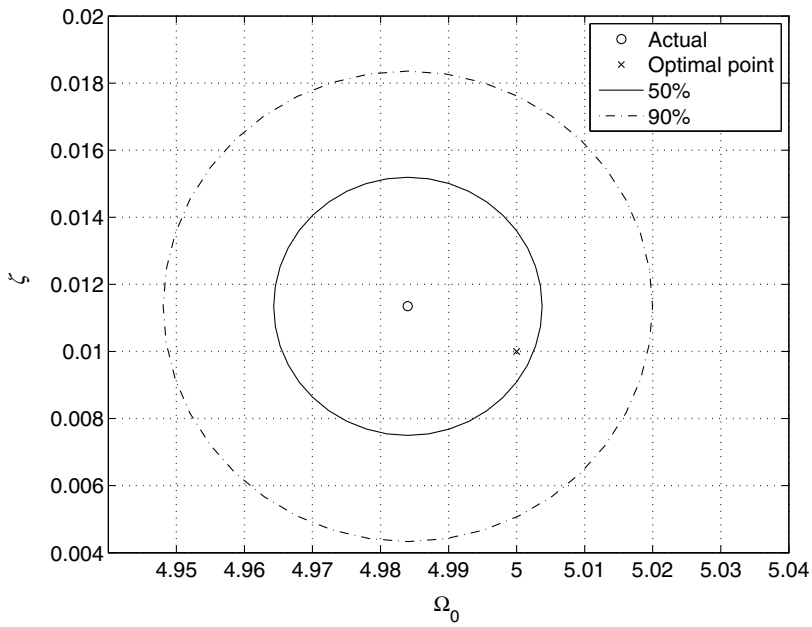

Figure 4.2 Contours of the marginal PDF of Ω and ζ

Figure 4.4 shows the conditional PDF of the natural frequency Ω and the damping ratio ζ with the spectral intensity and the prediction-error variance fixed at their optimal values. The conditional PDFs by the Bayesian time-domain approach and the Gaussian approximation are plotted with solid lines and dashed lines, respectively. The two groups of curves are on top on each other, indicating that the Gaussian approximation is accurate when the number of data points is sufficiently large. This can be used to represent the posterior PDF, e.g., for statistical moments computation.

www.EngineeringBooksPdf.com

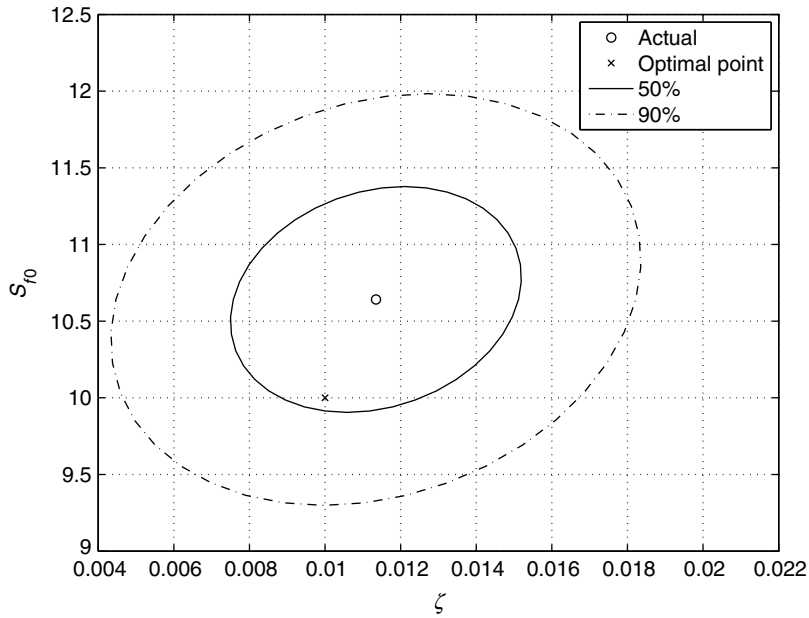

Figure 4.3 Contours of the marginal PDF of ζ and S_{f0}

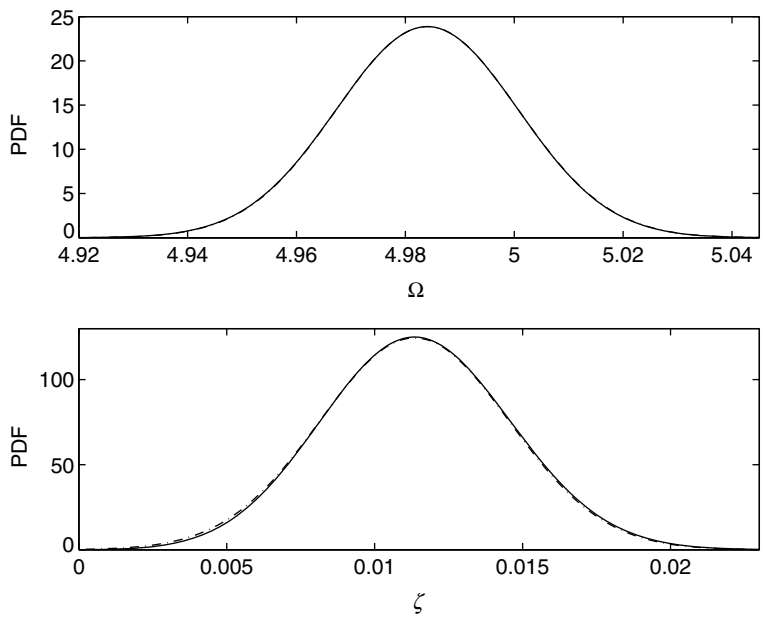

Figure 4.4 Conditional PDFs of Ω and ζ

www.EngineeringBooksPdf.com

Another case is investigated with a very short period of measurement, namely $T = 5$ s, so it contains less than four fundamental periods of the oscillator. The Bayesian time-domain method is used for the identification and Figure 4.5 shows the conditional PDF of Ω and ζ with all other parameters fixed at their optimal values. The solid lines show the conditional posterior PDF obtained by the Bayesian method and the dashed lines show the Gaussian approximation. It is clear that the posterior PDF is non-Gaussian. This confirms that the Bayesian time-domain approach is capable to offer the correct inference without assuming the type of the posterior PDF. In the case of a non-Gaussian posterior PDF, statistical moments, such as the variances of the estimates, can be computed by direct Monte Carlo simulation. The results are shown in Table 4.2 in the same fashion as Table 4.1. The computed uncertainty obtained here is reasonable by judging the normalized distance of the estimates.

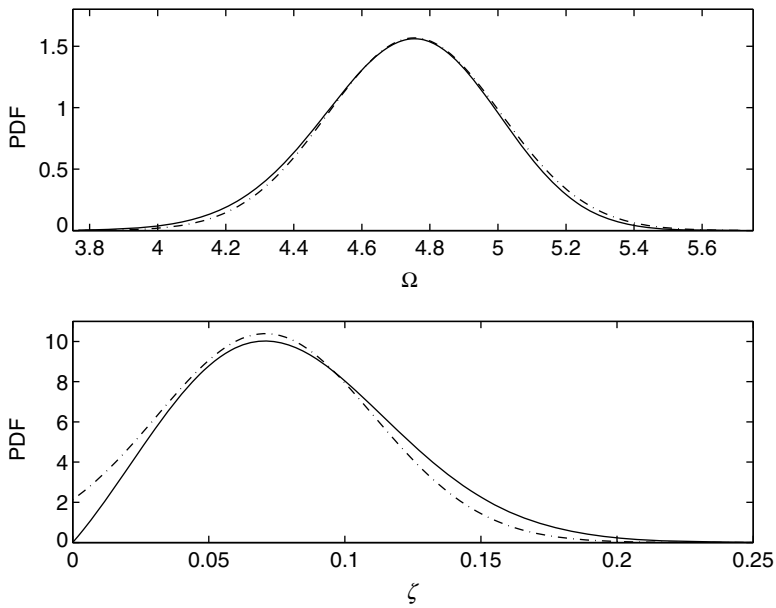

Figure 4.5 Conditional PDFs of Ω and ζ for $T = 5$ s

Table 4.2 Identification results using one set of response data with $T = 5$ s

Parameter	Actual $\tilde{\theta}$	Optimal $\theta^\star$	S.D. σ	COV	$\lvert\tilde{\theta} - \theta^\star\rvert/\sigma$
Ω	5.0000	4.7544	0.2599	0.052	0.94
ζ	0.0100	0.0708	0.0460	4.602	1.32
S_{f0}	10.000	14.078	5.0632	0.506	0.81
σ_ϵ	0.3545	0.4078	0.0329	0.093	1.62

To confirm the convergence of the Bayesian time-domain method, a response measurement of $T = 10\,000$ s is generated. Then, the time-domain method is applied for identification with different duration of the measurement: $T = 200, 210, 220, \ldots, 10\,000$ s. The identified

www.EngineeringBooksPdf.com

natural frequency and damping ratio are shown with solid lines in Figure 4.6. Furthermore, the dashed lines show the optimal plus and minus three standard deviations for different durations of measurement. In other words, the intervals between the dashed lines are the $\pm 3\sigma$ confidence intervals, including approximately a probability of 99.7%. It is not surprising that the uncertainty of the parameter estimates decreases with increasing durations of the measurement. Furthermore, the identified parameters are converging to the actual values when the measured duration increases.

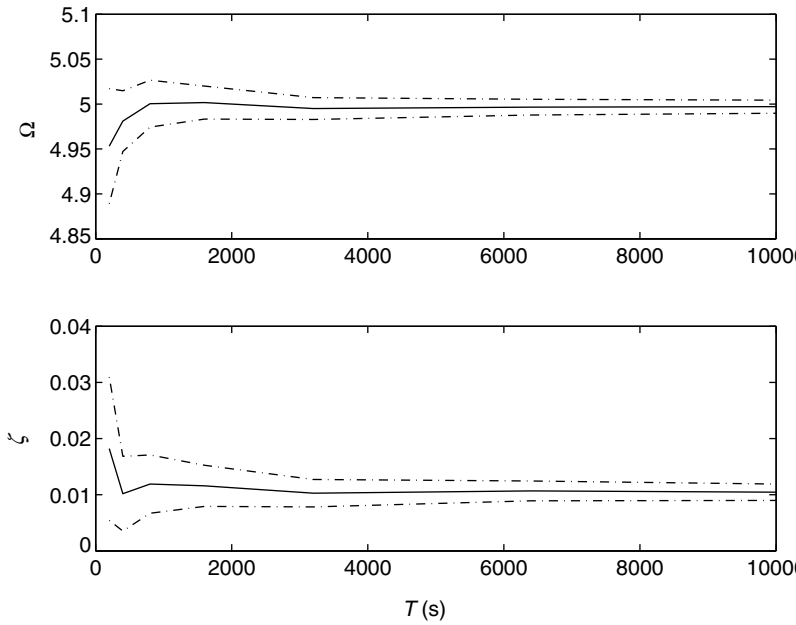

Figure 4.6 Identified natural frequency and damping ratio with different durations of measurement

4.6 Application to Model Updating with Unmeasured Earthquake Ground Motion

4.6.1 Transient Response of a Linear Oscillator

This example considers the identification of a single-degree-of-freedom (SDOF) system with the simulated noisy transient displacement measurement shown in Figure 4.7. Here, the modulating function A in Equation (4.20) is taken as the Heaviside unit step function, and the input f is zero-mean stationary Gaussian white noise with spectral intensity S_{f0}. The actual parameters in $\tilde{\boldsymbol{\theta}} = [\tilde{\Omega}, \tilde{\zeta}, \tilde{S}_{f0}, \tilde{\sigma}_\epsilon]^T$ used to generate the simulated data are $\tilde{\Omega} = 5.0$ rad/s, $\tilde{\zeta} = 0.01$, $\tilde{S}_{f0} = 10$ cm^{2}s^{-3} and $\tilde{\sigma}_\epsilon = 0.2285$ cm. The chosen value of $\tilde{\sigma}_\epsilon$ corresponds to a 10% rms prediction-error level, i.e., the noise is 10% of the rms of the noise-free response. The time step used to generate the data is 0.005s. However, a much larger sampling time step was chosen ($\Delta t = 0.05$s) and the total time interval is $T = 50$s, so that the number of data points is $N = 1000$.

www.EngineeringBooksPdf.com

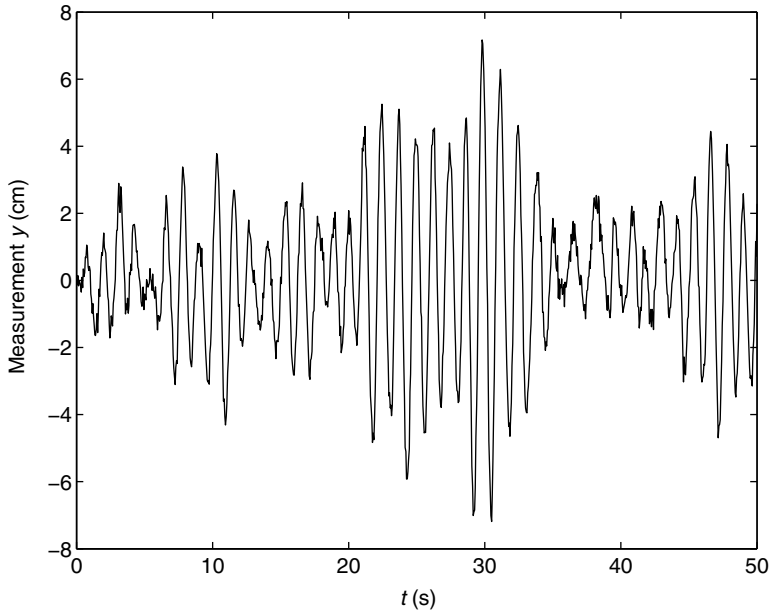

Figure 4.7 Measured time history of the transient response

Table 4.3 refers to the identification results using a single set of displacement measurements $\mathcal{D}$. It shows the most probable values of the parameters, the calculated standard deviations σ_Ω, σ_ζ, $\sigma_{S_{f0}}$ and σ_{σ_ϵ}, the coefficient of variation for each parameter and the normalized distance for each parameter, defined in the same way as in the previous example. Here, the value $N_p = 25$ (corresponding to approximately one period of the oscillator) was used in Equation (4.34). Repeating the identification with a value of $N_p = 50$ yields identical results, verifying that using $N_p = 2\pi/\Omega\Delta t$ is sufficient.

Table 4.3 Identification results for one set of data

Parameter	Actual $\tilde{\theta}$	Optimal $\theta^\star$	S.D. σ	COV	$\lvert\tilde{\theta} - \theta^\star\rvert/\sigma$
Ω	5.0000	5.0065	0.0513	0.010	0.13
ζ	0.0100	0.0237	0.0105	1.051	1.31
S_{f0}	10.000	10.712	1.1110	1.111	0.64
σ_ϵ	0.2285	0.2188	0.0058	0.026	1.65

Figure 4.8 shows contours in the (Ω, ζ) plane of the marginal updated PDF $p(\Omega, \zeta|\mathcal{D}, \mathcal{C})$ calculated for the set of simulated data used for Table 4.3. The estimates of Ω and ζ can be considered uncorrelated since the major axes of the ellipses are horizontal. Figure 4.9 shows the contours in the (ζ, S_{f0}) plane of the marginal updated PDF $p(\zeta, S_{f0}|\mathcal{D}, \mathcal{C})$. The estimates of the damping ratio and the spectral intensity are considered highly correlated. This is in contrast to the case of Ω and ζ. The Gaussian approximation for the updated PDF can be

www.EngineeringBooksPdf.com

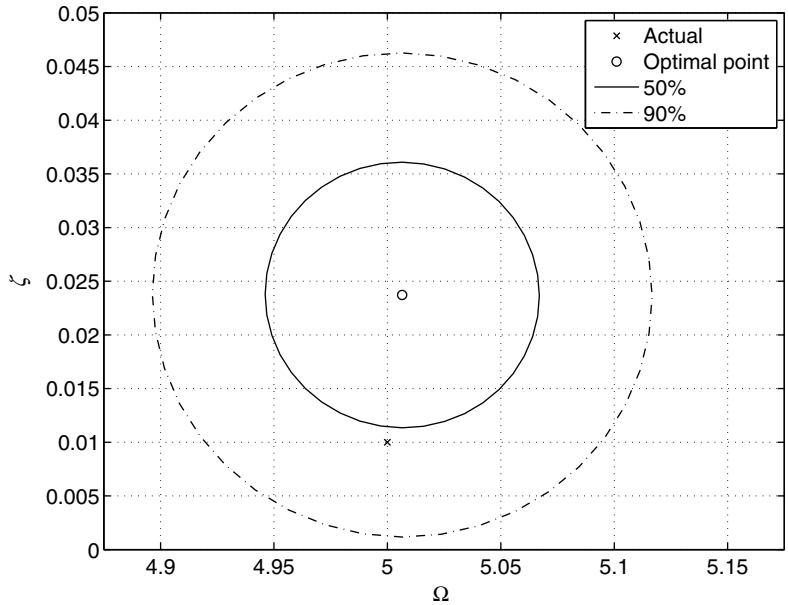

Figure 4.8 Contours of the marginal updated joint PDF of the natural frequency and damping ratio

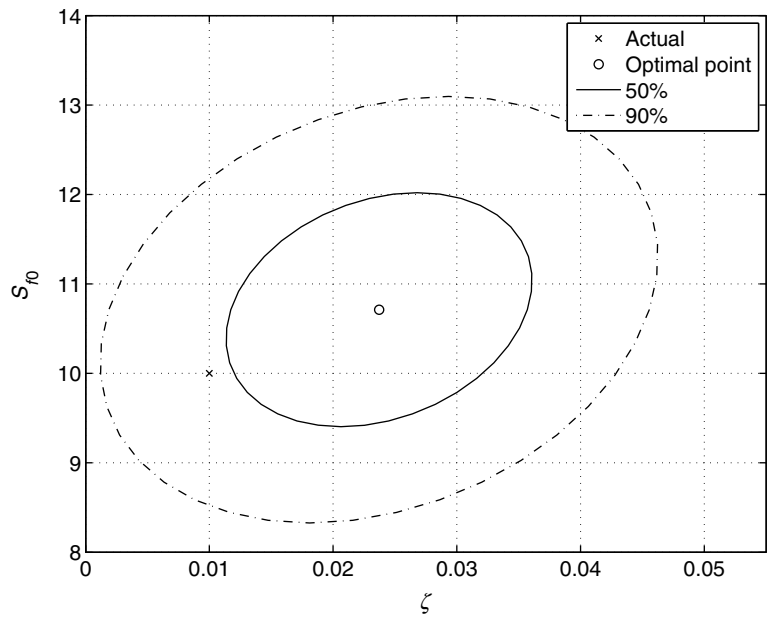

Figure 4.9 Contours of the marginal updated joint PDFs of the damping ratio and spectral intensity

www.EngineeringBooksPdf.com

verified by examining the conditional PDFs. Figure 4.10 shows the conditional PDFs of the natural frequency and the damping ratios: $p(\Omega|\mathcal{D}, \zeta^\star, S^\star_{f0}, \sigma^\star_\epsilon, \mathcal{C})$ and $p(\zeta|\mathcal{D}, \Omega^\star, S^\star_{f0}, \sigma^\star_\epsilon, \mathcal{C})$. The solid lines are obtained by using the Bayesian time-domain approach without assuming the type of posterior distributions and the dashed lines are obtained by using the Gaussian approximation. It can be seen that the two sets of curves are on top of each other, implying that the Gaussian approximation is accurate.

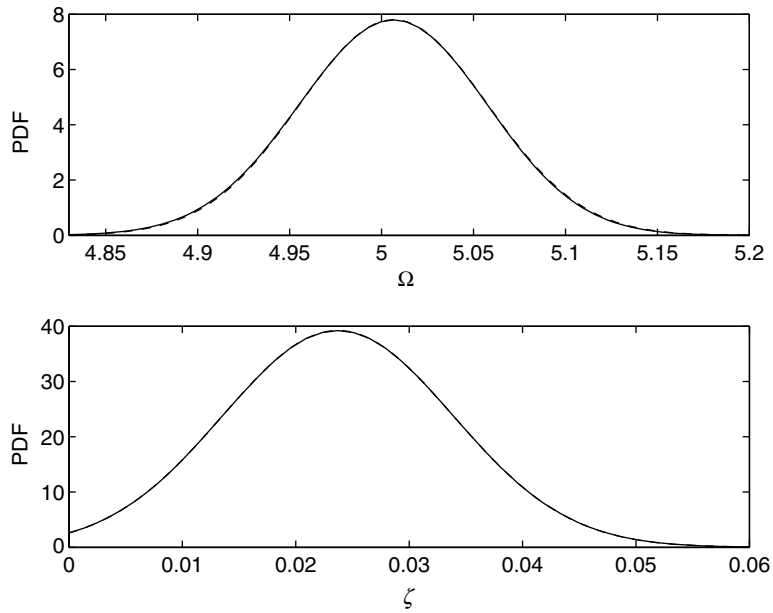

Figure 4.10 Conditional PDFs of the natural frequency and damping ratio

4.6.2 Building Subjected to Nonstationary Ground Excitation

This example uses the simulated response data from the 10-story shear building shown in Figure 4.11. This building has a uniformly distributed floor mass and interstory stiffness over its height and the stiffness to mass ratio is chosen to be 4637 s^{-2} so that the first four modal frequencies are 1.62 Hz, 4.82 Hz, 7.92 Hz and 10.84 Hz. The damping ratios are 2% for all modes. Displacements at the fourth and top floors were measured over a time interval of $T = 40$ s, with a sampling time interval $\Delta t = 0.02$ s. Therefore, the total number of measured time steps is $N = 2000$ and corresponds to approximately 65 fundamental periods of the building. Note that a much smaller time step (0.002 s) was used for the data simulation so that the signal contains a frequency content higher than the Nyquist frequency in order to simulate a realistic situation.

The structure is subjected to base acceleration given by stationary white noise with spectral intensity $S_{\ddot{g}0} = 0.25\ \mathrm{m^2 s^{-3}}$ modulated by:

$$A(t) = A_0 t e^{1-A_0 t} U(t) \tag{4.50}$$

www.EngineeringBooksPdf.com

where $U(.)$ denotes the Heaviside unit step function. This modulating function has its maximum $\max_t A(t) = 1$ at time $t = 1/A_0$ and the value of A_0 is taken to be 0.1s^{-1}. The forcing distributing matrix in Equation (4.12) for the ground motion is given by $\mathbf{T}_0 = \mathbf{M}[-1, -1, \ldots, -1]^T$. The measurement noise for the response is taken to be 10%, i.e., the rms of the measurement noise for a particular channel of measurement is equal to 10% of the rms of the noise-free response at the corresponding DOF. Modal updating using the Bayesian time-domain approach is carried out for the lowest three modes of the structure. The number of previous points used for the conditional PDF is $N_p = 30$ and corresponds to using previous data points roughly over one fundamental period as the conditioning information at each time step in Equation (4.34).

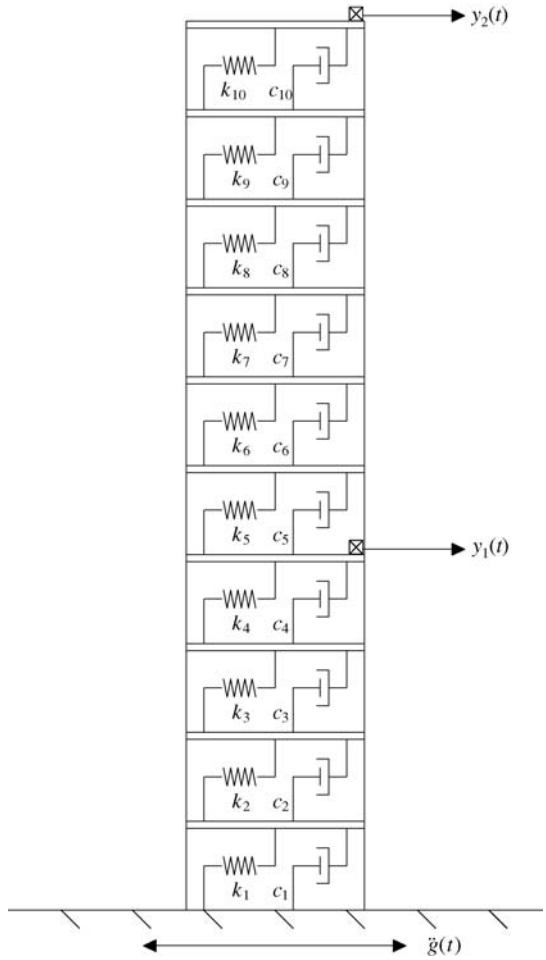

Figure 4.11 Ten-story shear building subjected to earthquake ground motion

Figure 4.12 shows the displacement time histories of the fourth and top floor. The structure was on a calm environment before the earthquake arrived. Then, it was excited and the maximum response was achieved approximately at the 8th second. The excitation level and hence the response dropped gradually after the 10th second. *Beating* was also observed in the

www.EngineeringBooksPdf.com

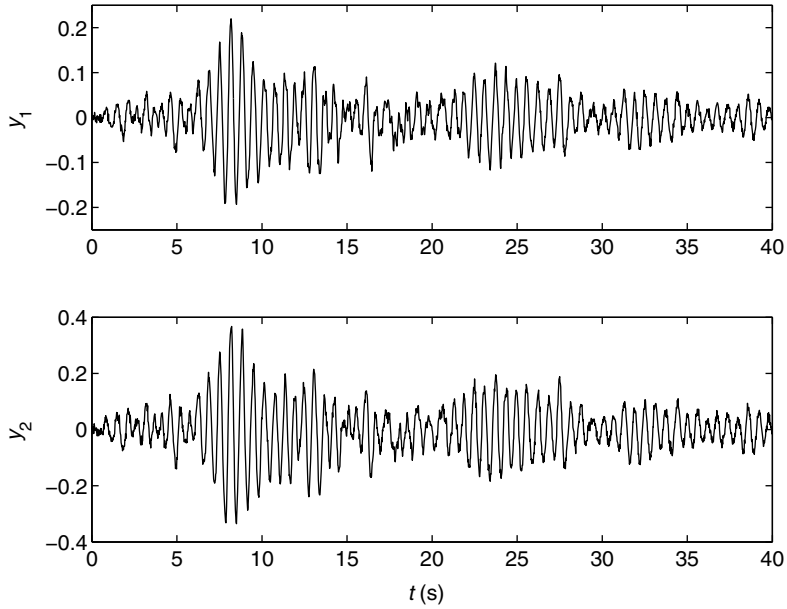

Figure 4.12 Measurements on the fourth and top floor

measurement. Figure 4.13 shows the corresponding Fourier amplitude spectrum. It is worth noting that the peak of the third mode at 7.92 Hz does not appear in the spectrum of the 4th floor and it is a challenging task to identify this mode. Table 4.4 shows the identification results for three modes. The second column corresponds to the actual values used for generation of the simulated measurement, the third and fourth columns correspond to the identified optimal parameters and the corresponding standard deviations, respectively, the fifth column lists the coefficients of variation for the parameters, and the last column shows the normalized distance described previously. The first group of rows in the table corresponds to modal frequencies, followed by the modal damping ratios, the ratios of the mode shape components between the fourth and the top floors, the elements of the modal forcing spectral matrix $\mathbf{S}_{f0}$, the elements of the prediction-error covariance matrix $\mathbf{\Sigma}_\epsilon$ and finally the amplitude of the modulating function A_0. In this case of ground excitation, the off-diagonal entities of the spectral intensity matrix are given by:

$$S_{f0}^{(m,m')} = \sqrt{S_{f0}^{(m,m)} S_{f0}^{(m',m')}}, \quad m, m' = 1, 2, \ldots, N_m \tag{4.51}$$

so only the diagonal elements of $\mathbf{S}_{f0}$ have to be identified. The mode shapes are normalized so that the mode shape components at the top floor are equal to unity for each of the modes considered.

As observed previously in the examples for the Bayesian spectral density approach, the coefficients of variation for the frequencies in all cases are much smaller than those of the damping ratios, indicating that frequencies are identified much better than damping ratios. An additional result observed, but not tabulated here, is that the modal damping ratios exhibit significant correlation with the corresponding modal forcing spectral intensities.

www.EngineeringBooksPdf.com

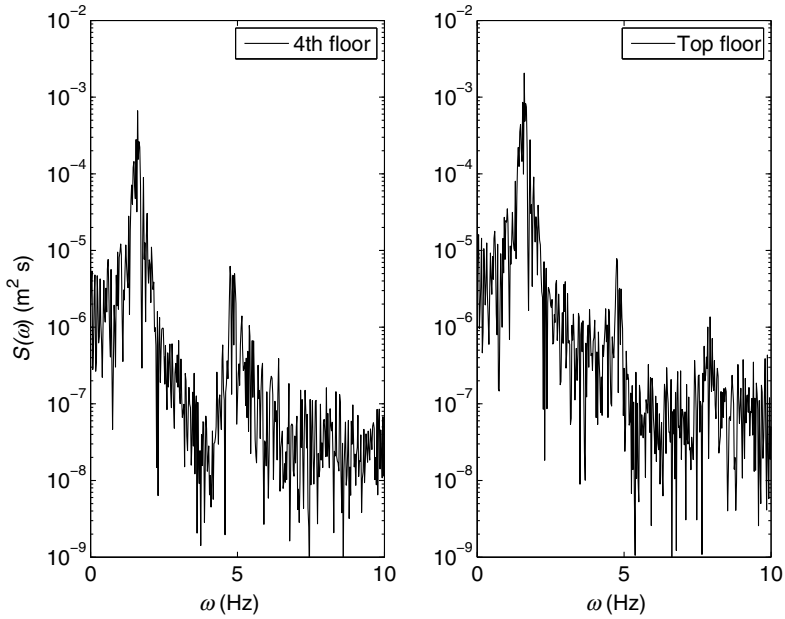

Figure 4.13 Response spectra of the measurements

Table 4.4 Identification results for the 10-story shear building

Parameter	Actual $\tilde{\theta}$	Optimal $\theta^\star$	S.D. σ	COV	$\lvert\tilde{\theta} - \theta^\star\rvert/\sigma$
$\Omega^{(1)}$	1.6198	1.6108	0.0141	0.009	0.64
$\Omega^{(2)}$	4.8232	4.8471	0.0261	0.005	0.91
$\Omega^{(3)}$	7.9189	7.8647	0.0760	0.010	0.71
$\zeta^{(1)}$	0.0200	0.0261	0.0076	0.382	0.80
$\zeta^{(2)}$	0.0200	0.0212	0.0030	0.152	0.40
$\zeta^{(3)}$	0.0200	0.0146	0.0086	0.432	0.62
$\phi_{10}^{(1)}/\phi_4^{(1)}$	1.7702	1.7638	0.0063	0.004	1.03
$\phi_{10}^{(2)}/\phi_4^{(2)}$	−1.0000	−0.9220	0.0598	0.060	1.30
$\phi_{10}^{(3)}/\phi_4^{(3)}$	6.2457	10.3763	3.8341	0.614	1.08
$S_{f0}^{(1,1)}$	0.1281	0.1268	0.0114	0.089	0.12
$S_{f0}^{(2,2)}$	0.0414	0.0399	0.0051	0.122	0.29
$S_{f0}^{(3,3)}$	0.0003	0.0002	0.0001	0.177	3.02
$\sigma_{\epsilon 1}$ (fourth floor)	0.0052	0.0053	0.0001	0.019	1.14
$\sigma_{\epsilon 2}$ (top floor)	0.0091	0.0093	0.0002	0.018	1.06
A_0	0.1000	0.0992	0.0032	0.032	0.24

www.EngineeringBooksPdf.com

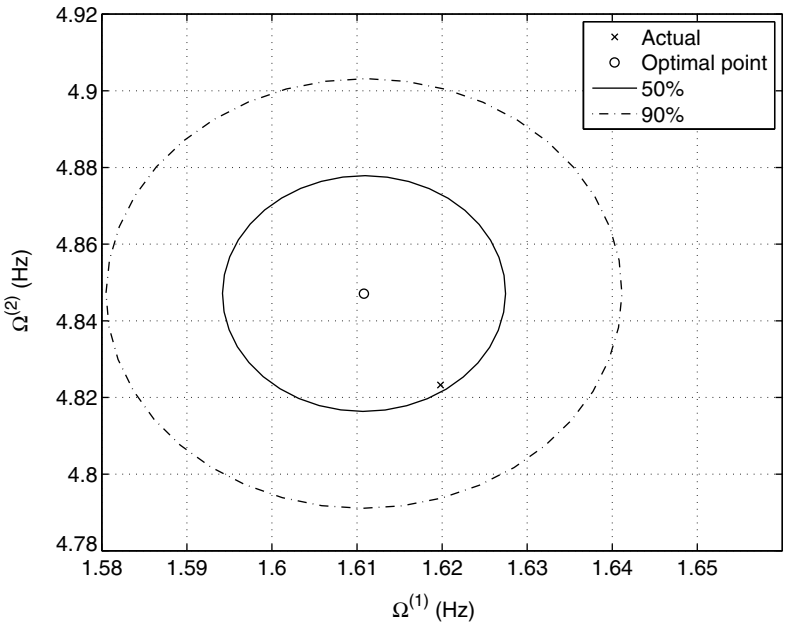

Figure 4.14 Contours of the marginal updated joint PDF of the first two modal frequencies

Figure 4.14 shows the contours in the $(\Omega^{(1)}, \Omega^{(2)})$ plane of the marginal updated PDF of $\Omega^{(1)}$ and $\Omega^{(2)}$. The actual parameters are at reasonable distances, measured in terms of the estimated standard deviations, from the identified optimal parameters. This verifies that the quantified uncertainty is reasonable. Figure 4.15 are typical plots showing comparisons between the conditional PDFs of $\Omega^{(1)}$ and $\Omega^{(2)}$ (keeping all the other parameters fixed at their optimal values) by the Bayesian time-domain method (solid lines) and the Gaussian approximation described at the end of Section 4.4 (dashed lines). It can be seen that the Gaussian approximation is very accurate as the two groups of lines are overlapping. Note that the Gaussian approximation is not an assumption of the method but the consequence of a large number of data points.

The same set of data is also analyzed by assuming a stationary response but the identification process is not converging for the third mode. This is not surprising by observing Figure 4.13, which shows no peak of the third mode of the building in its Fourier amplitude spectrum. Furthermore, the model parameters of the other modes are also biased. Therefore, consideration of the nonstationarity of the response is important when there is obvious evidence for the response to exhibit nonstationary behavior [285].

4.7 Concluding Remarks

In this chapter, the Bayesian time-domain approach was introduced for identification of the model parameters and stochastic excitation parameters of linear multi-degree-of-freedom systems using noisy stationary or nonstationary response measurements. The direct exact formulation was presented but it turned out to be computationally prohibited for a large number of data points. Then, an approximated likelihood function expansion was proposed to resolve this obstacle. For a globally identifiable case with a large number of data points, the updated PDF

www.EngineeringBooksPdf.com

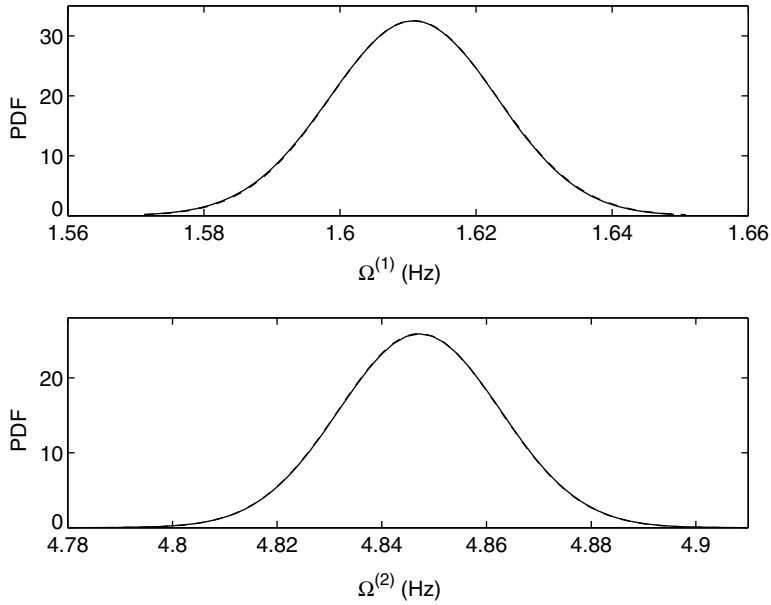

Figure 4.15 Conditional PDFs of the lower two modal frequencies

of the model parameters can be accurately approximated by a multi-variate Gaussian distribution. Note that this is not an assumption of the method but the consequence of the central limit theorem. The mean and covariance matrix of the updated distribution offer an estimate of the most probable values of the parameters and their associated uncertainties. Quantification of the uncertainty in the identified model parameters is useful for post-processing, such as damage detection or reliability analysis.

The presented methodology simultaneously utilizes response histories at all measured degrees of freedom, although only one observed degree of freedom is necessary to identify the modal frequencies and damping ratios. Measurement noise is considered explicitly. Computation of the uncertainty does not require parametric identification from a number of different data sets and then calculating the statistics of these estimates. Instead, it follows directly from the methodology applied to a single set of measurements. Finally, the Bayesian time-domain methodology is expected to lead to improved model identification using ambient vibration data where nonstationarity is evident.

4.8 Comparison of Spectral Density Approach and Time-domain Approach

Chapter 3 and Chapter 4 presented the Bayesian spectral density approach and Bayesian time-domain approach. The comparison can be summarized as follows:

1. *Computational efficiency*
   The Bayesian spectral density approach, which is a frequency-domain method, transforms the response time history to the spectral density estimator by discrete Fourier transform.

www.EngineeringBooksPdf.com

The spectral density estimators at different frequencies possess tractable properties so that they follow independent Wishart distributions in a certain frequency band, regardless of the distribution of the signal in the time domain. The method is efficient in the sense that most of the information from the data for identification of the model parameters, especially those related to the frequency structure, concentrates in a very limited bandwidth around the peaks in the spectrum. Therefore, the number of frequencies involved in the computation of the posterior PDF is significantly smaller than the total number of frequencies in the spectrum, i.e., INT($N/2$)+1. However, computation of the inverse and determinant of the matrices $E[\mathbf{S}_{y,N}(\omega_k)] \in \mathbb{C}^{N_o \times N_o}$, $k \in \mathcal{K}$, is required for each frequency included in establishing the likelihood function. For the case of a large number of observed degrees of freedom N_o, this will considerably slow down the methodology.

The Bayesian time-domain approach utilizes the Bayes' theorem repeatedly to factorize the likelihood function into the product of a joint PDF and conditional PDFs:

$$p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) \approx p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_p}|\boldsymbol{\theta}, \mathcal{C}) \prod_{n=N_p+1}^{N} p(\mathbf{y}_n|\boldsymbol{\theta}; \mathbf{y}_{n-N_p}, \mathbf{y}_{n-N_p+1}, \ldots, \mathbf{y}_{n-1}, \mathcal{C}) \tag{4.52}$$

This approximation resolves the computational difficulty encountered in the direct exact formulation that requires repeated computations of the solution of linear simultaneous algebraic equations and determinants of the matrices with huge dimensions. The efficiency in the approximated expansion is gained by the appreciation that the conditioning information can be truncated within one period of the system only. For linear systems, the expressions for the reduced-order likelihood function $p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_p}|\boldsymbol{\theta}, \mathcal{C})$ and the conditional PDFs $p(\mathbf{y}_n|\boldsymbol{\theta}; \mathbf{y}_{n-N_p}, \mathbf{y}_{n-N_p+1}, \ldots, \mathbf{y}_{n-1}, \mathcal{C})$ are available since they are Gaussian and the correlation functions are known in closed forms regardless of the stationarity of the response. For stationary response, the method is very efficient in the sense that evaluation of all the conditional PDFs $p(\mathbf{y}_n|\boldsymbol{\theta}; \mathbf{y}_{n-N_p}, \mathbf{y}_{n-N_p+1}, \ldots, \mathbf{y}_{n-1}, \mathcal{C})$ requires the inverse and determinant of two relatively small matrices only.

Even though the spectral density approach requires computation of the inverse and determinant of a number of matrices, the size of these matrices is only $N_o \times N_o$. They are significantly smaller than the $N_oN_p \times N_oN_p$ matrix $\boldsymbol{\Sigma}_{Y_{1,N_p}}(= \boldsymbol{\Sigma}_{22})$ required in the time-domain approach. Comparison of the computational efficiency between the two methods depends on the number of the elements in the frequency index set and the number of data points in a fundamental period. The ratio of the computations required by the Bayesian spectral density approach and the Bayesian time-domain approach can be approximated by:

$$\frac{C_{BSDA}}{C_{BTDA}} \approx \frac{N_\omega N_o}{N_p^3} \tag{4.53}$$

where N_ω is the number of points in the frequency index set in the Bayesian spectral density approach, N_o is the number of observed degrees of freedom and N_p is the number of previous time points to be included in the conditional PDF for the Bayesian time-domain approach. In a typical situation, the Bayesian spectral density approach is more efficient. For example, if $N_o = 10$, $N_\omega = 1000$, and $N_p = 50$, then $C_{BSDA}/C_{BTDA} \approx 2/25$.

www.EngineeringBooksPdf.com

2. *Accuracy*
   The Bayesian spectral density approach approximates the spectral density matrix estimators as Wishart distributed random matrices. This is the consequence of the special structure of the covariance matrix of the real and imaginary parts of the discrete Fourier transforms in Equation (3.53) [295]. Another approximation is made on the independency of the spectral density matrix estimators at different frequencies. These two approximations were verified to be accurate at the frequencies around the peaks of the spectrum. The spectral density estimators in the frequency range with small spectral values will become dependent since aliasing and leakage effects have a greater impact on their values. Therefore, the likelihood function is constructed to include the spectral density estimators in a limited bandwidth only. In particular, the loss of information due to the exclusion of some of the frequencies affects the estimation of the prediction-error variance but not the parameters that govern the time-frequency structure of the response, e.g., the modal frequencies or stiffness of a structure.

   The only approximation made in the Bayesian time-domain approach is that the system response at a particular time step estimated by its entire history is essentially the same as conditioning on a significantly smaller number of previous time steps. In practice, the time-domain approach provides virtually an *exact solution* in the sense that the Bayesian approach utilizes the complete information inherited in the measurement. Therefore, the Bayesian time-domain approach provides more accurate statistical inference of the model parameters with the information in the data.
3. *Applicability/Robustness*
   The spectral density estimator is Wishart distributed regardless of the distribution of the original time-domain signal. Therefore, the expressions for the likelihood function in Chapter 3 are valid even for nonlinear systems. The only challenge is on the computation of the mean spectrum but this may be accommodated by simulations. However, the Bayesian spectral density approach is not applicable for nonstationary response measurements.

   In the Bayesian time-domain method, the joint PDF $p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_p}|\boldsymbol{\theta}, \mathcal{C})$ and the conditional PDFs $p(\mathbf{y}_n|\mathbf{y}_{n-N_p}, \mathbf{y}_{n-N_p+1}, \ldots, \mathbf{y}_{n-1}, \boldsymbol{\theta}, \mathcal{C})$ are known in closed-forms for the stationary or nonstationary responses of linear multi-degree-of-freedom systems. However, its extension for nonlinear systems is not a trivial task since their responses are in general non-Gaussian. Significant efforts have been made in solving the Fokker–Planck equation for the non-Gaussian response PDF of nonlinear systems [39, 74, 96, 158, 208, 211]. However, these methods compute the PDF at a particular time instant but not the joint PDF of the response at different time steps, required by the time-domain approach. Up to the author's knowledge, an efficient computation method of the joint PDF $p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_p}|\boldsymbol{\theta}, \mathcal{C})$ and the conditional PDFs $p(\mathbf{y}_n|\mathbf{y}_{n-N_p}, \mathbf{y}_{n-N_p+1}, \ldots, \mathbf{y}_{n-1}, \boldsymbol{\theta}, \mathcal{C})$ for general nonlinear systems is not available at the current stage of the literature and may not be even in the near future. Therefore, an extension of the Bayesian time-domain approach to nonlinear systems seems not achievable, at least in the foreseeable future.

4.9 Extended Readings

Chapters 3 and 4 presented two Bayesian methods to handle output-only measurements if the excitation can be modeled by stochastic process with a prescribed parametric model.

www.EngineeringBooksPdf.com

Other Bayesian methods were developed by the author and his co-workers in recent years for different applications. In this section, several selected methods are briefly introduced. Note that the symbols are modified here to be consistent with this book so some of them are slightly different from the original articles.

The Bayesian fast Fourier transform approach uses the statistical properties of discrete Fourier transforms, instead of the spectral density estimators, to construct the likelihood function and the updated PDF of the model parameters [292]. It does not rely on the approximation of the Wishart distributed spectrum. Expressions of the covariance matrix of the real and imaginary parts of the discrete Fourier transform were given. The only approximation was made on the independency of the discrete Fourier transforms at different frequencies. Therefore, the Bayesian fast Fourier transform approach is more accurate than the spectral density approach in the sense that one of the two approximations in the latter is released. However, since the fast Fourier transform approach considers the real and imaginary parts of the discrete Fourier transform, the corresponding covariance matrices are $2N_o \times 2N_o$, instead of $N_o \times N_o$ in the spectral density approach. Therefore, the spectral density approach is computationally more efficient than the fast Fourier transform approach.

Instead of output-only response measurements, measured input–output data was considered in Yuen and Katafygiotis [291]. Unlike other existing methods, it considers explicitly the measurement noise in both the input and output data. Let N denote the total number of observed time steps. Using the Bayes' theorem, the updated PDF of the parameters $\boldsymbol{\theta}$ given the measured input $\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_N$ of the excitation $\mathbf{F}$ and the measured response $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ is given by:

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N, \mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_N, \mathcal{C}) = \\ \kappa_0 p(\boldsymbol{\theta}|\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_N, \mathcal{C}) p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N|\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_N, \boldsymbol{\theta}, \mathcal{C}) \end{aligned} \tag{4.54}$$

where κ_0 is a normalizing constant. The probability distribution $p(\boldsymbol{\theta}|\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_N, \mathcal{C})$ represents the information from the measured input only. It can be approximated by the prior PDF $p(\boldsymbol{\theta}|\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_N, \mathcal{C}) \approx p(\boldsymbol{\theta}|\mathcal{C})$ since the measured input alone does not have much say on the model parameters (though it contains some information of the prediction-error variance, e.g., the rms of the measurement noise should be less than the rms of the measurement). The likelihood function $p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N|\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_N, \boldsymbol{\theta}, \mathcal{C})$ reflects the contribution of the measured data $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ and $\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_N$ in establishing the updated PDF of the model parameters. Since $p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N|\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_N, \boldsymbol{\theta}, \mathcal{C})$ is jointly Gaussian, direct calculation of this PDF encounters similar computational problems as in the exact formulation shown in Section 4.2. Therefore, a similar approximated likelihood expansion is introduced:

$$\begin{aligned} p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N|\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_N, \boldsymbol{\theta}, \mathcal{C}) \approx p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_p}|\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_N, \boldsymbol{\theta}, \mathcal{C}) \\ \times \prod_{n=N_p+1}^{N} p(\mathbf{y}_n|\mathbf{y}_{n-N_p}, \mathbf{y}_{n-N_p+1}, \ldots, \mathbf{y}_{n-1}, \mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_N, \boldsymbol{\theta}, \mathcal{C}) \end{aligned} \tag{4.55}$$

Again, the conditional PDFs are conditional on the last N_p time steps only. In order to obtain the reduced-order joint PDF and the conditional PDFs, the differential equation of the state space model was converted to a difference equation. By considering the input measurement noise,

www.EngineeringBooksPdf.com

the usually observed problem of underestimating the parametric uncertainty can be resolved. For more details, please refer to Yuen and Katafygiotis [291].

In Yuen *et al.* [287], the Bayesian unified approach was proposed to handle the more general case with incomplete input and incomplete output noisy measurements. This feature avoids the possible underestimation on the parametric uncertainty in practice. The Bayesian unified method opens a wide range of applications, including the special cases of ambient vibration surveys, and measured input–output noisy data. An application was presented for a building subjected to wind and ground excitation simultaneously. The unmeasured wind pressure was modeled as a stochastic process with uncertain parameters but the ground excitation was observed with measurement noise. The structural response was also observed at a limited number of DOFs only. The unified method was applied successfully for model updating and damage detection purpose.

In the literature and the previously mentioned identification methods, the input is either measured or modeled as a prescribed parametric stochastic model (even though the parameters may be unknown). This seems to be a necessary condition for model identification purpose. For example, consider a linear single-degree-of-freedom system. In the frequency domain, the response $\mathcal{X}$ is equal to the input $\mathcal{F}$, magnified by the transfer function of the oscillator $\mathcal{H}$:

$$\mathcal{X}(\omega) = \mathcal{H}(\omega)\mathcal{F}(\omega) \tag{4.56}$$

Therefore, if the time-frequency model of the input is completely unknown, the output measurement does not have any say on the transfer function of the oscillator since there exists a set of inputs to match with the measured outputs:

$$\mathcal{F}(\omega) = \mathcal{X}(\omega)/\mathcal{H}(\omega) \tag{4.57}$$

In other words, identification of the model parameters is impossible in this case since all model parameters (as long as the associated transfer function is non-zero) give the same likelihood (perfect match) to the measurement.

However, if two or more DOFs are measured in a multi-story building subjected to ground motion, the ratio between the response of different DOFs will be constrained by the prescribed class of structural models, such as shear building models. In Yuen and Katafygiotis [294], a frequency-domain method for unknown inputs was proposed and it does not assume any time-frequency model for the inputs. The method takes the advantage that when the number of the measured channels is larger than the number of independent external excitations, there are mathematical constraints among the responses of different degrees of freedom. For example, when the building is subjected to earthquake ground motion, the number of independent inputs is one. If the responses of two or more degrees of freedom are measured for this building, there is information to infer the structural properties even though no assumption is made on the time-frequency content of the excitation. Specifically, the data is partitioned into two parts:

$$\begin{aligned} \mathcal{D}^{\mathrm{A}} &= [\mathbf{y}_0^{\mathrm{A}}, \mathbf{y}_1^{\mathrm{A}}, \ldots, \mathbf{y}_{N-1}^{\mathrm{A}}] \\ \mathcal{D}^{\mathrm{B}} &= [\mathbf{y}_0^{\mathrm{B}}, \mathbf{y}_1^{\mathrm{B}}, \ldots, \mathbf{y}_{N-1}^{\mathrm{B}}] \end{aligned} \tag{4.58}$$

where $\mathbf{y}_n^{\mathrm{A}}$ and $\mathbf{y}_n^{\mathrm{B}}$, $n = 0, 1, \ldots, N-1$, are the measurements corresponding to the first N_F and last $N_o - N_F$ DOFs, respectively. The number N_F denotes the dimension of the input $\mathbf{F}$.

www.EngineeringBooksPdf.com

Then, the likelihood function can be expanded by the Bayes' theorem:

$$p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) = p(\mathcal{D}^{\mathrm{A}}, \mathcal{D}^{\mathrm{B}}|\boldsymbol{\theta}, \mathcal{C}) = p(\mathcal{D}^{\mathrm{B}}|\boldsymbol{\theta}, \mathcal{D}^{\mathrm{A}}, \mathcal{C})p(\mathcal{D}^{\mathrm{A}}|\boldsymbol{\theta}, \mathcal{C}) \tag{4.59}$$

It turns out that the likelihood function $p(\mathcal{D}^{\mathrm{A}}|\boldsymbol{\theta}, \mathcal{C})$ does not depend on $\boldsymbol{\theta}$ for this partitioning arrangement and the conditional likelihood function $p(\mathcal{D}^{\mathrm{B}}|\boldsymbol{\theta}, \mathcal{D}^{\mathrm{A}}, \mathcal{C})$ can be constructed in the frequency domain. For details, please refer to Yuen and Katafygiotis [294]. However, caution must be made on the 'ill-posedness' of the inverse problem since this relaxation of the assumption on the input stochastic model breaks the bonding between different frequencies and induces a higher degree of ill conditioning of the problem. Therefore, the Bayesian framework is important in order to indicate if it is globally identifiable to avoid misleading results.

If the mathematical model for the system of concern has too many uncertain parameters, the measurement will not provide sufficient mathematical constraints/equations to uniquely identify the uncertain parameters. However, experienced engineers can identify the critical substructures for monitoring. Then, a free body diagram can be drawn to focus on these critical substructures only. Note that the internal forces on the boundary of the substructures are unknown and difficult to measure, so they are treated as an *uncertain input to the substructure*. Furthermore, these internal forces share the dominant frequencies of the structure so they cannot be modeled arbitrarily as white noise or other prescribed colored noise. However, with the same idea as in Yuen and Katafygiotis [294], these interface forces can be treated as unknown inputs without assuming their time-frequency content [289]. This enables a large number of possible applications in structural health monitoring and also enhances the computational efficiency since one does not need to consider the whole system.

www.EngineeringBooksPdf.com

5

Model Updating Using Eigenvalue–Eigenvector Measurements

Keywords: Bayesian inference; damage detection; eigenvalue problem; iterative algorithm; modal data; mode shape expansion; model updating; structural health monitoring; substructure; system identification

In this chapter, a Bayesian methodology is presented for dynamical model updating using noisy incomplete modal data corresponding to eigenvalues (modal frequencies) and partial eigenvectors (mode shapes) of some of the modes of a dynamical system. The procedure can be used to find the most probable model within a specified class of dynamical models as well as the most probable values of the system modal frequencies and the full system mode shapes. The method does not require matching measured modes with corresponding modes from the dynamical model, which is in contrast to many existing methods. To find the most probable values of the dynamical model parameters and system modal parameters, the method uses an iterative scheme involving a series of coupled linear optimization problems. Furthermore, it does not require solving the eigenvalue problem of any dynamical model; instead, the eigen equations appear in the prior probability distribution to provide soft constraints. The method is computationally efficient and robust, judging from its successful application to noisy simulated data for a twelve-story building model and for a three-dimensional six-story braced-frame model. This latter example is also used to demonstrate an application to structural health monitoring.

5.1 Introduction

A generalized eigenvalue problem can be defined as follows:

$$\mathbf{A}\boldsymbol{\phi} = \lambda\mathbf{B}\boldsymbol{\phi} \tag{5.1}$$

Bayesian Methods for Structural Dynamics and Civil Engineering Ka-Veng Yuen
© 2010 John Wiley & Sons (Asia) Pte Ltd

www.EngineeringBooksPdf.com

where $\mathbf{A} \in \mathbb{R}^{N_d \times N_d}$ and $\mathbf{B} \in \mathbb{R}^{N_d \times N_d}$ are square matrices to govern the behavior of a system, and λ and $\boldsymbol{\phi}$ ($\neq \mathbf{0}$) are called the eigenvalue and eigenvector, respectively, of the generalized eigenvalue problem. In the case if $\mathbf{B}$ is invertible, the generalized eigenvalue problem deduces to a standard eigenvalue problem:

$$(\mathbf{B}^{-1}\mathbf{A})\boldsymbol{\phi} = \lambda\boldsymbol{\phi} \tag{5.2}$$

for the matrix $\mathbf{B}^{-1}\mathbf{A}$.

For structural/mechanical dynamics problems, the equation of motion of a linear system with N_d degrees of freedom (DOFs) is given by:

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{T}_0\mathbf{F}(t) \tag{5.3}$$

where $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$ are the mass, damping and stiffness matrix, respectively, and $\mathbf{F}$ is the external excitation with $\mathbf{T}_0$ being a force distributing matrix. If $\mathbf{A} = \mathbf{K}$ and $\mathbf{B} = \mathbf{M}$ are taken, then the generalized eigenvalue problem in Equation (5.1) will be associated to the solution for free vibration of the undamped system. The eigenvalues are the squared modal frequencies:

$$\lambda^{(m)} = \Omega^{(m)^2}, \; m = 1, 2, \ldots, N_d \tag{5.4}$$

and the eigenvectors, $\boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}^{(2)}, \ldots, \boldsymbol{\phi}^{(N_d)}$, are the mode shapes.

Most existing global structural health monitoring methods use dynamical model updating to determine local loss of stiffness by minimizing a measure of the difference between the modal frequencies and mode shapes measured in dynamic tests and those calculated from a finite-element model of the structure. The measured modal parameters are those estimated from dynamic test data using some modal identification procedure. A generic form of the goodness-of-fit function to be minimized is:

$$J_g(\boldsymbol{\theta}) = \sum_{m=1}^{N_m} w_m \left[\lambda^{(m)}(\boldsymbol{\theta}) - \hat{\lambda}^{(m)}\right]^2 + \sum_{m=1}^{N_m} w'_m \left|\left|\boldsymbol{\phi}^{(m)}(\boldsymbol{\theta}) - \hat{\boldsymbol{\phi}}^{(m)}\right|\right|^2 \tag{5.5}$$

where $\hat{\lambda}^{(m)}$ and $\hat{\boldsymbol{\phi}}^{(m)}$ are the measured eigenvalue and eigenvector of the mth mode, $\lambda^{(m)}(\boldsymbol{\theta})$ and $\boldsymbol{\phi}^{(m)}(\boldsymbol{\theta})$ are the eigenvalue and eigenvector, respectively, of the mth mode from the dynamical model with parameter vector $\boldsymbol{\theta}$ that determines the stiffness and mass matrix, and w_m and w'_m, $m = 1, 2, \ldots, N_m$, are chosen weightings that depend on the specific method. One major difficulty is that mode matching is required, i.e., it is necessary to determine which model mode matches which measured mode. If only measurements of partial mode shapes are available, this will not be a trivial task. Another major difficulty is that the N_m observed modes in dynamic tests might not necessarily be the N_m lowest-frequency modes in practice. In other words, some lower modes might not be detected. For example, some torsional modes are not excited. Furthermore, in the case where there is damage in the structure, the order of the modes might switch because the local loss of stiffness from damage may affect some modal frequencies more than others, making the mode matching even more challenging.

Recently, methods have been proposed for solving this model updating problem which avoid mode matching [18, 51, 52, 267]. This is accomplished by employing the concept of system mode shapes that are used to represent the actual mode shapes of the structural system at all degrees of freedom corresponding to those of the dynamical model, but they are distinct from the

www.EngineeringBooksPdf.com

mode shapes of the dynamical model, as will be seen more clearly later. Bayesian probabilistic methods are then used to update the dynamical model parameters and the system mode shapes based on the available modal data. Furthermore, Rayleigh quotient frequencies, which are based on the dynamical model and the system mode shapes, are employed instead of the modal frequencies of the dynamical model, so the eigenvalue problem needs never be solved.

First, let's introduce the key idea of the proposed approach with a simple ideal example. Consider the generalized eigen equation for the mth mode of the structure:

$$\mathbf{K}\boldsymbol{\phi}^{(m)} = \Omega^{(m)^2}\mathbf{M}\boldsymbol{\phi}^{(m)} \tag{5.6}$$

where $\Omega^{(m)}$ and $\boldsymbol{\phi}^{(m)}$ are the modal frequency and mode shape of the mth mode, respectively, the mass matrix $\mathbf{M} \in \mathbb{R}^{N_d \times N_d}$ is assumed known and the stiffness matrix $\mathbf{K} = \mathbf{K}(\boldsymbol{\theta}) \in \mathbb{R}^{N_d \times N_d}$ is parameterized by the uncertain parameters in $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_{N_\theta}]^T \in \mathbb{R}^{N_\theta}$:

$$\mathbf{K} = \mathbf{K}_0 + \sum_{l=1}^{N_\theta} \theta_l \mathbf{K}_l \tag{5.7}$$

where the subsystem stiffness matrices $\mathbf{K}_l$, $l = 0, 1, \ldots, N_\theta$, are prescribed up to a scaling factor. Then, by plugging Equation (5.7) into Equation (5.6), the eigenvalue problem becomes:

$$\left[\mathbf{K}_1\boldsymbol{\phi}^{(m)}, \mathbf{K}_2\boldsymbol{\phi}^{(m)}, \ldots, \mathbf{K}_{N_\theta}\boldsymbol{\phi}^{(m)}\right]\boldsymbol{\theta} = \left(\Omega^{(m)^2}\mathbf{M} - \mathbf{K}_0\right)\boldsymbol{\phi}^{(m)} \tag{5.8}$$

Therefore, if $N_\theta = N_d$ and the set $\{\mathbf{K}_1\boldsymbol{\phi}^{(m)}, \mathbf{K}_2\boldsymbol{\phi}^{(m)}, \ldots, \mathbf{K}_{N_\theta}\boldsymbol{\phi}^{(m)}\}$ is linearly independent, the stiffness parameter vector $\boldsymbol{\theta}$ is readily obtained:

$$\boldsymbol{\theta} = \left[\mathbf{K}_1\boldsymbol{\phi}^{(m)}, \mathbf{K}_2\boldsymbol{\phi}^{(m)}, \ldots, \mathbf{K}_{N_\theta}\boldsymbol{\phi}^{(m)}\right]^{-1}\left(\Omega^{(m)^2}\mathbf{M} - \mathbf{K}_0\right)\boldsymbol{\phi}^{(m)} \tag{5.9}$$

Otherwise, the problem is unidentifiable and multiple solutions exist. In this case, the *smallest least-squares solution* may be obtained by using the *Penrose generalized inverse*:

$$\boldsymbol{\theta}^{SLS} = \left[\mathbf{K}_1\boldsymbol{\phi}^{(m)}, \mathbf{K}_2\boldsymbol{\phi}^{(m)}, \ldots, \mathbf{K}_{N_\theta}\boldsymbol{\phi}^{(m)}\right]^{\dagger}\left(\Omega^{(m)^2}\mathbf{M} - \mathbf{K}_0\right)\boldsymbol{\phi}^{(m)} \tag{5.10}$$

where † denotes the Penrose generalized inverse of a matrix [189]. In this case, measurement of more modes or simplification of the parameterization will be necessary for a unique solution.

The nice feature of this idea is that the order of the measured mode m is not necessarily known since there is no matching between the measured and model modes in contrast to Equation (5.5). This is important for practical situations where it is difficult to judge whether the N_m measured modes are the lowest N_m modes. This is because some of the modes may not be excited in the modal testing and they are missing in modal identification. However, this formulation requires complete mode shape measurements, which are not available in practice. Furthermore, the measurements are contaminated by measurement noise but this formulation does not have an explicit treatment on it.

In this chapter, a Bayesian model updating method using incomplete modal data is presented with applications to structural health monitoring. As reported in the literature [18, 51, 52, 267], the realistic assumption is made that only the modal frequencies and partial mode shapes of some modes are measured; system mode shapes are also introduced, which avoid mode matching between the measured modes and those of the dynamical model. The novel feature

www.EngineeringBooksPdf.com

in this work is that system frequencies are also introduced as parameters to be identified in order to represent actual modal frequencies of the dynamical system (assuming that the dynamical behavior is well approximated by linear dynamics; otherwise they should be interpreted in the equivalent linear sense). The eigen equations of the dynamical model are used only in the prior probability distribution to provide soft constraints. Furthermore, to calculate the most probable values of the model parameters based on the modal data, an efficient iterative procedure is used that involves a series of coupled linear optimization problems, rather than directly solving the challenging nonlinear optimization problem by some general algorithm that may give convergence difficulties in the high-dimensional parameter space.

In the next section, the proposed updating approach is presented which provides estimates of the system modal frequencies and system mode shapes, as well as estimates of the stiffness model parameters, based on incomplete modal data. Examples with a twelve-story building and a three-dimensional braced frame will be used to demonstrate the method with applications to structural health monitoring.

5.2 Formulation

A class of dynamical models $\mathcal{C}$ is considered with N_d DOFs that has a known mass matrix $\mathbf{M} \in \mathbb{R}^{N_d \times N_d}$ (which is assumed to be established with sufficient accuracy from the engineering drawings of the structure) and the stiffness matrix $\mathbf{K}(\boldsymbol{\theta}) \in \mathbb{R}^{N_d \times N_d}$ is parameterized by $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_{N_\theta}]^T \in \mathbb{R}^{N_\theta}$ as follows:

$$\mathbf{K}(\boldsymbol{\theta}) = \mathbf{K}_0 + \sum_{l=1}^{N_\theta} \theta_l \mathbf{K}_l \tag{5.11}$$

where the subsystem stiffness matrices $\mathbf{K}_l$, $l = 0, 1, \ldots, N_\theta$, are specified, e.g., by a finite-element model of the structure. The scaling parameters in $\boldsymbol{\theta}$ allow the nominal model matrix given by $\boldsymbol{\theta} = \boldsymbol{\theta}^\eta$ in Equation (5.11) to be updated based on dynamic test data from the system.

Assume that $N_m (\leq N_d)$ modes of the system are measured (not necessarily the first N_m lowest frequency modes), which have eigenvalues $\lambda^{(m)}, m = 1, 2, \ldots, N_m$, and real eigenvector components $\boldsymbol{\phi}^{(m)} \in \mathbb{R}^{N_d}$, $m = 1, 2, \ldots, N_m$. It is not assumed that these modal parameters satisfy exactly the eigenvalue problem of any given dynamical model $(\mathbf{M}, \mathbf{K}(\boldsymbol{\theta}))$ because there are always modeling approximations and errors. The quantities $\lambda^{(m)}$ and $\boldsymbol{\phi}^{(m)}$ are referred to as the system eigenvalue and eigenvector, respectively of the mth mode to distinguish them from the corresponding modal parameters given by any dynamical model specified by $\boldsymbol{\theta}$. Given a parameter vector $\boldsymbol{\theta}$, the model can be defined in $\mathcal{C}$. The prior probability density function (PDF) for $\boldsymbol{\lambda} = \left[\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(N_m)}\right]^T$ and $\boldsymbol{\phi} = \left[\boldsymbol{\phi}^{(1)\,T}, \boldsymbol{\phi}^{(2)\,T}, \ldots, \boldsymbol{\phi}^{(N_m)\,T}\right]^T$ is chosen as:

$$p(\boldsymbol{\lambda}, \boldsymbol{\phi} | \boldsymbol{\theta}, \mathcal{C}) = \kappa_0 \exp\left[-\frac{1}{2} J_g(\boldsymbol{\lambda}, \boldsymbol{\phi}; \boldsymbol{\theta})\right] \tag{5.12}$$

where κ_0 is a normalizing constant and the goodness-of-fit function is given by:

$$J_g(\boldsymbol{\lambda}, \boldsymbol{\phi}; \boldsymbol{\theta}) = \begin{bmatrix} (\mathbf{K}(\boldsymbol{\theta}) - \lambda^{(1)}\mathbf{M})\boldsymbol{\phi}^{(1)} \\ (\mathbf{K}(\boldsymbol{\theta}) - \lambda^{(2)}\mathbf{M})\boldsymbol{\phi}^{(2)} \\ \vdots \\ (\mathbf{K}(\boldsymbol{\theta}) - \lambda^{(N_m)}\mathbf{M})\boldsymbol{\phi}^{(N_m)} \end{bmatrix}^T \boldsymbol{\Sigma}_{eq}^{-1} \begin{bmatrix} (\mathbf{K}(\boldsymbol{\theta}) - \lambda^{(1)}\mathbf{M})\boldsymbol{\phi}^{(1)} \\ (\mathbf{K}(\boldsymbol{\theta}) - \lambda^{(2)}\mathbf{M})\boldsymbol{\phi}^{(2)} \\ \vdots \\ (\mathbf{K}(\boldsymbol{\theta}) - \lambda^{(N_m)}\mathbf{M})\boldsymbol{\phi}^{(N_m)} \end{bmatrix} \tag{5.13}$$

www.EngineeringBooksPdf.com

This prior PDF is based on choosing a Gaussian PDF as a probability model for the eigen equation errors, where the prior covariance matrix $\mathbf{\Sigma}_{eq}$ controls the size of these equation errors. The uncertainty in the equation errors for each mode are modeled as independent and identically distributed, so:

$$\mathbf{\Sigma}_{eq} = \sigma_{eq}^2 \mathbf{I}_{N_d N_m} \tag{5.14}$$

where $\mathbf{I}_{N_d N_m}$ denotes the $N_d N_m \times N_d N_m$ identity matrix and σ_{eq}^2 is a prescribed equation-error variance. The usage of this variance parameter allows for explicit treatment of modeling error as the parametric models for the stiffness matrix, and hence the eigen equation, is never exact in practice. If this error level can be estimated, the mathematical constraint given by the eigen equation will become a *soft constraint* instead of a *rigid constraint*. In other words, errors of the eigen equation in the level corresponding to σ_{eq} is allowed. In the numerical examples presented later, the value of σ_{eq}^2 is chosen to be very small so that the eigen equations are nearly satisfied. This means that the system modal frequencies and mode shapes will correspond closely to modal parameters of the identified dynamical model. For modal data from a real structure, this would be a reasonable strategy to start with. If the measured modal parameters did not agree well with those corresponding to the identified (most probable) dynamical model, implying considerable modeling errors, then σ_{eq}^2 could be increased. This procedure allows explicit control of the inherent trade-off between how well the measured modal parameters are matched and how well the eigen equations of the identified dynamical model are satisfied. This additional modeling flexibility is an appealing feature of this method.

The prior PDF $p(\boldsymbol{\lambda}, \boldsymbol{\phi}|\boldsymbol{\theta}, \mathcal{C})$ implies that, given a class of dynamical models and before using the dynamic test data, the most probable values of $\boldsymbol{\lambda}$ and $\boldsymbol{\phi}$ are those that minimize the Euclidean norm (2-norm) of the error in the eigen equation for the dynamical model. This implies that the prior most probable values of $\boldsymbol{\lambda}$ and $\boldsymbol{\phi}$ are the squared modal frequencies and mode shapes of a dynamical model, but these values are never explicitly required. This prior PDF will have multiple peaks because there is no implied ordering of the modes here.

The prior PDF for all the unknown parameters is given by:

$$p(\boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\theta}|\mathcal{C}) = p(\boldsymbol{\lambda}, \boldsymbol{\phi}|\boldsymbol{\theta}, \mathcal{C}) p(\boldsymbol{\theta}|\mathcal{C}) \tag{5.15}$$

where the prior PDF $p(\boldsymbol{\theta}|\mathcal{C})$ can be taken as a Gaussian distribution with mean $\boldsymbol{\theta}^{\eta}$ representing the nominal values of the model parameters and with covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\theta}}$. In the examples later, the prior covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\theta}}$ is taken to be diagonal with large variances, giving virtually a non-informative prior.

To construct the likelihood function, the measurement error $\boldsymbol{\epsilon}$ is introduced:

$$\begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\psi}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda} \\ \mathbf{L}_o \boldsymbol{\phi} \end{bmatrix} + \boldsymbol{\epsilon} \tag{5.16}$$

and a Gaussian probability model is chosen for $\boldsymbol{\epsilon} \in \mathbb{R}^{N_m(N_o+1)}$ with zero mean and covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\epsilon}}$, which can be obtained by Bayesian modal identification methods [136, 285, 290, 291], $\hat{\boldsymbol{\psi}} = \left[\hat{\boldsymbol{\psi}}^{(1)\,T}, \hat{\boldsymbol{\psi}}^{(2)\,T}, \ldots, \hat{\boldsymbol{\psi}}^{(N_m)\,T}\right]^T$ and $\hat{\boldsymbol{\lambda}} = [\hat{\lambda}^{(1)}, \hat{\lambda}^{(2)}, \ldots, \hat{\lambda}^{(N_m)}]^T$, where $\hat{\boldsymbol{\psi}}^{(m)} \in \mathbb{R}^{N_o}$ gives the observed components of the system eigenvector of the mth mode and $\hat{\lambda}^{(m)}$ gives the corresponding observed system eigenvalue from dynamic test data. Finally,

www.EngineeringBooksPdf.com

$\mathbf{L}_o$ is an $N_m N_o \times N_m N_d$ observation matrix of '1s' or '0s' that picks the components of $\boldsymbol{\phi}$ corresponding to the N_o measured DOFs. The likelihood function is therefore:

$$p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} | \boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathcal{C}\right) = p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} | \boldsymbol{\lambda}, \boldsymbol{\phi}\right) = \mathcal{G}\left(\left[\hat{\boldsymbol{\lambda}}^T, \hat{\boldsymbol{\psi}}^T\right]^T; \left[\boldsymbol{\lambda}^T, (\boldsymbol{\Gamma}\boldsymbol{\phi})^T\right]^T, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\right) \tag{5.17}$$

that is, a Gaussian distribution with mean $\left[\boldsymbol{\lambda}^T, (\mathbf{L}_o\boldsymbol{\phi})^T\right]^T$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$.

The posterior PDF for the unknown parameters is given by the Bayes' theorem:

$$\begin{aligned} p\left(\boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\theta} | \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \mathcal{C}\right) &= \kappa_1 p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} | \boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathcal{C}\right) p(\boldsymbol{\lambda}, \boldsymbol{\phi} | \boldsymbol{\theta}, \mathcal{C}) p(\boldsymbol{\theta} | \mathcal{C}) \\ &= \kappa_1 p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} | \boldsymbol{\lambda}, \boldsymbol{\phi}\right) p(\boldsymbol{\lambda}, \boldsymbol{\phi} | \boldsymbol{\theta}, \mathcal{C}) p(\boldsymbol{\theta} | \mathcal{C}) \end{aligned} \tag{5.18}$$

The most probable values of the unknown parameters can be found by maximizing this PDF. To proceed, the objective function is defined as [286]:

$$\begin{aligned} J(\boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\theta}) = &\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^\eta)^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}^\eta) + \frac{1}{2\sigma_{eq}^2} \sum_{m=1}^{N_m} \left|\left| \left(\mathbf{K}(\boldsymbol{\theta}) - \lambda^{(m)} \mathbf{M}\right) \boldsymbol{\phi}^{(m)} \right|\right|^2 \\ &+ \frac{1}{2} \begin{bmatrix} \hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda} \\ \hat{\boldsymbol{\psi}} - \mathbf{L}_o \boldsymbol{\phi} \end{bmatrix}^T \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1} \begin{bmatrix} \hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda} \\ \hat{\boldsymbol{\psi}} - \mathbf{L}_o \boldsymbol{\phi} \end{bmatrix} \end{aligned} \tag{5.19}$$

which is the negative logarithm of the posterior PDF without including the constant that does not depend on the uncertain parameters. Here, $||.||$ denotes the Euclidean norm. Then, the function $J(\boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\theta})$ is minimized instead of maximizing the posterior PDF. The objective function J is not a quadratic function for the uncertain parameters. However, this function is quadratic for any of the uncertain parameter vectors of $\boldsymbol{\lambda}$, $\boldsymbol{\phi}$ or $\boldsymbol{\theta}$ if the other two are fixed. Therefore, the original nonlinear optimization problems can be done iteratively through a sequence of linear optimization problems, as shown in the next section.

5.3 Linear Optimization Problems

The mode shapes are usually measured with incomplete components, i.e., with missing DOFs but the modal frequencies are measured with relatively high accuracy. Therefore, the sequence of optimization starts from computing the missing components of the mode shapes. First set the updated model parameters at their nominal values:

$$\boldsymbol{\theta}^\star = \boldsymbol{\theta}^\eta \tag{5.20}$$

and the eigenvalues at their measured values:

$$\boldsymbol{\lambda}^\star = \hat{\boldsymbol{\lambda}} \tag{5.21}$$

www.EngineeringBooksPdf.com

Then, perform a sequence of iterations comprised of the following linear optimization problems:

$$\begin{aligned}\boldsymbol{\phi}^\star &= \arg\min_{\boldsymbol{\phi}} J(\boldsymbol{\lambda}^\star, \boldsymbol{\phi}, \boldsymbol{\theta}^\star)\\ \boldsymbol{\lambda}^\star &= \arg\min_{\boldsymbol{\lambda}} J(\boldsymbol{\lambda}, \boldsymbol{\phi}^\star, \boldsymbol{\theta}^\star)\\ \boldsymbol{\theta}^\star &= \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\lambda}^\star, \boldsymbol{\phi}^\star, \boldsymbol{\theta})\end{aligned} \tag{5.22}$$

until the prescribed convergence criterion is satisfied. Each of these three optimization problems is explained in more detail in the next three subsections.

5.3.1 Optimization for Mode Shapes

By minimizing the objective function $J(\boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\theta})$ in Equation (5.19) with respect to $\boldsymbol{\phi}$, the optimal vector $\boldsymbol{\phi}^\star$ can be obtained:

$$\boldsymbol{\phi}^\star = [\sigma_{eq}^{-2}\mathbf{G}_\phi + \mathbf{L}_o^T(\boldsymbol{\Sigma}_\epsilon^{-1})_{22}\mathbf{L}_o]^{-1}\mathbf{L}_o^T[(\boldsymbol{\Sigma}_\epsilon^{-1})_{21}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^\star) + (\boldsymbol{\Sigma}_\epsilon^{-1})_{22}\hat{\boldsymbol{\psi}}] \tag{5.23}$$

where $(\boldsymbol{\Sigma}_\epsilon^{-1})_{21}$ and $(\boldsymbol{\Sigma}_\epsilon^{-1})_{22}$ are the $N_m N_o \times N_m$ left bottom and $N_m N_o \times N_m N_o$ right bottom sub-matrices of $\boldsymbol{\Sigma}_\epsilon^{-1}$, and the symmetric matrix $\mathbf{G}_\phi$ is given by:

$$\mathbf{G}_\phi = \begin{bmatrix} (\lambda^{(1)\star}\mathbf{M} - \mathbf{K}^\star)^2 & & & \mathbf{0} \\ & (\lambda^{(2)\star}\mathbf{M} - \mathbf{K}^\star)^2 & & \\ & & \ddots & \\ \mathbf{0} & & & (\lambda^{(N_m)\star}\mathbf{M} - \mathbf{K}^\star)^2 \end{bmatrix}_{N_d N_m \times N_d N_m} \tag{5.24}$$

where the updated stiffness matrix $\mathbf{K}^\star = \mathbf{K}(\boldsymbol{\theta}^\star)$.

5.3.2 Optimization for Modal Frequencies

By minimizing the objective function $J(\boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\theta})$ in Equation (5.19) with respect to $\boldsymbol{\lambda}$, the updated parameter vector $\boldsymbol{\lambda}^\star$ is given by:

$$\boldsymbol{\lambda}^\star = [\mathbf{G}_\lambda + (\boldsymbol{\Sigma}_\epsilon^{-1})_{11}]^{-1}\left(\sigma_{eq}^{-2}\begin{bmatrix} \boldsymbol{\phi}^{(1)\star T}\mathbf{M}\mathbf{K}^\star\boldsymbol{\phi}^{(1)\star} \\ \boldsymbol{\phi}^{(2)\star T}\mathbf{M}\mathbf{K}^\star\boldsymbol{\phi}^{(2)\star} \\ \vdots \\ \boldsymbol{\phi}^{(N_m)\star\, T}\mathbf{M}\mathbf{K}^\star\boldsymbol{\phi}^{(N_m)\star} \end{bmatrix} + (\boldsymbol{\Sigma}_\epsilon^{-1})_{11}\hat{\boldsymbol{\lambda}} + (\boldsymbol{\Sigma}_\epsilon^{-1})_{12}(\hat{\boldsymbol{\psi}} - \mathbf{L}_o\boldsymbol{\phi}^\star)\right) \tag{5.25}$$

www.EngineeringBooksPdf.com

where $(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1})_{11}$ and $(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1})_{12}$ are the $N_m \times N_m$ left top and $N_m \times N_m N_o$ right top sub-matrices of $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1}$, and the matrix $\mathbf{G}_\lambda$ is given by:

$$\mathbf{G}_\lambda = \sigma_{eq}^{-2} \begin{bmatrix} \boldsymbol{\phi}^{(1)\star T}\mathbf{M}^2\boldsymbol{\phi}^{(1)\star} & & & \mathbf{0} \\ & \boldsymbol{\phi}^{(2)\star T}\mathbf{M}^2\boldsymbol{\phi}^{(2)\star} & & \\ & & \ddots & \\ \mathbf{0} & & & \boldsymbol{\phi}^{(N_m)\star\, T}\mathbf{M}^2\boldsymbol{\phi}^{(N_m)\star} \end{bmatrix}_{N_m \times N_m} \tag{5.26}$$

5.3.3 *Optimization for Model Parameters*

By minimizing Equation (5.19) with respect to $\boldsymbol{\theta}$, the updated model parameter vector $\boldsymbol{\theta}^\star$ is given by:

$$\boldsymbol{\theta}^\star = (\sigma_{eq}^{-2}\mathbf{G}_\theta^T\mathbf{G}_\theta + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1})^{-1}(\sigma_{eq}^{-2}\mathbf{G}_\theta^T\mathbf{b} + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\theta}^\eta) \tag{5.27}$$

where the matrix $\mathbf{G}_\theta$ is given by:

$$\mathbf{G}_\theta = \begin{bmatrix} \mathbf{K}_1\boldsymbol{\phi}^{(1)\star} & \mathbf{K}_2\boldsymbol{\phi}^{(1)\star} & \cdots & \mathbf{K}_{N_{\boldsymbol{\theta}}}\boldsymbol{\phi}^{(1)\star} \\ \mathbf{K}_1\boldsymbol{\phi}^{(2)\star} & \mathbf{K}_2\boldsymbol{\phi}^{(2)\star} & \cdots & \mathbf{K}_{N_{\boldsymbol{\theta}}}\boldsymbol{\phi}^{(2)\star} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_1\boldsymbol{\phi}^{(N_m)\star} & \mathbf{K}_2\boldsymbol{\phi}^{(N_m)\star} & \cdots & \mathbf{K}_{N_{\boldsymbol{\theta}}}\boldsymbol{\phi}^{(N_m)\star} \end{bmatrix}_{N_d N_m \times N_{\boldsymbol{\theta}}} \tag{5.28}$$

and the vector $\mathbf{b} \in \mathbb{R}^{N_d N_m}$ is given by:

$$\mathbf{b} = \begin{bmatrix} (\lambda^{(1)\star}\mathbf{M} - \mathbf{K}_0)\boldsymbol{\phi}^{(1)\star} \\ (\lambda^{(2)\star}\mathbf{M} - \mathbf{K}_0)\boldsymbol{\phi}^{(2)\star} \\ \vdots \\ (\lambda^{(N_m)\star}\mathbf{M} - \mathbf{K}_0)\boldsymbol{\phi}^{(N_m)\star} \end{bmatrix}_{N_d N_m \times 1} \tag{5.29}$$

5.4 Iterative Algorithm

The three optimization problems are coupled. In other words, the expressions of the optimal vectors depend on the optimal values of other parameters. In order to search for the overall optimal parameters efficiently, the proposed methodology updates the system modal frequencies, system mode shapes and stiffness scaling parameters in an iterative manner by using successively the optimization results of Section 5.3. The iterative procedure consists of the following steps:

1. Take the initial values of the model parameters as the nominal values, $\boldsymbol{\theta}^\star = \boldsymbol{\theta}^\eta$, and the eigenvalues as the measured values, $\boldsymbol{\lambda}^\star = \hat{\boldsymbol{\lambda}}$. Then, $\mathbf{K}^\star = \mathbf{K}(\boldsymbol{\theta}^\star)$.

www.EngineeringBooksPdf.com

2. Update the estimates of the system eigenvectors $\boldsymbol{\phi}^{(m)\star}$, $m = 1, 2, \ldots, N_m$, using Equation (5.23).
3. Update the estimates of the system eigenvalues (squared modal frequencies) $\lambda^{(m)\star}$, $m = 1, 2, \ldots, N_m$, using Equation (5.25).
4. Update the estimates of the model parameters $\boldsymbol{\theta}^\star$ by using Equation (5.27).
5. Iterate the previous Steps 2, 3 and 4 until the model parameters in $\boldsymbol{\theta}^\star$ satisfy some convergence criterion, thereby giving the most probable values of the model parameters based on the modal data.

5.5 Uncertainty Estimation

The posterior PDF in Equation (5.18) can be well approximated by a Gaussian distribution centered at the optimal (most probable) parameters $(\boldsymbol{\lambda}^\star, \boldsymbol{\phi}^\star, \boldsymbol{\theta}^\star)$ and with covariance matrix $\boldsymbol{\Gamma}(\boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\theta})$ equal to the inverse of the Hessian of the objective function $J(\boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\theta}) = -\ln p(\boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\theta}|\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \mathcal{C})$ calculated at the optimal parameters [19]. This covariance matrix is given by:

$$\boldsymbol{\Gamma}(\boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\theta}) = \begin{bmatrix} \sigma_{eq}^{-2}\mathbf{G}_\lambda + (\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1})_{11} & -2\sigma_{eq}^{-2}\mathbf{L}_1 + (\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1})_{12}\mathbf{L}_o & -\sigma_{eq}^{-2}\mathbf{L}_2 \\ & \sigma_{eq}^{-2}\mathbf{G}_\phi + \mathbf{L}_o^T(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1})_{22}\mathbf{L}_o & \sigma_{eq}^{-2}\mathbf{L}_3 \\ \text{sym} & & \sigma_{eq}^{-2}\mathbf{G}_\theta^T\mathbf{G}_\theta + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \end{bmatrix}^{-1} \tag{5.30}$$

where the matrix $\mathbf{L}_1$ is given by:

$$\mathbf{L}_1 = \begin{bmatrix} \boldsymbol{\phi}^{(1)\star\,T}\mathbf{M}(\mathbf{K}^\star - \lambda^{(1)\star}\mathbf{M}) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{\phi}^{(N_m)\star\,T}\mathbf{M}(\mathbf{K}^\star - \lambda^{(N_m)\star}\mathbf{M}) \end{bmatrix}_{N_m \times N_d N_m} \tag{5.31}$$

The lth column of $\mathbf{L}_2$ is given by:

$$(\mathbf{L}_2)_{l^{th}col} = \begin{bmatrix} \boldsymbol{\phi}^{(1)\star T}\mathbf{M}\mathbf{K}_l\boldsymbol{\phi}^{(1)\star} \\ \boldsymbol{\phi}^{(2)\star T}\mathbf{M}\mathbf{K}_l\boldsymbol{\phi}^{(2)\star} \\ \vdots \\ \boldsymbol{\phi}^{(N_m)\star T}\mathbf{M}\mathbf{K}_l\boldsymbol{\phi}^{(N_m)\star} \end{bmatrix}_{N_m \times 1} \tag{5.32}$$

The lth column of $\mathbf{L}_3$ is given by:

$$(\mathbf{L}_3)_{l^{th}col} = \begin{bmatrix} [(\mathbf{K}^\star - \lambda^{(1)\star}\mathbf{M})\mathbf{K}_l + \mathbf{K}_l(\mathbf{K}^\star - \lambda^{(1)\star}\mathbf{M})]\boldsymbol{\phi}^{(1)\star} \\ [(\mathbf{K}^\star - \lambda^{(2)\star}\mathbf{M})\mathbf{K}_l + \mathbf{K}_l(\mathbf{K}^\star - \lambda^{(2)\star}\mathbf{M})]\boldsymbol{\phi}^{(2)\star} \\ \vdots \\ [(\mathbf{K}^\star - \lambda^{(N_m)\star}\mathbf{M})\mathbf{K}_l + \mathbf{K}_l(\mathbf{K}^\star - \lambda^{(N_m)\star}\mathbf{M})]\boldsymbol{\phi}^{(N_m)\star} \end{bmatrix}_{N_d N_m \times 1} \tag{5.33}$$

www.EngineeringBooksPdf.com

Summarizing, by letting $\boldsymbol{\vartheta} = [\boldsymbol{\lambda}^T, \boldsymbol{\phi}^T, \boldsymbol{\theta}^T]^T \in \mathbb{R}^{N_m(N_d+1)+N_\theta}$ denote all the uncertain model parameters, the posterior PDF $p(\boldsymbol{\vartheta}|\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}})$ can be approximated by $\mathcal{G}(\boldsymbol{\vartheta}; \boldsymbol{\vartheta}^\star, \boldsymbol{\Gamma})$ where the mean $\boldsymbol{\vartheta}^\star$ consists of the most probable values of $\boldsymbol{\vartheta}$ and the covariance matrix $\boldsymbol{\Gamma}$ of dimension $N_m(N_d + 1) + N_\theta$ is given above.

5.6 Applications to Structural Health Monitoring

5.6.1 Twelve-story Shear Building

In this simulated-data example, a twelve-story shear building is considered. It is assumed that this building has uniformly distributed floor mass and uniform stiffness across the height. The mass per floor is taken to be 100 metric tons, while the interstory stiffness is chosen to be $\tilde{K} = 202.767$ MN/m so that the first five modal frequencies are 0.900, 2.686, 4.429, 6.103 and 7.680 Hz. The covariance matrix $\boldsymbol{\Sigma}_\epsilon$ is diagonal with the variances corresponding to a 1.0% coefficient of variation of the measurement error of the squared modal frequencies and mode shapes for all modes, a reasonable value based on typical modal test results. For the simulated modal data, a sample of zero-mean Gaussian noise with covariance matrix $\boldsymbol{\Sigma}_\epsilon$ was added to the exact modal frequencies and mode shapes.

The nominal value of each story stiffness parameter is selected from a uniform distribution over $2\tilde{K}$ to $3\tilde{K}$, where $\tilde{K}$ is the actual interstory stiffness, so the nominal values $\boldsymbol{\theta}^\eta$ are significantly overestimated and the variation between the different interstory stiffness values is very substantial. Recall that the nominal stiffness values are taken as the initial values in the optimization to find the most probable values based on the modal data and it affects the prior distribution of the stiffness parameters. The subsystem stiffness matrices are given by $\mathbf{K}_0 = \mathbf{0}_{12\times 12}$ since the problem is linear:

$$\mathbf{K}_1 = \begin{bmatrix} 1 & \mathbf{0}_{1\times 11} \\ \mathbf{0}_{11\times 1} & \mathbf{0}_{11\times 11} \end{bmatrix} \text{MN/m} \tag{5.34}$$

for the first story, and

$$\mathbf{K}_l = \begin{bmatrix} & \mathbf{0}_{(l-2)\times 12} & & \\ \mathbf{0}_{1\times(l-2)} & 1 & -1 & \mathbf{0}_{1\times(12-l)} \\ \mathbf{0}_{1\times(l-2)} & -1 & 1 & \mathbf{0}_{1\times(12-l)} \\ & \mathbf{0}_{(12-l)\times 12} & & \end{bmatrix} \text{MN/m} \tag{5.35}$$

for other stories, i.e., $l = 2, 3, \ldots, 12$.

Consider first the case where complete measurements (all 12 DOFs) are utilized. Table 5.1 shows the initial (nominal) values and the final identified (most probable) values of the stiffness parameters by applying the proposed updating method using different numbers of measured modes; recall that the actual value is 202.767 MN/m for all these parameters. Even for such a high level of modeling error in the nominal model (100–200%), the proposed methodology successfully corrects the error and the identified stiffness parameters are distributed around the actual value. By using Equation (5.30), the coefficients of variation (COV) can be estimated and they are shown in parentheses. It is not surprising that the uncertainty reduces when the number of measured modes increases.

www.EngineeringBooksPdf.com

Table 5.1 Identification results for the most probable stiffness parameters using different number of measured modes (with coefficients of variation in parentheses)

Parameter	Initial values	2 modes	3 modes	4 modes	5 modes
θ_1	412.2	198.9 (0.018)	200.7 (0.011)	204.9 (0.009)	204.4 (0.007)
θ_2	499.2	204.2 (0.027)	204.2 (0.019)	205.3 (0.017)	203.6 (0.016)
θ_3	561.8	191.6 (0.029)	193.0 (0.027)	196.7 (0.014)	199.4 (0.008)
θ_4	460.0	202.9 (0.045)	207.7 (0.026)	202.0 (0.010)	203.9 (0.008)
θ_5	523.6	184.2 (0.056)	204.6 (0.018)	203.1 (0.012)	204.2 (0.010)
θ_6	595.3	207.8 (0.057)	202.0 (0.016)	201.0 (0.014)	202.3 (0.008)
θ_7	488.7	199.8 (0.039)	202.3 (0.017)	204.1 (0.010)	202.6 (0.008)
θ_8	476.2	204.0 (0.030)	205.6 (0.023)	205.6 (0.011)	204.3 (0.008)
θ_9	435.6	209.1 (0.028)	204.3 (0.021)	199.6 (0.011)	201.4 (0.007)
θ_{10}	467.3	198.9 (0.025)	209.0 (0.017)	205.6 (0.011)	205.2 (0.008)
θ_{11}	464.4	201.9 (0.029)	203.5 (0.015)	201.1 (0.009)	201.4 (0.006)
θ_{12}	464.1	193.2 (0.035)	201.1 (0.015)	203.7 (0.008)	202.6 (0.006)

Consider next the incomplete mode shape measurements where only six sensors on the first, fourth, fifth, seventh, tenth and top floors are available. The results presented in Table 5.2 are based on five measured modes and show the initial values, final most probable values, standard deviations and COVs of the stiffness parameters, which are comparable to the COV of the modal data. Figure 5.1 shows the iterative history for the most probable values of the stiffness parameters, with convergence occurring in about 120 iterations. Again, the same set of nominal stiffness values is used so the nominal model overestimated the interstory stiffnesses by 100 to 200%. The parameters converge very quickly even for such an unsatisfactory set of initial values. The CPU time for 200 iterations is about 0.8 s with a conventional dual CPU 3.0 GHz personal computer under the MATLAB® environment [171]. Figure 5.2 shows the comparison between the identified system mode shapes (solid lines) and the actual mode shapes (dashed lines) for the first five modes but the two sets of curves are on top of each other. Of course, it is no wonder that the mode shape components of the observed degrees of freedom are estimated better than the unobserved ones.

Table 5.2 Identification results for the most probable stiffness parameters with measurements from five sensors and five modes

Parameter	Initial values	Identified $\theta^\star$	S.D. σ_θ	COV
θ_1	412.2	199.9	2.01	0.010
θ_2	499.2	196.5	3.90	0.019
θ_3	561.8	206.0	3.28	0.016
θ_4	460.0	204.2	2.06	0.010
θ_5	523.6	204.3	2.35	0.012
θ_6	595.3	205.9	3.01	0.015
θ_7	488.7	202.1	2.22	0.011
θ_8	476.2	205.9	2.28	0.011
θ_9	435.6	198.2	2.87	0.014
θ_{10}	467.3	206.3	2.95	0.015
θ_{11}	464.4	201.2	1.42	0.007
θ_{12}	464.1	202.9	1.95	0.010

www.EngineeringBooksPdf.com

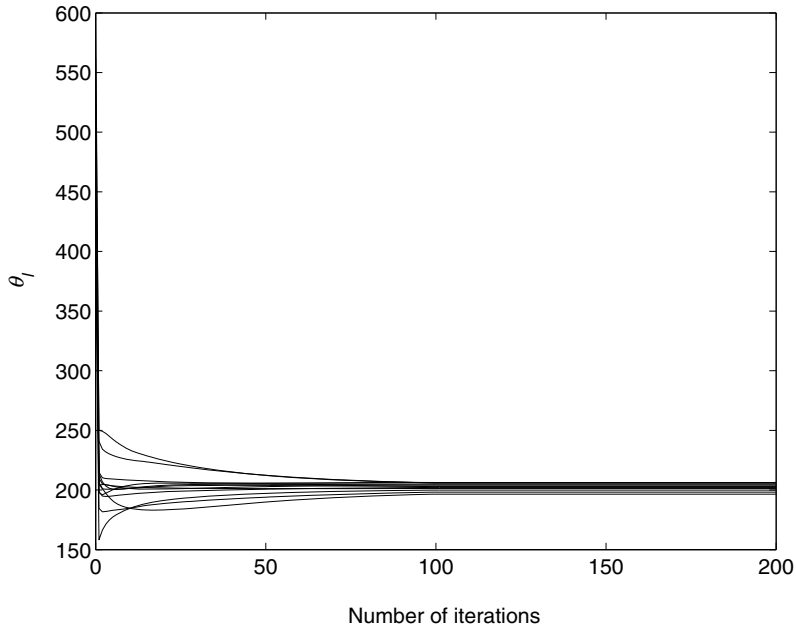

Figure 5.1 Iteration history for the most probable values of the stiffness parameters with incomplete measurement of mode shapes

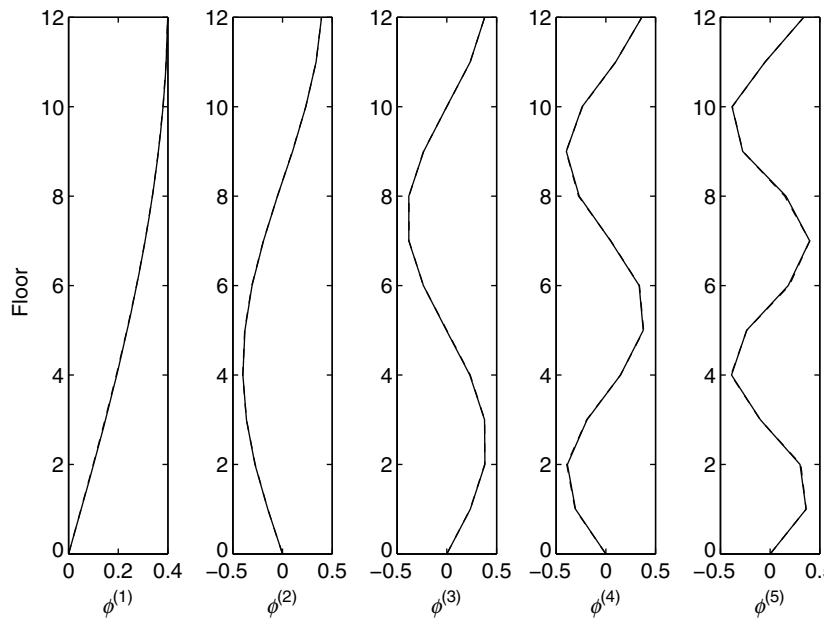

Figure 5.2 Comparison of the identified system mode shapes and actual mode shapes

www.EngineeringBooksPdf.com

5.6.2 Three-dimensional Six-story Braced Frame

In this example, the Bayesian model updating method is applied to update the finite-element model of a three-dimensional six-story braced frame, which is based on a model of an actual laboratory test structure. It is square in plan with width $a = 5$ m. There are four columns for each floor, one at each corner. Each of them have interstory stiffnesses of 10 MN/m and 15 MN/m in the x and y directions, respectively. Furthermore, each face in each floor is stiffened by a brace and its stiffness is taken to be 20 MN/m. As a result, the interstory stiffness is 80 MN/m and 100 MN/m in the x and y directions, respectively. The floor mass is taken to be 10 metric tons for each floor. As a result, the first five modal frequencies of the structure are 3.432, 3.837, 6.305, 10.10 and 11.29 Hz.

In order to locate the face(s) that sustain damage, four stiffness parameters are used for each story to give twenty-four stiffness parameters, $\theta_{4(l-1)+1} = K_{l,+x}$, $\theta_{4(l-1)+2} = K_{l,+y}$, $\theta_{4(l-1)+3} = K_{l,-x}$ and $\theta_{4(l-1)+4} = K_{l,-y}$, $l = 1, 2, \ldots, 6$, where the index l represents the story number and '$+x$', '$+y$', '$-x$' and '$-y$' represent the direction of the outward normal of the face. The actual values of these stiffnesses are $K_{l,+x} = K_{l,-x} = 50$ MN/m and $K_{l,+y} = K_{l,-y} = 40$ MN/m, $l = 1, 2, \ldots, 6$. In other words, $\theta_1 = \theta_3 = \cdots = \theta_{23} = 50$ MN/m and $\theta_2 = \theta_4 = \cdots = \theta_{24} = 40$ MN/m. The floor plan is shown in Figure 5.3. The point $\mathbf{O}'_l(\bar{x}_l, \bar{y}_l)$ is the stiffness center of floor l, where $\bar{x}_l$ and $\bar{y}_l$, $l = 1, 2, \ldots, 6$, are given by:

$$\bar{x}_l = \frac{a(K_{l,+x} - K_{l,-x})}{2(K_{l,+x} + K_{l,-x})}; \quad \bar{y}_l = \frac{a(K_{l,+y} - K_{l,-y})}{2(K_{l,+y} + K_{l,-y})} \tag{5.36}$$

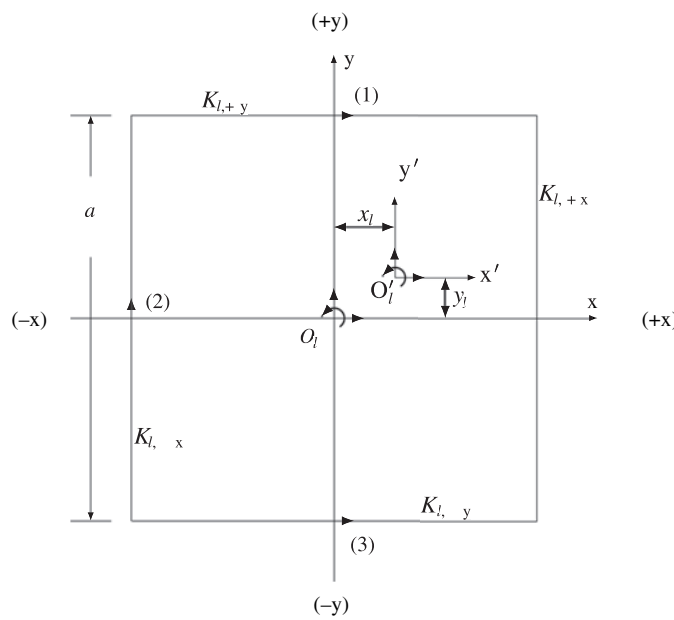

Figure 5.3 Floor plan for the 18-DOF model for identification

www.EngineeringBooksPdf.com

The element stiffness matrix for the lth story with respect to the $\mathbf{O}'_l$ coordinates is:

$$\mathbf{K}'_l = \begin{bmatrix} K_{l,+y} + K_{l,-y} & 0 & 0 & -K_{l,+y} - K_{l,-y} & 0 & 0 \\ 0 & K_{l,+x} + K_{l,-x} & 0 & 0 & -K_{l,+x} - K_{l,-x} & 0 \\ 0 & 0 & K_{lt} & 0 & 0 & -K_{lt} \\ -K_{l,+y} - K_{l,-y} & 0 & 0 & K_{l,+y} + K_{l,-y} & 0 & 0 \\ 0 & -K_{l,+x} - K_{l,-x} & 0 & 0 & K_{l,+x} + K_{l,-x} & 0 \\ 0 & 0 & -K_{lt} & 0 & 0 & K_{lt} \end{bmatrix} \tag{5.37}$$

where K_{lt} is given by:

$$K_{lt} = \left(\frac{a}{2} - \bar{x}_l\right)^2 K_{l,+x} + \left(\frac{a}{2} - \bar{y}_l\right)^2 K_{l,+y} + \left(\frac{a}{2} + \bar{x}_l\right)^2 K_{l,-x} + \left(\frac{a}{2} + \bar{y}_l\right)^2 K_{l,-y} \tag{5.38}$$

Here, the first three DOFs and the last three DOFs correspond to the lower floor and the upper floor of the story, respectively. Each of these sets of three DOFs correspond to the x-translational, y-translational and torsional motion.

The element stiffness matrix for the lth story with respect to the DOFs (1), (2) and (3) in Figure 5.3 of the upper and lower floors is given by:

$$\mathbf{K}_l = \bar{\mathbf{T}}^T \mathbf{K}'_l \bar{\mathbf{T}} \tag{5.39}$$

where $\bar{\mathbf{T}}$ is given by:

$$\bar{\mathbf{T}} = \begin{bmatrix} \mathbf{T}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^{-1} \end{bmatrix}; \quad \mathbf{T} = \begin{bmatrix} 1 & 0 & -\frac{a}{2} + \bar{y}_l \\ 0 & 1 & -\frac{a}{2} - \bar{x}_l \\ 1 & 0 & \frac{a}{2} + \bar{y}_l \end{bmatrix} \tag{5.40}$$

The stiffness matrix for the 18-DOF structural model is assembled from those of the floors. The DOFs for this stiffness matrix are (1), (2) and (3) shown in Figure 5.3 for each floor. However, this stiffness matrix is not linear in the stiffness parameters θ_l. In this case, the relationship between the stiffness matrix and the stiffness parameters can be linearized:

$$\mathbf{K} = \mathbf{K}_0 + \sum_{l=1}^{24} \theta_l \mathbf{K}_l \tag{5.41}$$

where:

$$\mathbf{K}_l = \frac{\partial \mathbf{K}}{\partial \theta_l}, \quad l = 1, 2, \ldots, 24 \tag{5.42}$$

and

$$\mathbf{K}_0 = \mathbf{K} - \sum_{l=1}^{24} \theta_l \mathbf{K}_l \tag{5.43}$$

where the matrices $\mathbf{K}_l$, $l = 0, 1, \ldots, 24$, will need to be updated in every iteration.

www.EngineeringBooksPdf.com

It is assumed that only the first three x-directional and y-directional modes are measured but not any of the torsional modes. This is done deliberately to simulate a common situation where some of the modes are not excited sufficiently to be able to observe. In the identification process, it is unknown that there are some missing modes. The six measured modes correspond to the 1st (3.432 Hz), 2nd (3.837 Hz), 4th (10.10 Hz), 5th (11.29 Hz), 7th (18.08 Hz) and 9th (21.31 Hz) modes. Sensors are placed on the $+y$ and $-y$ faces of the 1st, 2nd, 5th and 6th floors, and the $-x$ face of all floors to measure the modal frequencies and mode shape components. The covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\epsilon}}$ is diagonal with 0.5% COV for the modal data. For the simulated modal data, a sample of zero-mean Gaussian noise with covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\epsilon}}$ was added to the exact modal frequencies and mode shapes. Initial values for all stiffness parameters are taken to be 100 MN/m, which overestimates the values by 100% and 150% for the $\pm x$ and $\pm y$ faces, respectively.

The iteration history for the most probable values of the stiffness parameters is shown in Figure 5.4 and the values essentially converge after 500 iterations. It took about 11 s of PC time to finish 1000 iterations. The stiffness parameters converge onto one of two values: one for the $\pm x$ faces (approximately 50 MN/m) and the other for the $\pm y$ faces (approximately 40 MN/m). Recall that the substructure stiffness matrices $\mathbf{K}_l$, $l = 0, 1, \ldots, 24$, depend on the uncertain stiffness parameters in this example, which slows down the convergence rate. Nevertheless, the methodology is still considered very efficient.

Table 5.3 shows the actual values, identified values, standard deviations and COVs of the stiffness parameters. The standard deviations are computed using Equation (5.30) and they

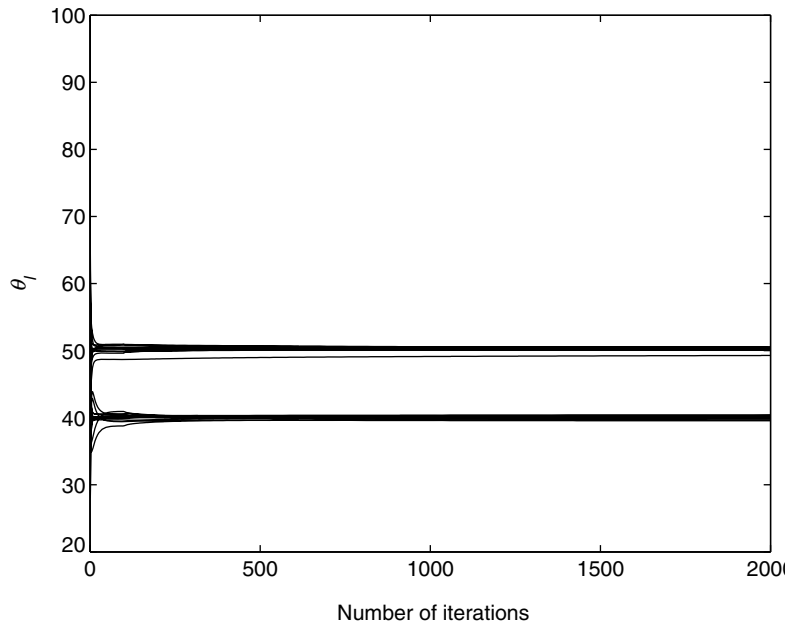

Figure 5.4 Iteration history for the most probable values of the twenty stiffness parameters of the undamaged structure

www.EngineeringBooksPdf.com

Table 5.3 Identification results for the undamaged structure

Parameter	Actual $\tilde{\theta}$	Identified $\theta^{\star}$	S.D. σ_{θ}	COV
$\theta_{1,+x}$	50	50.10	0.21	0.0042
$\theta_{1,+y}$	40	40.13	0.22	0.0054
$\theta_{1,-x}$	50	50.20	0.12	0.0025
$\theta_{1,-y}$	40	40.28	0.21	0.0052
$\theta_{2,+x}$	50	49.28	0.57	0.0113
$\theta_{2,+y}$	40	40.00	0.25	0.0061
$\theta_{2,-x}$	50	50.47	0.36	0.0072
$\theta_{2,-y}$	40	39.83	0.27	0.0068
$\theta_{3,+x}$	50	50.57	0.17	0.0033
$\theta_{3,+y}$	40	39.55	0.45	0.0111
$\theta_{3,-x}$	50	50.45	0.10	0.0020
$\theta_{3,-y}$	40	40.37	0.38	0.0096
$\theta_{4,+x}$	50	50.42	0.31	0.0062
$\theta_{4,+y}$	40	39.92	0.23	0.0057
$\theta_{4,-x}$	50	50.10	0.19	0.0038
$\theta_{4,-y}$	40	40.37	0.20	0.0049
$\theta_{5,+x}$	50	50.45	0.22	0.0043
$\theta_{5,+y}$	40	40.15	0.38	0.0096
$\theta_{5,-x}$	50	50.06	0.14	0.0027
$\theta_{5,-y}$	40	39.63	0.36	0.0089
$\theta_{6,+x}$	50	50.23	0.19	0.0038
$\theta_{6,+y}$	40	40.15	0.16	0.0040
$\theta_{6,-x}$	50	50.30	0.11	0.0022
$\theta_{6,-y}$	40	40.09	0.16	0.0039

vary up to about 1%. The identified values are all close to the actual values. The difference between the actual and identified values are of similar order to the corresponding estimated standard deviations.

5.6.2.1 Application to Structural Health Monitoring

The same three-dimensional braced-frame structure is assumed to be damaged on the $+x$ face of the first story and the $+y$ face of the fourth story, giving a stiffness reduction of one third and one quarter, respectively, of the original brace stiffnesses for these two faces. They correspond to 15% and 10% stiffness reductions for the stiffness parameters corresponding to these faces, respectively. It is noted that there is no sensor on the $+x$ face. The first five modal frequencies of the damaged structure are 3.419, 3.788, 6.251, 10.04 and 11.16 Hz and the damage alters the order of the modes compared with the undamaged structure. Furthermore, the translational and torsional modes are no longer purely translational and torsional, respectively, especially for the higher modes. Here, six modes, which behave closer to translational, are measured and they are the 1st (3.419 Hz), 2nd (3.788 Hz), 4th (10.04 Hz), 5th (11.16 Hz), 6th (16.14 Hz), and 7th (17.90 Hz) modes. Initial values for the stiffness parameters are again taken to be 100 MN/m. Noisy modal data is used to identify this damaged structure, as in the case of the undamaged structure.

www.EngineeringBooksPdf.com

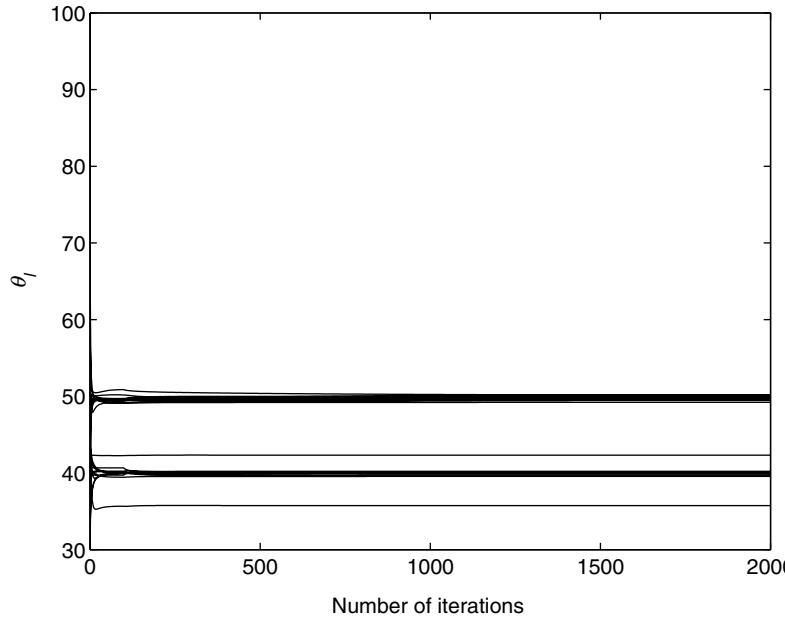

Figure 5.5 Iteration history for the most probable values of the twenty stiffness parameters of the damaged structure

The iteration history is shown in Figure 5.5. It took 23 s of PC time to finish 2000 iterations, with convergence occurring in about 1000 iterations. The most probable values of the stiffness parameters converge onto one of four values, as expected: one for the $+y$ face of the first story, one for the other $\pm y$ faces, one for the $+x$ face of the third story and one for the other $\pm x$ faces. The identification results are shown in Table 5.4, which has a similar format to Table 5.3. It is clearly seen that the $+y$ face of the first story and the $+x$ face of the third story have had substantial stiffness reductions compared to the undamaged structure.

The discussion above uses the most probable values of the stiffness parameters based on modal data from the undamaged and damaged structure. In order to further portray the damage, these are the most probable values and the standard deviations for the stiffness parameters are used to compute the probability that a given stiffness parameter θ_l has been reduced by a certain fraction d compared to the undamaged state of the structure [18, 268]. An asymptotic Gaussian approximation [197] is used for the integrals involved to give:

$$\begin{aligned} P_l^{dam}(d) &= P(\theta_l^{pd} < (1-d)\theta_l^{ud}|\mathcal{C}) \\ &= \int_{-\infty}^{\infty} P(\theta_l^{pd} < (1-d)\theta_l^{ud}|\theta_l^{ud}, \mathcal{C})\, p(\theta_l^{ud}|\mathcal{C})\mathrm{d}\theta_l^{ud} \\ &\approx \Phi\left[\frac{(1-d)\theta_l^{\star\, ud} - \theta_l^{\star\, pd}}{\sqrt{(1-d)^2(\sigma_l^{ud})^2 + (\sigma_l^{pd})^2}}\right] \end{aligned} \tag{5.44}$$

www.EngineeringBooksPdf.com

Table 5.4 Identification results for the damaged structure

Parameter	Actual $\tilde{\theta}$	Identified $\theta^\star$	S.D. σ_θ	COV
$\theta_{1,+x}$	42.5	42.32	0.19	0.0043
$\theta_{1,+y}$	40	40.13	0.19	0.0048
$\theta_{1,-x}$	50	49.65	0.16	0.0032
$\theta_{1,-y}$	40	40.09	0.20	0.0049
$\theta_{2,+x}$	50	50.17	0.74	0.0147
$\theta_{2,+y}$	40	39.55	0.53	0.0131
$\theta_{2,-x}$	50	49.91	0.35	0.0070
$\theta_{2,-y}$	40	39.84	0.54	0.0135
$\theta_{3,+x}$	50	49.47	0.20	0.0039
$\theta_{3,+y}$	40	40.05	0.34	0.0084
$\theta_{3,-x}$	50	49.55	0.13	0.0027
$\theta_{3,-y}$	40	39.96	0.34	0.0086
$\theta_{4,+x}$	50	49.21	0.35	0.0070
$\theta_{4,+y}$	36	35.73	0.30	0.0082
$\theta_{4,-x}$	50	49.98	0.24	0.0047
$\theta_{4,-y}$	40	39.62	0.34	0.0085
$\theta_{5,+x}$	50	50.14	0.28	0.0056
$\theta_{5,+y}$	40	39.98	0.30	0.0075
$\theta_{5,-x}$	50	49.56	0.14	0.0028
$\theta_{5,-y}$	40	39.94	0.28	0.0071
$\theta_{6,+x}$	50	49.49	0.24	0.0047
$\theta_{6,+y}$	40	40.22	0.19	0.0048
$\theta_{6,-x}$	50	49.76	0.12	0.0023
$\theta_{6,-y}$	40	40.06	0.19	0.0047

where $\Phi(\cdot)$ is the cumulative distribution function of the standard Gaussian random variable, $\theta_l^{\star ud}$ and $\theta_l^{\star pd}$ denote the most probable values of the stiffness parameters for the undamaged and (possibly) damaged structures respectively, and σ_l^{ud} and σ_l^{pd} are the corresponding standard deviations of the stiffness parameters.

The probabilities of damage for the twenty-four θ_l are shown in Figure 5.6. It can be clearly seen that the $+x$ face of the first story and the $+y$ face of the fourth story have damage with a probability of almost unity. The means of the damage are 15.5% and 10.5% while their actual values are 15% and 10%, respectively. Furthermore, the COVs of these estimates are 0.6% and 1.0%, respectively. These types of plots can be interpreted as follows. Consider the curve for the probability of damage of the $+x$ face of the first story. The damage is 13.5% or more with a probability of almost 1.0 and has a probability of only 0.05 to exceed 16.3%. For the twenty-two faces that are undamaged, the difference between the mean stiffness parameter values of the undamaged and damaged structures is less than 2.5%, which is within the expected uncertainty level.

5.7 Concluding Remarks

A Bayesian model updating methodology is presented with application to structural health monitoring. The method utilizes noisy incomplete modal data, i.e., there can be missing mode

www.EngineeringBooksPdf.com

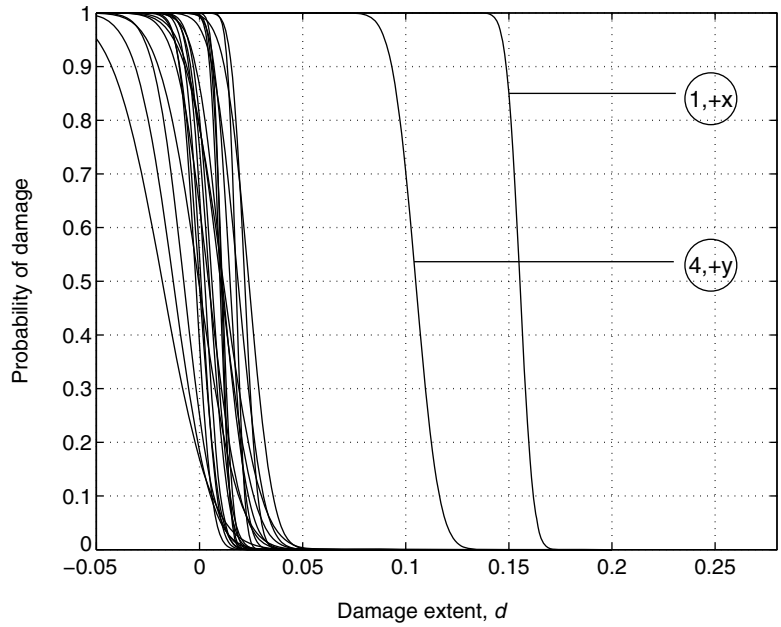

Figure 5.6 Probability of damage of different faces of the frame

shape components and missing modes. It does not require any matching between measured and calculated modes from a structural model, so there is no need to solve the eigenvalue problem corresponding to the structural model. Instead of solving directly the nonlinear optimization problem, an efficient algorithm of a sequence of linear optimizations is introduced. The illustrative examples confirm the effectiveness of the proposed approach, showing it to be both computationally efficient and robust. Furthermore, the method successfully traces back to the correct parameter values even though the initial values are very far from the actual values. The second example is a very difficult problem with large number of unknown stiffness parameters and mode shape components. It shows that the method updates the structural models successfully without knowing that there are missing modes in the modal data set and that the ordering of the modes switches due to damage. However, the method does not require any information about the switching and missing modes. The measured modes for the undamaged and damaged structure correspond to different orders of the structural modes and this information is also unknown in the updating process. This is very important in practical usage.

www.EngineeringBooksPdf.com

www.EngineeringBooksPdf.com

6

Bayesian Model Class Selection

Keywords: asymptotic expansion; evidence; information entropy; Markov Chain Monte Carlo simulation; modal identification; Ockham factor; regression problem; robustness; seismic attenuation

6.1 Introduction

The most commonly encountered problem in system identification is to find the best/optimal model in a specified class of models, e.g., the class of shear building models or class of bilinear hysteretic models. The problem is referred to as *parametric identification* and has been discussed in the previous chapters. The more general problem of model class selection has not been well explored in system identification. In this problem, the concern is on the selection of a suitable class of models for parametric identification. It is obvious that a more complicated model can fit the data better than a less complicated one which has fewer adjustable/uncertain parameters. Therefore, if the optimal model class is chosen by minimizing some norm of the error between the measured data and the corresponding predictions of the optimal model in each model class, the optimal model class will always be the most complicated one since it covers the largest output space. For example, in modal identification, using a twenty-mode model would always be better than using a ten-mode model because the former one would fit the data better, although the improvement might be negligible. This approach is therefore likely to lead to over-fitted model classes. When an over-fitted model is used for future prediction, it is likely to lead to poor results because the identified model depends too much on the details of the data and the measurement noise in the data has an important role in the data fitting. Therefore, in model class selection, it is necessary to penalize more complicated model classes. This point was first recognized by H. Jeffreys who did pioneering work on the application of Bayesian methods [121]. He pointed out the need for a quantitative expression of the very old philosophy expounded by William of Ockham (or Occam in Latin) and known as *Ockham's razor*, which is roughly translated from Latin as: 'It is vain to do with more what can be done with fewer' [242]. In the present context, this philosophy implies that simpler model classes are more preferable than unnecessarily complicated ones. In other words, the selected class of models

Bayesian Methods for Structural Dynamics and Civil Engineering Ka-Veng Yuen
© 2010 John Wiley & Sons (Asia) Pte Ltd

www.EngineeringBooksPdf.com

should agree closely with the observed behavior of the system but otherwise be as simple as possible. Box and Jenkins (1970) also emphasized the same principle when they referred to the need for parsimonious models in time-series forecasting, although they did not give a quantitative expression of their principle of parsimony [30]. Akaike recognized that maximum likelihood estimation is insufficient for model order selection in time-series forecasting using ARMA models and came up with another term to be added to the logarithm of the likelihood function that penalizes against complicated parameterization of the models [2]. This was later modified by Akaike [3] and by Schwarz [232].

In recent years, there has been a re-appreciation of the work of Jeffreys on the application of Bayesian methods [121], especially due to the expository publications of E.T. Jaynes [118, 120]. In particular, the Bayesian approach to model class selection has been further developed by showing that the *evidence* for each model class provided by the data (that is, the probability of getting the data based on the whole model class) automatically enforces a quantitative expression of a *principle of model parsimony* or of Ockham's razor [98, 164, 242]. There is no need to introduce any ad-hoc penalty term as was done in some of the earlier work on this problem.

Influenced by the mind of forward modeling problems, it is easily directed to adopt complicated model classes so as to capture various complex physical mechanisms. However, the more complicated the model class is utilized, the more uncertain parameters are normally induced unless extra mathematical constraints are imposed. In the former case, the model output may not necessarily be accurate even if the model well characterizes the physical system since the combination of the many small errors from each uncertain parameter can induce a large output error. In the latter case, it is possible that the extra constraints induce substantial errors. Therefore, it is important to use a proper model class for system identification purpose. In this chapter, the Bayesian model class selection approach is introduced and applied to select the most plausible/suitable class of mathematical models representing a static or dynamical (structural, mechanical, atmospheric, ...) system (from some specified model classes) by using its response measurements. This approach has been shown to be promising in several research areas, such as artificial neural networks [164, 297], structural dynamics and model updating [23], damage detection [150] and fracture mechanics [151], etc.

Example. Does Small Posterior Uncertainty Imply Good Data Fitting?
From the above discussion, it is clear that a good model class to use for identification should be capable to fit the data well and provide sufficient robustness to modeling error and measurement noise. However, the definition of the degree of data fitting is not yet defined. This example investigates the relationship between small posterior uncertainty and good data fitting.

Consider two single-parameter model classes with the likelihood functions shown in Figure 6.1 for the same set of measurement. This figure shows also the log-likelihood functions. In this case, $p(\mathcal{D}|\theta, \mathcal{C}_2) = 4p(\mathcal{D}|\theta, \mathcal{C}_1)$ so $\ln p(\mathcal{D}|\theta, \mathcal{C}_2) = 2\ln 2 + \ln p(\mathcal{D}|\theta, \mathcal{C}_1)$. With the same prior distribution, the posterior PDFs and uncertainty of the parameter of these two model classes are identical since the difference between the log-likelihood functions is a constant. However, model class $\mathcal{C}_2$ provides better fitting to the data as its maximum likelihood value is four times of that for $\mathcal{C}_1$:

$$p(\mathcal{D}|\theta^\star, \mathcal{C}_2) = 4p(\mathcal{D}|\theta^\star, \mathcal{C}_1) \tag{6.1}$$

www.EngineeringBooksPdf.com

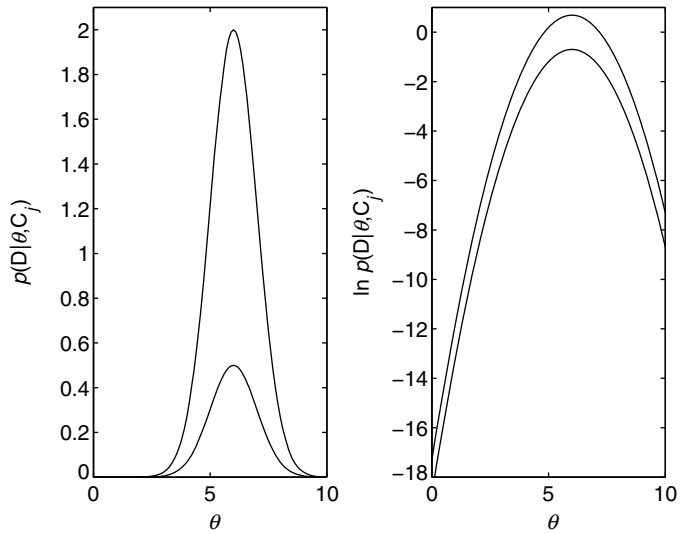

Figure 6.1 Likelihood and log-likelihood functions

Therefore, the same level of posterior uncertainty can be associated with different levels of data fitting. This example demonstrates that the degree of data fitting has no direct relationship with the posterior uncertainty of the parameters. Furthermore, in this example, the normalizing constant of $\mathcal{C}_1$ will be four times of that with $\mathcal{C}_2$ and it will be demonstrated in the following sections that the normalizing constant plays an important role in model class selection.

Example. Dow Jones Industrial Average
Another issue to address is whether good data fitting implies good modeling. With the data of 2008, let's attempt to use the following empirical model for the daily closing values of the Dow Jones Industrial Average (DJIA):

$$\ln DOW(n) = b_0 + b_1 \exp(\sin n) + b_2 \exp(\sin 2n) + \cdots + b_{252} \exp(\sin 252n) \qquad (6.2)$$

where $DOW(n)$ is the closing value of the index on the nth trading day starting from 2/1/2008 and there were 253 trading days in this year. The logarithm used on the left hand side guarantees a positive value of the model output. In this case, the unknown coefficients $b_l, l = 0, 1, \ldots, 252$, can be determined by the 253 data points and they are $b_0 = 22.28$, $b_1 = -0.0169$, $b_2 = -0.0157$, $b_3 = 0.0035$, $b_4 = -0.0053, \ldots$, etc. This curve fits perfectly the closing values of all trading days in 2008 since there exists a one-to-one linear mapping from the 253 coefficients to the 253 closing values. In other words, the fitting error is zero and the maximum likelihood value is infinity with a perfect match. Furthermore, the posterior variances for all coefficients are zero. However, this model over-fits the data. If it is used to regenerate the closing values in any day of 2008, it will give perfect results. However, if this model is used for *prediction*, the story will be very different and the results for 2009 are shown in logarithmic scale in Figure 6.2 (This part was written in April 2009). The prediction goes all the way to 6.3×10^8 in mid April and then drops down to less than 1 point in mid June. Obviously, the prediction

www.EngineeringBooksPdf.com

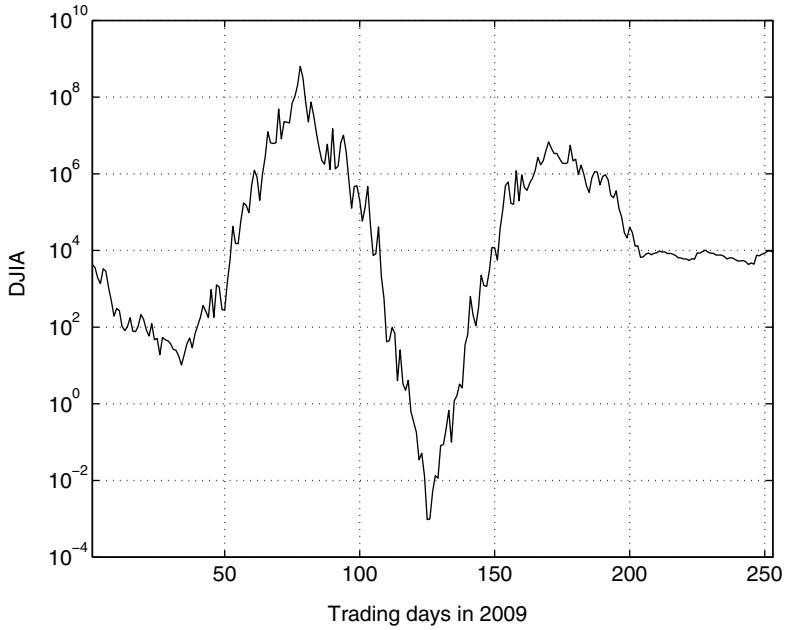

Figure 6.2 *Prediction* of the Dow Jones Industrial Average for 2009

by this model is not realistic even though it matches perfectly the data in the past. Assuming that the closing value of a day is changed for some reason, the identified coefficients and hence the prediction will be altered due to the change but the new model still fits perfectly the measurements. Figure 6.3 shows the prediction by the model trained with the measurement of 2008 with the only difference on the 202nd trading day with 2% reduction. With this mild perturbation of the measurements (only 2% change in one data point out of the 253 data points), a visible change in the prediction can be observed, especially in the neighborhoods of the local maxima and minima. In other applications where measurement noise exists in all data points, the measurement error will contribute substantially to the identification results of the model parameters and hence the prediction of the over-fitted model. In conclusion, a large maximum likelihood value implies good data fitting but not necessarily a suitable model class for identification if the identified model is to be used for future prediction. Therefore, it is not appropriate to select the class of models for future prediction simply by choosing the one with the largest maximum likelihood value.

6.1.1 Sensitivity, Data Fitness and Parametric Uncertainty

Posterior uncertainty is a measure of the spread of the product of the prior distribution and the likelihood function. As shown previously in the 4th example of Section 2.1.6 (Chapter 2) a small posterior uncertainty is possible for poor data fitting. In this case, the normalizing constant will be large.

www.EngineeringBooksPdf.com

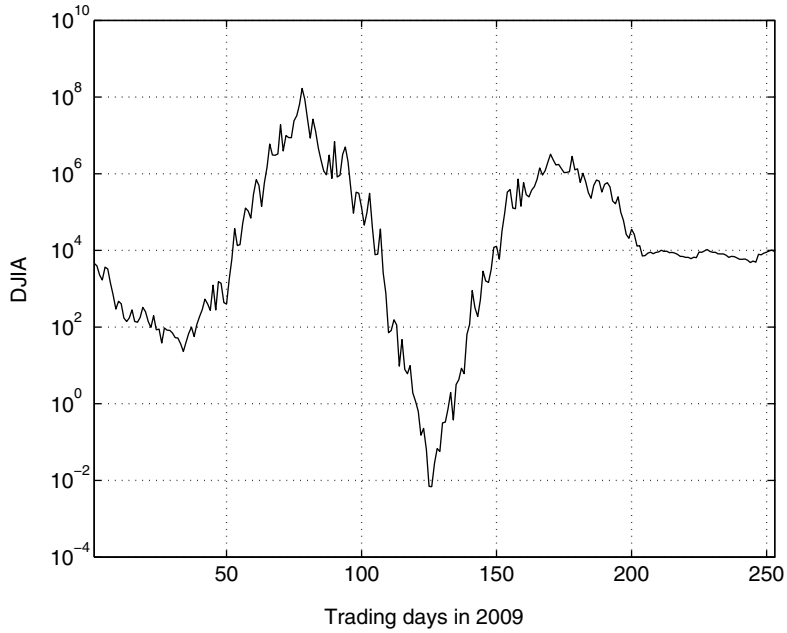

Figure 6.3 *Prediction* of the Dow Jones Industrial Average for 2009 with small perturbations of the data

On the other hand, the sensitivity is the change of model output $\mathbf{X}$ due to parameter perturbation and it can be quantified by the following matrix:

$$\mathcal{J} = \frac{\partial \mathbf{X}}{\partial \boldsymbol{\theta}} \tag{6.3}$$

The slope of the posterior PDF depends on the sensitivity, which controls the rate of the change of the model output due to perturbation of the parameters. On the other hand, the posterior uncertainty of the parameters is controlled by the decaying rate (slope) of the posterior PDF in the neighborhood around the optimal point. Therefore, it is particularly important to investigate the sensitivity of a model around the optimal parameters.

If a model class is utilized for future prediction, it is desirable to obtain a robust model class that fits the data well even with errors in the parameters. In this case, the maximum likelihood value has to be large and the likelihood value remains large in a sufficiently large neighborhood of the most plausible parameter values. Therefore, the topology of the likelihood function around the maximum is flat and the sensitivity is low. Figure 6.4 shows schematically two likelihood functions to demonstrate these relationships. The two likelihood functions are assumed to be obtained with the same set of measurements but with different model classes. Model class $\mathcal{C}_1$ obviously has a larger maximum likelihood value than $\mathcal{C}_2$, indicating that the optimal model in $\mathcal{C}_1$ fits the data better than that of $\mathcal{C}_2$. Furthermore, the posterior uncertainty of the parameter in $\mathcal{C}_1$ is also smaller than that of $\mathcal{C}_2$ if the same prior distribution is used. However, if the optimal models are to be used for future prediction, model class $\mathcal{C}_2$ is more reliable since it has a larger parameter range that provides satisfactory fitting to the data so it

www.EngineeringBooksPdf.com

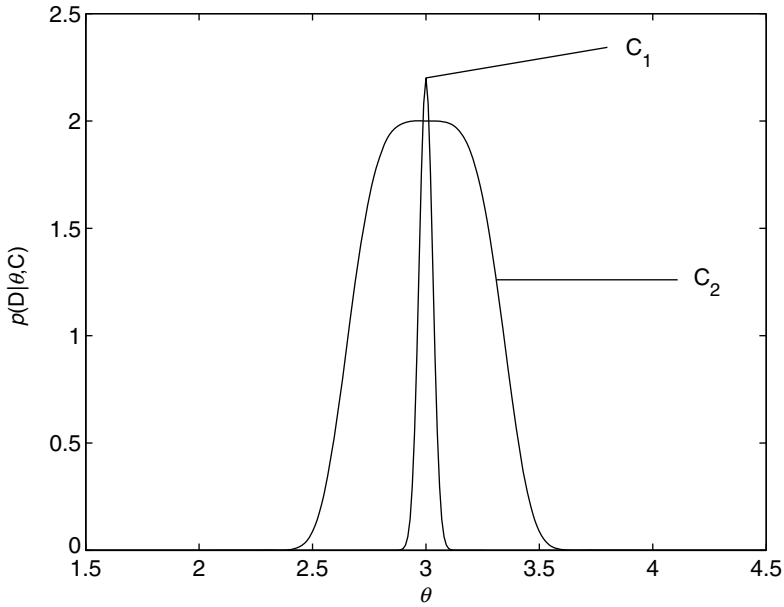

Figure 6.4 Schematic plot for demonstration of sensitivity

is more robust. On the other hand, the maximum likelihood margin of $\mathcal{C}_1$ may not be sufficient to compensate if the model parameter has a tiny error. Therefore, a reliable model class should have a reasonable tradeoff between the data fitting capability and the robustness to model error. This example also demonstrates that even for the same number of parameters, the model class with a larger maximum likelihood value is not necessarily the better for identification. Table 6.1 shows the four combinations of data fitting capability and posterior uncertainty. It is desirable to use a model class that has a high value of the maximum likelihood and a large posterior uncertainty so that efficient and robust identification performance can be expected but this is difficult to achieve. Data fitting capability can be easily enhanced by adding free parameters but this will inevitably degrade the robustness of the model class. Therefore, a suitable model class should possess a reasonable balance between these two important properties.

Table 6.1 Identification results for one set of data and frequency range $(0, 1.2\tilde{\Omega}]$

Posterior uncertainty\ Maximum likelihood value	High	Low
Large	Efficient and robust	Poor modeling
Small	Efficient but fragile	Poor modeling

In the next section, the Bayesian model class selection method is introduced for quantification and selection of model classes. It will be discussed for the globally identifiable case and the general case. The Ockham factor is introduced and it serves as the penalty for a complicated model, which appears naturally from the evidence. Computational issues will be discussed and

www.EngineeringBooksPdf.com

the transitional Markov Chain Monte Carlo (TMCMC) simulation method will be introduced. Linear and nonlinear regression problems are presented in Section 6.3. Formulae will be given for two special but popular choices of the prior distribution of the uncertain parameters. In Sections 6.4 and 6.5, illustrative examples are presented for the determination of the number of modes for modal identification of a building and the decision on the seismic attenuation empirical relationship. Finally, prior distribution will be revisited and an interesting and inspirational example will be given to conclude the book.

6.2 Bayesian Model Class Selection

Let $\mathcal{D}$ denote the input–output or output-only data from a physical system or phenomenon. The goal is to use $\mathcal{D}$ to select the most plausible/suitable class of models representing the system out of N_C given classes of models $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_{N_C}$. Since probability may be interpreted as a measure of plausibility based on specified information [63], the probability of a class of models conditional on the set of dynamic data $\mathcal{D}$ is required. This can be obtained by using the Bayes' theorem as follows:

$$P(\mathcal{C}_j|\mathcal{D}, \mathcal{U}) = \frac{p(\mathcal{D}|\mathcal{C}_j, \mathcal{U})P(\mathcal{C}_j|\mathcal{U})}{p(\mathcal{D}|\mathcal{U})}, \quad j = 1, 2, \ldots, N_C \tag{6.4}$$

where $p(\mathcal{D}|\mathcal{U})$ is the denominator is given by the law of total probability:

$$p(\mathcal{D}|\mathcal{U}) = \sum_{j=1}^{N_C} p(\mathcal{D}|\mathcal{C}_j, \mathcal{U})P(\mathcal{C}_j|\mathcal{U}) \tag{6.5}$$

and $\mathcal{U}$ expresses the user's judgement on the initial plausibility of the model classes in terms of the prior plausibility $P(\mathcal{C}_j|\mathcal{U})$ on the model classes $\mathcal{C}_j$, $j = 1, 2, \ldots, N_C$. The prior plausibilities are normalized in the same way as probabilities:

$$\sum_{j=1}^{N_C} P(\mathcal{C}_j|\mathcal{U}) = 1 \tag{6.6}$$

In the context of decision theory, this prior plausibility can be utilized to assign different weightings for model classes to take into account of other non-modeling factors, e.g., to penalize model classes with heavier computational demands. However, this is not considered in the scope of the book. The factor $p(\mathcal{D}|\mathcal{C}_j, \mathcal{U})$ is called the *evidence* for the model class $\mathcal{C}_j$ provided by the data $\mathcal{D}$. It expresses how likely the data are obtained if the model class $\mathcal{C}_j$ is assumed and plays an important role on model class selection. Since there is a certain randomness in the measurement, the data are more typical if the evidence is large. In this case, the identification results are more reliable by using the more representative data set. The user's preference $\mathcal{U}$ is irrelevant in $p(\mathcal{D}|\mathcal{C}_j, \mathcal{U})$ and so it can be dropped from the notation because it is assumed that $\mathcal{C}_j$ alone specifies the probability density function (PDF) for the data:

$$p(\mathcal{D}|\mathcal{C}_j, \mathcal{U}) = p(\mathcal{D}|\mathcal{C}_j) \tag{6.7}$$

In other words, $\mathcal{C}_j$ specifies not only a class of deterministic dynamic models but also the probability descriptions for the prediction error and initial plausibility for each model in the

www.EngineeringBooksPdf.com

model class $\mathcal{C}_j$ [19]. Equation (6.4) shows that the most plausible model class is the one that maximizes $p(\mathcal{D}|\mathcal{C}_j)P(\mathcal{C}_j|\mathcal{U})$ with respect to j.

The plausibility $P(\mathcal{C}_j|\mathcal{D},\mathcal{U})$ can be used not only for selection of the most plausible class of models, but also for response prediction based on all the model classes. Let Q denote a quantity to be predicted, e.g., first story drift. Then, the PDF of Q given the data $\mathcal{D}$ can be calculated from the law of total probability as follows:

$$p(Q|\mathcal{D},\mathcal{U}) \approx \sum_{j=1}^{N_C} p(Q|\mathcal{D},\mathcal{C}_j)P(\mathcal{C}_j|\mathcal{D},\mathcal{U}) \tag{6.8}$$

rather than using only the most plausible model for prediction. This approximation is exact only if the N_C model classes provides independent predictions for the variable Q. For example, if two of the model classes are identical (say $\mathcal{C}_1 = \mathcal{C}_2$), then its contribution will be 'double counted.' However, it will be computationally expensive to take into account of the correlations among all the model classes, especially for large N_C. However, this approximation still performs better than using the most plausible model class alone in a normal situation.

On the other hand, if the plausibility $P(\mathcal{C}^\star|\mathcal{D},\mathcal{U})$ for the most plausible model class $\mathcal{C}^\star$ is much larger than the plausibility of the others, then the above expression is approximated by:

$$p(Q|\mathcal{D},\mathcal{U}) \approx p(Q|\mathcal{D},\mathcal{C}^\star) \tag{6.9}$$

and it is sufficient to just use the most plausible model class for predicting the system behavior.

The evidence for $\mathcal{C}_j$ provided by the data $\mathcal{D}$ is given by the law of total probability:

$$p(\mathcal{D}|\mathcal{C}_j) = \int_{\boldsymbol{\Theta}_j} p(\mathcal{D}|\boldsymbol{\theta},\mathcal{C}_j)p(\boldsymbol{\theta}|\mathcal{C}_j)\mathrm{d}\boldsymbol{\theta}, \quad j = 1, 2, \ldots, N_C \tag{6.10}$$

where $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\Theta}_j)$ is the parameter vector in the parameter space $\boldsymbol{\Theta}_j \subset \mathbb{R}^{N_j}$ and defines each model in $\mathcal{C}_j$. The parameter vector $\boldsymbol{\theta}$ depends on the model class $\mathcal{C}_j$ even though it is not explicitly reflected in the symbol. This is done only for the purpose of simplifying the notation. The prior PDF $p(\boldsymbol{\theta}|\mathcal{C}_j)$ is specified by the user and $p(\mathcal{D}|\boldsymbol{\theta},\mathcal{C}_j)$ is the likelihood function. The parameter vectors $\boldsymbol{\theta}$ are different for different model classes but this is not shown explicitly in the symbol for a notation simplification purpose. Note that the evidence is equal to the reciprocal of the normalizing constant in establishing the posterior PDF. Let's revisit the 4th example in Section 2.1.6 (Chapter 2). In this case, the normalizing constant is

$$\kappa_0 = 2\sqrt{\pi}\sigma_0 \exp\left[\frac{(\hat{Q}_1 - \hat{Q}_2)^2}{4\sigma_0^2}\right]$$

with $\sigma_0^2 = 0.01$ and $\kappa_0 = 1.85 \times 10^{173}$. If another model class $\mathcal{C}_2$ is used with σ_0^2 reduced to 0.005, then $\kappa_0 = 6.83 \times 10^{346}$. Therefore, the first model class has a smaller normalizing constant than the second one. Since the normalizing constant is the reciprocal of the evidence, a smaller value of the normalizing constant associates with a larger evidence. Furthermore, it implies that with this model class the probability of obtaining such data is higher. Therefore, the identification result with this more representative data is more reliable.

In practice, direct calculation/numerical integration of the evidence integral in Equation (6.10) is often computationally prohibitive. Consider a model class with only

www.EngineeringBooksPdf.com

10 uncertain parameters. For each parameter, its range is partitioned into 20 intervals so there are $20^{10} = 1024 \times 10^{10}$ hyper-cubes in the parameter space for evaluation. For each cube, computation of the likelihood function of a given set of parameters requires one system analysis. If it takes 10^{-4} s for each run, it requires 1024×10^6 s ≈ 32 years for the numerical evaluation of the evidence integral. For practical problems, there are usually more uncertain parameters and a finer discretization is necessary, so the computational demand will grow drastically. Therefore, a more computationally feasible approach is needed.

6.2.1 Globally Identifiable Case

In *globally identifiable* cases [19], the posterior/updated PDF for $\boldsymbol{\theta}$ given a large amount of data $\mathcal{D}$ may be approximated accurately by a Gaussian distribution, so the evidence $p(\mathcal{D}|\mathcal{C}_j)$ can be approximated by using Laplace's method for asymptotic expansion [197]:

$$p(\mathcal{D}|\mathcal{C}_j) \approx p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j)p(\boldsymbol{\theta}^\star|\mathcal{C}_j)(2\pi)^{\frac{N_j}{2}}|\mathcal{H}_j(\boldsymbol{\theta}^\star)|^{-\frac{1}{2}}, \quad j = 1, 2, \ldots, N_C \tag{6.11}$$

where N_j is the number of uncertain parameters for the model class $\mathcal{C}_j$. The optimal parameter vector $\boldsymbol{\theta}^\star$ maximizes $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)$ in the interior of $\boldsymbol{\Theta}_j$ and $\mathcal{H}_j(\boldsymbol{\theta}^\star)$ is the Hessian matrix of the objective function:

$$J(\boldsymbol{\theta}) = -\ln\left[p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}_j)p(\boldsymbol{\theta}|\mathcal{C}_j)\right] \tag{6.12}$$

with respect to $\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta}^\star$. Specifically, the (l, l') component of the Hessian matrix $\mathcal{H}_j(\boldsymbol{\theta}^\star)$ is given by:

$$\mathcal{H}_j^{(l,l')}(\boldsymbol{\theta}^\star) = \frac{\partial^2}{\partial\theta_l\theta_{l'}}J(\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \tag{6.13}$$

where θ_l is the lth component of the parameter vector $\boldsymbol{\theta}$. This matrix can be computed by the finite difference method (Appendix A). Note that the solution in Equation (6.11) is exact if the posterior PDF is Gaussian.

The maximum likelihood value $p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j)$ in Equation (6.11) will be larger for those model classes $\mathcal{C}_j$ that make the probability of the data $\mathcal{D}$ higher, that is, that give a better fitting to the data. For example, if the likelihood function is Gaussian, then the largest value of $p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j)$ will be given by the model class $\mathcal{C}_j$ that gives the smallest least-squares fitting to the data. As mentioned earlier, this likelihood function favors model classes with more uncertain parameters. If the number of data points N in $\mathcal{D}$ is large, the likelihood value will be the dominant one in Equation (6.11) because it increases exponentially with N, while the other factors behave as $N^{-0.5}$, as shown below. The remaining factors in Equation (6.11) are called the *Ockham factor* by Gull [98]:

$$\mathcal{O}_j = p(\boldsymbol{\theta}^\star|\mathcal{C}_j)(2\pi)^{\frac{N_j}{2}}|\mathcal{H}_j(\boldsymbol{\theta}^\star)|^{-\frac{1}{2}} \tag{6.14}$$

This represents a penalty against complicated parameterization [98, 164], as demonstrated in the following discussion.

Next, it is attempted to show that the Ockham factor decreases exponentially with the number of uncertain parameters in the model class. For this purpose, consider an alternative

www.EngineeringBooksPdf.com

expression for it, derived as follows. It is known that for a large number N of data points in $\mathcal{D}$, the updated PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)$ is well approximated by a Gaussian PDF with mean $\boldsymbol{\theta}^\star$ and covariance matrix given by the inverse of the Hessian matrix $\mathcal{H}_j(\boldsymbol{\theta}^\star)$. The principal posterior variances for $\boldsymbol{\theta}$, denoted by σ_l^2 with $l = 1, 2, \ldots, N_j$, are therefore the inverse of the eigenvalues of this Hessian matrix [19]. The determinant factor $|\mathcal{H}_j(\boldsymbol{\theta}^\star)|^{-0.5}$ in the Ockham factor can therefore be expressed as the product of all the σ_l for $l = 1, 2, \ldots, N_j$. Assume that the prior PDF $p(\boldsymbol{\theta}|\mathcal{C}_j)$ is Gaussian with mean $\boldsymbol{\theta}^\eta$ (most probable value *a priori*) and a diagonal covariance matrix with variances $\left(\sigma_l^\eta\right)^2$ with $l = 1, 2, \ldots, N_j$. The logarithm of the Ockham factor for the model class $\mathcal{C}_j$ can therefore be expressed as:

$$\begin{aligned} \ln \mathcal{O}_j &= \ln \left[p(\boldsymbol{\theta}^\star|\mathcal{C}_j)(2\pi)^{\frac{N_j}{2}} |\mathcal{H}_j(\boldsymbol{\theta}^\star)|^{-\frac{1}{2}} \right] \\ &\approx -\sum_{l=1}^{N_j} \ln \frac{\sigma_l^\eta}{\sigma_l} - \frac{1}{2} \sum_{l=1}^{N_j} \left(\frac{\theta_l^\star - \theta_l^\eta}{\sigma_l^\eta} \right)^2 \end{aligned} \tag{6.15}$$

Since the prior variances will always be greater than the posterior variances if the data provide any information about the model parameters in the model class $\mathcal{C}_j$, all the terms in the first summation in Equation (6.15) will be positive and so will the terms in the second summation unless the posterior most probable value $\theta_l^\star$ just happens to coincide with the prior most probable value θ_l^η. Thus, one might expect that the log-Ockham factor $\ln \mathcal{O}_j$ will decrease if the number of parameters N_j for the model class $\mathcal{C}_j$ is increased. This expectation is confirmed by noting that the posterior variances are inversely proportional to the number of data points N in $\mathcal{D}$, so the dependence of the log-Ockham factor is:

$$\ln \mathcal{O}_j = -\frac{1}{2} N_j \ln N + R_j \tag{6.16}$$

where the remainder R_j depends primarily on the choice of prior PDF and is of order 1 for large N. It is not difficult to show that this result holds for even more general forms of the prior PDF than the Gaussian PDF used here.

In the special case if the likelihood function does not depend on a particular parameter θ_L, the posterior uncertainty of this parameter will be controlled by its prior distribution. This parameter is independent of other parameters for the prior distribution assumed so it represents also one principle random variable and its posterior variance is equal to its prior variance $(\sigma_L^\eta)^2$. By Equation (6.15), the terms associated with θ_L are zero:

$$-\ln \frac{\sigma_L^\eta}{\sigma_L} - \frac{1}{2} \left(\frac{\theta_L^\star - \theta_L^\eta}{\sigma_L^\eta} \right)^2 = 0$$

since the updated value of this parameter is equal to the most probable value in the prior distribution: $\theta_L^\star = \theta_L^\eta$. In this case, the inclusion of this 'dummy' parameter in the model class has no effect on the Ockham factor or the likelihood function so there is no penalty for this extra parameter in hampering the usage of this model class. If this model class is used for future prediction, the extra 'dummy' parameter does not affect the model output even though it is 'model-unidentifiable.' Therefore, it is reasonable that the Bayesian approach penalizes complicated model classes according to the level of parametric uncertainty reduction but not just the number of parameters.

www.EngineeringBooksPdf.com

On the other hand, it follows from the Bayes' Theorem that the exact expression for the evidence is given by:

$$p(\mathcal{D}|\mathcal{C}_j) = \frac{p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j)p(\boldsymbol{\theta}^\star|\mathcal{C}_j)}{p(\boldsymbol{\theta}^\star|\mathcal{D}, \mathcal{C}_j)} \tag{6.17}$$

A comparison of this equation and Equation (6.11) shows that the Ockham factor is approximately equal to the ratio $p(\boldsymbol{\theta}^\star|\mathcal{C}_j)/p(\boldsymbol{\theta}^\star|\mathcal{D}, \mathcal{C}_j)$ which is always less than unity if the data provide any information about the model parameters in the model class $\mathcal{C}_j$. Indeed, for large N, the negative logarithm of this ratio is an asymptotic approximation of the information about $\boldsymbol{\theta}$ provided by data $\mathcal{D}$ [147]. Therefore, the log-Ockham factor $\ln \mathcal{O}_j$ removes the amount of information about $\boldsymbol{\theta}$ provided by $\mathcal{D}$ from the log-likelihood $\ln\ p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j)$ to give the log-evidence, $\ln\ p(\mathcal{D}|\mathcal{C}_j)$.

The Ockham factor can also be interpreted as a measure of robustness of a model class. If the updated PDF for the model parameters for a given model class is very 'peaked,' then the ratio $p(\boldsymbol{\theta}^\star|\mathcal{C}_j)/p(\boldsymbol{\theta}^\star|\mathcal{D}, \mathcal{C}_j)$, and so the Ockham factor, is very small. However, a narrow peak implies that response predictions using this model class will depend too sensitively on the optimal parameters in $\boldsymbol{\theta}^\star$. Small errors in the parameter estimation will lead to large errors in the response predictions. Therefore, a class of models with a small Ockham factor will not be robust to measurement noise during parameter estimation, i.e., during selection of the optimal model within the class.

To summarize, in the Bayesian approach to model selection, the model classes are ranked according to $p(\mathcal{D}|\mathcal{C}_j)P(\mathcal{C}_j|\mathcal{U})$ for $j = 1, 2, \ldots, N_C$, where the most plausible class of models representing the system is the one which gives the largest value of this quantity. The evidence $p(\mathcal{D}|\mathcal{C}_j)$ can be calculated for each class of models using Equation (6.11) where the likelihood $p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j)$ is evaluated using the methods presented in Chapters 2–5. The prior distribution $P(\mathcal{C}_j|\mathcal{U})$ over all the model classes $\mathcal{C}_j$, $j = 1, 2, \ldots, N_C$, can be used for other concerns, such as computational demand. However, it is out of the scope of this book and uniform prior plausibilities are chosen, leaving the Ockham factor alone to penalize the model classes.

Although the Bayesian method can be used to avoid a statistical trap in selecting a model class, the most important issue is still on the fundamental understanding of the underlying system or phenomenon so that good model class candidates can be constructed. Otherwise, satisfactory identification results can never be anticipated among the poor candidates.

In practice, the likelihood value, Ockham factor and evidence may associate with a large order for large N. Computational problems (i.e., giving either zero or infinity) may be encountered for direct calculation and/or normalization of the plausibilities. To resolve this problem, one first calculates the log-likelihood and the log-Ockham factor and hence the log-evidence, denoted by $\ln\ p(\mathcal{D}|\mathcal{C}_1), \ln\ p(\mathcal{D}|\mathcal{C}_2), \ldots, \ln\ p(\mathcal{D}|\mathcal{C}_{N_C})$. Instead of taking the exponential of the log-evidence and then normalizing the plausibility, the maximum log-evidence is subtracted from the log-evidence of each model class and then taking the exponential of this array. This operation does not affect the relative plausibility between different model classes. Finally, the plausibility of a model class can be obtained by normalizing this array:

$$P(\mathcal{C}_j|\mathcal{D}, \mathcal{U}) = \frac{\exp\left(\ln\ p(\mathcal{D}|\mathcal{C}_j) - M\right)}{\sum_{j=1}^{N_C} \exp\left(\ln\ p(\mathcal{D}|\mathcal{C}_j) - M\right)}, \quad j = 1, 2, \ldots, N_C \tag{6.18}$$

where $M = \max_j \ln\ p(\mathcal{D}|\mathcal{C}_j)$ is the maximum log-evidence.

www.EngineeringBooksPdf.com

6.2.1.1 Comparison with Akaike's Approaches

The Akaike information criterion (AIC) [2] states that the best model class among the $\mathcal{C}_j$, $j = 1, 2, \ldots, N_C$, is chosen by maximizing an objective function $\text{AIC}(\mathcal{C}_j|\mathcal{D})$ over j that is defined by:

$$\text{AIC}(\mathcal{C}_j|\mathcal{D}) = \ln\ p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j) - N_j \tag{6.19}$$

where the maximum log-likelihood value is of order N, which is the number of data points in $\mathcal{D}$, while the penalty term is taken to be N_j, the number of adjustable parameters in the model class $\mathcal{C}_j$. Akaike actually stated his criterion as minimizing $-2\ln\ p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j) + 2N_j$ but the equivalent form in Equation (6.19) is more appropriate here. When the number of data points is large, the first term will dominate. Akaike (1976) and Schwarz (1978) later developed independently another version of the objective function, referred to as the Bayesian information criterion (BIC), that is defined by [3, 232]:

$$\text{BIC}(\mathcal{C}_j|\mathcal{D}) = \ln\ p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j) - \frac{1}{2}N_j \ln\ N \tag{6.20}$$

where now the penalty term increases with the number of data points N. Similarly, the original version was to minimize $-2\ln p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j) + N_j \ln\ N$.

The BIC can be compared directly with the logarithm of the evidence from Equation (6.11):

$$\ln\ p(\mathcal{D}|\mathcal{C}_j) \approx \ln\ p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j) + \ln\ \mathcal{O}_j \tag{6.21}$$

where the log-Ockham factor $\ln\ \mathcal{O}_j$ is given by Equation (6.16). This shows that for large N, the BIC agrees with the leading order terms in the log-evidence and so in this case it is equivalent to the Bayesian approach using equal prior plausibilities for all of the $P(\mathcal{C}_j|\mathcal{U})$. However, in practice, the Bayesian model class selection is especially useful when N is not large so the selection of model class is difficult by the user's judgement. Therefore, the BIC cannot replace the Bayesian model class selection method if N is not sufficiently large since the residual term has an important contribution. For example, consider a model class $\mathcal{C}_1$ that under-estimates the response of a system by no less than 1%. Another model class $\mathcal{C}_2$ is constructed by duplicating the output from $\mathcal{C}_1$ but having an extra term and an extra parameter θ_{ex}:

$$Q_2 = Q_1(1 + 0.001U[|\theta_{ex}|])$$

where $U(.)$ is the Heaviside unit step function that is unity if the argument is positive and zero otherwise. Therefore, the extra factor in $\mathcal{C}_2$ rectifies the under-estimation of $\mathcal{C}_1$ although the improvement is small. If this contributes to an increment of 0.1 of the maximum log-likelihood value, both the AIC and BIC will prefer $\mathcal{C}_1$ as the improvement of the maximum log-likelihood value in $\mathcal{C}_2$ is not sufficiently large to compensate the penalty induced by the extra parameter (1.0 for AIC and $0.5\ln\ N$ for BIC). However, if the selected model class is used for future prediction, $\mathcal{C}_2$ should be preferred since it improves the prediction even though the improvement is small and the extra parameter is unidentifiable. The unidentifiability of θ_{ex} implies that the model output is insensitive to θ_{ex} so it is robust to the perturbation of this parameter. If the Bayesian model class selection method is used, $\mathcal{C}_2$ is preferred. The improvement of data fitting by this extra factor increases the maximum likelihood value and also reduces the posterior

www.EngineeringBooksPdf.com

uncertainty of the parameters (except for θ_{ex}). Note that there is no information gained from the measurement to trim down the uncertainty of the extra parameter from its prior PDF. On the other hand, the increment of the maximum likelihood value overcomes the reduction of the Ockham factor since they are proportional to $\left(\sigma_\epsilon^\star\right)^{-N}$ and $\left(\sigma_\epsilon^\star\right)^{N_j-1}$, respectively. Here, N_j is the number of uncertain parameters including θ_{ex}, and $N > N_j$ in a normal situation.

6.2.2 General Case

For the general case where the posterior PDF may not be approximated by Gaussian distribution, the asymptotic expansion in Equation (6.11) is not valid. Since the posterior PDF is normalized, the log-evidence can be rewritten as [55, 56]:

$$\begin{aligned} \ln\ p(\mathcal{D}|\mathcal{C}_j) &= \left[\ln\ p(\mathcal{D}|\mathcal{C}_j)\right] \int_{\boldsymbol{\Theta}_j} p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\Theta}_j} \left[\ln\ p(\mathcal{D}|\mathcal{C}_j)\right] p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} \end{aligned} \tag{6.22}$$

By using the Bayes' theorem, the log-evidence can be expanded:

$$\ln\ p(\mathcal{D}|\mathcal{C}_j) = \ln\ p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}_j) - \ln\ \frac{p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)}{p(\boldsymbol{\theta}|\mathcal{C}_j)} \tag{6.23}$$

Then, Equation (6.22) can be expressed as the difference of two terms [55, 56]:

$$\ln\ p(\mathcal{D}|\mathcal{C}_j) = \int_{\boldsymbol{\Theta}_j} [\ln\ p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}_j)] p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} - \int_{\boldsymbol{\Theta}_j} \left[\ln \frac{p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)}{p(\boldsymbol{\theta}|\mathcal{C}_j)}\right] p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} \tag{6.24}$$

The first term is a measure of the average log-goodness of fit of the model class $\mathcal{C}_j$. It accounts for the log-goodness of fit for different combinations of the parameters, weighted by the posterior PDF, instead of the optimal parameters alone. An ideal model class should fit the data well even with a reasonably small perturbation of the parameters from their optimal values. In the special case if the likelihood function is of the Gaussian type and the prior PDF is relatively flat, the posterior PDF is approximately Gaussian and the log-likelihood function takes the following form:

$$\ln\ p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}_j) = \ln\ p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \left[\mathcal{H}_j(\boldsymbol{\theta}^\star) - \mathcal{H}_j^\eta(\boldsymbol{\theta}^\star)\right] (\boldsymbol{\theta} - \boldsymbol{\theta}^\star) \tag{6.25}$$

where $\mathcal{H}_j^\eta(\boldsymbol{\theta}^\star)$ is the Hessian matrix of the negative logarithm of the prior distribution evaluated at the posterior most probable parameters, but not the peak of the prior. The second term

www.EngineeringBooksPdf.com

vanishes when $\boldsymbol{\theta} = \boldsymbol{\theta}^\star$. Then, the first term in Equation (6.24) is given by:

$$\begin{aligned}\int_{\boldsymbol{\Theta}_j} [\ln\ p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}_j)] p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j) \mathrm{d}\boldsymbol{\theta} \\ = \ln\ p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j) - \frac{1}{2}\int_{\boldsymbol{\Theta}_j} (\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \mathcal{H}_j(\boldsymbol{\theta}^\star)(\boldsymbol{\theta} - \boldsymbol{\theta}^\star) p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j) \mathrm{d}\boldsymbol{\theta} \\ + \frac{1}{2}\int_{\boldsymbol{\Theta}_j} (\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \mathcal{H}_j^\eta(\boldsymbol{\theta}^\star)(\boldsymbol{\theta} - \boldsymbol{\theta}^\star) p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j) \mathrm{d}\boldsymbol{\theta}\end{aligned} \tag{6.26}$$

In order to evaluate the integrals on the right side, a transformation $\mathbf{V} : \boldsymbol{\psi} \to \boldsymbol{\theta}$ is introduced:

$$\boldsymbol{\theta} - \boldsymbol{\theta}^\star = \mathbf{V}\boldsymbol{\psi} \tag{6.27}$$

where $\mathbf{V} \in \mathbb{R}^{N_j \times N_j}$ is a squared matrix that diagonalizes the Hessian matrix:

$$\mathbf{V}^{-1}\mathcal{H}_j(\boldsymbol{\theta}^\star)\mathbf{V} = \mathbf{D} \tag{6.28}$$

with a diagonal matrix $\mathbf{D}$. Since the Hessian matrix is symmetric, a solution of Equation (6.28) exists and a rigid rotation matrix $\mathbf{V}$ can be selected such that $\mathbf{V}^{-1} = \mathbf{V}^T$ and its determinant is $|\mathbf{V}| = 1$. By Equation (6.27), it can be shown that $\boldsymbol{\psi}$ is a zero-mean Gaussian with the covariance matrix $\mathbf{V}^{-1}\mathcal{H}(\boldsymbol{\theta}^\star)^{-1}\mathbf{V} = \mathbf{D}^{-1}$, which is diagonal. In other words, different components ψ_l and $\psi_{l'}$, $l \neq l'$, are independent Gaussian random variables with variances $1/D^{(l,l)}$ and $1/D^{(l',l')}$, respectively.

Since $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} = p(\boldsymbol{\psi}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\psi}$, the first integral on the right side of Equation (6.26) can be evaluated:

$$\begin{aligned}\int_{\boldsymbol{\Theta}_j} (\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \mathcal{H}_j(\boldsymbol{\theta}^\star)(\boldsymbol{\theta} - \boldsymbol{\theta}^\star) p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} &= \int_{\boldsymbol{\Psi}_j} \boldsymbol{\psi}^T \mathbf{V}^T \mathcal{H}_j(\boldsymbol{\theta}^\star)\mathbf{V}\boldsymbol{\psi}\, p(\boldsymbol{\psi}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\psi} \\ &= \int_{\boldsymbol{\Psi}_j} \boldsymbol{\psi}^T \mathbf{D}\boldsymbol{\psi}\, p(\boldsymbol{\psi}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\psi} \\ &= \sum_{l=1}^{N_j} D^{(l,l)} \int_{\boldsymbol{\Psi}_j} \psi_l^2\, p(\boldsymbol{\psi}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\psi} \\ &= \sum_{l=1}^{N_j} D^{(l,l)}/D^{(l,l)} \\ &= N_j\end{aligned} \tag{6.29}$$

Therefore, the first term on right hand side of Equation (6.24) is given by:

$$\int_{\boldsymbol{\Theta}_j} [\ln\ p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}_j)] p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} = \ln\ p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j) - \frac{N_j - I^\eta}{2} \tag{6.30}$$

www.EngineeringBooksPdf.com

where the quantity I^η is defined as:

$$I^\eta \equiv \int_{\boldsymbol{\Theta}_j} (\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \mathcal{H}_j^\eta(\boldsymbol{\theta}^\star)(\boldsymbol{\theta} - \boldsymbol{\theta}^\star) p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} \tag{6.31}$$

Note that $I^\eta < N_j$ if the data contain any information for the model parameters. Otherwise, $I^\eta = N_j$.

The second term in Equation (6.24) is the relative entropy between the posterior and prior PDFs [148]. The relative entropy is a measure of the information gained about the parameters given the data $\mathcal{D}$. The log-evidence is equal to the average log-goodness of fit of the model class, penalized by the measure of the information gained from the data. Small uncertainty (e.g., small standard deviation) of the model parameters is the consequence that the model output is highly sensitive to the model parameters and this is not necessarily a good model class for future modeling/prediction. This term depends highly on the choice of the prior distribution so there is no general closed-form solution for this integral. However, in the special case if both the prior and posterior PDFs are Gaussian, it can be simplified as follows:

$$\begin{aligned}
&\int_{\boldsymbol{\Theta}_j} \left[\ln \frac{p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)}{p(\boldsymbol{\theta}|\mathcal{C}_j)}\right] p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} \\
&\quad = \frac{1}{2}\ln\frac{\mathcal{H}_j(\boldsymbol{\theta}^\star)}{\mathcal{H}_j^\eta(\boldsymbol{\theta}^\eta)} - \frac{1}{2}\int_{\boldsymbol{\Theta}_j} \left(\boldsymbol{\theta} - \boldsymbol{\theta}^\star\right)^T \mathcal{H}_j(\boldsymbol{\theta}^\star)\left(\boldsymbol{\theta} - \boldsymbol{\theta}^\star\right) p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} \\
&\qquad + \frac{1}{2}\int_{\boldsymbol{\Theta}_j} \left(\boldsymbol{\theta} - \boldsymbol{\theta}^\eta\right)^T \mathcal{H}_j^\eta(\boldsymbol{\theta}^\star)\left(\boldsymbol{\theta} - \boldsymbol{\theta}^\eta\right) p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} \\
&\quad = \frac{1}{2}\ln\frac{\mathcal{H}_j(\boldsymbol{\theta}^\star)}{\mathcal{H}_j^\eta(\boldsymbol{\theta}^\eta)} - \frac{N_j}{2} + \frac{1}{2}\int_{\boldsymbol{\Theta}_j} \left(\boldsymbol{\theta} - \boldsymbol{\theta}^\eta\right)^T \mathcal{H}_j^\eta(\boldsymbol{\theta}^\star)\left(\boldsymbol{\theta} - \boldsymbol{\theta}^\eta\right) p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta}
\end{aligned} \tag{6.32}$$

where $\boldsymbol{\theta}^\eta$ is the prior most probable parameter vector. Note that the Hessian matrices do not depend on the point for evaluation for Gaussian distributions so $\mathcal{H}_j(\boldsymbol{\theta}^\star) = \mathcal{H}_j$ and $\mathcal{H}_j^\eta(\boldsymbol{\theta}^\eta) = \mathcal{H}_j^\eta(\boldsymbol{\theta}^\star) = \mathcal{H}_j^\eta$. The last integral can be simplified as follows:

$$\begin{aligned}
&\int_{\boldsymbol{\Theta}_j} \left(\boldsymbol{\theta} - \boldsymbol{\theta}^\eta\right)^T \left(\mathcal{H}_j^\eta\right)\left(\boldsymbol{\theta} - \boldsymbol{\theta}^\eta\right) p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} \\
&\quad = \int_{\boldsymbol{\Theta}_j} \left(\boldsymbol{\theta} - \boldsymbol{\theta}^\star\right)^T \left(\mathcal{H}_j^\eta\right)\left(\boldsymbol{\theta} - \boldsymbol{\theta}^\star\right) p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} \\
&\qquad + 2\int_{\boldsymbol{\Theta}_j} \left(\boldsymbol{\theta} - \boldsymbol{\theta}^\star\right)^T \left(\mathcal{H}_j^\eta\right)\left(\boldsymbol{\theta}^\star - \boldsymbol{\theta}^\eta\right) p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} \\
&\qquad + \int_{\boldsymbol{\Theta}_j} \left(\boldsymbol{\theta}^\star - \boldsymbol{\theta}^\eta\right)^T \left(\mathcal{H}_j^\eta\right)\left(\boldsymbol{\theta}^\star - \boldsymbol{\theta}^\eta\right) p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)\mathrm{d}\boldsymbol{\theta} \\
&\quad = I^\eta + \left(\boldsymbol{\theta}^\star - \boldsymbol{\theta}^\eta\right)^T \left(\mathcal{H}_j^\eta\right)\left(\boldsymbol{\theta}^\star - \boldsymbol{\theta}^\eta\right)
\end{aligned} \tag{6.33}$$

www.EngineeringBooksPdf.com

Therefore, the relative entropy is given by:

$$\int_{\boldsymbol{\Theta}_j} \left[\ln \frac{p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)}{p(\boldsymbol{\theta}|\mathcal{C}_j)} \right] p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j) \mathrm{d}\boldsymbol{\theta} = \frac{1}{2} \ln \frac{\mathcal{H}_j}{\mathcal{H}_j^\eta} - \frac{N_j - I^\eta}{2} + \frac{1}{2} \left(\boldsymbol{\theta}^\star - \boldsymbol{\theta}^\eta\right)^T \left(\mathcal{H}_j^\eta\right) \left(\boldsymbol{\theta}^\star - \boldsymbol{\theta}^\eta\right) \tag{6.34}$$

By comparing Equations (6.30) and (6.34), the term $(N_j - I^\eta)/2$ appears in both and will be cancelled. The penalty consists of two parts. The first one is represented by $0.5 \ \ln \mathcal{H}_j/\mathcal{H}_j^\eta$ and it is a measure of the reduction of the principle variances of the parameters due to the updating process. For large N, this term is proportional to $\ln \ N$ since the Hessian matrix is proportional to N. The second part is $0.5 \left(\boldsymbol{\theta}^\star - \boldsymbol{\theta}^\eta\right)^T \left(\mathcal{H}_j^\eta\right) \left(\boldsymbol{\theta}^\star - \boldsymbol{\theta}^\eta\right)$ and represents the distance between the updated parameters and the nominal values. The expected value of this term is $N_j/2$ and does not have a trend with N.

6.2.3 *Computational Issues: Transitional Markov Chain Monte Carlo Method*

For general updated PDFs that may have complex topology (e.g., a model unidentifiable case), asymptotic expansion is not applicable and numerical simulation is necessary to evaluate the evidence integral. With a large number of data points, the problem may not necessarily be globally identifiable but the posterior PDFs are often concentrated in a region with small volume. Therefore, direct application of the Markov Chain Monte Carlo (MCMC) simulation algorithm encounters difficulties of having too many repeating samples. In order to resolve this problem, the transitional Markov Chain Monte Carlo simulation algorithm [54] was developed by extending the concept of the adaptive Markov Chain Monte Carlo simulation procedure introduced in Chapter 2 [17]. This is a method for sampling the posterior PDF of a model class $\mathcal{C}$ in an adaptive manner. First, sampling is performed for a sequence of un-normalized kernel sampling densities, $\mathcal{P}^{(s)}$, $s = 0, 1, \ldots, s_0$, given by:

$$\mathcal{P}^{(s)}(\boldsymbol{\theta}) = p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})^{\beta_s} p(\boldsymbol{\theta}|\mathcal{C}) \tag{6.35}$$

where the tempering parameter β_s increases monotonically with s such that the starting value is $\beta_0 = 0$ and the target value is $\beta_{s_0} = 1$. It is not necessary for normalization since the MCMC algorithm requires only the relative probability density between two different parameter vectors. Therefore, the intermediate PDFs evolve gradually from the prior PDF to the posterior PDF. Samples of the model parameters are generated according to each intermediate PDF, $\mathcal{P}^{(s)}$, $s = 0, 1, \ldots, s_0$. The algorithm proceeds as follows.

The resampling weighting for each sample is the ratio of the intermediate PDFs for the sth and $(s-1)$th levels, evaluated at $\boldsymbol{\theta}_n^{(s-1)}$:

$$w\left(\boldsymbol{\theta}_n^{(s-1)}\right) = \frac{p(\mathcal{D}|\boldsymbol{\theta}_n^{(s-1)}, \mathcal{C})^{\beta_s} p(\boldsymbol{\theta}_n^{(s-1)}|\mathcal{C})}{p(\mathcal{D}|\boldsymbol{\theta}_n^{(s-1)}, \mathcal{C})^{\beta_{s-1}} p(\boldsymbol{\theta}_n^{(s-1)}|\mathcal{C})} = p(\mathcal{D}|\boldsymbol{\theta}_n^{(s-1)}, \mathcal{C})^{\beta_s - \beta_{s-1}} \tag{6.36}$$

www.EngineeringBooksPdf.com

For large N, the samples $\boldsymbol{\theta}_n^{(s-1)}$, $n = 1, 2, \ldots, N$, are distributed according to the intermediate PDF for the $(s-1)$th level that is given by normalizing $\mathcal{P}^{(s-1)}$. Use $\hat{W}_s$ to denote the expectation of the weighting $w(\boldsymbol{\theta}^{(s-1)})$ and it can be estimated by the sample average of $w(\boldsymbol{\theta}_n^{(s-1)})$, $n = 1, 2, \ldots, N$:

$$\frac{\int_{\boldsymbol{\Theta}} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})^{\beta_s} p(\boldsymbol{\theta}|\mathcal{C})\mathrm{d}\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})^{\beta_{s-1}} p(\boldsymbol{\theta}|\mathcal{C})\mathrm{d}\boldsymbol{\theta}} \approx \frac{1}{N}\sum_{n=1}^{N} w\left(\boldsymbol{\theta}_n^{(s-1)}\right) \equiv \hat{W}_s \tag{6.37}$$

Finally, the evidence $p(\mathcal{D}|\mathcal{C})$ can be estimated by the product of $\hat{W}_s$, $s = 1, 2, \ldots, s_0$:

$$\begin{aligned} p(\mathcal{D}|\mathcal{C}) &= \frac{\int_{\boldsymbol{\Theta}} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}) p(\boldsymbol{\theta}|\mathcal{C})\mathrm{d}\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}|\mathcal{C})\mathrm{d}\boldsymbol{\theta}} \\ &= \frac{\int_{\boldsymbol{\Theta}} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})^{\beta_{s_0}} p(\boldsymbol{\theta}|\mathcal{C})\mathrm{d}\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})^{\beta_0} p(\boldsymbol{\theta}|\mathcal{C})\mathrm{d}\boldsymbol{\theta}} \\ &= \frac{\int_{\boldsymbol{\Theta}} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})^{\beta_1} p(\boldsymbol{\theta}|\mathcal{C})\mathrm{d}\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})^{\beta_0} p(\boldsymbol{\theta}|\mathcal{C})\mathrm{d}\boldsymbol{\theta}} \frac{\int_{\boldsymbol{\Theta}} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})^{\beta_2} p(\boldsymbol{\theta}|\mathcal{C})\mathrm{d}\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})^{\beta_1} p(\boldsymbol{\theta}|\mathcal{C})\mathrm{d}\boldsymbol{\theta}} \cdots \frac{\int_{\boldsymbol{\Theta}} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})^{\beta_{s_0}} p(\boldsymbol{\theta}|\mathcal{C})\mathrm{d}\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C})^{\beta_{s_0-1}} p(\boldsymbol{\theta}|\mathcal{C})\mathrm{d}\boldsymbol{\theta}} \\ &\approx \prod_{s=1}^{s_0} \hat{W}_s \end{aligned} \tag{6.38}$$

6.3 Model Class Selection for Regression Problems

In Chapter 2, Section 2.4, parametric identification was introduced for linear and nonlinear regression problems. In this section, the Bayesian model class selection is applied to these problems. In order to smooth the presentation, some of the equations from Section 2.4 are repeated in this section.

6.3.1 Linear Regression Problems

A general linear regression problem can be represented by the following relationship:

$$Q(\mathbf{x}; \mathbf{b}, \mathcal{C}_j) = \sum_{l=1}^{N_b} b_l x_l \tag{6.39}$$

where N_b is the total number of uncertain coefficients, b_ls. The measurement of Q is denoted by y and is modeled as:

$$y = Q(\mathbf{x}; \mathbf{b}, \mathcal{C}_j) + \epsilon \tag{6.40}$$

where ϵ is a zero-mean Gaussian random variable with variance σ_ϵ^2 and is used to represent the measurement noise and modeling error. The uncertain parameters in $\boldsymbol{\theta} = [\mathbf{b}^T, \sigma_\epsilon^2]^T$ include

www.EngineeringBooksPdf.com

the coefficients b_ls and the prediction-error variance σ_ϵ^2 so:

$$N_j = N_b + 1 \tag{6.41}$$

The variables $x_1, x_2, \ldots, x_{N_b}$ in $\mathbf{x}$ are the measured variables in establishing the regression relationship.

The data $\mathcal{D}$ include the measurement of $\mathbf{x}$ and the corresponding values of y. By assuming that the prediction errors in different records are statistically independent, the likelihood function is obtained:

$$p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}_j) = (2\pi)^{-\frac{N}{2}} \sigma_\epsilon^{-N} \exp\left[-\frac{N}{2\sigma_\epsilon^2} J_g(\mathbf{b}; \mathcal{D}, \mathcal{C}_j)\right] \tag{6.42}$$

where N is the total number of measured records. The goodness-of-fit function $J_g(\mathbf{b}; \mathcal{D}, \mathcal{C}_j)$ represents the degree of data fitting and is given by:

$$J_g(\mathbf{b}; \mathcal{D}, \mathcal{C}_j) = \frac{1}{N} \sum_{n=1}^{N} \left[y(n) - \sum_{l=1}^{N_b} b_l x_l(n) \right]^2 \tag{6.43}$$

A smaller value of this function implies better fitting to the data. Two special but popular choices of prior distributions of the uncertain parameters are now discussed.

6.3.1.1 Independent Uniform Prior

With a uniform prior PDF of the coefficients, the optimal coefficient vector $\mathbf{b}^\star$ can be obtained by minimizing $J_g(\mathbf{b}; \mathcal{D}, \mathcal{C}_j)$ if the range of the uniform distribution is sufficiently large. This can be done by solving the linear algebraic equation $\partial J_g(\mathbf{b}; \mathcal{D}, \mathcal{C}_j)/\partial \mathbf{b} = \mathbf{0}$, and the updated coefficient vector is readily obtained:

$$\mathbf{b}^\star = \mathbf{A}^{-1} \begin{bmatrix} \frac{1}{N} \sum_{n=1}^{N} x_1(n) y(n) \\ \frac{1}{N} \sum_{n=1}^{N} x_2(n) y(n) \\ \vdots \\ \frac{1}{N} \sum_{n=1}^{N} x_{N_b}(n) y(n) \end{bmatrix} \tag{6.44}$$

where $\mathbf{A}$ is an $N_b \times N_b$ symmetric matrix given by:

$$\mathbf{A} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}(n) \mathbf{x}(n)^T \tag{6.45}$$

Furthermore, the updated fitting-error variance $\sigma_\epsilon^{2\star}$ can be obtained by maximizing the product of the prior PDF and the likelihood function, which is proportional to the posterior PDF. It is simply given by the minimum 'goodness of fit value:'

$$\sigma_\epsilon^{2\star} = \min_{\mathbf{b}} J_g(\mathbf{b}; \mathcal{D}, \mathcal{C}_j) = J_g(\mathbf{b}^\star; \mathcal{D}, \mathcal{C}_j) \tag{6.46}$$

www.EngineeringBooksPdf.com

Therefore, the maximum likelihood value is:

$$p(\mathcal{D}|\boldsymbol{\theta}^\star, \mathcal{C}_j) = (2\pi\sigma_\epsilon^{2\star})^{-\frac{N}{2}} \exp\left(-\frac{N}{2}\right) \tag{6.47}$$

For large N, the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)$ is approximately Gaussian centered at the optimal parameter $\boldsymbol{\theta}^\star = [\mathbf{b}^{\star T}, \sigma_\epsilon^{2\star}]^T$. Therefore, the uncertainty of the parameter estimates can be represented by the covariance matrix given by $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \mathcal{H}_j(\boldsymbol{\theta}^\star)^{-1}$, where the Hessian matrix $\mathcal{H}_j(\boldsymbol{\theta}^\star)$ is given by:

$$\mathcal{H}_j(\boldsymbol{\theta}^\star) = N \begin{bmatrix} (\sigma_\epsilon^{2\star})^{-1}\mathbf{A} & \mathbf{0}_{N_b\times 1} \\ \mathbf{0}_{1\times N_b} & \frac{1}{2}(\sigma_\epsilon^{2\star})^{-2} \end{bmatrix} \tag{6.48}$$

Finally, by using the asymptotic expansion in Equation (6.11), the evidence $p(\mathcal{D}|\mathcal{C}_j)$ can be approximated by:

$$p(\mathcal{D}|\mathcal{C}_j) \approx p(\boldsymbol{\theta}^\star|\mathcal{C}_j) \exp\left(-\frac{N}{2}\right) \frac{\pi^{\frac{N_j-N}{2}} (\sqrt{2}\sigma_\epsilon^\star)^{N_j+1-N}}{\sqrt{N^{N_j}|\mathbf{A}|}} \tag{6.49}$$

It has to be noted that the above formula is correct if the uniform prior distribution has a sufficiently large range to cover all the significant region of the likelihood function. In the special case if the number of uncertain coefficients is equal to the number of data points, i.e., $N_b = N$, over-fitting is anticipated. In this case, $N_j - N = 1$ as the prediction-error variance is also an uncertain parameter. Since the optimal model in this model class matches with the data perfectly, $\sigma_\epsilon^2 = 0$ and the evidence is zero:

$$p(\mathcal{D}|\mathcal{C}_j) \approx p(\boldsymbol{\theta}^\star|\mathcal{C}_j) \exp\left(-\frac{N}{2}\right) \frac{2\sqrt{\pi}\sigma_\epsilon^{2\star}}{\sqrt{N^{N+1}|\mathbf{A}|}} = 0$$

By using Equation (6.14) with Equation (6.48), the Ockham factor is given by:

$$\mathcal{O}_j = p(\boldsymbol{\theta}^\star|\mathcal{C}_j)|\mathbf{A}|^{-\frac{1}{2}} \left(\frac{\pi}{N}\right)^{\frac{N_j}{2}} \left(\sqrt{2}\sigma_\epsilon^\star\right)^{N_j+1} \tag{6.50}$$

6.3.1.2 Gaussian Prior for the Coefficients and Inverse Gamma Distribution for the Prediction-error Parameter

Another popular choice of the prior PDF for the uncertain parameters is a Gaussian distribution for the coefficients and is discussed in this section. First, the prior PDF is assumed *separable*:

$$p(\boldsymbol{\theta}|\mathcal{C}_j) = p\left(\mathbf{b}, \sigma_\epsilon^2|\mathcal{C}_j\right) = p(\mathbf{b}|\mathcal{C}_j)p\left(\sigma_\epsilon^2|\mathcal{C}_j\right) \tag{6.51}$$

For the uncertain coefficients in $\mathbf{b}$, Gaussian prior PDFs are taken with mean $\mathbf{b}^\eta$ and covariance matrix $\mathbf{V}$:

$$p(\mathbf{b}|\mathcal{C}_j) = (2\pi)^{-\frac{N_b}{2}} |\mathbf{V}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{b}-\mathbf{b}^\eta)^T\mathbf{V}^{-1}(\mathbf{b}-\mathbf{b}^\eta)\right] \tag{6.52}$$

www.EngineeringBooksPdf.com

On the other hand, the prior distribution for prediction-error variance is taken to be the conjugate prior and is the inverse Gamma distribution:

$$p(\sigma_\epsilon^2|\mathcal{C}_j) = \frac{\beta_0^{\alpha_0}}{\sigma_\epsilon^{2(\alpha_0+1)}\Gamma(\alpha_0)} \exp\left(-\frac{\beta_0}{\sigma_\epsilon^2}\right) \tag{6.53}$$

Then, the product of the prior PDF and likelihood function is given by:

$$p(\boldsymbol{\theta}|\mathcal{C}_j)p(\mathcal{D}|\boldsymbol{\theta},\mathcal{C}_j) = (2\pi)^{-\frac{N_b+N}{2}}|\mathbf{V}|^{-\frac{1}{2}}\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\sigma_\epsilon^{-2(\alpha_0+1)-N}\exp\left(-\frac{\beta_0}{\sigma_\epsilon^2}\right)$$
$$\times \exp\left[-\frac{1}{2}(\mathbf{b}-\mathbf{b}^\eta)^T\mathbf{V}^{-1}(\mathbf{b}-\mathbf{b}^\eta) - \frac{1}{2\sigma_\epsilon^2}\sum_{n=1}^{N}\left(y(n)-\sum_{l=1}^{N_b}b_l x_l(n)\right)^2\right] \tag{6.54}$$

Given any positive value of σ_ϵ^2, the conditional optimal coefficient vector $\mathbf{b}^\star$ can be obtained by minimizing the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{C}_j)$, or equivalently $p(\boldsymbol{\theta}|\mathcal{C}_j)p(\mathcal{D}|\boldsymbol{\theta},\mathcal{C}_j)$ conditional on the value of σ_ϵ^2. This can be done by solving the linear algebraic equation $\partial p(\boldsymbol{\theta}|\mathcal{C}_j)p(\mathcal{D}|\boldsymbol{\theta},\mathcal{C}_j)/\partial\mathbf{b} = \mathbf{0}$ and the solution is given by:

$$\mathbf{b}^\star\left(\sigma_\epsilon^2\right) = \left(N\sigma_\epsilon^{-2}\mathbf{A}+\mathbf{V}^{-1}\right)^{-1}\left(N\sigma_\epsilon^{-2}\begin{bmatrix}\frac{1}{N}\sum_{n=1}^{N}x_1(n)y(n)\\ \frac{1}{N}\sum_{n=1}^{N}x_2(n)y(n)\\ \vdots \\ \frac{1}{N}\sum_{n=1}^{N}x_{N_b}(n)y(n)\end{bmatrix}+\mathbf{V}^{-1}\mathbf{b}^\eta\right) \tag{6.55}$$

where the matrix $\mathbf{A}$ has the same form as Equation (6.45). It can be easily shown that the matrix $N\sigma_\epsilon^{-2}\mathbf{A}+\mathbf{V}^{-1}$ is invertible.

For a given coefficient vector $\mathbf{b}$, the conditional optimal prediction-error variance $\sigma_\epsilon^{2\star}$ can be found by solving $\partial p(\boldsymbol{\theta}|\mathcal{C}_j)p(\mathcal{D}|\boldsymbol{\theta},\mathcal{C}_j)/\partial\sigma_\epsilon^2 = 0$ and is readily obtained:

$$\sigma_\epsilon^{2\star}(\mathbf{b}) = \frac{NJ_g(\mathbf{b};\mathcal{D},\mathcal{C}_j)+2\beta_0}{N+2(\alpha_0+1)} \tag{6.56}$$

The availability of the conditional optimal expressions in Equations (6.55) and (6.56) allows one to search for the optimal parameters by the following iterative algorithm. First, take $\sigma_\epsilon^2 = 0$ in Equation (6.55) to compute the conditional optimal values for the uncertain coefficients. Then, by using Equation (6.56), the conditional prediction-error variance can be computed. By using this value of the prediction-error variance, the conditional optimal for the uncertain coefficients can be updated. By repeating the last two steps, the updated uncertain parameters can be obtained until convergence is achieved.

The posterior PDF has some tractable properties that can be used for evaluation of the evidence integral. The conditional PDF $p(\mathbf{b}|\mathcal{D},\mathcal{C}_j,\sigma_\epsilon^2)$ is Gaussian with mean $\mathbf{b}^\star(\sigma_\epsilon^2)$ and covariance matrix $(N\sigma_\epsilon^{-2}\mathbf{A}+\mathbf{V}^{-1})^{-1}$. Furthermore, the conditional PDF $p(\sigma_\epsilon^2|\mathcal{D},\mathcal{C}_j,\mathbf{b})$ follows the inverse Gamma distribution $IG(\alpha,\beta)$, where the shape and scale parameters are given by $\alpha = N/2+\alpha_0$ and $\beta = NJ_g(\mathbf{b};\mathcal{D},\mathcal{C}_j)/2+\beta_0$. Furthermore, for large N, the

www.EngineeringBooksPdf.com

following approximation can be taken: $(N\sigma_\epsilon^{-2}\mathbf{A} + \mathbf{V}^{-1})^{-1} \approx \sigma_\epsilon^2/N\mathbf{A}^{-1}$ and the marginal distribution of the prediction-error variance is approximately the inverse Gamma distribution $IG(\alpha^\star, \beta^\star)$ where the shape and scale parameters are given by:

$$\begin{aligned} \alpha^\star &= \frac{N - N_b}{2} + \alpha_0 \\ \beta^\star &= \frac{N}{2} J_g(\mathbf{b}^\star; \mathcal{D}, \mathcal{C}) + \beta_0 \end{aligned} \tag{6.57}$$

These properties can be used to obtain the closed-form solution of the evidence integral for the linear regression model classes. By using Equation (6.51), the evidence integral in Equation (6.10) can be rewritten as:

$$p(\mathcal{D}|\mathcal{C}_j) = \int_0^\infty \left[\int_B p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}_j) p(\mathbf{b}|\mathcal{C}_j) \mathrm{d}\mathbf{b} \right] p(\sigma_\epsilon^2|\mathcal{C}_j) \mathrm{d}\sigma_\epsilon^2 \tag{6.58}$$

where B is the domain for the formula coefficient vector $\mathbf{b}$. Use $I_1(\sigma_\epsilon^2; \mathcal{C}_j)$ to denote the integral on B:

$$I_1(\sigma_\epsilon^2; \mathcal{C}_j) \equiv \int_B p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}_j) p(\mathbf{b}|\mathcal{C}_j) \mathrm{d}\mathbf{b} \tag{6.59}$$

This integral has a closed-form solution:

$$\begin{aligned} I_1(\sigma_\epsilon^2; \mathcal{C}_j) = (\sqrt{2\pi}\sigma_\epsilon)^{-N} &\left| N\sigma_\epsilon^{-2}\mathbf{A}\mathbf{V} + \mathbf{I}_{N_b} \right|^{-\frac{1}{2}} \\ &\times \exp\left[-\frac{1}{2}(\mathbf{b}^\star - \mathbf{b}^\eta)^T \mathbf{V}^{-1} (\mathbf{b}^\star - \mathbf{b}^\eta) - \frac{N}{2\sigma_\epsilon^2} J_g(\mathbf{b}^\star; \mathcal{D}, \mathcal{C}_j) \right] \end{aligned} \tag{6.60}$$

Then, numerical integration can be applied to evaluate the one dimensional integral of the evidence:

$$p(\mathcal{D}|\mathcal{C}_j) = \int_0^\infty I_1(\sigma_\epsilon^2; \mathcal{C}_j) p(\sigma_\epsilon^2|\mathcal{C}_j) \mathrm{d}\sigma_\epsilon^2 \tag{6.61}$$

However, for large N, the posterior PDF will be dominated by the likelihood function and the approximation $N\sigma_\epsilon^{-2}\mathbf{A}\mathbf{V} + \mathbf{I}_{N_b} \approx N\sigma_\epsilon^{-2}\mathbf{A}\mathbf{V}$ is accurate. Then, $I_1(\sigma_\epsilon^2; \mathcal{C}_j)$ can be approximated by:

$$I_1(\sigma_\epsilon^2; \mathcal{C}_j) \approx \frac{\sigma_\epsilon^{-N+N_b}}{\sqrt{(2\pi)^N N^{N_b} |\mathbf{A}| \cdot |\mathbf{V}|}} \exp\left[-\frac{1}{2}(\mathbf{b}^\star - \mathbf{b}^\eta)^T \mathbf{V}^{-1} (\mathbf{b}^\star - \mathbf{b}^\eta) - \frac{N}{2\sigma_\epsilon^2} J_g(\mathbf{b}^\star; \mathcal{D}, \mathcal{C}_j) \right] \tag{6.62}$$

Finally, the evidence of model class $\mathcal{C}_j$ is readily obtained by integrating the prediction-error variance:

$$p(\mathcal{D}|\mathcal{C}_j) \approx \frac{(\beta_0)^{\alpha_0} \Gamma(\alpha^\star)}{\sqrt{(2\pi)^N N^{N_b} |\mathbf{A}| \cdot |\mathbf{V}|} (\beta^\star)^{\alpha^\star} \Gamma(\alpha_0)} \exp\left[-\frac{1}{2}(\mathbf{b}^\star - \mathbf{b}^\eta)^T \mathbf{V}^{-1} (\mathbf{b}^\star - \mathbf{b}^\eta) \right] \tag{6.63}$$

where $\alpha^\star$ and $\beta^\star$ are given by Equation (6.57).

www.EngineeringBooksPdf.com

6.3.2 Nonlinear Regression Problems

With the same fashion of Equation (6.39), the predictive formula in a nonlinear regression model class takes the following general form:

$$Q(\mathbf{x}; \mathbf{b}, \mathbf{n}, \mathcal{C}_j) = f(\mathbf{x}; \mathbf{n}) + \sum_{l=1}^{N_b} b_l f_l(\mathbf{x}; \mathbf{n}) \tag{6.64}$$

where f and $f_l, l = 1, 2, \ldots, N_b$, are prescribed nonlinear functions with uncertain coefficients $\mathbf{n} \in \mathbb{R}^{N_n}$. Again, the measurement is assumed to be different from its model predicted value due to measurement noise and modeling error: $y = Q(\mathbf{x}; \mathbf{b}, \mathbf{n}, \mathcal{C}_j) + \epsilon$. The random variable is also modeled as a zero-mean Gaussian random variable with variance σ_ϵ^2. Given a number of measurements of $\mathbf{x}$ and y, the likelihood function takes the same form as in Equation (6.42) but the goodness-of-fit function is given by:

$$J_g(\mathbf{b}; \mathcal{D}, \mathcal{C}_j) = \frac{1}{N} \sum_{n=1}^{N} \left[y(n) - f(\mathbf{x}; \mathbf{n}) - \sum_{l=1}^{N_b} b_l f_l(\mathbf{x}; \mathbf{n}) \right]^2 \tag{6.65}$$

The optimal parameters can be solved by the iterative method introduced in Chapter 2, Section 2.4.2. By observing Equations (6.64) and (6.65), the integrand of the evidence integral has a complex topology for the nonlinear model classes. Since a closed-form solution is not available in general, the evidence integral has to be evaluated numerically, e.g., by Monte Carlo simulation. However, direct Monte Carlo simulation is inefficient for large N and N_j since the region of significant probability may be very concentrated with a complicated topology. Therefore, the following procedure is introduced to reduce the dimension of the Monte Carlo simulation and this increases the volume of the high probability density region compared to the sample space.

Here, a *separable* prior distribution is considered:

$$p(\boldsymbol{\theta}|\mathcal{C}_j) = p(\mathbf{b}, \mathbf{n}, \sigma_\epsilon^2|\mathcal{C}_j) = p(\mathbf{b}|\mathcal{C}_j) p(\mathbf{n}|\mathcal{C}_j) p(\sigma_\epsilon^2|\mathcal{C}_j) \tag{6.66}$$

Then, the evidence integral for a nonlinear model class is given by:

$$p(\mathcal{D}|\mathcal{C}_j) = \int_{\mathcal{N}} I_2(\mathbf{n}; \mathcal{C}_j) p(\mathbf{n}|\mathcal{C}_j) \mathrm{d}\mathbf{n} \tag{6.67}$$

where $\mathcal{N}$ is the integration domain of the uncertain parameters in $\mathbf{n}$ and $I_2(\mathbf{n}; \mathcal{C}_j)$ is defined as follows:

$$I_2(\mathbf{n}; \mathcal{C}_j) = \int_0^\infty \int_B p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}_j) p(\mathbf{b}, \sigma_\epsilon^2|\mathcal{C}_j) \mathrm{d}\mathbf{b} \mathrm{d}\sigma_\epsilon^2 \tag{6.68}$$

where B is the integration domain of the vector $\mathbf{b}$. For a uniform prior PDF, this integral can be evaluated with the same approach as in Equation (6.49):

$$I_2(\mathbf{n}; \mathcal{C}_j) \approx p(\boldsymbol{\theta}^\star|\mathcal{C}_j) \exp\left(-\frac{N}{2}\right) \frac{(\sqrt{2\pi}\sigma_\epsilon^\star)^{N_j - N}}{\sqrt{2N^{N_j}|\mathbf{A}|}} \tag{6.69}$$

www.EngineeringBooksPdf.com

In the case with a Gaussian prior for $\mathbf{b}$ and an inverse Gamma prior for σ_ϵ^2, this integral can be evaluated with the same approach as in Equation (6.63):

$$I_2(\mathbf{n};\mathcal{C}_j) \approx \frac{(\beta_0)^{\alpha_0}\Gamma(\alpha^\star)}{\sqrt{(2\pi)^N N^{N_b}|\mathbf{A}|\cdot|\mathbf{V}|}(\beta^\star)^{\alpha^\star}\Gamma(\alpha_0)} \exp\left[-\frac{1}{2}(\mathbf{b}^\star-\mathbf{b}^\eta)^T\mathbf{V}^{-1}(\mathbf{b}^\star-\mathbf{b}^\eta)\right] \quad (6.70)$$

Since the analytic solutions for $I_2(\mathbf{n};\mathcal{C}_j)$ are available, the dimension of the original evidence integral is reduced from $N_b + N_n + 1$ to N_n and can be estimated by direct Monte Carlo simulation:

$$p(\mathcal{D}|\mathcal{C}_j) = E_{p(\mathbf{n}|\mathcal{C}_j)}\left[I_2(\mathbf{n};\mathcal{C}_j)\right] = \frac{1}{N_s}\sum_{s=1}^{N_s} I_2(\mathbf{n}^{(s)};\mathcal{C}_j) \quad (6.71)$$

with independent samples $\mathbf{n}^{(s)}$, $s = 1, 2, \ldots, N_s$, simulated according to the prior PDF $p(\mathbf{n}|\mathcal{C}_j)$.

6.4 Application to Modal Updating

A twenty-story building and its response measurements are considered in this example. The Bayesian model class selection approach is applied to choose the optimal number of modes for a linear model. It is assumed that this building has a uniformly distributed floor mass and interstory stiffness over its height. The stiffness to mass ratios $\tilde{K}_l/M_l$, $l = 1, 2, \ldots, 20$, are chosen to be 1500 s^{-2} so that the fundamental frequency of the building is 0.4722 Hz. Rayleigh damping is assumed, i.e., the damping matrix $\mathbf{C}$ is given by:

$$\mathbf{C} = \alpha_1\mathbf{M} + \alpha_2\mathbf{K} \quad (6.72)$$

where $\alpha_1 = 0.044\,48\ \mathrm{s}^{-1}$ and $\alpha_2 = 0.001\,69$ s so that the damping ratios for the first two modes are 1.00%. The structure is assumed to be subjected to a wide-band random ground motion, which can be adequately modeled as Gaussian white noise with spectral intensity $S_{f0} = 10^{-5}$ $\mathrm{m}^2\ \mathrm{s}^{-3}$. The governing equation is given by:

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = -[M_1, M_2, \ldots, M_{20}]^T\ddot{x}_g \quad (6.73)$$

where $\ddot{x}_g$ is the ground acceleration, and M_l is the lth floor mass, $l = 1, 2, \ldots, 20$.

The data $\mathcal{D}$ consist of the simulated absolute acceleration at the top floor with 5% noise added over a time interval $T = 60$ s, using a sampling interval $\Delta t = 0.01$ s. The added noise is simulated using spectral intensity $S_{\epsilon 0} = 4.15 \times 10^{-8}$ $\mathrm{m}^2\ \mathrm{s}^{-3}$. Figure 6.5 shows the response spectrum obtained by Equation (3.25) and five modes clearly show up. However, it is difficult to determine the number of significant modes in the spectrum. Therefore, modal models with five to nine modes are considered. Each model class $\mathcal{C}_j$, $j = 1, 2, \ldots, 5$, consists of a linear modal model with $m = j + 4$ modes and the uncertain parameters are the modal frequencies, the two Rayleigh damping coefficients, the modal participation factors, and the spectral intensity $S_{\epsilon 0}$ of the prediction error at the observed degree of freedom. Even though the mass and stiffness matrices are unknown, the damping ratios satisfy the following relationship for the Rayleigh damping model [58]:

$$\zeta_m = \frac{\alpha_1}{2\Omega_m} + \frac{\alpha_2\Omega_m}{2} \quad (6.74)$$

www.EngineeringBooksPdf.com

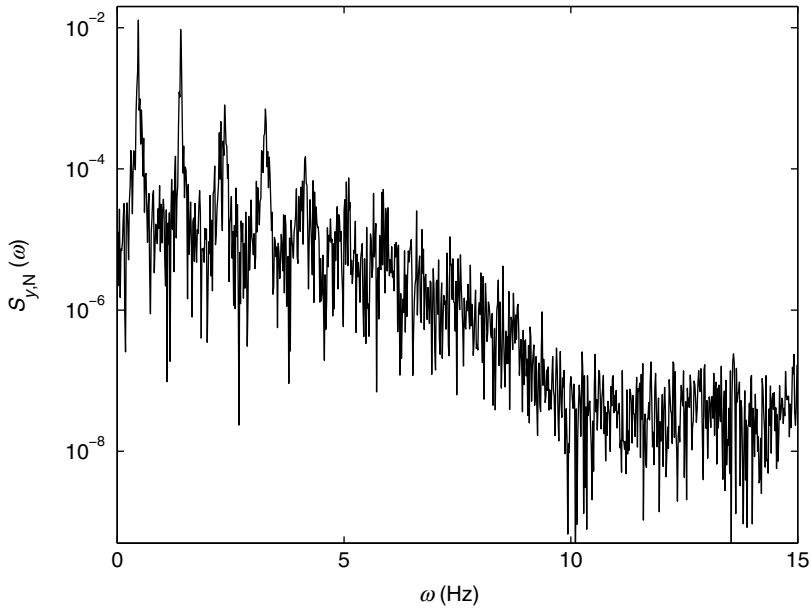

Figure 6.5 Measured spectrum of the building

Independent prior distributions for the parameters are taken as follows: a Gaussian distribution for the natural frequencies with mean $0.5(2m-1)$ Hz and a coefficient of variation 10% for the mth mode. Furthermore, the Rayleigh coefficients, the modal participation factor and the spectral intensity of the prediction error are assumed to be uniformly distributed over a sufficiently wide range to let the likelihood function determine their values. Note that the ranges of these distributions do not affect the model class selection results since they influence all modal models in the same way. Therefore, the computation of the Ockham factor and plausibility will exclude the prior PDF of these parameters but the one for the modal frequencies will still be included.

The Bayesian spectral density approach in Chapter 3 is used for parametric identification. The spectral density estimator is utilized up to 8 Hz to include all the peaks so $N_\omega = 480$. Table 6.2 shows the optimal modal frequencies for model classes with different number of modes. There is in general no difficulty in identifying the first five modes but it is not the

Table 6.2 Optimal modal frequencies (in Hz) of the building

j, m	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6	Ω_7	Ω_8	Ω_9
Exact	0.4722	1.414	2.347	3.267	4.167	5.043	5.889	6.701	7.474
$j=1, m=5$	0.4895	1.418	2.357	3.276	4.169	—	—	—	—
$j=2, m=6$	0.4910	1.417	2.352	3.274	4.189	5.731	—	—	—
$j=3, m=7$	0.4913	1.413	2.359	3.283	4.166	5.063	5.916	—	—
$j=4, m=8$	0.4861	1.418	2.350	3.272	4.167	5.063	5.892	7.389	—
$j=5, m=9$	0.4883	1.418	2.359	3.275	4.166	5.046	5.865	6.611	7.429

www.EngineeringBooksPdf.com

Table 6.3 Plausibilities of models with different number of modes based on data

j, m	$\ln\ p(\mathcal{D}\|\boldsymbol{\theta}^\star, \mathcal{C}_j)$	$\ln\ \mathcal{O}_j$	$\ln\ p(\mathcal{D}\|\mathcal{C}_j)$	$P(\mathcal{C}_j\|\mathcal{D}, \mathcal{U})$
$j = 1, m = 5$	4821.7	−88.66	4733.0	0.00
$j = 2, m = 6$	4837.1	−100.0	4737.1	0.00
$j = 3, m = 7$	4873.6	−117.1	4756.5	1.00
$j = 4, m = 8$	4883.6	−130.6	4753.0	0.00
$j = 5, m = 9$	4895.5	−150.5	4745.0	0.00

case for the higher modes in some of the modal models. Table 6.3 shows the values of the log-maximum likelihood value, the log-Ockham factor, the log-evidence and the plausibility of each model class, calculated from Equation (6.4) using the evidence for each model from Equation (6.11) and equal prior plausibilities $P(\mathcal{C}_j|\mathcal{U}) = 1/5$. It is no wonder that the maximum log-likelihood value increases with an increasing number of modes used in a model class but the log-Ockham factor goes in an opposite direction. It turns out that using seven modes is optimal. Figure 6.6 shows the spectrum estimated directly from the data (zigzag curve) and the optimal model spectrum using seven modes (smoother curve). The optimal model using seven modes fits the measured spectrum very well. Furthermore, all the seven optimal modal frequencies are very close to their target values, but it is not the case for using five or six modes. It was found that if the AIC is used, a model class with an extra parameter is preferred only if its log-likelihood is increased by no less than 1.0. In this case, nine modes are optimal because the penalty term is too small compared to the changing of the log-likelihood term in Equation (6.19). On the other hand, according to the BIC in Equation (6.20), a model class with

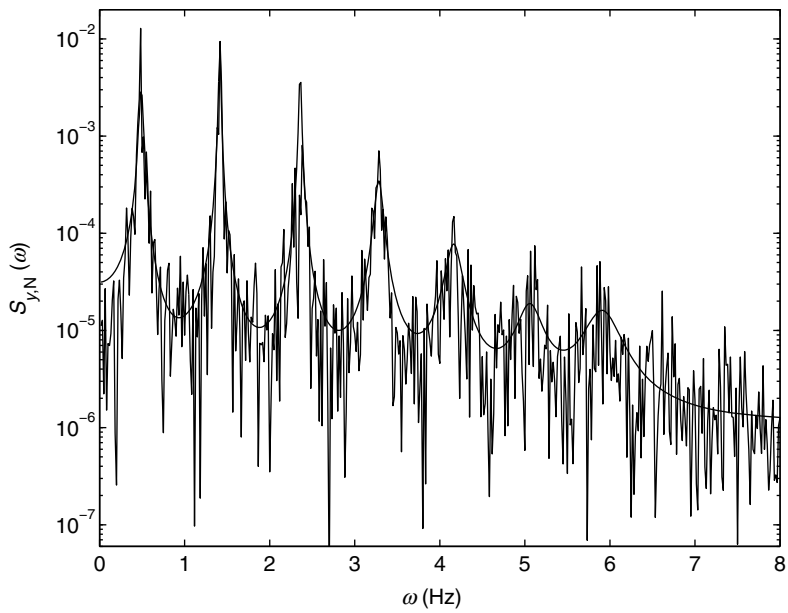

Figure 6.6 Spectrum estimated from the measurements and from the optimal model with seven modes

www.EngineeringBooksPdf.com

an extra parameter is preferred only if its log-likelihood is increased by no less than $0.5 \ln N$, which is 3.087 in this case. The BIC also favors the nine-mode model class. However, judging from the measured spectrum, it is more reasonable to use the seven-mode model as the eighth and ninth modes are not evident.

6.5 Application to Seismic Attenuation Empirical Relationship

6.5.1 Problem Description

Prediction of peak ground acceleration (PGA) has received great attention in the societies of civil engineering, earthquake engineering and seismology in the previous decades. A significant amount of work can be found in predicting the PGA by using the magnitudes of earthquakes, station-to-hypocenter distances, and the properties of the site foundations [27–29, 33, 34, 125, 126, 175]. In particular, the Boore–Joyner–Fumal seismic attenuation formula is a well-known regression model for estimation of the PGA and is given by [27, 28]:

$$\log_{10} PGA = b_1 + b_2(M - M_0) + b_3(M - M_0)^2 + b_4 r + b_5 \log_{10} r + b_6 G_B + b_7 G_C \tag{6.75}$$

where M is the moment magnitude of an earthquake [99], M_0 is a shifting constant and $M_0 = 6$ was used in Boore *et al.* [27, 28], r is the observation station-to-hypocenter distance (in km), G_B and G_C are site foundation classification variables, $G_B = 1$ for class B and 0 otherwise, and $G_C = 1$ for class C and 0 otherwise [27, 28] (the classification of foundation will be discussed further in Section 6.5.3.1), and ϵ is a zero-mean Gaussian random variable to model the predictive/fitting error. In this model, the parameters to be identified are the coefficients b_ls and the standard deviation σ_ϵ for the predictive/fitting error ϵ and they can be estimated by the maximum likelihood method with strong-motion records.

In the literature Shi and Shen [235], Wang *et al.* [270], Wong *et al.* [273] and Zheng and Wong [302], the Crouse–McGuire model [65] was used for the attenuation model with the strong-motion records from different areas of China and this model has essentially a different functional form from Equation (6.75):

$$\ln PGA = b_1 + b_2 M + b_3 \ln [r + n_1 \exp (n_2 M)] \tag{6.76}$$

where the variables M and r are defined in the same way as in Equation (6.75). Note the predictive models in Equations (6.75) and (6.76) are both empirical models. It is reasonable to speculate if better results can be obtained by adding/erasing terms on these models or even by considering a different functional form. A good predictive model class should have good capability of fitting the strong-motion records and otherwise be insensitive to observation and modeling error. If the number of free parameters in a model class is very large, this model class possesses a powerful data fitting capability but may become very sensitive to the measurement noise, referred to as *over-fitting*. This is because the identified parameters of this model class depend highly on the details of the data (including the measurement noise and modeling errors) and it may cause large errors in future predictions. In this section, the regression formula given in Boore *et al.* [27, 28] is examined. Thirty two model classes are constructed by including subsets of terms in the regression equation (Equation (6.75)). In this section, the Bayesian model class selection approach is used to select among these thirty two model classes. It

www.EngineeringBooksPdf.com

considers the plausibility of each model class conditional on the database. The plausibility of a model class can be factorized as a product of the maximum likelihood and a measure of the robustness of the model class (namely the Ockham factor). The Ockham factor penalizes those model classes that are over-parameterized and therefore prone to error due to the noise in the data set. Therefore, the most plausible model class should have the optimal tradeoff between the data fitting capability and the robustness to measurement noise and modeling error. A database of 249 strong-motion records from the China Earthquake Data Center [50] is utilized for this study.

6.5.2 Selection of the Predictive Model Class

In order to examine the suitability of the predictive model class in Equation (6.75) and to propose the most suitable one, model class candidates are constructed. Here, predictive model classes are considered in a similar functional form of Equation (6.75) but to include different combinations of terms in different model classes. First of all, the constant b_1 is necessary to serve as a scaling factor of the PGA for any model class so all model classes contain b_1. As will be discussed later in Section 6.5.3.1 that only strong-motion records with moment magnitude $M > 3.5$ are utilized, $M_0 = 3.5$ is taken in this study. For the site properties, the terms $b_6 G_B$ and $b_7 G_C$ are either included or excluded together in a model class. Therefore, there are $2^5 = 32$ model class candidates, namely $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_{32}$.

Let $\mathcal{D}$ denote the data that include the measured PGA and the corresponding earthquake magnitude, observation station-to-hypocenter distance and site foundation properties.

The uncertain model parameter vector is:

$$\boldsymbol{\theta} = [\mathbf{b}^T, \sigma_\epsilon^2]^T \tag{6.77}$$

that includes the uncertain coefficients and the prediction-error variance. Given a predictive model class $\mathcal{C}_j$, the optimal parameter vector $\boldsymbol{\theta}^\star$ for a given model class is the one that maximizes the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)$, which provides a measure of the relative plausibility of the values of the parameters in $\boldsymbol{\theta}$. By the Bayes' theorem, the posterior PDF for the parameters $\boldsymbol{\theta}$ is given by:

$$p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j) = \kappa_1 p(\boldsymbol{\theta}|\mathcal{C}_j)(2\pi)^{-\frac{N}{2}} \sigma_\epsilon^{-N} \exp\left[-\frac{N}{2\sigma_\epsilon^2} J_g(\mathbf{b}; \mathcal{D}, \mathcal{C}_j)\right] \tag{6.78}$$

where κ_1 is a normalizing constant, N is the number of strong-motion records in $\mathcal{D}$ and $p(\boldsymbol{\theta}|\mathcal{C}_j)$ is the prior PDF of the model parameters expressing the user's judgment about the relative plausibility of the values of the predictive model parameters without using the data. In this study, a uniform prior PDF is used for not biasing the parametric identification results prior to the data. The goodness-of-fit function $J_g(\mathbf{b}; \mathcal{D}, \mathcal{C}_j)$ is given by:

$$J_g(\mathbf{b}; \mathcal{D}, \mathcal{C}_j) = \frac{1}{N} \sum_{n=1}^{N} \left[\log_{10} PGA_o(n) - \log_{10} PGA(\mathbf{x}(n); \mathbf{b}, \mathcal{C}_j)\right]^2 \tag{6.79}$$

where $PGA(\mathbf{x}(n); \mathbf{b}, \mathcal{C}_j)$ is the model predicted PGA of the nth record and $PGA_o(n)$ is its corresponding observed value. The most probable model parameter vector $\boldsymbol{\theta}^\star$ is obtained by maximizing the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)$ in Equation (6.78). This is equivalent to minimizing

www.EngineeringBooksPdf.com

the goodness-of-fit function $J_g(\mathbf{b}; \mathcal{D}, \mathcal{C}_j)$ over all the parameters in $\mathbf{b}$ because a uniform prior PDF is used. A predictive formula in any model class $\mathcal{C}_j$, which includes subsets of terms of Equation (6.75), can be written as:

$$\log_{10} PGA(\mathbf{x}; \mathbf{b}, \mathcal{C}_j) = \sum_{l=1}^{N_b} b_l x_l \tag{6.80}$$

where $b_1, b_2, \ldots, b_{N_b}$ are the uncertain coefficients for identification and $x_1, x_2, \ldots, x_{N_b}$ are the corresponding data, i.e., some of the following variables: $1, M - M_0, (M - M_0)^2, r, \log_{10} r$, G_B and G_C.

By solving the following simultaneous linear algebraic equations $\partial J_g(\mathbf{b}; \mathcal{D}, \mathcal{C}_j)/\partial b_l = 0$, $l = 1, 2, \ldots, N_b$, the closed-form solution of the most probable coefficient vector $\mathbf{b}^\star$ can be obtained:

$$\mathbf{b}^\star = \mathbf{A}^{-1} \begin{bmatrix} \frac{1}{N}\sum_{n=1}^{N} x_1(n) \log_{10} PGA_o(n) \\ \frac{1}{N}\sum_{n=1}^{N} x_2(n) \log_{10} PGA_o(n) \\ \vdots \\ \frac{1}{N}\sum_{n=1}^{N} x_{N_b}(n) \log_{10} PGA_o(n) \end{bmatrix} \tag{6.81}$$

where the matrix $\mathbf{A} \in \mathbb{R}^{N_b \times N_b}$ has components $A^{(l,l')}$ given by $A^{(l,l')} = \frac{1}{N}\sum_{n=1}^{N} x_l(n) x_{l'}(n)$. Similarly, the most plausible value of the predictive/fitting error variance can be found by solving $\partial p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)/\partial \sigma_\epsilon = 0$ and is readily obtained by:

$$\sigma_\epsilon^{2\star} = \min_{\mathbf{b}} J_g(\mathbf{b}; \mathcal{D}, \mathcal{C}_j) = J_g(\mathbf{b}^\star; \mathcal{D}, \mathcal{C}_j) \tag{6.82}$$

For large N, the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)$ is approximately Gaussian centered at the optimal parameter $\boldsymbol{\theta}^\star$. Therefore, the uncertainty of the parameter estimates can be represented by its covariance matrix given by $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \mathcal{H}_j(\boldsymbol{\theta}^\star)^{-1}$, i.e., the inverse of the Hessian matrix evaluated at the optimal point. The diagonal elements of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ are the marginal variances of the corresponding elements of $\boldsymbol{\theta}$ and the quantification of the uncertainty of the model parameters can be used for the uncertainty analysis of the predicted PGA.

It is attempted to use $\mathcal{D}$ to select the most suitable predictive model class among the aforementioned model class candidates $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_{32}$. Since probability may be interpreted as a measure of plausibility [63], the most suitable model class is the one with the highest plausibility $P(\mathcal{C}_j|\mathcal{D}, \mathcal{U})$. The prior plausibility is taken as a uniform prior:

$$P(\mathcal{C}_j|\mathcal{U}) = 1/32, \quad j = 1, 2, \ldots, 32 \tag{6.83}$$

Therefore, the plausibility of a model class is proportional to the evidence:

$$P(\mathcal{C}_j|\mathcal{D}) \propto p(\mathcal{D}|\mathcal{C}_j) \tag{6.84}$$

In the current situation that the goodness-of-fit function is quadratic for the uncertain parameters (except for σ_ϵ^2), the conditional posterior PDF $p(\mathbf{b}|\mathcal{D}, \mathcal{C}_j, \sigma_\epsilon^2)$ is Gaussian and the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{C}_j)$ is approximately Gaussian for large N. According to Section 6.3.1.1, the evidence

www.EngineeringBooksPdf.com

$p(\mathcal{D}|\mathcal{C}_j)$ can be approximated by:

$$p(\mathcal{D}|\mathcal{C}_j) \approx p(\boldsymbol{\theta}^\star|\mathcal{C}_j)\exp\left(-\frac{N}{2}\right)\frac{\pi^{\frac{N_j-N}{2}}(\sqrt{2}\sigma_\epsilon^\star)^{N_j+1-N}}{\sqrt{N^{N_j}|\mathbf{A}|}} \tag{6.85}$$

where $N_j = N_b + 1$ is the number of uncertain parameters in the predictive model class $\mathcal{C}_j$, i.e., including the unknown coefficients to be determined in the predictive formula and the unknown predictive/fitting error parameter, and $\boldsymbol{\theta}^\star$ is the optimal parameter vector found by maximizing the posterior PDF, which is equivalent to minimizing its negative logarithm. By Equations (6.14) and (6.85), the Ockham factor is given by:

$$\mathcal{O}_j = p(\boldsymbol{\theta}^\star|\mathcal{C}_j)|\mathbf{A}|^{-\frac{1}{2}}\left(\frac{\pi}{N}\right)^{\frac{N_j}{2}}(\sqrt{2}\sigma_\epsilon^\star)^{N_j+1} \tag{6.86}$$

It serves as a measure of the robustness of a model class. The model classes that depend too sensitively on their parameters are penalized more heavily because their robustness is low (sensitive to noise and modeling error) and over-fitting/over-parameterization may occur. In this case, small changes of the data points may substantially affect the parameter identification results. This is not preferable since future predictions will be sensitive to the measurement noise and modeling error of the data.

6.5.3 Analysis with Strong Ground Motion Measurements

6.5.3.1 Description of the Database

A database of strong-motion records is obtained from the China Earthquake Data Center [50]. In this study, the horizontal components of the PGA records are considered and only the records with $M > 3.5$ are utilized so $M_0 = 3.5$ is taken in Equation (6.75). Table 6.4 summarizes the details of the records in this database. There are 249 records, observed from 32 stations in the Tangshan and Xinjiang regions. Figure 6.7 shows the site-to-hypocenter distances versus the magnitudes of the earthquakes for each record in both normal and logarithm scales for r. The symbols 'o' and 'x' denote the records in the Tangshan and Xinjiang region, respectively. Figure 6.8 shows the measured values of $\log_{10} PGA$ versus M and r. On the left subplot, there are two circles (Tangshan) around the lower right corner, implying small PGA values with large M. This is because the two records associate with large values of r as can be seen from the two circles near the lower right corner on the right subplot.

Table 6.4 Information of the database

				Site class		
Region	Records	Earthquakes	Stations	A	B	C
Tangshan	94	18	19	8	6	5
Xinjiang	155	125	13	4	6	3

www.EngineeringBooksPdf.com

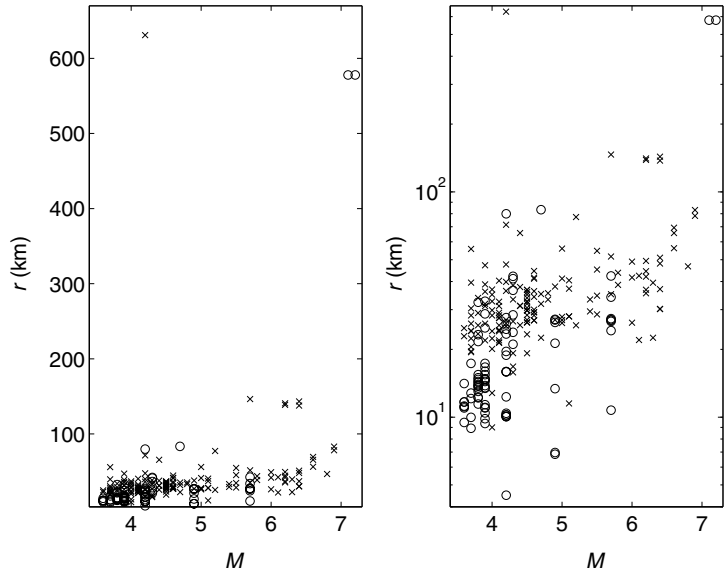

Figure 6.7 Distribution of M and r in the database

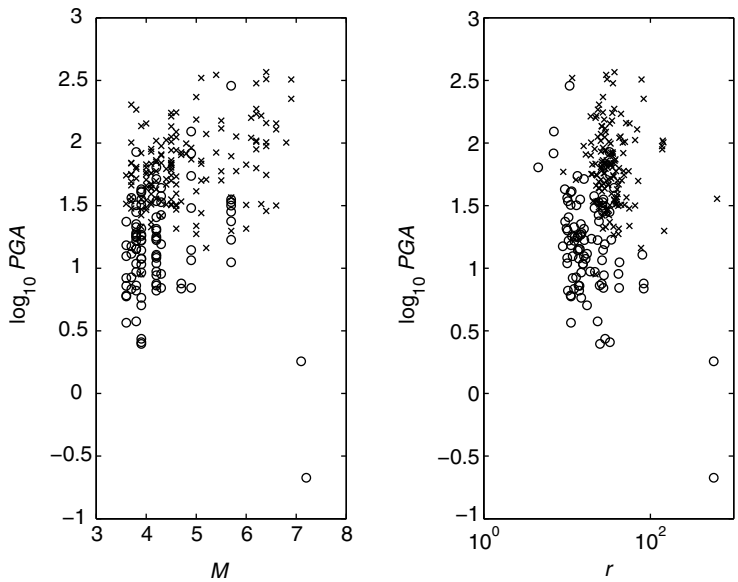

Figure 6.8 Measured $\log_{10}$ PGA versus M and r

www.EngineeringBooksPdf.com

The classification of the site foundation depends on its stiffness and the averaged shear velocity over the upper 30 m is used as its measure [27]. Since only the soil/rock types are provided by the data center, the site class is defined in a slightly different way from the original definition by Boore *et al.* [27]. Specifically, granite, sandstone, bedrock, siltstone, and conglomerate are classified as class A. Alluvium, diluvium, and weathering conglomerate are included in class B. Soft soil, clay and subclay are classified as class C. This problem is region-dependent because geographical and geological conditions are important factors that govern the traveling waves from the earthquake source to the site. Therefore, independent studies of the predictive model class selection are performed separately using the data of the Tangshan and Xinjiang regions.

Note that the range of $\log_{10} PGA$ in the data set lies in the interval of $[-1, 3]$. The prior PDF for b_1 in Equation (6.75) is taken to be a uniform distribution in the range from -1 to 3. For the other parameters (if they are included in a predictive model class), their prior PDFs are also taken as uniform distribution in order to let the measurement to infer the parameter values. By considering the previous work in Boore *et al.* [27–29], Brillinger and Preisler [33, 34], Joyner and Boore [125, 126] and Melchers [175], the range of the uniform distribution is taken as $[-1, 1]$ for b_2, b_3, b_5, b_6 and b_7. For b_4, it was observed that it is much smaller from a previous study. It is also expected to happen in this case because r is much larger than $\log_{10} r$ in the range of the data (Figure 6.7). Specifically, the range for b_4 is taken to be $[-0.01, 0.01]$. Note that as long as the range is sufficiently large, the values of the bounds do not influence the parametric identification results of $\boldsymbol{\theta}^\star$.

6.5.3.2 Tangshan Region

Tangshan city is located in the North China Plain. Table 6.5 shows the model class selection results of the predictive formula using the horizontal records of the Tangshan region. The first column shows the ranking of each model class. A smaller number of the ranking corresponds to a higher plausibility of the model class. The second column shows the parameters being included in that model class, e.g., '1 2 3 5' denotes a model class with free parameters b_1, b_2, b_3, and b_5 in Equation (6.75). The third and fourth columns show the variance of the fitting error and the corresponding maximum log-likelihood value. The fifth column shows the value of the log-Ockham factor $\ln \mathcal{O}_j$ in Equation (6.16), that indicates the robustness of the model class, and the last column shows the plausibility of each model class. The full model class with all seven parameters has the largest maximum likelihood value. This is intuitive because the full model class has the largest output space so it is capable to fit the data at least as well as any other model classes with less free parameters considered in this study. However, its robustness is not as good as the other model classes and its Ockham factor is 5.79×10^{-10}, which is the smallest among all the model class candidates. Table 6.6 shows the most probable coefficients of each model class and Table 6.7 shows the posterior standard deviation calculated using the Bayesian probabilistic approach.

From Table 6.6, it is surprising that the values of b_6 are always larger than b_7 but the terms $b_6 G_B$ and $b_7 G_C$ are not included in the optimal model of $\mathcal{C}_1$. By observing the model classes that include these two terms, the optimal values of these parameters are of similar order of its posterior standard deviation. This implies that the data do not provide evidence for such terms to be statistically significant. Furthermore, some of the values of b_7 (in some of the model

www.EngineeringBooksPdf.com

Table 6.5 Model class selection results (Tangshan)

Number	Parameters	$\sigma_\epsilon^{2\star}$	$\ln\ p(\mathcal{D}\|\boldsymbol{\theta}^\star, \mathcal{C}_j)$	$\ln\ \mathcal{O}_j$	$P(\mathcal{C}_j\|\mathcal{D}, \mathcal{U})$
1	1 2 3 5	0.088	−19.0	−15.1	7×10^{-1}
2	1 2 4 5	0.092	−21.0	−14.6	1.6×10^{-1}
3	1 2 3 4 5	0.088	−19.0	−16.9	1.1×10^{-1}
4	1 2 3 5 6 7	0.086	−18.3	−20.0	1.1×10^{-2}
5	1 2 5	0.105	−27.4	−11.7	4.5×10^{-3}
6	1 2 4 5 6 7	0.091	−20.6	−19.3	2×10^{-3}
7	1 2 3 4 5 6 7	0.086	−18.3	−21.8	1.7×10^{-3}
8	1 3 4 5	0.100	−25.1	−15.2	1.4×10^{-3}
9	1 2 4	0.106	−27.9	−12.6	1.1×10^{-3}
10	1 3 4	0.109	−29.0	−13.4	1.7×10^{-4}
11	1 2 3 4	0.106	−27.9	−14.9	1.1×10^{-4}
12	1 2 5 6 7	0.104	−26.9	−16.4	7.1×10^{-5}
13	1 3 4 5 6 7	0.099	−24.8	−19.8	1.8×10^{-5}
14	1 2 4 6 7	0.105	−27.5	−17.3	1.6×10^{-5}
15	1 5	0.128	−36.7	−8.75	8.1×10^{-6}
16	1 4 5	0.123	−34.9	−11.2	4.2×10^{-6}
17	1 3 4 6 7	0.108	−28.7	−18.0	2.2×10^{-6}
18	1 4	0.129	−37.0	−9.69	2.2×10^{-6}
19	1 3 5	0.122	−34.4	−12.4	2×10^{-6}
20	1 2 3 4 6 7	0.105	−27.5	−19.6	1.6×10^{-6}
21	1 2 3	0.125	−35.8	−12.4	5×10^{-7}
22	1 5 6 7	0.125	−35.7	−13.3	2.4×10^{-7}
23	1 4 5 6 7	0.120	−33.8	−15.8	1.3×10^{-7}
24	1 4 6 7	0.126	−36.0	−14.2	7×10^{-8}
25	1 3 5 6 7	0.120	−33.7	−16.9	4.5×10^{-8}
26	1 2 3 6 7	0.123	−34.7	−16.9	1.7×10^{-8}
27	1	0.175	−51.4	−6.35	3.7×10^{-11}
28	1 3	0.166	−48.9	−10.2	9.7×10^{-12}
29	1 2	0.174	−51.3	−8.94	3×10^{-12}
30	1 6 7	0.173	−50.8	−10.5	1.1×10^{-12}
31	1 3 6 7	0.162	−47.9	−14.4	3.8×10^{-13}
32	1 2 6 7	0.172	−50.6	−13.1	9.4×10^{-14}

classes with low plausibility) are negative. This seems to contradict with physics at the first glance. However, this is only due to the large posterior uncertainty of this parameter, i.e., the data do not have much say on the value of this coefficient.

Another noteworthy point is on the model class with parameters b_1 and b_2 only, i.e., the twenty eighth model class. The optimal value for b_2 is negative and it seems to imply that the larger the moment magnitude of an earthquake, the smaller the PGA. However, this is only due to the non-uniform distribution of M and r in the data set. Specifically, the correlation coefficient between the M and r is 0.64 in this data set. This implies that a large value of M in a record often associates with a large value of r, that reduces the PGA. Furthermore, the correlation coefficient between the $\log_{10} PGA$ and M is -0.0511 and hence the optimal coefficient b_2 is negative. Therefore, a model class with too few free parameters may cause under-fitting to the data.

www.EngineeringBooksPdf.com

Table 6.6 Optimal parameters of each predictive model class (Tangshan)

Number	b_1	b_2	b_3	b_4	b_5	b_6	b_7
1	1.9	0.76	−0.19	—	−0.86	—	—
2	1.8	0.34	—	−0.0025	−0.66	—	—
3	1.9	0.73	−0.17	−0.000 31	−0.83	—	—
4	1.8	0.76	−0.19	—	−0.85	0.081	0.0033
5	2.3	0.28	—	—	−1.1	—	—
6	1.8	0.33	—	−0.0024	−0.66	0.06	0.0082
7	1.8	0.75	−0.18	−0.000 11	−0.83	0.08	0.0031
8	1.8	—	0.12	−0.0038	−0.49	—	—
9	1.1	0.27	—	−0.0041	—	—	—
10	1.2	—	0.11	−0.0051	—	—	—
11	1.1	0.28	−0.005	−0.0041	—	—	—
12	2.3	0.27	—	—	−1.1	0.074	0.037
13	1.8	—	0.12	−0.0037	−0.49	0.058	0.012
14	1.1	0.27	—	−0.0041	—	0.062	−0.016
15	2.0	—	—	—	−0.68	—	—
16	1.7	—	—	−0.0014	−0.39	—	—
17	1.2	—	0.11	−0.005	—	0.056	−0.008
18	1.3	—	—	−0.0026	—	—	—
19	2.3	—	0.049	—	−0.93	—	—
20	1.1	0.30	−0.018	−0.0039	—	0.064	−0.017
21	0.86	0.83	−0.29	—	—	—	—
22	2.0	—	—	—	−0.69	0.12	0.027
23	1.7	—	—	−0.0014	−0.39	0.12	0.0094
24	1.2	—	—	−0.0027	—	0.11	−0.0064
25	2.2	—	0.046	—	−0.91	0.10	0.040
26	0.84	0.83	−0.29	—	—	0.11	−0.024
27	1.2	—	—	—	—	—	—
28	1.2	—	−0.042	—	—	—	—
29	1.2	−0.03	—	—	—	—	—
30	1.1	—	—	—	—	0.10	0.026
31	1.2	—	−0.046	—	—	0.13	0.014
32	1.2	−0.039	—	—	—	0.11	0.024

In order to balance the data fitting capability and robustness, a relatively simple model class is chosen by the Bayesian model class selection approach and its optimal model is given by:

$$\log_{10} PGA = 1.9 + 0.76(M - 3.5) - 0.19(M - 3.5)^2 - 0.86 \log_{10} r \tag{6.87}$$

where PGA is in cm/s^2. The term $-0.86 \log_{10} r$ indicates that the PGA decreases with an increasing site-to-hypocenter distance. The PGA decreases by 45% ($\approx (1 - 2^{-0.86}) \times 100\%$) if the value of r is doubled and all the other factors remain the same. This predictive model class has a plausibility over 0.7 by the Bayesian model selection approach. It is noted that:

$$\frac{\partial \log_{10} PGA}{\partial M} = -0.38M + 2.09 \tag{6.88}$$

www.EngineeringBooksPdf.com

Table 6.7 Uncertainty of the parameters for each predictive model class (Tangshan)

Number	b_1	b_2	b_3	b_4	b_5	b_6	b_7
1	0.17	0.13	0.043	—	0.14	—	—
2	0.19	0.06	—	0.000 66	0.17	—	—
3	0.19	0.20	0.083	0.0012	0.19	—	—
4	0.18	0.13	0.043	—	0.13	0.07	0.08
5	0.15	0.06	—	—	0.14	—	—
6	0.20	0.06	—	0.000 66	0.17	0.07	0.08
7	0.19	0.20	0.083	0.0012	0.19	0.07	0.08
8	0.20	—	0.026	0.000 84	0.17	—	—
9	0.05	0.06	—	0.000 53	—	—	—
10	0.04	—	0.026	0.000 73	—	—	—
11	0.09	0.19	0.082	0.000 99	—	—	—
12	0.16	0.06	—	—	0.14	0.08	0.09
13	0.20	—	0.026	0.000 84	0.17	0.08	0.08
14	0.06	0.06	—	0.000 53	—	0.08	0.09
15	0.15	—	—	—	0.12	—	—
16	0.22	—	—	0.000 74	0.19	—	—
17	0.05	—	0.027	0.000 73	—	0.08	0.09
18	0.04	—	—	0.000 45	—	—	—
19	0.19	—	0.022	—	0.16	—	—
20	0.10	0.19	0.082	0.0010	—	0.08	0.09
21	0.08	0.15	0.048	—	—	—	—
22	0.15	—	—	—	0.12	0.08	0.09
23	0.22	—	—	0.000 73	0.19	0.08	0.09
24	0.06	—	—	0.000 45	—	0.08	0.10
25	0.19	—	0.023	—	0.16	0.08	0.09
26	0.08	0.15	0.048	—	—	0.08	0.09
27	0.04	—	—	—	—	—	—
28	0.05	—	0.019	—	—	—	—
29	0.06	0.06	—	—	—	—	—
30	0.06	—	—	—	—	0.10	0.11
31	0.06	—	0.019	—	—	0.10	0.11
32	0.08	0.06	—	—	—	0.10	0.11

so it becomes negative if $M > 5.5$. This is because there are only two data points with a moment magnitude larger than 6.0 but the associated hypocenter-station distances are large (Figure 6.7) so the resultant PGAs are small (Figure 6.8). Therefore, the predictive model in Equation (6.87) is reliable only for a moment magnitude between 3.5 and 5.5. More records with $M > 5.5$ are needed to expand the applicable range.

Figure 6.9 summarizes the result of Equation (6.87) for site class A. The six curves (from top to bottom) show the $\log_{10} PGA$ versus the magnitude of earthquake M for six levels of distance, namely 10, 50, 100, 250, 500 and 1000 km. This chart can be used for prediction of peak ground acceleration in the design process. One first estimates the closest fault from the site so the corresponding curve can be selected. This curve shows the estimated peak ground acceleration under different levels of earthquake, which may be estimated by neural networks [194].

www.EngineeringBooksPdf.com

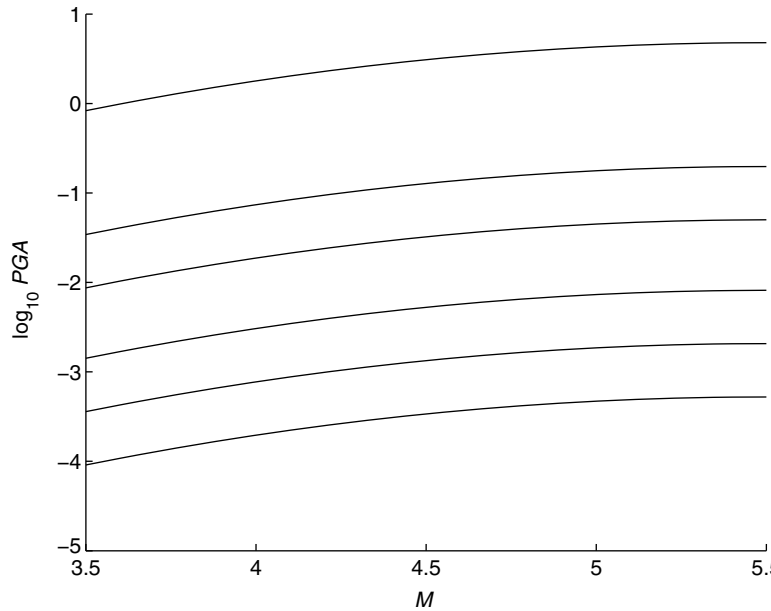

Figure 6.9 Design chart for site class A (Tangshan)

6.5.3.3 Xinjiang Region

Xinjiang is in the North West region of China. In the same fashion as in Tables 6.5 and 6.6, Tables 6.8 and 6.9 show the predictive model class selection results and the optimal parameters of each model class. From Table 6.9, the optimal predictive model for the PGA is given by:

$$\log_{10} PGA = 1.9 + 0.22(M - 3.5) - 0.4 \log_{10} r + 0.24 G_B + 0.06 G_C \tag{6.89}$$

Note that the first four model classes possess similar plausibility, implying that the Bayesian model selection method does not have a strong preference on the most plausible model class. This is in contrast to the previous case in the Tangshan region, in which the plausibility of the optimal model class is over 0.7. With the data of Xinjiang, a multi-model predictive formula can be used as follows:

$$\log_{10} PGA = \sum_{j=1}^{4} \kappa (\log_{10} PGA)_j P(\mathcal{C}_j|\mathcal{D}) \tag{6.90}$$

where $(\log_{10} PGA)_j$ is the prediction of PGA using the jth most plausible model class and $\kappa = 1/\sum_{j=1}^{4} P(\mathcal{C}_j|\mathcal{D})$ is a normalizing constant. In other words, the PGA is estimated using four model classes with the weightings being the relative plausibility given the data. Specifically, this multi-model is given by:

$$\begin{aligned}\log_{10} PGA = {} & 2.2 + 0.12(M - 3.5) + 0.033(M - 3.5)^2 + 0.00071 r - 0.56 \log_{10} r \\ & + 0.26 G_B + 0.076 G_C\end{aligned} \tag{6.91}$$

www.EngineeringBooksPdf.com

Table 6.8 Model class selection results (Xinjiang)

Number	Parameters	$\sigma_\epsilon^{2\star}$	$\ln\ p(\mathcal{D}\|\boldsymbol{\theta}^\star, \mathcal{C}_j)$	$\ln\ \mathcal{O}_j$	$P(\mathcal{C}_j\|\mathcal{D}, \mathcal{U})$
1	1 2 5 6 7	0.067	−10.0	−18.6	2.3×10^{-1}
2	1 2 4 5 6 7	0.064	−7.54	−21.2	2×10^{-1}
3	1 3 5 6 7	0.066	−8.88	−19.8	2×10^{-1}
4	1 3 4 5 6 7	0.064	−6.43	−22.4	1.7×10^{-1}
5	1 2 5	0.072	−16.5	−13.6	4.9×10^{-2}
6	1 3 5	0.072	−15.6	−14.8	3.5×10^{-2}
7	1 2 3 5 6 7	0.066	−8.65	−21.9	3×10^{-2}
8	1 2 3 4 5 6 7	0.063	−6.07	−24.6	2.9×10^{-2}
9	1 2 4 5	0.071	−14.7	−16.1	2.4×10^{-2}
10	1 3 4 5	0.070	−13.7	−17.4	1.8×10^{-2}
11	1 2 6 7	0.072	−15.9	−16.4	5.5×10^{-3}
12	1 2 3 5	0.072	−15.6	−16.9	4.4×10^{-3}
13	1 3 6 7	0.071	−15.1	−17.6	3.6×10^{-3}
14	1 2 3 4 5	0.070	−13.7	−19.5	2.3×10^{-3}
15	1 2	0.078	−22.5	−11.5	9.9×10^{-4}
16	1 2 4 6 7	0.071	−14.8	−19.4	8.3×10^{-4}
17	1 3 4 6 7	0.070	−13.9	−20.7	5.7×10^{-4}
18	1 2 3 6 7	0.071	−15.0	−19.7	5×10^{-4}
19	1 3	0.078	−22.0	−12.7	4.8×10^{-4}
20	1 2 4	0.077	−21.1	−14.5	2.1×10^{-4}
21	1 3 4	0.076	−20.5	−15.7	1.1×10^{-4}
22	1 2 3 4 6 7	0.070	−13.7	−22.7	8.8×10^{-5}
23	1 2 3	0.078	−22.0	−14.8	6.3×10^{-5}
24	1 2 3 4	0.076	−20.5	−17.7	1.5×10^{-5}
25	1	0.090	−33.6	−7.85	5.8×10^{-7}
26	1 5	0.090	−33.4	−9.81	1×10^{-7}
27	1 4	0.090	−33.4	−10.7	4.1×10^{-8}
28	1 4 5	0.090	−33.3	−12.2	1×10^{-8}
29	1 6 7	0.090	−33.0	−12.6	9.4×10^{-9}
30	1 5 6 7	0.089	−32.3	−14.5	2.9×10^{-9}
31	1 4 6 7	0.089	−32.7	−15.4	7.4×10^{-10}
32	1 4 5 6 7	0.089	−32.3	−16.8	2.8×10^{-10}

and the associated uncertainty of the parameters b_1, b_2, ..., b_7 are 0.21, 0.02, 0.006, 0.000 47, 0.15, 0.08 and 0.097, respectively. Of course, one may consider to include five or even more model classes in Equation (6.90) but the results will be virtually the same since the plausibilities of the fifth or below model classes are small. One important point is that this multi-mode predictive formula includes all seven parameters but it is not the *optimal model* in the full model class (eighth model class in Table 6.9) as their parameter values are not the same. This multi-model predictive formula (along with the posterior uncertainty of the parameters) does not have the most powerful data fitting capability but it possesses a higher level of robustness than the optimal model in the full model class. In the same fashion as Figure 6.9, Figure 6.10 summarizes the result of Equation (6.91) for site class A. Again, the six curves (from top to bottom) show the $\log_{10} PGA$ versus the magnitude of the earthquake M for six levels of distance, namely 10, 50, 100, 250, 500 and 1000 km.

www.EngineeringBooksPdf.com

Table 6.9 Optimal parameters of each predictive model class (Xinjiang)

Number	b_1	b_2	b_3	b_4	b_5	b_6	b_7
1	1.9	0.22	—	—	−0.40	0.24	0.060
2	2.4	0.24	—	0.0015	−0.74	0.28	0.11
3	2.1	—	0.067	—	−0.40	0.24	0.043
4	2.5	—	0.073	0.0015	−0.75	0.27	0.092
5	2.2	0.17	—	—	−0.40	—	—
6	2.3	—	0.052	—	−0.41	—	—
7	2.0	0.066	0.049	—	−0.41	0.24	0.046
8	2.5	0.082	0.05	0.0015	−0.76	0.27	0.097
9	2.6	0.19	—	0.0013	−0.68	—	—
10	2.7	—	0.057	0.0013	−0.70	—	—
11	1.4	0.19	—	—	—	0.20	−0.030
12	2.3	0.032	0.042	—	−0.41	—	—
13	1.5	—	0.057	—	—	0.20	−0.046
14	2.7	0.044	0.045	0.0013	−0.71	—	—
15	1.6	0.12	—	—	—	—	—
16	1.4	0.19	—	−0.000 62	—	0.20	−0.018
17	1.6	—	0.059	−0.000 65	—	0.20	−0.035
18	1.5	0.053	0.042	—	—	0.20	−0.045
19	1.7	—	0.038	—	—	—	—
20	1.7	0.13	—	−0.000 71	—	—	—
21	1.7	—	0.04	−0.000 73	—	—	—
22	1.5	0.051	0.044	−0.000 65	—	0.20	−0.033
23	1.7	0.02	0.032	—	—	—	—
24	1.7	0.018	0.035	−0.000 73	—	—	—
25	1.8	—	—	—	—	—	—
26	1.9	—	—	—	−0.083	—	—
27	1.8	—	—	−0.000 32	—	—	—
28	1.9	—	—	−0.000 16	−0.053	—	—
29	1.7	—	—	—	—	0.093	0.12
30	1.9	—	—	—	−0.14	0.10	0.16
31	1.7	—	—	−0.000 38	—	0.093	0.13
32	2.0	—	—	0.000 13	−0.17	0.10	0.16

6.5.4 Concluding Remarks

The Bayesian model class selection approach was applied for selection of the seismic attenuation model. In order to balance between the capability of data fitting and the robustness to noise and modeling error, the predictive model class is selected by maximizing the plausibility of the model class given the strong-motion records. A database of 249 strong-motion records in the Tangshan and Xinjiang regions of China was utilized for the analysis. It turned out that the most plausible model class is not the full model class even though the latter gives the smallest fitting error. The most plausible model class is less prone to noise so a more reliable future prediction can be anticipated. If several predictive model classes possess similar plausibilities given the measurements, the multi-model predictive formula can be considered as in the case of the Xinjiang region. The Bayesian approach allows us to obtain not only the optimal

www.EngineeringBooksPdf.com

Table 6.10 Uncertainty of the parameters for each predictive model class (Xinjiang)

Number	b_1	b_2	b_3	b_4	b_5	b_6	b_7
1	0.17	0.03	—	—	0.11	0.08	0.10
2	0.25	0.24	—	0.000 68	0.19	0.08	0.10
3	0.17	—	0.009	—	0.11	0.08	0.10
4	0.25	—	0.009	0.000 67	0.19	0.08	0.10
5	0.16	0.03	—	—	0.11	—	—
6	0.17	—	0.008	—	0.11	—	—
7	0.18	0.10	0.029	—	0.11	0.08	0.10
8	0.25	0.10	0.029	0.000 67	0.19	0.08	0.10
9	0.25	0.03	—	0.000 69	0.19	—	—
10	0.25	—	0.0085	0.000 68	0.19	—	—
11	0.09	0.03	—	—	—	0.08	0.10
12	0.17	0.10	0.03	—	0.11	—	—
13	0.08	—	0.009	—	—	0.08	0.10
14	0.26	0.10	0.03	0.000 68	0.19	—	—
15	0.04	0.03	—	—	—	—	—
16	0.09	0.03	—	0.000 41	—	0.08	0.10
17	0.08	—	0.009	0.000 41	—	0.08	0.10
18	0.10	0.10	0.03	—	—	0.08	0.10
19	0.03	—	0.008	—	—	—	—
20	0.04	0.03	—	0.000 43	—	—	—
21	0.03	—	0.008	0.000 43	—	—	—
22	0.10	0.10	0.03	0.000 41	—	0.08	0.10
23	0.07	0.11	0.032	—	—	—	—
24	0.07	0.10	0.031	0.000 43	—	—	—
25	0.02	—	—	—	—	—	—
26	0.17	—	—	—	0.11	—	—
27	0.03	—	—	0.000 46	—	—	—
28	0.25	—	—	0.000 73	0.18	—	—
29	0.09	—	—	—	—	0.09	0.10
30	0.20	—	—	—	0.12	0.09	0.11
31	0.09	—	—	0.000 46	—	0.09	0.10
32	0.29	—	—	0.000 76	0.21	0.09	0.11

parameters within a model class but also the associated uncertainties of the parameter values. The quantified uncertainty can be further used for uncertainty analysis of the prediction.

6.6 Prior Distributions – Revisited

Prior distribution does not significantly affect the parametric identification (both identified values and associated uncertainty) results if it is sufficiently flat in the range with significant likelihood values. Therefore, it is common to absorb the prior distribution into the normalizing constant and the results are equivalent to the maximum likelihood solution. However, it is not appropriate to absorb the prior distribution into the normalizing constant for model class

www.EngineeringBooksPdf.com

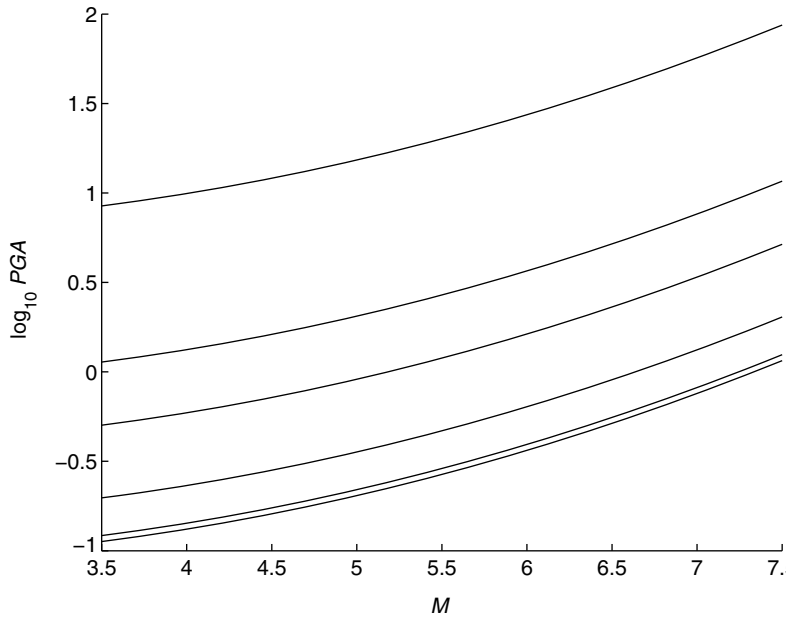

Figure 6.10 Design chart for site class A (Xinjiang)

selection. The evidence integral is the inner product of the prior PDF and the likelihood function:

$$p(\mathcal{D}|\mathcal{C}_j) = \int_{\Theta} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}_j) p(\boldsymbol{\theta}|\mathcal{C}_j) \mathrm{d}\boldsymbol{\theta} \tag{6.92}$$

It is clear that the prior distribution affects the modal class selection results. Therefore, the choice of prior distribution is important for model class selection because it offers a reference for comparison in quantifying the information gained from the data. The prior distribution expresses how much previous experience or information a user has about a model class. A more informative prior distribution is used if the user has more experience of the model class. The evidence of such a model class is surplus due to the lifting of the prior PDF. However, inappropriate previous information on the parameters will be penalized by the small value of the inner product of the prior distribution and the likelihood function. In general, it is more difficult to give the prior distribution for empirical models since the physical meaning of the parameters are not as obvious as physical models. More investigations are needed to explore further in this direction.

For a given likelihood function, the bounds are given by:

$$\inf_{\boldsymbol{\theta}} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}_j) \leq p(\mathcal{D}|\mathcal{C}_j) \leq \sup_{\boldsymbol{\theta}} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}_j) \tag{6.93}$$

Note that $\inf_{\boldsymbol{\theta}} p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{C}_j) = 0$ if the range of any of the parameters is infinite. Equalities hold if the prior distribution is a multi-dimensional Dirac delta function centered at the parameter vector where the likelihood function is maximized or minimized. However, it is unusual to achieve the bounds since the prior distribution is too informative.

www.EngineeringBooksPdf.com

In general the Bayesian model selection approach favors model classes with physical meaning because the prior distributions of the physical parameters can be obtained with smaller uncertainties. As a result, the ratio between the posterior and prior uncertainties will be relatively larger so the penalty by the Ockham factor will be smaller. For example, consider a finite-element model for a building and its uncertain stiffness parameters. Before the data are obtained, it is possible to estimate these parameters from the structural drawings and the properties of the materials. Although there is an uncertainty due to the discrepancy of the materials properties, modeling error and possible damages, the prior uncertainties of these parameters can be kept at a reasonably small level. On the other hand, it is difficult to estimate, prior to the data, the uncertain parameters of an empirical model so the prior distribution has to cover a relatively large range. Therefore, the penalty by the Ockham factor will be heavier since the ratio between the posterior and prior uncertainty will be smaller.

There is no direct relationship between the penalty by the Ockham factor and the number of degrees of freedom of a model. For a model with a large number of degrees of freedom, a large number of uncertain parameters normally come along and this will indirectly increase the penalty. However, it is also possible to impose extra mathematical constraints to reduce the number of uncertain parameters. For example, consider a structural truss model with 5000 members. The number of uncertain parameters can be reduced by grouping the members into several categories only and use one uncertain stiffness parameter for each of them. However, if the extra constraints by equating some of the uncertain parameters are inaccurate, the model class will be penalized by the likelihood function, but not the Ockham factor.

Note that it is appropriate to select the prior distribution and its boundaries by previous experience or physical laws but not by observing the measurements. In other words, it is inappropriate to select the prior distribution by observing the bounds of the measurements.

6.7 Final Remarks

The book 'The Black Swan: The Impact of the Highly Improbable' [257] received great attention due to the global financial crisis in the Fall of 2008. There is an example about a turkey with its 1001 days. Here, this example is slightly modified for discussion in the Bayesian point of view. A *mathematically gifted goose*, called Goody, was raised by a farmer. Goody observed that there were 1.0% of the geese being killed in the farm every day and he started to estimate the probability of survival for the next day. The initial probability was 0.99 due to his observation and this was his prior information for estimating the *death rate* in the Poisson model used in Chapter 2. Goody decided to use the conjugate prior distribution, i.e., the Gamma distribution in this case, with a shape parameter equal to 2 and a scale parameter equal to 0.005:

$$p(\lambda|\mathcal{C}) = 200^2 \lambda \mathrm{e}^{-200\lambda} \tag{6.94}$$

This distribution has the most probable value (mode) 0.005, mean 0.01 and standard deviation $\sqrt{2} \times 0.005$. After N days of survival, the likelihood function was given by:

$$p(\mathcal{D}|\lambda, \mathcal{C}) = N\mathrm{e}^{-N\lambda} \tag{6.95}$$

as shown in Chapter 2. Then, Goody updated the rate of death by the Bayes' theorem:

$$p(\lambda|\mathcal{D}, \mathcal{C}) = \kappa_0 p(\lambda|\mathcal{C}) p(\mathcal{D}|\lambda, \mathcal{C}) = 200^2 N \kappa_0 \lambda \mathrm{e}^{-(N+200)\lambda} \tag{6.96}$$

www.EngineeringBooksPdf.com

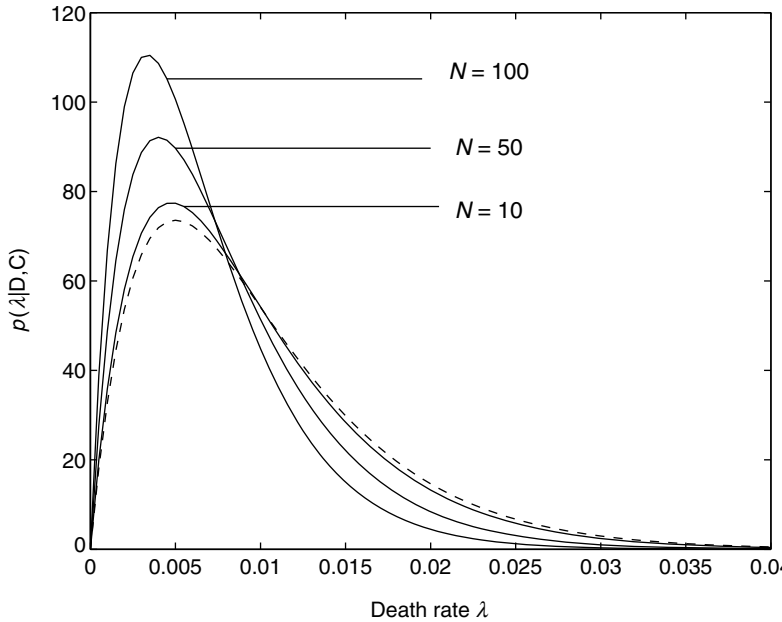

Figure 6.11 Prior and posterior PDFs of the death rates after 10, 50 and 100 days

The posterior PDFs were Gamma distributed and they are shown in Figure 6.11 for $N = 10, 50,$ and 100. The prior PDF is also shown with a dashed line for reference. The most probable value of the distribution in Equation (6.96) is given by:

$$\lambda^{\star} = \frac{1}{N + 200} \tag{6.97}$$

The mean value is $2/(N + 200)$ and the standard deviation is $\sqrt{2}/(N + 200)$. Therefore, the coefficient of variation was always $1/\sqrt{2}$. As time went by, the updated death rate kept on decreasing with a decreasing standard deviation of the estimation.

Meanwhile, the farmer also estimated the day for killing Goody. At the beginning, his prior information was also based on the data obtained from his experience in the farm. Since Goody was raised for selling to the market, the farmer would kill it if its weight reached a prescribed level w'. Also, he monitored the weight of Goody everyday, denoted by $w_0, w_1, \ldots, w_N$, so the growth on the nth day was defined as:

$$\Delta w_n = w_n - w_{n-1}, \; n = 1, 2, 3, \ldots \tag{6.98}$$

The increment of weight was a discrete stochastic process and it was modeled as a uniformly distributed i.i.d. with range $[0, U]$, where the upper bound U is uncertain. The farmer updated the posterior PDF on the nth day as follows. First, the likelihood function is given by:

$$p(\Delta w_1, \Delta w_2, \ldots, \Delta w_n | U, \mathcal{C}) = \begin{cases} U^{-n}, & \text{if } \Delta w_k \leq U, \; \forall k = 1, 2, \ldots, n \\ 0, & \text{otherwise} \end{cases} \tag{6.99}$$

www.EngineeringBooksPdf.com

To simplify the problem, a non-informative prior distribution for U was used so the updated PDF for U was:

$$p(U|\Delta w_1, \Delta w_2, \ldots, \Delta w_n, \mathcal{C}) = \begin{cases} \kappa_1 U^{-n}, & \text{if } U \geq \delta_n \\ 0, & \text{otherwise} \end{cases} \tag{6.100}$$

where δ_n was the maximum daily growth up to the nth day:

$$\delta_n \equiv \max_{k=1,2,\ldots,n} \Delta w_k \tag{6.101}$$

The normalizing constant is:

$$\kappa_1 = (n-1)\delta_n^{n-1}, \; n = 2, 3, 4, \ldots \tag{6.102}$$

and so the mean is equal to:

$$E(U|\Delta w_1, \Delta w_2, \ldots, \Delta w_n, \mathcal{C}) = \frac{n-1}{n-2}\delta_n \tag{6.103}$$

For a given value of U, the PDF of Goody's weight on any future day could be readily obtained. By integrating from w' to infinity, the probability of killing Goody on or before that day could be obtained. Then, by using the updated PDF of U in Equation (6.100) and the law of total probability, the updated cumulative distribution of killing Goody on or before a particular day could be computed and the most probable killing day and the confidence interval were readily obtained. This procedure could be proceeded daily to update the results.

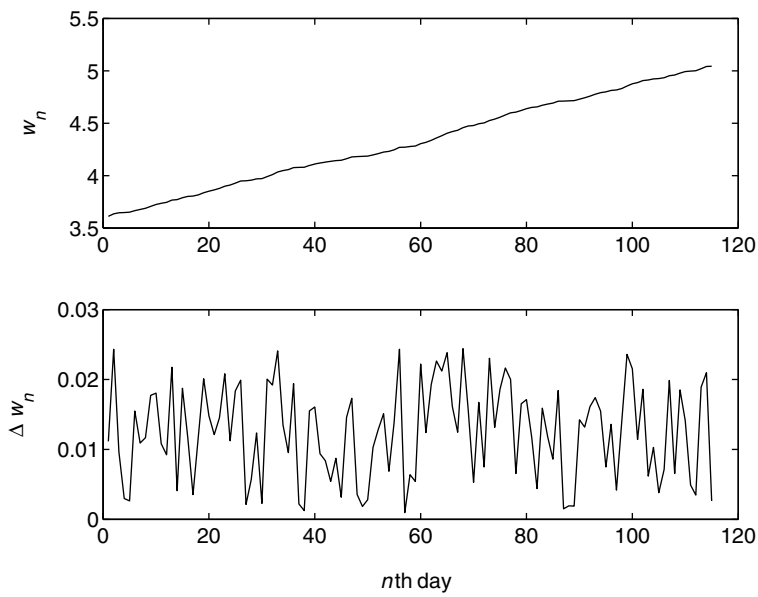

Figure 6.12 Weight of Goody and daily growth

www.EngineeringBooksPdf.com

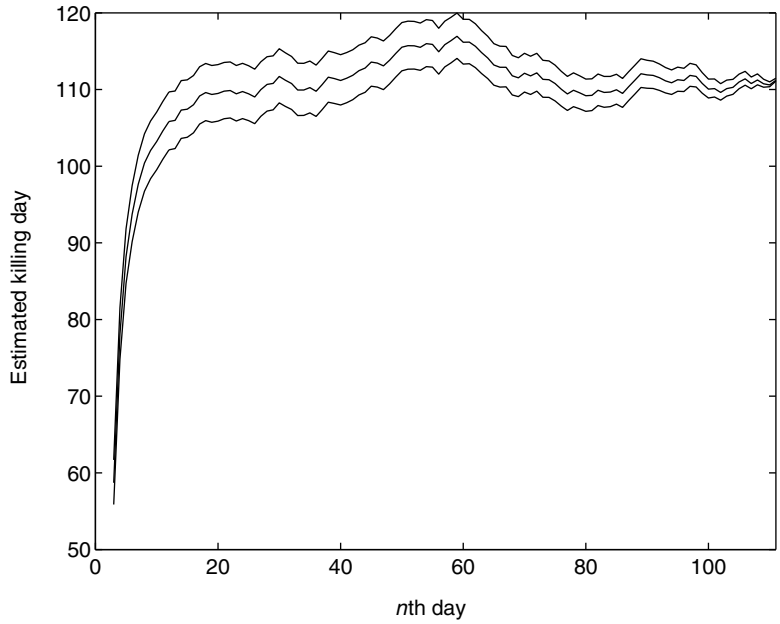

Figure 6.13 Prediction by the farmer

Here, the problem is simplified for demonstration. Assume that the farmer simply updated daily the upper bound of the uniform distribution U but treated it as a fixed value instead of a random variable. In other words, the updated value of U on the nth day was obtained:

$$U_n = \frac{n-1}{n-2}\delta_n \tag{6.104}$$

Now, the weight of Goody on a future day $n' > n$ could be estimated with mean:

$$E(w_{n'}|\Delta w_1, \Delta w_2, \ldots, \Delta w_n, \mathcal{C}) = w_n + \frac{n-1}{2(n-2)}(n'-n)\delta_n \tag{6.105}$$

and standard deviation:

$$\sigma_{w_n} = \frac{n-1}{n-2}\sqrt{\frac{n'-n}{12}}\delta_n \tag{6.106}$$

The expected day for killing Goody was obtained by solving the following equation:

$$w_n + \frac{n-1}{2(n-2)}(n'-n)\delta_n = w' \tag{6.107}$$

and is given by:

$$n'^{\star} = n + \frac{2(n-2)(w'-w_n)}{(n-1)\delta_n} \tag{6.108}$$

www.EngineeringBooksPdf.com

A typical simulated time history of Goody's weight is shown in Figure 6.12. The upper subplot shows the weight of Goody versus the number of days and the lower subplot shows the growth of each day. The corresponding estimation by the farmer can be found in Figure 6.13. The upper and lower bounds of the confidence interval represent the days with probabilities 75% and 25%, respectively, that Goody achieves the target weight on or before. Therefore, the confidence interval covers a range with 50% probability. In this example, Goody was killed on the 107th day when it obtained the lowest ever probability of being killed. On the other hand, on the 106th day, the farmer was almost sure to kill Goody on the next day.

Did Goody make any wrong calculation? Why was Goody killed when it obtained the lowest ever probability of being killed? Is our logic the same as Goody in our matters? Why do more people consider insurance policies after major earthquakes?

www.EngineeringBooksPdf.com

Appendix A

Relationship between the Hessian and Covariance Matrix for Gaussian Random Variables

Consider a Gaussian random vector $\boldsymbol{\theta}$ with mean $\boldsymbol{\theta}^\star$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ so its joint probability density function (PDF) is given by:

$$p(\boldsymbol{\theta}) = (2\pi)^{-\frac{N_{\boldsymbol{\theta}}}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star)\right] \tag{A.1}$$

The objective function can be defined as its negative logarithm:

$$J(\boldsymbol{\theta}) \equiv -\ln p(\boldsymbol{\theta}) = \frac{N_{\boldsymbol{\theta}}}{2} \ln 2\pi + \frac{1}{2} \ln |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}| + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star) \tag{A.2}$$

which is a quadratic function of the components in $\boldsymbol{\theta}$. By taking partial differentiations with respect to θ_l and $\theta_{l'}$, the (l, l') component of the Hessian matrix can be obtained:

$$\mathcal{H}^{(l,l')}(\boldsymbol{\theta}^\star) = \left.\frac{\partial^2 J(\boldsymbol{\theta})}{\partial\theta_l \partial\theta_{l'}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} = (\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1})^{(l,l')} \tag{A.3}$$

so the Hessian matrix is equal to the inverse of the covariance matrix:

$$\mathcal{H}(\boldsymbol{\theta}^\star) = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \tag{A.4}$$

For Gaussian random variables, the second derivatives of the objective function are constant for all $\boldsymbol{\theta}$ because the objective function is a quadratic function of $\boldsymbol{\theta}$. Therefore, the Hessian matrix can be computed without obtaining the mean vector $\boldsymbol{\theta}^\star$.

The elements in the Hessian matrix carry the conditional information of the random vector because they are obtained by fixing all other parameters. The diagonal elements are the curvature of the objective function in the corresponding direction. The reciprocals of these diagonal

Bayesian Methods for Structural Dynamics and Civil Engineering Ka-Veng Yuen
© 2010 John Wiley & Sons (Asia) Pte Ltd

www.EngineeringBooksPdf.com

elements are the conditional variances of the uncertain parameters in $\boldsymbol{\theta}$. However, the diagonal elements in the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ are the marginal variances of the parameters.

In many applications, the objective function is known only implicitly so the components of the Hessian matrix has to be computed numerically, e.g., by the finite difference method, and the diagonal elements are given by:

$$\begin{aligned}
\mathcal{H}^{(l,l)}(\boldsymbol{\theta}^\star) &= \left[\frac{\partial}{\partial\theta_l}\left(\frac{\partial J(\boldsymbol{\theta})}{\partial\theta_l}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \\
&\approx \frac{1}{\Delta\theta_l}\left[\left.\frac{\partial J(\boldsymbol{\theta})}{\partial\theta_l}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_l/2} - \left.\frac{\partial J(\boldsymbol{\theta})}{\partial\theta_l}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_l/2}\right] \\
&\approx \frac{1}{\Delta\theta_l}\left[\frac{J(\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_l)-J(\boldsymbol{\theta}^\star)}{\Delta\theta_l} - \frac{J(\boldsymbol{\theta}^\star)-J(\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_l)}{\Delta\theta_l}\right] \\
&= \frac{J(\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_l)-2J(\boldsymbol{\theta}^\star)+J(\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_l)}{(\Delta\theta_l)^2}
\end{aligned} \tag{A.5}$$

where $\Delta\boldsymbol{\theta}_l$ is a vector with all elements being zero except the lth element equal to a properly selected step $\Delta\theta_l$ (> 0):

$$\Delta\boldsymbol{\theta}_l = [0, \ldots, 0, \Delta\theta_l, 0, \ldots, 0]^T \tag{A.6}$$

Furthermore, the off-diagonal elements can be computed as follows:

$$\begin{aligned}
\mathcal{H}^{(l,l')}(\boldsymbol{\theta}^\star) &= \left[\frac{\partial}{\partial\theta_{l'}}\left(\frac{\partial J(\boldsymbol{\theta})}{\partial\theta_l}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \\
&\approx \frac{1}{2\Delta\theta_{l'}}\left[\left.\frac{\partial J(\boldsymbol{\theta})}{\partial\theta_l}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_{l'}} - \left.\frac{\partial J(\boldsymbol{\theta})}{\partial\theta_l}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_{l'}}\right] \\
&\approx \frac{1}{2\Delta\theta_{l'}}\left[\frac{J(\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_l+\Delta\boldsymbol{\theta}_{l'})-J(\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_l+\Delta\boldsymbol{\theta}_{l'})}{2\Delta\theta_l}\right. \\
&\qquad\qquad \left. - \frac{J(\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_l-\Delta\boldsymbol{\theta}_{l'})-J(\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_l-\Delta\boldsymbol{\theta}_{l'})}{2\Delta\theta_l}\right] \\
&= \frac{1}{4\Delta\theta_l\Delta\theta_{l'}}[J(\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_l+\Delta\boldsymbol{\theta}_{l'}) - J(\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_l-\Delta\boldsymbol{\theta}_{l'}) \\
&\qquad\qquad - J(\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_l+\Delta\boldsymbol{\theta}_{l'}) + J(\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_l-\Delta\boldsymbol{\theta}_{l'})]
\end{aligned} \tag{A.7}$$

where $\Delta\boldsymbol{\theta}_l$ and $\Delta\boldsymbol{\theta}_{l'}$ are vectors with zero elements except the lth and l'th elements equal to $\Delta\theta_l$ and $\Delta\theta_{l'}$, respectively.

Example. Gaussian Random Variable
Assume that θ is a Gaussian random variable with mean μ and variance σ^2 so its PDF is:

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(\theta-\mu)^2}{2\sigma^2}\right] \tag{A.8}$$

www.EngineeringBooksPdf.com

Then, the objective function is given by:

$$J(\theta) \equiv -\ln\ p(\theta) = \frac{1}{2}\ln(2\pi) + \ln\ \sigma + \frac{(\theta-\mu)^2}{2\sigma^2} \tag{A.9}$$

which is a quadratic function of θ.

It is assumed that the objective function is known only implicitly so the finite difference method is used to estimate the Hessian and, hence, the variance of this random variable. By Equation (A.5), the Hessian can be evaluated at an arbitrary point θ':

$$\begin{aligned}
\mathcal{H}(\theta') &= \tfrac{J(\theta'+\Delta\theta)-2J(\theta')+J(\theta'-\Delta\theta)}{(\Delta\theta)^2} \\
&= \tfrac{1}{(\Delta\theta)^2}\left\{\left[\tfrac{1}{2}\ln(2\pi)+\ln\ \sigma+\tfrac{(\theta'+\Delta\theta-\mu)^2}{2\sigma^2}\right]-2\left[\tfrac{1}{2}\ln(2\pi)+\ln\ \sigma+\tfrac{(\theta'-\mu)^2}{2\sigma^2}\right]\right. \\
&\qquad\left.+\left[\tfrac{1}{2}\ln(2\pi)+\ln\ \sigma+\tfrac{(\theta'-\Delta\theta-\mu)^2}{2\sigma^2}\right]\right\} \\
&= \tfrac{1}{(\Delta\theta)^2}\left[\tfrac{(\Delta\theta)^2}{\sigma^2}\right] \\
&= \sigma^{-2}
\end{aligned} \tag{A.10}$$

It does not depend on the point θ' and the step size $\Delta\theta$ and this solution is exact. However, for other distributions, Equation (A.4) is not correct but it provides a good approximation if the uncertainty is small. Nevertheless, the Hessian calculated by the finite difference method depends on the point of evaluation and the step size. The point of evaluation can be fixed at the most probable value of the distribution and selection of the step size has to be carefully handled. This will be discussed in the next example.

Example. Gamma Random Variable
Assume that the positive-valued random variable θ is Gamma distributed with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$. The PDF of θ is given by:

$$p(\theta) = \frac{\theta^{\alpha-1}\exp(-\theta/\beta)}{\beta^{\alpha}\Gamma(\alpha)},\ \theta > 0 \tag{A.11}$$

where Γ is the Gamma function. The mean and variance of this distribution are given by $\alpha\beta$ and $\alpha\beta^2$, respectively. If $\alpha = 1$, this distribution is deduced to the exponential distribution and the most probable value is $\theta^\star = 0$. If $\alpha > 1$, the most probable parameter is given by:

$$\theta^\star = (\alpha-1)\beta \tag{A.12}$$

Furthermore, the objective function can be defined as the negative logarithm of the PDF:

$$J(\theta) \equiv -\ln p(\theta) = -(\alpha-1)\ln\theta + \frac{\theta}{\beta} + \alpha\ln\beta + \ln\Gamma(\alpha),\ \theta > 0 \tag{A.13}$$

www.EngineeringBooksPdf.com

Again, it is assumed that the objective function is known only implicitly. By Equation (A.5), the Hessian can be computed:

$$\begin{aligned}\mathcal{H}(\theta^\star) &= \frac{1}{(\Delta\theta)^2}\left\{\left[-(\alpha-1)\ln(\theta^\star+\Delta\theta)+\frac{\theta^\star+\Delta\theta}{\beta}\right]-2\left[-(\alpha-1)\ln\theta^\star+\frac{\theta^\star}{\beta}\right]\right.\\ &\qquad\left.+\left[-(\alpha-1)\ln(\theta^\star-\Delta\theta)+\frac{\theta^\star-\Delta\theta}{\beta}\right]\right\}\\ &= \frac{(\alpha-1)}{(\Delta\theta)^2}\left[-\ln(\theta^\star+\Delta\theta)+2\ln\theta^\star-\ln(\theta^\star-\Delta\theta)\right]\\ &= -\frac{(\alpha-1)}{(\Delta\theta)^2}\left[\ln\left(1+\frac{\Delta\theta}{\theta^\star}\right)+\ln\left(1-\frac{\Delta\theta}{\theta^\star}\right)\right]\end{aligned} \tag{A.14}$$

In contrast to the Gaussian random variable, this solution depends on the point for evaluation and also the finite-difference step size.

For a random variable with a small uncertainty, the variance can be approximated by $\mathcal{H}(\theta^\star)^{-1}$. To demonstrate this, let $\alpha = \mu/\beta$ for Equation (A.11) so the mean is μ and the variance is $\mu\beta$. For small β, the Hessian estimated by finite difference method can be approximated by:

$$\begin{aligned}\mathcal{H}(\theta^\star) &= -\frac{(\alpha-1)}{(\Delta\theta)^2}\left[\ln\left(1+\frac{\Delta\theta}{\theta^\star}\right)+\ln\left(1-\frac{\Delta\theta}{\theta^\star}\right)\right]\\ &\approx -\frac{(\alpha-1)}{(\Delta\theta)^2}\left[\frac{\Delta\theta}{\theta^\star}-\frac{1}{2}\left(\frac{\Delta\theta}{\theta^\star}\right)^2+\cdots-\frac{\Delta\theta}{\theta^\star}-\frac{1}{2}\left(\frac{\Delta\theta}{\theta^\star}\right)^2+\cdots\right]\\ &\approx \frac{(\alpha-1)}{(\Delta\theta)^2}\left(\frac{\Delta\theta}{\theta^\star}\right)^2\\ &= \frac{1}{(\mu-\beta)\beta}\\ &\approx \frac{1}{\mu\beta}\end{aligned} \tag{A.15}$$

Therefore, the Hessian estimated variance provides a good approximation: $\mathcal{H}(\theta^\star)^{-1} \approx \mu\beta$. In the following, two cases are demonstrated and they correspond to a large variance and a small variance.

Case 1: $\alpha = 10$ and $\beta = 0.1$

This distribution corresponds to the more spread PDF in Figure A.1. In this case, the mean and variance are $\alpha\beta = 1$ and $\alpha\beta^2 = 0.1$, respectively. Therefore, the coefficient of variation is $1/\sqrt{10} \approx 32\%$ and it represents a case of large uncertainty. By using Equation (A.14), the Hessian and the estimated variance are shown in Figure A.2 for different finite-difference step sizes up to 0.5. It is clearly seen that the estimation depends on the step size. The correct value of the variance is 0.1 but the estimation is 10% off for this random variable with a large coefficient of variation.

www.EngineeringBooksPdf.com

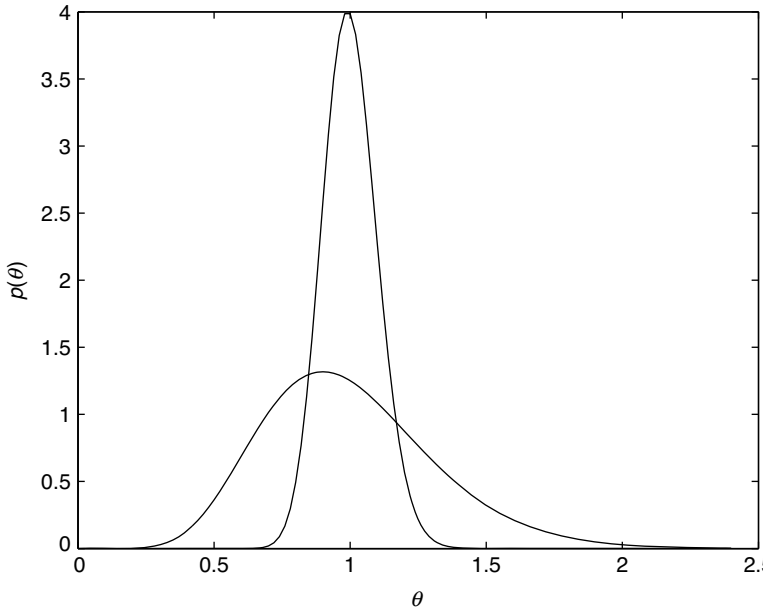

Figure A.1 Gamma distributions

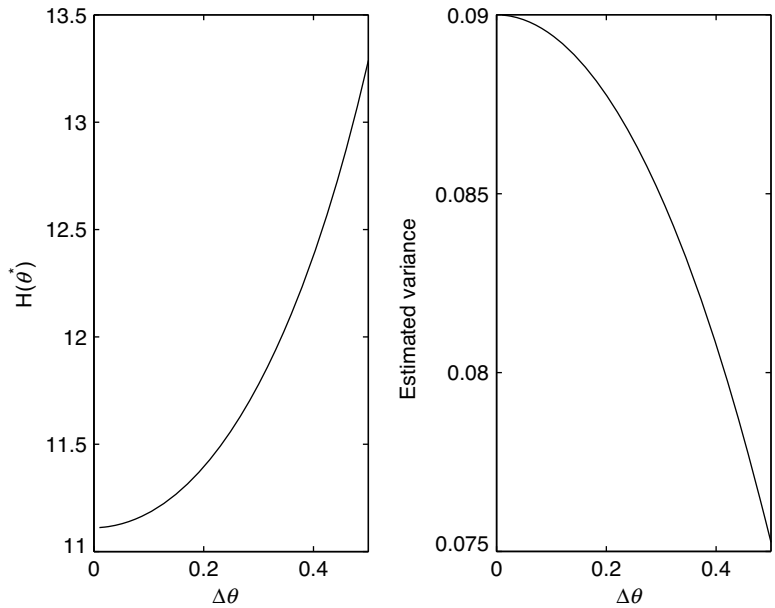

Figure A.2 Finite-difference estimated Hessian and variance ($\alpha = 10$, $\beta = 0.1$)

www.EngineeringBooksPdf.com

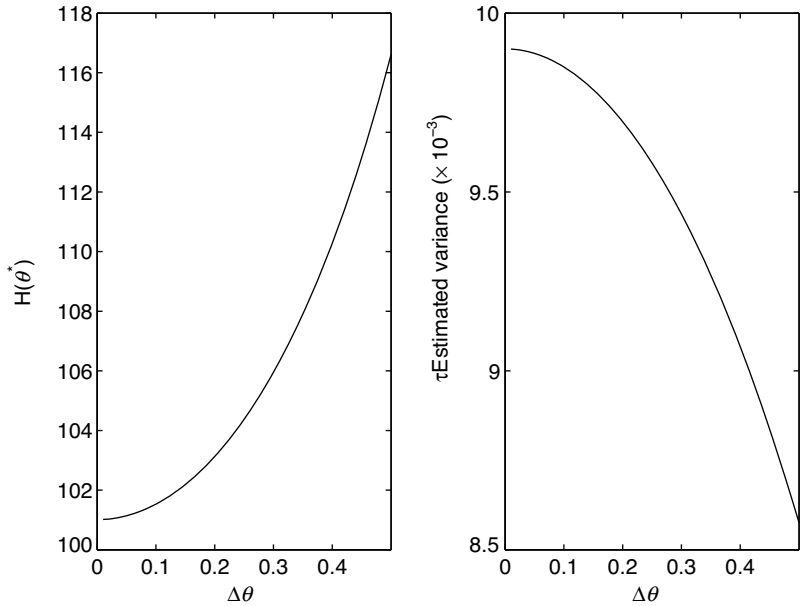

Figure A.3 Finite-difference estimated Hessian and variance ($\alpha = 100$, $\beta = 0.01$)

Case 2: $\alpha = 100$ and $\beta = 0.01$
This distribution corresponds to the more concentrated PDF in Figure A.1. In this case, the mean and variance are $\alpha\beta = 1$ and $\alpha\beta^2 = 0.01$, respectively, so the coefficient of variation is 10%. This random variable has the same mean as Case 1 but a smaller variance. Figure A.3 shows the Hessian and the estimated variance using different finite-difference step sizes. Again, the estimation depends on the step size. The correct value of the variance is 0.01 and the error of the estimation is about 1% when the step size is small. In general, the finite-difference estimated variance is more accurate for a smaller coefficient of variation of the distribution because the local topology of the PDF is more representative for the global distribution in this case. This example shows that this approximation is acceptable up to 10% COV.

www.EngineeringBooksPdf.com

Appendix B

Contours of Marginal PDFs for Gaussian Random Variables

Two methods are introduced in this appendix for drawing the contours of marginal PDFs.

Method 1. Eigenvalue Problem Method

Consider a vector of two Gaussian random variables $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ with mean $\boldsymbol{\theta}^\star = [\theta_1^\star, \theta_2^\star]^T$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$. The goal here is to obtain the parametric form of the joint PDF contour that covers an area with a prescribed probability. First, define a vector of two new random variables $\mathbf{y} = [y_1, y_2]^T$ by the following transformation:

$$\mathbf{y} = \mathbf{P}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star) \tag{B.1}$$

where $\mathbf{P} \in \mathbb{R}^{2\times 2}$ is a fixed invertible matrix. Therefore, $\mathbf{y}$ is also Gaussian with zero mean:

$$E[\mathbf{y}] = E[\mathbf{P}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star)] = \mathbf{P}(E[\boldsymbol{\theta}] - \boldsymbol{\theta}^\star) = \mathbf{0} \tag{B.2}$$

and covariance matrix:

$$\boldsymbol{\Sigma}_{\mathbf{y}} \equiv E[\mathbf{y}\mathbf{y}^T] = E[\mathbf{P}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star)(\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T\mathbf{P}^T] = \mathbf{P}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{P}^T \tag{B.3}$$

If the matrix $\mathbf{P}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{P}^T$ is diagonal, say $\mathbf{D}$, the Gaussian random variables y_1 and y_2 are uncorrelated and, hence, statistically independent. Then, it is an easy task to draw the PDF contours in the $y_1 - \boldsymbol{\theta}^\star - y_2$ coordinate system. In order to obtain a solution for the matrix $\mathbf{P}$ to fulfill this goal, consider the eigenvalue problem of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$:

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{V} = \mathbf{V}\mathbf{D} \tag{B.4}$$

Since the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is symmetric, the eigenvector matrix $\mathbf{V}$ can be normalized such that $\mathbf{V}^{-1} = \mathbf{V}^T$. In particular, this matrix takes the following form:

$$\mathbf{V} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}, \ \alpha \in \mathbb{R} \tag{B.5}$$

Bayesian Methods for Structural Dynamics and Civil Engineering Ka-Veng Yuen
© 2010 John Wiley & Sons (Asia) Pte Ltd

www.EngineeringBooksPdf.com

which represents rigid body rotation. Therefore, Equation (B.4) can be rewritten as follows:

$$\mathbf{V}^T \mathbf{\Sigma}_{\boldsymbol{\theta}} \mathbf{V} = \mathbf{D} \tag{B.6}$$

By letting $\mathbf{P} = \mathbf{V}^T$, the covariance matrix of $\mathbf{y}$ in Equation (B.3) is diagonal: $\mathbf{\Sigma}_{\mathbf{y}} = \mathbf{D}$. In other words, the random variables y_1 and y_2 in the vector $\mathbf{y}$ are independent. Furthermore, the diagonal elements, D_1 and D_2, of the matrix $\mathbf{D}$ are their variances. Therefore, the joint PDF of y_1 and y_2 is:

$$p(y_1, y_2) = \frac{1}{2\pi\sqrt{D_1 D_2}} \exp\left(-\frac{y_1^2}{2D_1} - \frac{y_2^2}{2D_2}\right) \tag{B.7}$$

The column vectors in the matrix $\mathbf{V}$ indicates the principal axes in the (θ_1, θ_2) plane, i.e., the direction of the axes of the rotated coordinate system.

In order to draw the contours of the joint PDF in the (θ_1, θ_2) plane, the rotated coordinate system is utilized. By Equation (B.7), the α-contour has the following parametric form:

$$\begin{aligned} y_1 &= \alpha \cos\beta \\ y_2 &= \alpha \sin\beta, \quad \beta \in [0, 2\pi) \end{aligned} \tag{B.8}$$

where α is a distance parameter that controls the size and probability of the area enclosed by the contour. By Equation (B.1) with $\mathbf{P} = \mathbf{V}^T$, the coordinates in the original coordinate system can be obtained:

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \boldsymbol{\theta}^\star + \mathbf{V} \begin{bmatrix} y_1\sqrt{D_1} \\ y_2\sqrt{D_2} \end{bmatrix} \tag{B.9}$$

The probability of the area enclosed by this elliptical α-contour can be obtained by integrating the joint PDF in Equation (B.7) over this area:

$$P(\alpha) = 1 - \exp\left(-\alpha^2/2\right) \tag{B.10}$$

On the other hand, a contour enclosing an area of probability P has the corresponding distance parameter α:

$$\alpha = \sqrt{-2\ln(1-P)} \tag{B.11}$$

Therefore, the area of the ellipse is proportional to $-\ln(1-P)$. In particular, if one wants to draw the contour that encloses the area of 50% probability, the value of α should be taken as $\alpha = \sqrt{2\ln 2} = 1.1774$. For $P = 0.9$, $\alpha = \sqrt{2\ln 10} = 2.1460$. For $P = 0.99$, $\alpha = 2\sqrt{\ln 10} = 3.0349$. Table B.1 summarizes the relationship between P and α.

Table B.1 Relationship between P and α

P	0.1	0.2	0.3	0.4	0.5	0.6	0.7
α	0.4590	0.6680	0.8446	1.0108	1.1774	1.3537	1.5518
P	0.8	0.85	0.9	0.95	0.99	0.999	0.9999
α	1.7941	1.9479	2.1460	2.4477	3.0349	3.7169	4.2919

www.EngineeringBooksPdf.com

Example. Consider two Gaussian random variables θ_1 and θ_2 with mean $\theta_1^\star = 3$ and $\theta_2^\star = 2$ and covariance matrix:

$$\mathbf{\Sigma}_{\boldsymbol{\theta}} = \begin{bmatrix} 4 & 1.5 \\ 1.5 & 1 \end{bmatrix}$$

The eigenvalues can be obtained by solving the characteristic equation:

$$\left| \mathbf{\Sigma}_{\boldsymbol{\theta}} - D_l \mathbf{I}_2 \right| = 0, \; l = 1, 2 \tag{B.12}$$

In this case, this equation becomes:

$$\begin{vmatrix} 4 - D_l & 1.5 \\ 1.5 & 1 - D_l \end{vmatrix} = 0$$

The eigenvalues are given by $D_1 = 4.6213$ and $D_2 = 0.3787$ and the corresponding eigenvector matrix is:

$$\mathbf{V} = \begin{bmatrix} 0.9239 & -0.3827 \\ 0.3827 & 0.9239 \end{bmatrix}$$

so $\cos\alpha = 0.9239$ and $\alpha = 0.3926$ rad $\approx 22.5^\circ$.

Figure B.1 shows the three contours that cover areas with 50%, 90% and 99% of probabilities. The random variables $y_1 = 0.9239(\theta_1 - 3) + 0.3827(\theta_2 - 2)$ and $y_2 = -0.3827(\theta_1 - 3) + 0.9239(\theta_2 - 2)$ are independent Gaussians with variances 4.6213 and 0.3787, respectively.

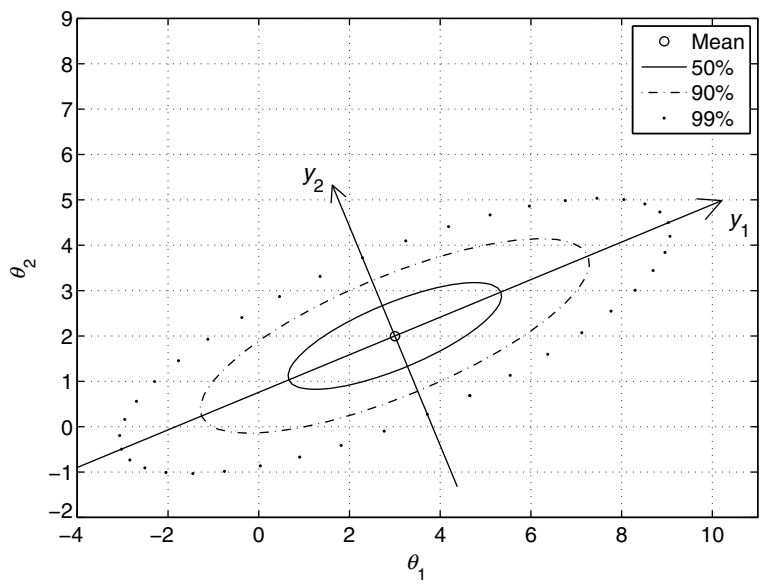

Figure B.1 Contours of the joint PDF

www.EngineeringBooksPdf.com

Method 2. Cholesky Decomposition Method

First, the covariance can be factorized by the Cholesky Decomposition since it is symmetric [189]:

$$\mathbf{\Sigma}_{\boldsymbol{\theta}} = \mathbf{L}\mathbf{L}^T$$

where $\mathbf{L}$ is a lower triangular matrix. In general, the matrix $\mathbf{L}$ can be computed by the function 'chol' in MATLAB® [171]. However, in this case with a 2×2 covariance matrix, the matrix $\mathbf{L}$ can be easily obtained:

$$\mathbf{L} = \begin{bmatrix} \sqrt{\mathbf{\Sigma}_{\boldsymbol{\theta}}^{(1,1)}} & 0 \\ \dfrac{\mathbf{\Sigma}_{\boldsymbol{\theta}}^{(2,1)}}{\sqrt{\mathbf{\Sigma}_{\boldsymbol{\theta}}^{(1,1)}}} & \sqrt{\mathbf{\Sigma}_{\boldsymbol{\theta}}^{(2,2)} - \dfrac{\left(\mathbf{\Sigma}_{\boldsymbol{\theta}}^{(2,1)}\right)^2}{\mathbf{\Sigma}_{\boldsymbol{\theta}}^{(1,1)}}} \end{bmatrix}$$

If y_1 and y_2 are independent standard Gaussian random variables, then the random variables θ_1 and θ_2 have means $\theta_1^\star$ and $\theta_2^\star$ and a covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\theta}}$ if they are defined as follows:

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} \theta_1^\star \\ \theta_2^\star \end{bmatrix} + \mathbf{L} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

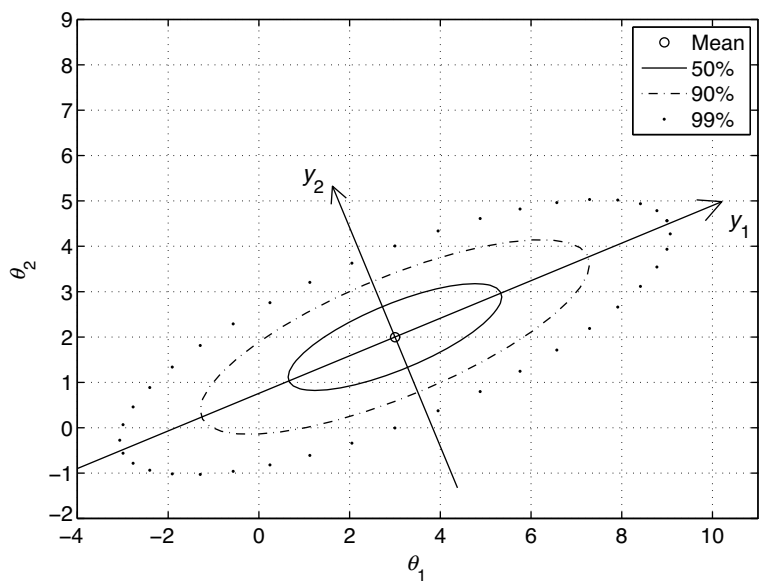

Figure B.2 Contours of the joint PDF

www.EngineeringBooksPdf.com

In a similar fashion as in the eigenvalue problem method, the α-contour can be found by taking y_1 and y_2 as:

$$\begin{aligned} y_1 &= \alpha \cos \beta \\ y_2 &= \alpha \sin \beta, \quad \beta \in [0, 2\pi) \end{aligned} \tag{B.13}$$

Then, Equation (B.13) will give the coordinates of the α-contour. Again, the probability–α relationship is the same as in Table B.1.

The example in the previous section is repeated using the Cholesky Decomposition Method and the matrix $\mathbf{L}$ is given by:

$$\mathbf{L} = \frac{1}{4} \begin{bmatrix} 8 & 0 \\ 3 & \sqrt{7} \end{bmatrix}$$

Three contours are shown in Figure B.2 and they are identical to the corresponding ones in Figure B.1. However, the transformations in these two methods are not the same. For example, the point $y_1 = 0$, $y_2 = \alpha$ will be transformed to the major/minor axis in Method 1 but this point will remain on the θ_2 axis in the second method.

www.EngineeringBooksPdf.com

www.EngineeringBooksPdf.com

Appendix C

Conditional PDF for Prediction

C.1 Two Random Variables

Consider two Gaussian random variables x_1 and x_2 with mean μ_1 and μ_2, respectively. Also, use σ_1^2 and σ_2^2 to denote their variances. The correlation coefficient is denoted by ρ, with $|\rho| \leq 1$.

If $|\rho| \neq 1$, the joint PDF of x_1 and x_2 can be written as:

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\} \tag{C.1}$$

Now, the problem of predicting x_1 with the measurement of x_2 is considered. First, by Equation (C.1), the conditional PDF $p(x_1|x_2)$ can be obtained by normalizing $p(x_1, x_2)$ with x_2 fixed at its given value:

$$\begin{aligned} p(x_1|x_2) &= \kappa_0 \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho x_1(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\} \\ &= \kappa_0' \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x_1^2}{\sigma_1^2} - \left(\frac{2\mu_1}{\sigma_1^2} + \frac{2\rho(x_2-\mu_2)}{\sigma_1\sigma_2}\right)x_1\right]\right\} \end{aligned} \tag{C.2}$$

where κ_0 and κ_0' are normalizing constants so that $\int_{-\infty}^{\infty} p(x_1|x_2)\mathrm{d}x_1 = 1$. By Equation (C.2), the conditional PDF is also Gaussian for x_1 since $\ln p(x_1|x_2)$ is a quadratic function of x_1. On the other hand, the PDF for a Gaussian random variable can be written as:

$$\begin{aligned} p(X) &= \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left[-\frac{(X-\mu_X)^2}{2\sigma_X^2}\right] \\ &= \kappa_1 \exp\left(-\frac{X^2 - 2\mu_X X}{2\sigma_X^2}\right) \end{aligned} \tag{C.3}$$

Bayesian Methods for Structural Dynamics and Civil Engineering Ka-Veng Yuen
© 2010 John Wiley & Sons (Asia) Pte Ltd

www.EngineeringBooksPdf.com

where $\kappa_1 = 1/(\sqrt{2\pi}\sigma_X)\exp\left(-\mu_X^2/2\sigma_X^2\right)$. By comparing the coefficients in the exponents of Equations (C.2) and (C.3), the following equations are obtained:

$$\begin{cases} \dfrac{1}{2\sigma_{x_1|x_2}^2} = \dfrac{1}{2(1-\rho^2)\sigma_1^2} \\ \dfrac{\mu_{x_1|x_2}}{\sigma_{x_1|x_2}^2} = \dfrac{\mu_1}{(1-\rho^2)\sigma_1^2} + \dfrac{\rho(x_2-\mu_2)}{(1-\rho^2)\sigma_1\sigma_2} \end{cases} \tag{C.4}$$

where $\mu_{x_1|x_2} \equiv E[x_1|x_2]$ and $\sigma_{x_1|x_2}^2 \equiv E\left[(x_1-\mu_{x_1|x_2})^2|x_2\right]$ are the conditional mean and variance, respectively. By solving these equations, the conditional variance $\sigma_{x_1|x_2}^2$ is given by:

$$\sigma_{x_1|x_2}^2 = (1-\rho^2)\sigma_1^2 \tag{C.5}$$

and the conditional mean $\mu_{x_1|x_2}$ is given by:

$$\mu_{x_1|x_2} = \mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2) \tag{C.6}$$

It is not surprising that a high absolute value of ρ facilitates the prediction in giving a small conditional variance.

Example. Consider two Gaussian random variables x_1 and x_2 with mean $\mu_1 = 3$ and $\mu_2 = 5$. The covariance matrix is taken as:

$$\mathbf{\Sigma_x} = \begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix}$$

Therefore, $\sigma_1 = 2$, $\sigma_2 = 3$, and $\rho = 1/2$. By Equation (C.6), the conditional mean $E[x_1|x_2]$ is given by:

$$E[x_1|x_2] = 3 + \frac{1/2 \times 2}{3}(x_2-5) = \frac{x_2+4}{3}$$

This can be used to predict x_1 if the measurement of x_2 is available. For example, if x_2 is measured to be 7, then $E[x_1|x_2 = 7] = 11/3$. Figure C.1 shows the joint PDF of x_1 and x_2. The plane $x_2 = 7$ is also shown in the figure. By Equation (C.5), the conditional variance can also be estimated:

$$\sigma_{x_1|x_2}^2 = \left[1-(1/2)^2\right]2^2 = 3$$

which does not depend on the value of x_2.

www.EngineeringBooksPdf.com

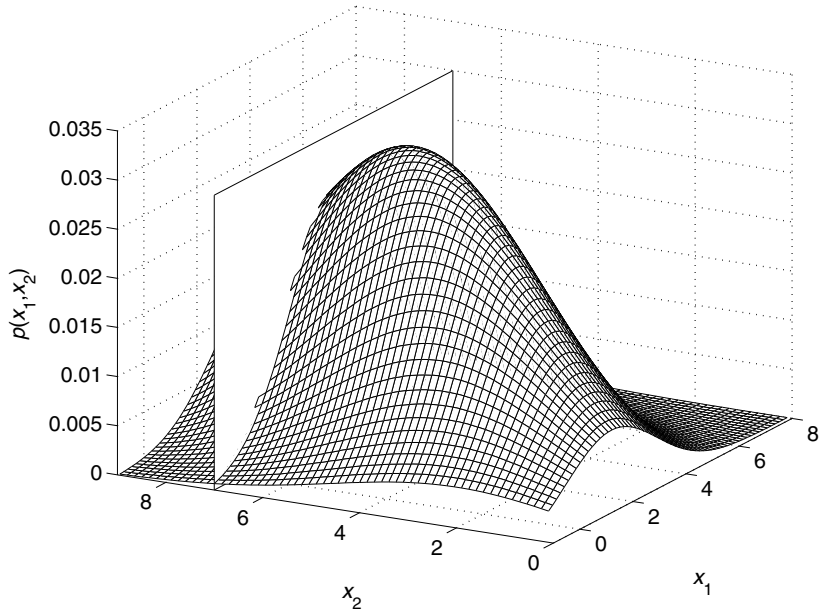

Figure C.1 Joint PDF

Figure C.2 shows the PDF $p(x_1, x_2 = 7)$ on the line $x_2 = 7$ and the peak occurs at $x_1 = 11/3$, which is the same as the conditional mean. Figure C.3 shows the contours of the joint PDF $p(x_1, x_2)$. The two solid lines show the major and minor axes of the elliptical contours. The '–.' curves show the contours enclosing the area with 50% and 90% probabilities. The horizontal

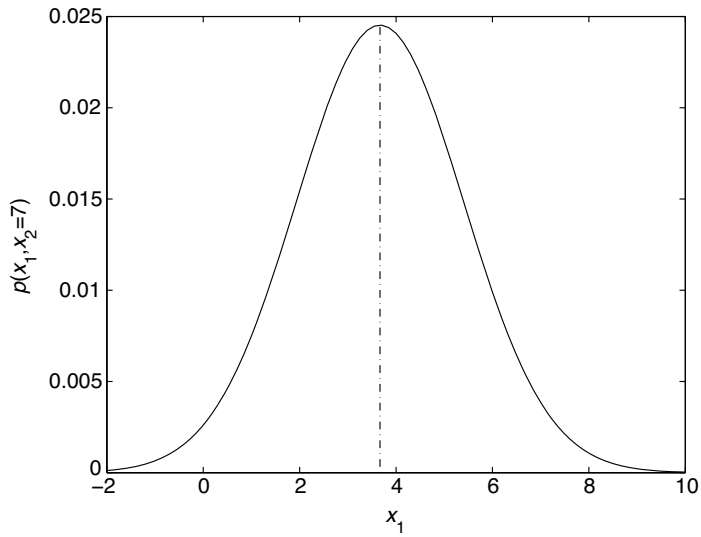

Figure C.2 Joint PDF on the line $x_2 = 7$

www.EngineeringBooksPdf.com

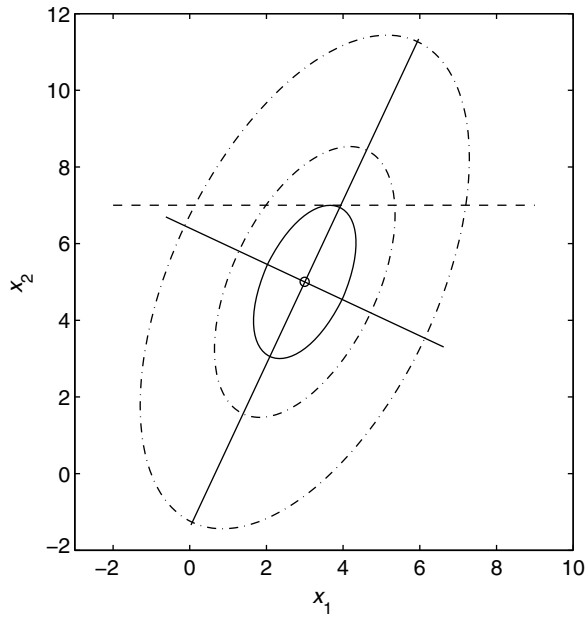

Figure C.3 Contours of the joint PDF

dashed line is $x_2 = 7$ and the solid ellipse is the contour that is tangential to $x_2 = 7$. It can be seen that the tangent point is at $x_1 = 11/3$, which is again the same as the conditional mean.

If $|\rho| = 1$, the covariance matrix $\mathbf{\Sigma_x} \equiv E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T] \in \mathbb{R}^{2\times 2}$ is singular, i.e., $|\mathbf{\Sigma_x}| = 0$, where $\mathbf{x} = [x_1, x_2]^T$ and $\boldsymbol{\mu} = [\mu_1, \mu_2]^T$. In this case, there exists a non-null vector $\boldsymbol{\phi} = [\phi_1, \phi_2]^T$ such that:

$$\mathbf{\Sigma_x}\boldsymbol{\phi} = \mathbf{0} \tag{C.7}$$

Then, a new random variable ξ is introduced:

$$\xi = \boldsymbol{\phi}^T(\mathbf{x} - \boldsymbol{\mu}) = \phi_1(x_1 - \mu_1) + \phi_2(x_2 - \mu_2) \tag{C.8}$$

It has zero mean $E[\xi] = 0$ and zero variance:

$$E[\xi^2] = E[\boldsymbol{\phi}^T(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T\boldsymbol{\phi}] = \boldsymbol{\phi}^T\mathbf{\Sigma_x}\boldsymbol{\phi} = 0 \tag{C.9}$$

In other words, $\xi = 0$ *almost everywhere (a.e.)* so:

$$\phi_1(x_1 - \mu_1) + \phi_2(x_2 - \mu_2) = 0 \tag{C.10}$$

www.EngineeringBooksPdf.com

except for a set of measure zero. Then, x_1 can be predicted with zero variance:

$$E[x_1|x_2] = \mu_1 - \frac{\phi_2}{\phi_1}(x_2 - \mu_2) \tag{C.11}$$

Example. Consider two Gaussian random variables x_1 and x_2 with mean $\mu_1 = 3$ and $\mu_2 = 5$. The covariance matrix is taken as

$$\mathbf{\Sigma_x} = \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix}$$

Therefore, $\sigma_1 = 2$, $\sigma_2 = 3$, and $\rho = 1$. One can easily find that:

$$\begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix} \begin{bmatrix} -3 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

so $\phi_1 = -3$ and $\phi_2 = 2$. This solution is non-unique and any vector parallel to it is also a solution. If x_2 is measured to be 7, then:

$$E[x_1|x_2] = 3 - \frac{2}{(-3)}(7 - 5) = 13/3$$

and the prediction has zero variance.

C.2 General Cases

Consider a Gaussian random vector $\mathbf{x} \in \mathbb{R}^{N_X}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma_x} \in \mathbb{R}^{N_X \times N_X}$. If $\mathbf{\Sigma_x}$ is non-singular, the joint PDF is given by:

$$p(\mathbf{x}) = (2\pi)^{-\frac{N_X}{2}} |\mathbf{\Sigma_x}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma_x}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right] \tag{C.12}$$

Then, the random vector $\mathbf{x}$ is partitioned as follows:

$$\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T \tag{C.13}$$

and prediction of $\mathbf{x}_1 \in \mathbb{R}^{N_1}$ given the measurement of $\mathbf{x}_2 \in \mathbb{R}^{N_2}$ is concerned. In the similar fashion as Equation (C.13), the covariance matrix is partitioned as:

$$\mathbf{\Sigma_x} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{12}^T & \mathbf{\Sigma}_{22} \end{bmatrix} \tag{C.14}$$

where $\mathbf{\Sigma}_{11} \in \mathbb{R}^{N_1 \times N_1}$, $\mathbf{\Sigma}_{12} \in \mathbb{R}^{N_1 \times N_2}$ and $\mathbf{\Sigma}_{22} \in \mathbb{R}^{N_2 \times N_2}$. In the same way, the inverse of the covariance matrix, i.e., the Hessian matrix, is partitioned:

$$\mathbf{\Sigma_x}^{-1} = \mathcal{H} = \begin{bmatrix} \mathcal{H}_{11} & \mathcal{H}_{12} \\ \mathcal{H}_{12}^T & \mathcal{H}_{22} \end{bmatrix} \tag{C.15}$$

www.EngineeringBooksPdf.com

where $\mathcal{H}_{11} \in \mathbb{R}^{N_1 \times N_1}$, $\mathcal{H}_{12} \in \mathbb{R}^{N_1 \times N_2}$ and $\mathcal{H}_{22} \in \mathbb{R}^{N_2 \times N_2}$. Then, the joint PDF of $\mathbf{x}$ in Equation (C.12) can be expanded:

$$p(\mathbf{x}) = (2\pi)^{-\frac{N_X}{2}} |\mathcal{H}|^{\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \mathcal{H}_{11}(\mathbf{x}_1 - \boldsymbol{\mu}_1) - \frac{1}{2}(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \mathcal{H}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2)\right.$$
$$\left. - \frac{1}{2}(\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \mathcal{H}_{12}^T(\mathbf{x}_1 - \boldsymbol{\mu}_1) - \frac{1}{2}(\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \mathcal{H}_{22}(\mathbf{x}_2 - \boldsymbol{\mu}_2)\right] \tag{C.16}$$

It is immediately followed that the conditional PDF $p(\mathbf{x}_1|\mathbf{x}_2)$ is given by:

$$\begin{aligned} p(\mathbf{x}_1|\mathbf{x}_2) &= \kappa_0 \exp\left[-\frac{1}{2}(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \mathcal{H}_{11}(\mathbf{x}_1 - \boldsymbol{\mu}_1) - \mathbf{x}_1^T \mathcal{H}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2)\right] \\ &= \kappa_0' \exp\left\{-\frac{1}{2}\mathbf{x}_1^T \mathcal{H}_{11}\mathbf{x}_1 + \mathbf{x}_1^T\left[\mathcal{H}_{11}\boldsymbol{\mu}_1 - \mathcal{H}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2)\right]\right\} \end{aligned} \tag{C.17}$$

where κ_0 and κ_0' are normalizing constants such that $\int p(\mathbf{x}_1|\mathbf{x}_2)\mathrm{d}\mathbf{x}_1 = 1$.

Use $\boldsymbol{\mu}_{\mathbf{x}_1|\mathbf{x}_1}$ and $\boldsymbol{\Sigma}_{\mathbf{x}_1|\mathbf{x}_1}$ to denote the conditional mean and conditional covariance matrix for $\mathbf{x}_1$ given $\mathbf{x}_2$. Then, a Gaussian conditional PDF can be written as:

$$\begin{aligned} p(\mathbf{x}_1|\mathbf{x}_2) &= \kappa_1 \exp\left[-\frac{1}{2}(\mathbf{x}_1 - \boldsymbol{\mu}_{\mathbf{x}_1|\mathbf{x}_2})^T \boldsymbol{\Sigma}_{\mathbf{x}_1|\mathbf{x}_2}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_{\mathbf{x}_1|\mathbf{x}_2})\right] \\ &= \kappa_1' \exp\left(-\frac{1}{2}\mathbf{x}_1^T \boldsymbol{\Sigma}_{\mathbf{x}_1|\mathbf{x}_2}^{-1}\mathbf{x}_1 + \mathbf{x}_1^T \boldsymbol{\Sigma}_{\mathbf{x}_1|\mathbf{x}_2}^{-1}\boldsymbol{\mu}_{\mathbf{x}_1|\mathbf{x}_2}\right) \end{aligned} \tag{C.18}$$

By comparing the exponents of Equations (C.17) and (C.18), the following equations are obtained:

$$\begin{cases} \boldsymbol{\Sigma}_{\mathbf{x}_1|\mathbf{x}_2}^{-1} = \mathcal{H}_{11} \\ \boldsymbol{\Sigma}_{\mathbf{x}_1|\mathbf{x}_2}^{-1}\boldsymbol{\mu}_{\mathbf{x}_1|\mathbf{x}_2} = \mathcal{H}_{11}\boldsymbol{\mu}_1 - \mathcal{H}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \end{cases} \tag{C.19}$$

It can be easily shown that:

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T)^{-1} & -(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T)^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T)^{-1} & (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1} \end{bmatrix} \tag{C.20}$$

with $\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T)^{-1} = [\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}]^T$. Therefore, $\mathcal{H}_{11}$ and $\mathcal{H}_{12}$ can be expressed in terms of $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$ and $\boldsymbol{\Sigma}_{22}$ as follows:

$$\mathcal{H}_{11} = (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T)^{-1} \tag{C.21}$$

and

$$\mathcal{H}_{12} = -(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T)^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \tag{C.22}$$

www.EngineeringBooksPdf.com

Substituting these equations into Equation (C.19), the conditional mean and its covariance matrix are given by:

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{x}_1|\mathbf{x}_2} \equiv E[\mathbf{x}_1|\mathbf{x}_2] &= \boldsymbol{\mu}_1 - \mathcal{H}_{11}^{-1}\mathcal{H}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)\end{aligned} \tag{C.23}$$

and

$$\boldsymbol{\Sigma}_{\mathbf{x}_1|\mathbf{x}_2} = \mathcal{H}_{11}^{-1} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T \tag{C.24}$$

If $\boldsymbol{\Sigma}_{\mathbf{x}}$ is singular, then one first diagonalizes $\boldsymbol{\Sigma}_{\mathbf{x}}$:

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \mathbf{V}\begin{bmatrix}\lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_{N_X}\end{bmatrix}\mathbf{V}^T \tag{C.25}$$

with $\mathbf{V}^T = \mathbf{V}^{-1}$. Since $\boldsymbol{\Sigma}_{\mathbf{x}}$ is singular, there is at least one zero eigenvalue. Note that the column vectors of $\mathbf{V}$ are mutually orthogonal since $\boldsymbol{\Sigma}_{\mathbf{x}}$ is symmetric. Define a partial eigenvector matrix $\mathbf{V}_0 \in \mathbb{R}^{N_X \times N_0}$ to include all eigenvectors associated with the zero eigenvalue, where N_0 is the *multiplicity* of the zero eigenvalue. It follows that the random vector:

$$\boldsymbol{\xi} = \mathbf{V}_0^T(\mathbf{x} - \boldsymbol{\mu}) \tag{C.26}$$

have zero mean and zero covariance matrix. Therefore, the linear algebraic equation:

$$\mathbf{V}_0^T(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{0} \tag{C.27}$$

gives the mathematical constraints for the random variables in $\mathbf{x}$ so the problem order can be reduced. Note that the matrix $\mathbf{V}_0$ has full rank, i.e., $\rho(\mathbf{V}_0) = N_0$. The following example is used to demonstrate this procedure.

Example. Consider four Gaussian random variables x_1, x_2, x_3, x_4 with means $1, 2, 3, 4$, respectively. The covariance matrix of $\mathbf{x} = [x_1, x_2, x_3, x_4]^T$ is:

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \begin{bmatrix}4 & 2 & 2 & 4 \\ 2 & 1 & 1 & 2 \\ 2 & 1 & 4 & 2 \\ 4 & 2 & 2 & 9\end{bmatrix}$$

The rank of this matrix is $\rho(\boldsymbol{\Sigma}_{\mathbf{x}}) = 3$ and the zero eigenvalue has a multiplicity of one. By using MATLAB® [171], its corresponding eigenvector is $\boldsymbol{\phi} = [0.4472, -0.8944, 0, 0]^T$ so

www.EngineeringBooksPdf.com

the matrix $\mathbf{V}_0^T$ in Equation (C.26) can be taken as:

$$\mathbf{V}_0^T = \begin{bmatrix} 0.4472 & -0.8944 & 0 & 0 \end{bmatrix}$$

This implies that the variable $\xi = 0.4472(x_1 - 1) - 0.8944(x_2 - 2)$ has zero mean and zero variance. In other words:

$$x_1 - 1 = 2(x_2 - 2)$$

at least almost everywhere (a.e.).

Case 1: Prediction of x_1 and x_2 by measurement of x_3 and x_4
Since x_1 and x_2 are fully correlated, the original problem with four random variables can be reduced to three random variables by predicting only x_2 with x_3 and x_4. The covariance matrix for $\mathbf{x}' = [x_2, x_3, x_4]^T$ is:

$$\mathbf{\Sigma}_{\mathbf{x}'} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 4 & 2 \\ 2 & 2 & 9 \end{bmatrix}$$

and non-singular. By using Equation (C.23), the conditional mean is:

$$E[x_2|x_3, x_4] = 2 + [1, 2] \begin{bmatrix} 4 & 2 \\ 2 & 9 \end{bmatrix}^{-1} \begin{bmatrix} x_3 - 3 \\ x_4 - 4 \end{bmatrix} = 2 + 0.1563(x_3 - 3) + 0.1875(x_4 - 4)$$

By using Equation (C.24), the conditional variance is 0.4688. Since x_1 is fully correlated with x_2, it can be predicted as follows:

$$E[x_1|x_3, x_4] = E[1 + 2(x_2 - 2)|x_3, x_4] = 1 + 0.3125(x_3 - 3) + 0.375(x_4 - 4)$$

with variance $2^2 \times 0.4688 \approx 1.875$. Finally, the conditional covariance matrix of x_1 and x_2 is:

$$\begin{bmatrix} 1.875 & 0.9375 \\ 0.9375 & 0.4688 \end{bmatrix}$$

Case 2: Prediction of x_1 and x_3 by measurement of x_2 and x_4
Since x_1 and x_2 are fully correlated, x_1 can be determined with zero variance: $x_1 = 1 + 2(x_2 - 2)$. To predict x_3 by x_2 and x_4, the covariance matrix is rearranged for the random vector $\mathbf{x}' = [x_3, x_2, x_4]^T$:

$$\mathbf{\Sigma}_{\mathbf{x}'} = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 1 & 2 \\ 2 & 2 & 9 \end{bmatrix}$$

By using Equation (C.23), the conditional mean is:

$$E[x_3|x_2, x_4] = 2 + [1, 2] \begin{bmatrix} 1 & 2 \\ 2 & 9 \end{bmatrix}^{-1} \begin{bmatrix} x_2 - 2 \\ x_4 - 4 \end{bmatrix} = 2 + (x_2 - 2)$$

www.EngineeringBooksPdf.com

By using Equation (C.24), the conditional variance is 3. Finally, The conditional covariance matrix for x_1 and x_3 is:

$$\begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}$$

Case 3: Prediction of x_3 and x_4 by measurement of x_1 and x_2
Since x_1 and x_2 are fully correlated, the conditioning information is redundant and x_2 can be omitted. In this case, the covariance matrix is rearranged for $\mathbf{x}' = [x_3, x_4, x_1]^T$:

$$\mathbf{\Sigma}_{\mathbf{x}'} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 9 & 4 \\ 2 & 4 & 4 \end{bmatrix}$$

By using Equation (C.23), the conditional mean is:

$$E\left[[x_3, x_4]^T \middle| x_1\right] = \begin{bmatrix} 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \end{bmatrix} \times 4^{-1}(x_1 - 1) = \begin{bmatrix} 3 + 0.5(x_1 - 1) \\ 4 + (x_1 - 1) \end{bmatrix}$$

By using Equation (C.24), the covariance matrix of the prediction error of x_3 and x_4 is:

$$\begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$$

www.EngineeringBooksPdf.com

www.EngineeringBooksPdf.com

References

[1] Agbabian, M. S. and Masri, S. F. *Proceedings of the International Workshop on Non-destructive Evaluation for Performance of Civil Structures*. Department of Civil Engineering, University of Southern California, Los Angeles, CA, 1988.

[2] Akaike, H. A new look at the statistical identification model. *IEEE Transactions on Automatic Control* **19**(6) (1974), 716–723.

[3] Akaike, H. On entropy maximization principle. In *Applications of Statistics*, P. R. Krishnaiah (Ed.), North Holland, Amsterdam (1976), pp. 27–41.

[4] Andersen, N. On the calculation of filter coefficients for maximum entropy spectral analysis. *Geophysics* **39**(1) (1974), 69–72.

[5] Andersen, P. and Kirkergaard, P. H. Statistical damage detection of civil engineering structures using ARMAV models. In *Proceedings of 16th International Modal Analysis Conference* (Santa Barbara, CA, 1998), pp. 356–362.

[6] Anh, V. V., Azzi, M., Duc, H., Johnson, G. M. and Tieng, Q. A reactive state-space model for prediction of urban air pollution. *Environmental Modelling and Software* **13**(3) (1998), 239–246.

[7] Asmussen, J. C., Ibrahim, S. R. and Brincker, R. Application of vector triggering random decrement. In *Proceedings of 15th International Modal Analysis Conference* (Orlando, FL, 1997), Vol. 2, pp. 1165–1171.

[8] Au, S. K. and Beck, J. L. A new adaptive importance sampling scheme. *Structural Safety* **21**(2) (1999), 135–158.

[9] Au, S. K. and Beck, J. L. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics* **16**(4) (2001), 263–277.

[10] Au, S. K. and Beck, J. L. First excursion probabilities for linear systems by very efficient importance sampling. *Probabilistic Engineering Mechanics* **16**(3) (2001), 193–208.

[11] Au, S. K. and Beck, J. L. Importance sampling in high dimensions. *Structural Safety* **25**(2) (2003), 139–163.

[12] Au, S. K., Papadimitriou, C. and Beck, J. L. Reliability of uncertain dynamical systems with multiple design points. *Structural Safety* **21**(2) (1999), 113–133.

[13] Barnard, G. A. Thomas Bayes's essay towards solving a problem in the doctrine of chances. *Biometrika* **45**(3–4) (1958), 293–315.

[14] Bayes, T. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* **53**(1) (1763), 370–418.

[15] Beck, J. L. *Determining Models of Structures from Earthquake Records*. Technical Report EERL 78-01, California Institute of Technology, Earthquake Engineering Research Laboratory, Pasadena, CA, 1978.

[16] Beck, J. L. Statistical system identification of structures. In *Structural Safety and Reliability*, ASCE, New York, NY (1990), pp. 1395–1402.

[17] Beck, J. L. and Au, S. K. Bayesian updating of structural models and reliability using Markov Chain Monte Carlo simulation. *Journal of Engineering Mechanics (ASCE)* **128**(4) (2002), 380–391.

[18] Beck, J. L., Au, S. K. and Vanik, M. W. Monitoring structural health using a probabilistic measure. *Computer-Aided Civil and Infrastructure Engineering* **16**(1) (2001), 1–11.

© 2010 John Wiley & Sons (Asia) Pte Ltd

www.EngineeringBooksPdf.com

[19] Beck, J. L. and Katafygiotis, L. S. Updating models and their uncertainties. I: Bayesian statistical framework. *Journal of Engineering Mechanics (ASCE)* **124**(4) (1998), 455–461.
[20] Beck, J. L., May, B. S. and Polidori, D. C. Determination of modal parameters from ambient vibration data for structural health monitoring. In *Proceedings of 1st World Conference on Structural Control* (Pasadena, CA, 1994), pp. TA3:3–TA3:12.
[21] Beck, J. L., Vanik, M. W., Polidori, D. C. and May, B. S. Ambient vibration surveys of a steel frame building in a healthy and damaged state. Technical Report EERL 97-03, California Institute of Technology, Earthquake Engineering Research Laboratory, Pasadena, CA, 1997.
[22] Beck, J. L. and Wu, Z. Special issue on structural health monitoring. *Computer-Aided Civil and Infrastructure Engineering* **21**(4) (2006).
[23] Beck, J. L. and Yuen, K.-V. Model selection using response measurements: Bayesian probabilistic approach. *Journal of Engineering Mechanics (ASCE)* **130**(2) (2004), 192–203.
[24] Berman, A. and Flannelly, W. G. Theory of incomplete models of dynamics structures. *AIAA Journal* **9**(8) (1971), 1481–1487.
[25] Bernal, D. Load vectors for damage localization. *Journal of Engineering Mechanics (ASCE)* **128**(1) (2002), 7–14.
[26] Bernal, D. and Beck, J. L. Special issue on Phase I of the IASC-ASCE structural health monitoring benchmark. *Journal of Engineering Mechanics (ASCE)* **130**(1) (2004).
[27] Boore, D. M., Joyner, W. B. and Fumal, T. E. Estimation of response spectra and peak accelerations from Western North American earthquakes: an interim report. Technical Report Open-File Report 93-509, US Geological Survey, Menlo Park, CA, 1993.
[28] Boore, D. M., Joyner, W. B. and Fumal, T. E. Estimation of response spectra and peak accelerations from Western North American earthquakes: an interim report, part 2. Technical Report Open-File Report 94-127, US Geological Survey, Menlo Park, CA, 1993.
[29] Boore, D. M., Joyner, W. B. and Fumal, T. E. Equations for estimating horizontal response spectra and peak acceleration from Western North American earthquakes: a summary of recent work (with 2005 erratum). *Seismology Research Letters* **68**(1) (1997), 128–153.
[30] Box, G. E. P. and Jenkins, G. M. *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, CA, 1970.
[31] Box, G. E. P. and Tiao, G. C. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA, 1973.
[32] Bradley, J. N. and Peterka, A. J. The hydraulic design of stilling basins: Hydraulic jumps on a horizontal apron (basin I). *Journal of the Hydraulics Division, Proceedings of ASCE 83(HY5)*, **83**(5) (1957), 1–24.
[33] Brillinger, D. R. and Preisler, H. K. An exploratory analysis of the Joyner–Boore attenuation data. *Bulletin of the Seismological Society of America* **74**(4) (1984), 1441–1450.
[34] Brillinger, D. R. and Preisler, H. K. Further analysis of the Joyner–Boore attenuation data. *Bulletin of the Seismological Society of America* **75**(2) (1985), 611–614.
[35] Brockwell, P. J. and Davis, R. A. *Time Series: Theory and Methods*. Springer-Verlag, New York, NY 1991.
[36] Brown, R. G. and Hwang, Y. C. *Introduction to Random Signals and Applied Kalman Filtering*. John Wiley & Sons, Inc., New York, NY, 1996.
[37] Bucher, C. G. Adaptive sampling – an iterative fast Monte Carlo procedure. *Structural Safety* **5**(2) (1988), 119–126.
[38] Buck, C. E., Cavanagh, W. G. and Litton, C. *Bayesian Approach to Intrepreting Archaeological Data*. John Wiley & Sons, Ltd, 1996.
[39] Cai, G. Q. and Lin, Y. K. A new approximate solution technique for randomly excited nonlinear oscillators. *International Journal of Nonlinear Mechanics* 23(5–6) (1988), 409–420.
[40] Calanni Fraccone, G., Ruzzene, M., Volovoi, V., Cento, P. and Vining, C. Assessment of uncertainty in response estimation for turbine engine bladed disks. *Journal of Sound and Vibration* 317(3–5) (2008), 625–645.
[41] Caughey, T. K. and O'Kelly, M. E. J. Classical normal modes in damped linear dynamic systems. *Journal of Applied Mechanics (ASME)*, **32**(12) (1965), 583–588.
[42] Caves, D. W., Herriges, J. A., Train, K. E. and Windle, R. J. A Bayesian approach to combining conditional demand and engineering models of electricity usage. *The Review of Economics and Statistics* **69**(3) (1987), 438–448.
[43] Chan, T. F. An optimal circulant preconditioner for Toeplitz systems. *SIAM Journal on Scientific Computing* **9**(4) (1988), 766–771.

www.EngineeringBooksPdf.com

[44] Chang, F. K. *Proceedings of 4th International Workshop on Structural Health Monitoring*. Stanford University, CA, 2003.
[45] Chang, S. W., Mok, K. M. and Yuen, K. V. Association of PM_{10} pollution episodes with the meteorological conditions in Macau. In *Proceedings of the 10th International Conference on Environmental Science and Technology* (Kos Island, Greece, 2007), pp. 90–95.
[46] Chaudhry, M. H. *Open Channel Flow*, Prentice-Hall, Inc., NJ, 1993. 2nd Edition. Springer, New York, NY, 2008.
[47] Chen, J. and Xu, Y. L. On modelling of typhoon-induced non-stationary wind speed for tall buildings. *The Structural Design of Tall and Special Buildings* **13**(2) (2004), 145–163.
[48] Chen, J. C., Peretti, L. F. and Garba, J. A. Spacecraft structural system identification by modal test. *Journal of Spacecraft and Rockets (AIAA)* **24**, January–February (1987), 90–94.
[49] Cheng, C. M., Jin, X. Q. and Sin, V. K. Stability of T. Chen's preconditioner from numerical range. *Numerical Mathematics, A Journal of Chinese Universities* (English Series) **16**(1) (2007), 28–36.
[50] China Earthquake Data Center. [http://smsd-iem.net/eqkview.asp].
[51] Ching, J. and Beck, J. L. Bayesian analysis of the Phase II IASC-ASCE structural health monitoring experimental benchmark data. *Journal of Engineering Mechanics (ASCE)* **130**(10) (2004), 1233–1244.
[52] Ching, J. and Beck, J. L. New Bayesian model updating algorithm applied to a structural health monitoring benchmark. *Structural Health Monitoring* **3**(4) (2004), 313–332.
[53] Ching, J., Beck, J. L., Porter, K. A. and Shaikhutdinov, R. Real-time Bayesian state estimation of uncertain dynamic system. Technical Report EERL 2004-01, Earthquake Engineering Research Laboratory, California Institute of Technology, Pasadena, CA, 2004.
[54] Ching, J. and Chen, Y. C. Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection and model averaging. *Journal of Engineering Mechanics (ASCE)* **133**(7) (2007), 816–832.
[55] Ching, J., Muto, M. and Beck, J. L. Bayesian linear structural model updating using Gibbs sampler with modal data. In *Proceedings of 9th Intertnational Conference on Structural Safety and Reliability* (Rome, Italy, 2005).
[56] Ching, J., Muto, M. and Beck, J. L. Structural model updating and health monitoring with incomplete modal data using Gibbs sampler. *Computer-Aided Civil and Infrastructure Engineering* **21**(4) (2006), 242–257.
[57] Choi, I. C., Mok, K. M. and Tam, S. C. Solving harmonic sea-level model with Kalman filter: a Macao case study. In *Proceedings of Carbonate Beaches 2000* (Reston, VA, 2002), pp. 38–52.
[58] Chopra, A. K. *Dynamics of Structures: Theory and Applications to Earthquake Engineering*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1995.
[59] Clough, R. W. and Penzien, J. *Dynamics of Structures*. McGraw-Hill, New York, NY, 1975.
[60] Conte, J. P. and Kumar, S. Statistical system identification of structures using ARMA models. In *Proceedings of 7th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability* (Worcester, MA, 1996), pp. 142–145.
[61] Cooper, J. E. Comparison of some time domain system identification techniques using approximate data correlations. *International Journal of Analytical and Experimental Modal Analysis* **4**(2) (1989), 51–57.
[62] Coppolino, R. N. A simultaneous frequency domain technique for estimation of modal parameters from measured data. SAE paper No. 811046, Aerospace Congress and Exposition (Anaheim, CA, 1981).
[63] Cox, R. T. *The Algebra of Probable Inference*. Johns Hopkins Press, Baltimore, MA, 1961.
[64] Craig, R. R., Kurdila, A. J. and Kim, H. M. State-space formulation of multi-shaker modal analysis. *Journal of Analytical and Experimental Modal Analysis* **5**(3) (1990), 169–183.
[65] Crouse, C. B. and McGuire, J. W. Site response studies for purpose of revising NEHRP seismic provisions. *Earthquake Spectra* **12**(3) (1996), 407–439.
[66] Curtis, L., Rea, W., Smith-Willis, P., Fenyves, E. and Pan, Y. Adverse health effects of outdoor air pollutants. *Environment International* **32**(6) (2006), 815–830.
[67] Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris, B. G. and Speizer, F. E. An association between air pollution and mortality in six US cities. *The New England Journal of Medicine* **329**(24) (1993), 1753–1759.
[68] Doebling, S. W., Farrar, C. R. and Prime, M. B. A review of damage identification methods that examine changes in dynamics properties. *Shock and Vibration Digest* **30**(2) (1998), 91–105.
[69] Doebling, S. W., Farrar, C. R., Prime, M. B. and Shevitz, D. W. Damage identification and health monitoring of structural and mechanical systems from changes in their vibrations characteristics: A literature review. Technical Report LA-13070-MS, Los Alamos National Laboratory, Los Alamos, NM, 1996.
[70] Doob, J. L. *Stochastic Processes*. John Wiley & Sons, Inc., New York, NY, 1953.

www.EngineeringBooksPdf.com

[71] Doyle, J. C., Francis, B. A. and Tannenbaum, A. R. *Feedback Control Theory*. Macmillan Publishing Company, New York, NY, 1992.
[72] Doyle, J. C., Glover, K., Khargonekar, P. P. and Francis, B. A. State-space solutions to standard h_2 and h_∞ control problems. *IEEE Transactions on Automatic Control* **34**(8) (1989), 831–847.
[73] Ebersbach, P. and Irretier, H. On the application of modal parameter estimation using frequency domain algorithms. *Journal of Analytical and Experimental Modal Analysis* **4**(4) (1989), 109–116.
[74] Er, G. K. and Iu, V. P. Probabilistic solutions to nonlinear random ship roll motion. *Journal of Engineering Mechanics (ASCE)* **125**(5) (1999), 570–574.
[75] Ewins, D. J. *Modal Testing – Theory, Practice and Applications*, 2nd Edition. Research Studies Press, Baldock, Hertfordshire, UK, 2000.
[76] Eykhoff, P. *System Identification*. John Wiley & Sons, Ltd, Chichester, UK, 1974.
[77] Farhat, C. and Hemez, F. M. Updating finite element dynamics models using element-by-element sensitivity methodology. *AIAA Journal* **31**(9) (1993), 1702–1711.
[78] Field, R. V., Hall, W. B. and Bergman, L. A. A matlab-based approach to the computation of probabilistic stability measures for controlled systems. In *Proceedings of 1st World Conference on Structural Control, International Association for Structural Control* (Pasadena, CA, 1994), Vol. 2, pp. TP4-13–TP4-22.
[79] Field, R. V., Voulgaris, P. G. and Bergman, L. A. Probabilistic stability robustness of structural systems. *Journal of Engineering Mechanics (ASCE)* **122**(10) (1996), 1012–1021.
[80] Field, R. V. J. Methods for model selection in applied science and engineering. Technical Report SAND 2004-5082, Sandia National Laboratories, Albu queque, NM, 2004.
[81] Flannelly, W. G. and Berman, A. The state of the art of system identification of vibrating structures. In *Proceedings of the ASME Winter Annual Meeting* (New York, NY, 1972), pp. 121–132.
[82] Frieden, B. R. *Science from Fisher Information*. Cambridge University Press, New York, NY, 2004.
[83] Gardner, M. W. and Dorling, S. R. Artificial neural networks (the multilayer perceptron) – a review of applications in the atmosphere sciences. *Atmospheric Environment* **32**(14) (1998), 2627–2636.
[84] Gelb, A. *Applied Optimal Estimation*. The MIT Press, Cambridge, UK, 1974.
[85] Gersch, W. and Foutch, D. A. Least squares estimates of structural system parameters using covariance function data. *Institute of Electrical and Electronics Engineers Transactions on Automatic Control* **AC-19**(6) (1974), 898–903.
[86] Gersch, W., Taoka, G. T. and Liu, R. Structural system parameter estimation by two-stage least squares method. *Journal of Engineering Mechanics (ASCE)* **102**(5) (1976), 883–899.
[87] Geyskens, P., Der Kiureghian, A. and Monteiro, P. Bayesian prediction of elastic modulus of concrete. *Journal of Structural Engineering (ASCE)* **124**(1) (1998), 89–95.
[88] Ghafar, A. A., Mossa, M. and Petrillo, A. Scour from flow downstream of a sluice gate after a horizontal apron. In *Proceedings of Sixth International Symposium on River Sedimentation* (New Delhi, India, 1995), pp. 1069–1088.
[89] Ghanem, R. and Shinozuka, M. Structural-system identification. I: theory. *Journal of Engineering Mechanics (ASCE)* **121**(2) (1995), 255–264.
[90] Ghanem, R. and Sture, S. Special issue on structural health monitoring. *Journal of Engineering Mechanics (ASCE)* **126**(7) (2000).
[91] Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*. Allison-Wesley, Boston, MA, 1989.
[92] Golub, G. H. and Van Loan, C. F. An analysis of total least squares problem. *SIAM Journal of Numerical Analysis* **17**(6) (1980), 883–893.
[93] Goodman, N. R. Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *The Annals of Mathematical Statistics* **34**(1) (1963), 152–177.
[94] Goodwin, G. C. and Sin, K. S. *Adaptive Filtering Prediction and Control*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1984.
[95] Goyal, P., Chan, A. T. and Jaiswal, N. Statistical models for the prediction of respirable suspended particulate matter in urban cities. *Atmospheric Environment* **40**(11) (2006), 2068–2077.
[96] Grigoriu, M. *Applied Non-Gaussian Processes: Examples, Theory, Simulation, Linear Random Vibration and MATLAB Solutions*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1995.
[97] Grivas, G. and Chaloulakou, A. Artificial neural network models for prediction of PM_{10} hourly concentrations, in the greater area of Athens, Greece. *Atmospheric Environment* **40**(7) (2006), 1216–1229.

www.EngineeringBooksPdf.com

[98] Gull, S. F. Bayesian inductive inference and maximum entropy. In *Maximum Entropy and Bayesian Methods*, J. Skilling (Ed.), Kluwer Academic Publishers, Boston, MA (1988), pp. 53–74.
[99] Hanks, T. and Kanomori, H. A moment magnitude scale. *Journal of Geophysical Research* **84**(B5) (1979), 2348–2350.
[100] Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**(1) (1970), 97–109.
[101] Hemez, F. M. and Farhat, C. Structural damage detection via a finite element model updating methodology. *International Journal of Analytical and Experimental Modal Analysis* **10**(3) (1995), 152–166.
[102] Heredia-Zavoni, E. and Esteva, L. Optimal instrumentation of uncertain structural systems subject to earthquake motions. *Earthquake Engineering and Structural Dynamics* **27**(4) (1998), 343–362.
[103] Hjelmstad, K. D. and Shin, S. Damage detection and assessment of structures from static response. *Journal of Engineering Mechanics (ASCE)* **123**(6) (1997), 568–576.
[104] Hoi, K.-I., Yuen, K.-V. and Mok, K.-M. Kalman filter based prediction system for Wintertime PM10 concentrations in Macau. *Global NEST (Network for Environmental Science and Technology) Journal* **10**(2) (2008), 140–150.
[105] Hoi, K.-I., Yuen, K.-V. and Mok, K.-M. Prediction of daily averaged PM10 concentrations by statistical time-varying model. *Atmospheric Environment* **43**(16) (2009), 2579–2581.
[106] Hoon, S. and Law, K. H. A Bayesian probabilistic approach for structure damage detection. *Earthquake Engineering and Structural Dynamics* **26**(12) (1997), 1259–1281.
[107] Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F. and Brasseur, O. A neural network forecast for daily average PM_{10} concentrations in Belgium. *Atmospheric Environment* **39**(18) (2005), 3279–3289.
[108] Hornik, K., Stinchcombe, M. and White, H. Multilayer feedforward networks are universal approximators. *Neural Network* **2**(5) (1989), 359–366.
[109] Hoshiya, M. Application on extended Kalman filter-WGI method in dynamic system identification. In *Stochastic Structural Dynamics, Progress in Theory and Application*, Elsevier, Oxford, UK (1988), pp. 103–124.
[110] Hoshiya, M. and Saito, E. Structural identification by extended Kalman filter. *Journal of Engineering Mechanics (ASCE)* **110**(12) (1984), 1757–1770.
[111] Hoshiya, M. and Yoshida, I. Identification of conditional stochastic Gaussian field. *Journal of Engineering Mechanics (ASCE)* **122**(2) (1996), 101–108.
[112] Housner, G. W., Bergman, L. A., Caughey, T. K., Chassiakos, A. G., Claus, R. O., Masri, S. F., Skelton, R. E., Soong, T. T., Spencer, B. F. and Yao, J. T. P. Special issue on structural control: past, present, and future. *Journal of Engineering Mechanics (ASCE)* **123**(9) (1997).
[113] Hoyt, J. W. and Sellin, R. H. J. Hydraulic jump as mixing layer. *Journal of Hydraulic Engineering (ASCE)* **115**(12) (1989), 1607–1614.
[114] Ibrahim, S. R. Double least squares approach for use in structural modal identification. *AIAA Journal* **24**(3) (1986), 499–503.
[115] Jaynes, E. T. Prior probabilities. *IEEE Transactions System Science and Cybernetics* **4**(3) (1968), 227–241.
[116] Jaynes, E. T. *Probability Theory with Applications in Science and Engineering*. Washington University, St Louis, MO, 1974.
[117] Jaynes, E. T. *Where do we Stand on Maximum Entropy?* MIT Press, Cambridge, MA, 1978.
[118] Jaynes, E. T. *Papers on Probability, Statistics and Statistical Physics*, R. Rosenkrantz (Ed.), Reidel, Dordrecht, The Netherlands, 1983.
[119] Jaynes, E. T. Bayesian methods: General background. In *Maximum Entropy and Bayesian Methods in Applied Statistics*, Cambridge University Press, Cambridge, UK (1986), pp. 1–25.
[120] Jaynes, E. T. *Probability Theory: The Logic of Science*, L. Bretthorst (Ed.), Cambridge University Press, Cambridge, UK 2003.
[121] Jeffreys, H. *Theory of Probability* (3rd Edition). Oxford Clarendon Press, Oxford, UK, 1961.
[122] Jiang, D., Zhang, Y., Hu, X., Zeng, Y., Tan, J. and Shao, D. Progress in developing an ANN model for air pollution index forecast. *Atmospheric Environment* **38**(40) (2004), 7055–7064.
[123] Jiang, H. and Deng, L. A Bayesian approach to the verification problem: applications to speaker verification. *IEEE Transactions on Speech and Audio Processing* **9**(8) (2001), 874–884.
[124] Jin, X. Q. Fast iterative solvers for symmetric Toeplitz systems – a survey and an extension. *Journal of Computational and Applied Mathematics* **66**(1–2) (1996), 315–321.

www.EngineeringBooksPdf.com

[125] Joyner, W. B. and Boore, D. M. Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake. *Bulletin of the Seismological Society of America* **71**(6) (1981), 2011–2038.

[126] Joyner, W. B. and Boore, D. M. Methods for regression analysis of strong motion data. *Bulletin of the Seismological Society of America* **83**(2) (1993), 469–487.

[127] Juang, J. N. and Suzuki, H. An eigen-system realization algorithm in frequency domain for modal parameter identification. *Journal of Vibration, Acoustics, Stress and Reliability in Design* **110**(1) (1988), 24–29.

[128] Kalman, R. E. A new approach to linear filtering and prediction problems. *Transactions of ASME, Journal of Basic Engineering* **82**(3) (1960), 35–45.

[129] Kalman, R. E. and Bucy, R. S. New results in linear filtering and prediction theory. *Transactions of ASME, Journal of Basic Engineering* **83** (1961), 95–108.

[130] Kanai, K. *Semi-empirical Formula for the Seismic Characteristics of the Ground*. Technical Report, Earthquake Research Institute, Tokyo, Japan, *University of Tokyo Bulletin*, 1957.

[131] Kareem, A. and Sun, W. J. Dynamic response of structures with uncertain damping. *Engineering Structures* **12**(1) (1990), 2–8.

[132] Katafygiotis, L. S. and Beck, J. L. Updating models and their uncertainties. II: Model identifiability. *Journal of Engineering Mechanics (ASCE)* **124**(4) (1998), 463–467.

[133] Katafygiotis, L. S. and Cabral, S. V. An adaptive importance sampling procedure for fitting multivariate Gaussian distributions. In *Proceedings of 13th ASCE Engineering Mechanics Conference* (Baltimore, MA, 1999), CD-ROM proceedings.

[134] Katafygiotis, L. S. and Cheung, S. H. Domain decomposition method for calculating the failure probability of linear dynamic systems subjected to Gaussian stochastic loads. *Journal of Engineering Mechanics (ASCE)* **132**(5) (2006), 475–486.

[135] Katafygiotis, L. S., Cheung, S. H. and Yuen, K.-V. Spherical subset simulation (S^3) for solving nonlinear dynamical reliability problems. *International Journal of Reliability and Safety* **4**(0) (2010), in press.

[136] Katafygiotis, L. S. and Yuen, K.-V. Bayesian spectral density approach for modal updating using ambient data. *Earthquake Engineering and Structural Dynamics* **30**(8) (2001), 1103–1123.

[137] Kermanshahi, B. *Design and Application of Neural Networks*, Chapter 3. Shokodo, Tokyo, Japan, 1999.

[138] Kim, H. M., Bartkowicz, T. J., Smith, S. W. and Zimmerman, D. Health monitoring of large structures. *Journal of Sound and Vibration* **29**(4) (1995), 18–21.

[139] Kim, H. M., Vanhorn, D. A. and Doiron, H. H. Free-decay time-domain modal identification for large space structures. *Journal of Guidance, Control and Dynamics (AIAA)* **17**(3) (1994), 513–519.

[140] Kim, K. J. and Durbin, P. A. Observations of the frequencies in a sphere wake and of drag increase by acoustic excitation. *Physics of Fluids* **31**(11) (1988), 3260–3265.

[141] Kitada, Y. Identification of nonlinear structural dynamic systems using wavelets. *Journal of Engineering Mechanics (ASCE)* **124**(10) (1998), 1059–1066.

[142] Koh, C. G., See, L. M. and Balendra, T. Damage detection of building: Numerical and experimental studies. *Journal of Structural Engineering (ASCE)* **121**(8) (1995), 1155–1160.

[143] Koh, C. G. and Shankar, K. Substructural identification method without interface measurement. *Journal of Engineering Mechanics (ASCE)* **129**(7) (2003), 769–776.

[144] Kolachalama, V. B., Bressloff, N. W. and Nair, P. B. Mining data from hemodynamic simulations via Bayesian emulation. *Biomedical Engineering Online* **6**(47) (2007).

[145] Kozin, F. and Natke, H. G. System identification techniques. *Structural Safety* **3**(3–4) (1986), 269–316.

[146] Krishnaiah, P. R. Some recent developments on complex multivariate distributions. *Journal of Multivariate Analysis* **6**(1) (1976), 1–30.

[147] Kullback, S. *Information Theory and Statistics*. Dover Publications Inc., Mineola, NY, 1968.

[148] Kullback, S., and Leibler, R. A. On information and sufficiency. *Annals of Mathematical Statistics* **22**(1) (1951), 79–86.

[149] Lam, H. F., Ko, J. M. and Wong, C. W. Localization of damaged structural connections based on experimental modal and sensitivity analysis. *Journal of Sound and Vibration* **210**(1) (1998), 91–115.

[150] Lam, H. F., Yuen, K.-V. and Beck, J. L. Structural health monitoring via measured Ritz vectors utilizing artificial neural networks. *Computer-Aided Civil and Infrastructure Engineering* **21**(4) (2006), 232–241.

[151] Lecampion, B. and Gunning, J. Model selection in fracture mapping from elastostatic data. *International Journal of Solids and Structures* **44**(5) (2007), 1391–1408.

www.EngineeringBooksPdf.com

[152] Lee, S. Y. *Structural Equation Modelling: A Bayesian Approach*. John Wiley & Sons, Inc., New York, NY, 2007.
[153] Levy, E. C. Complex curve fitting, *IEEE Transactions on Automatic Control* **AC-4**(1) (1959), 37–43.
[154] Li, Q. S., Wong, C. K., Fang, J. Q., Jeary, A. P. and Chow, Y. W. Field measurements of wind and structural responses of a 70-storey tall building under typhoon conditions. *The Structural Design of Tall Buildings* **9**(5) (2000), 325–342.
[155] Li, Q. S., Xiao, Y. Q., Wu, J. R., Fu, J. Y. and Li, Z. N. Typhoon effects on super-tall buildings. *Journal of Sound and Vibration* **313**(3–5) (2008), 581–602.
[156] Lin, J. S. and Zhang, Y. Nonlinear structural identification using extended Kalman filter. *Computers and Structures* **52**(4) (1994), 757–764.
[157] Lin, Y. K. *Probabilistic Theory of Structural Dynamics*. Robert E. Krieger Publishing Company, Malabar, FL, 1976.
[158] Lin, Y. K. and Cai, G. Q. *Probabilistic Structural Dynamics: Advanced Theory and Applications*. McGraw-Hill, Inc., New York, NY, 1995.
[159] Ljung, L. *System Identification: Theory for the User*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1987.
[160] Loh, C.-H. and Tsaur, Y.-H. Time domain estimation of structural parameters. *Engineering Structures* **10**(2) (1988), 95–105.
[161] Lutes, L. D. and Sarkani, S. *Stochastic Analysis of Structural and Mechanical Vibrations*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1997.
[162] Macau Meteorological and Geophysical Bureau. *General ambient monitoring station and its surrounding environment. [http://www.smg.gov.mo/]* – retrieved October 1, 2007.
[163] Macau Statistics and Census Bureau. *GDP per capita of Macau between 2002 and 2004. [http://www.dsec.gov.mo/]* – retrieved June 18, 2008.
[164] Mackay, D. J. C. Bayesian interpolation. *Neural Computation* **4**(3) (1992), 415–447.
[165] Madsen, H., Rosbjerg, D. and Harremooes, P. Application of the Bayesian approach in regional analysis of extreme rainfalls. *Stochastic Hydrology and Hydraulics* **9**(1) (1995), 77–88.
[166] Mahadevan, S. and Rebba, R. Validation of reliability computational models using Bayes networks. *Reliability Engineering and System Safety* **87**(2) (2005), 223–232.
[167] Mahadevan, S., Zhang, R. and Smith, N. Bayesian networks for system reliability reassessment. *Structural Safety* **23**(3) (2001), 231–251.
[168] Majumder, L. and Manohar, C. S. A time-domain approach for damage detection in beam structures using vibration data with a moving oscillator as an excitation source. *Journal of Sound and Vibration* **268**(4) (2003), 699–716.
[169] Mark, J. L. Regularization in the selection of radial basis function centers. *Neural Computation* **7**(3) (1995), 606–623.
[170] Marrison, C. I. and Stengel, R. F. Stochastic robustness synthesis applied to a benchmark control problem. *International Journal of Robust and Nonlinear Control* **5**(1) (1995), 13–31.
[171] MATLAB®. *MATLAB® User's Guide*. The MathWorks, Inc., Natick, MA, 1994.
[172] May, B. S. and Beck, J. L. Probabilistic control for the active mass driver benchmark structural model. *Earthquake Engineering and Structural Dynamics* **27**(11) (1998), 1331–1346.
[173] Mayes, R. L. and Johansen, D. D. A modal parameter extraction algorithm using best-fit reciprocal vectors. In *Proceedings of 16th International Modal Analysis Conference* (Santa Barbara, CA, 1998), pp. 517–521.
[174] Mazurek, D. F. and DeWolf, J. T. Experimental study of bridge monitoring technique. *Journal of Structural Engineering (ASCE)* **116**(9) (1990), 2532–2549.
[175] McLaughin, K. L. Maximum likelihood estimation of strong-motion attenuation relationships. *Earthquake Spectra* **7**(2) (1991), 267–279.
[176] Melchers, R. E. Importance sampling in structural systems. *Structural Safety* **6**(1) (1989), 3–10.
[177] Mendenhall, W., Beaver, R. J. and Beaver, B. M. *Introduction to Probability and Statistics*. Duxbury Press, Belmont, CA, 2005.
[178] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. and Teller, A. H. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**(6) (1953), 1087–1092.
[179] Mikami, A. and Sawada, T. Simultaneous identification of time and space variant dynamic soil properties during the 1995 Hyogoken-Nanbu earthquake. *Soil Dynamics and Earthquake Engineering* **25**(1) (2005), 69–77.

www.EngineeringBooksPdf.com

[180] Mockus, J., Eddy, W., Mockus, A., Mockus, L. and Reklaitis, G. *Bayesian Heuristic Approach to Dsicrete and Global Optimization: Algorithm, Visualization, Software, and Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[181] Modi, V. J. and Slater, J. E. Unsteady aerodynamics and vortex-induced aeroelastic instability of a structural angle section. *Journal of Wind Engineering and Industrial Aerodynamics* **116**(4) (1994), 449–456.

[182] Mok, K. M. and Hoi, K. I. Effect of meteorological conditions on PM_{10} concentrations – a study in Macau. *Environmental Monitoring and Assessment* **102**(1–3) (2005), 201–223.

[183] Mok, K. M. and Ieong, K. K. Turbulence generation and surface fluctuation in weak hydrauluc jumps. In *Advances in Fluid Modeling and Turbulence Measurements.* World Scientific Publishing Company. Pte Ltd, Singapore, (2002), pp. 79–86.

[184] Mok, K. M. and Tam, S. C. Short-term prediction of SO_2 concentration in Macau with artificial neural network. *Energy and Buildings* **28**(3) (1998), 279–286.

[185] Mottershead, J. E. and Friswell, M. I. Model updating in structural dynamics: a survey. *Journal of Sound and Vibration* **167**(2) (1993), 347–375.

[186] Mottershead, J. E. and Friswell, M. I. Special issue on inverse methods in structural dynamics. *Mechanical Systems and Signal Processing* **15**(1) (2001).

[187] Natke, H. G. and Yao, J. T. P. *Proceedings of the Workshop on Structural Safety Evaluation Based on System Identification Approaches*. Vieweg and Sohn, Wiesbaden, Germany, 1988.

[188] Naus, D. J. The effect of elevated temperature on concrete materials and structures – a literature review. Technical Report NUREG/CR-6900, US Nuclear Regulatory Commission, Washington, DC, 2006.

[189] Naylor, A. W. and Sell, G. R. *Linear Operator Theory in Engineering and Science*. Springer-Verlag, Berlin, 1982.

[190] Ng, C. N. and Yan, T. L. Recursive estimation of model parameters with sharp discontinuity in non-stationary air quality data. *Environmental Modelling and Software* **19**(1) (2004), 19–25.

[191] Nunnari, G., Dorling, S., Schlink, U., Cawley, G., Foxall, R. and Chattarton, T. Modelling SO_2 concentrations at a point with statistical approaches. *Environmental Modelling and Software* **19**(10) (2004), 887–905.

[192] Ojha S., Coutinho J. and Kumar, A. Developing systems to forecast ozone and particulate matter levels. *Environmental Progress* **21**(2) (2002), J7–J12.

[193] Ordieres, J. B., Vergara, E. P., Capuz, R. S. and Salazar, R. E. Neural network prediction model for fine particulate matter (PM2.5) on the US–Mexico border in El Paso (Texas) and Ciudad Juarez (Chihuahua). *Environmental Modelling and Software* **20**(5) (2005), 547–559.

[194] Panakkat, A. and Adeli, H. Recent efforts in earthquake prediction (1990–2007). *Natural Hazards Review* **9**(2) (2008), 70–80.

[195] Pandey, A. K. and Biswas, M. Experimental verification of flexibility difference method for locating damage in structures. *Journal of Sound and Vibration* **184**(2) (1995), 311–328.

[196] Papadimitriou, C., Beck, J. L. and Au, S. K. Entropy-based optimal sensor location for structural model updating. *Journal of Vibration and Control* **6**(5) (2000), 781–800.

[197] Papadimitriou, C., Beck, J. L. and Katafygiotis, L. S. Asymptotic expansions for reliability and moments of uncertain systems. *Journal of Engineering Mechanics (ASCE)* **123**(12) (1997), 1219–1229.

[198] Papadimitriou, C., Beck, J. L. and Katafygiotis, L. S. Updating robust reliability using structural test data. *Probabilistic Engineering Mechanics* **16**(2) (2001), 103–113.

[199] Papadimitriou, C. and Katafygiotis, L. S. Bayesian modeling and updating. In *Engineering Design Reliability Handbook*, N. Nikolaidis, D. M. Ghiocel and S. Singhal (Eds), CRC Press, Boca Raton, FL (2004), Chapter 22, pp. 22-1–22-20.

[200] Papadimitriou, C., Katafygiotis, L. S. and Au, S. K. Effects of structural uncertainties on TMD design: A reliability-based approach. *Journal of Structural Control* **4**(1) (1997), 65–88.

[201] Pappa, R. S., Doebling, S. W. and Kholwad, T. D. Online database of vibration-based damage detection experiments. *Journal of Sound and Vibration* **34**(1) (2000), 28–33.

[202] Pappa, R. S. and Juang, J. N. Galileo spacecraft modal identification using an eigensystem realization algorithm. *Journal of Astronautical Sciences* **33**(1) (1985), 95–118.

[203] Parkinson, G. V. Wind-induced instability of structures. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **269**(1199) (1971), 395–413.

[204] Peng, C. Y. and Iwan, W. D. An identification methodology for a class of hysteretic structures. *Earthquake Engineering and Structural Dynamics* **21**(8) (1992), 695–712.

www.EngineeringBooksPdf.com

[205] Peregrine, D. H. and Svendsen, I. A. Spilling breakers, bores and hydraulic jumps. In *Proceedings of 16th Conference on Coastal Engineering* (Hamburg, Germany, 1978), pp. 540–550.
[206] Perez, P. and Reyes, J. An integrated neural network model for PM_{10} forecasting. *Atmospheric Environment* **40**(16) (2006), 2845–2851.
[207] Pi, Y. L. and Mickleborough, N. C. Modal identification of vibrating structures using ARMA models. *Journal of Engineering Mechanics (ASCE)* **115**(10) (1989), 2232–2250.
[208] Polidori, D. C. and Beck, J. L. Approximate solutions for nonlinear random vibration problems. *Probabilistic Engineering Mechanics* **11**(3) (1996), 179–185.
[209] Pradlwarter, H. J. and Schuëller, G. I. Assessment of low probability events of dynamical systems by controlled Monte Carlo simulation. *Probabilistic Engineering Mechanics* **14**(3) (1999), 213–227.
[210] Pradlwarter, H. J., Schuëller, G. I., Koutsourelakis, P. S. and Charmpis, D. C. Reliability of structures in high dimensions, part I: algorithms and applications. *Probabilistic Engineering Mechanics* **19**(4) (2004), 409–417.
[211] Proppe, C. Exact stationary probability density functions for nonlinear systems under Poisson white noise excitation. *International Journal of Nonlinear Mechanics* **38**(4) (2003), 557–564.
[212] Quek, S. T., Wang, W. and Koh, C. G. System identification of linear MDOF structures under ambient excitation. *Earthquake Engineering and Structural Dynamics* **28**(1) (1999), 61–77.
[213] Rachev, S. T., Hsu, J. S. J., Bagasheva, B. S. and Fabozzi, F. J. *Bayesian Methods in Finance*. John Wiley & Sons, Inc., New York, NY, 2008.
[214] Raiffa, H. and Schlaifer, R. *Applied Statistical Decision Theory*. Division of Research, Graduate School of Business Administration, Harvard University, Boston, MA, 1961.
[215] Rajaratnam, N. Profile equation of the hydraulic jump. *Water Power* **14** (1962), 324–327.
[216] Rajaratnam, N. Hydraulic jumps. *Advances in Hydroscience* **4** (1967), 197–280.
[217] Raymer, J. and Willekens, F. *International Migration in Europe: Data, Models and Estimates*. John Wiley & Sons, Ltd, Chichester, UK, 2008.
[218] Rice, S. O. Mathematical analysis of random noise. *Bell System Technology Journal* **23**(3) (1944), 282–332.
[219] Richardson, M. and Formenti, D. L. Parameter estimation from frequency response measurements using rational fraction polynomials. In *Proceedings of 1st International Modal Analysis Conference* (Orlando, Florida, 1982), pp. 167–182.
[220] Rivas, T., Matias, J. M., Taboada, J. and Arguelles, A. Application of Bayesian networks to the evaluation of roofing slate quality. *Journal of Engineering Geology* **94**(1–2) (2007), 27–37.
[221] Roberts, J. B. First passage probability for nonlinear oscillators. *Journal of Engineering Mechanics (ASCE)* **102**(5) (1976), 851–866.
[222] Roberts, J. B., Dunne, J. F. and Debonos, A. A spectral method for estimation for nonlinear system parameters from measured response. *Probabilistic Engineering Mechanics* **10**(4) (1995), 199–207.
[223] Rossi, P. E., Allenby, G. M. and McCulloch, R. *Bayesian Statistics and Marketing*. John Wiley & Sons, Ltd, Chichester, UK, 2006.
[224] Rouse, H., Siao, T. T. and Nagaratnam, S. Turbulence characteristics of the hydraulic jump. *Journal of the Hydraulics Division, Proceedings of ASCE* **124**(HY1) (1959), 926–950.
[225] Rubinstein, R. Y. *Simulation and the Monte-Carlo Method*. John Wiley & Sons, Inc., New York, NY, 1981.
[226] Safak, E. Adaptive modeling, identification, and control of dynamical structural systems I: Theory. *Journal of Engineering Mechanics (ASCE)* **115**(11) (1989), 2386–2405.
[227] Salawu, O. S. Detection of structural damage through changes in frequency: a review. *Engineering Structures* **19**(9) (1997), 718–723.
[228] Sanayei, M., McClain, J. A. S., Wadia-Fascetti, S. and Santini, E. M. Parameter estimation incorporating modal data and boundary conditions. *Journal of Structural Engineering (ASCE)* **125**(9) (1999), 1048–1055.
[229] Sato, T. and Takei, K. Real time robust identification algorithm for structural systems with time-varying dynamic characteristics. In *Proceedings of SPIE 4th Annual Symposium on Smart Structures and Materials* (Bellingham, WA, 1997), pp. 393–404.
[230] Schoukens, J. and Pintelon, R. *Identification of Linear Systems: A Practical Guideline for Accurate Modeling*. Pergamon Press, London, 1991.
[231] Schuëller, G. I. and Stix, R. A critical appraisal of methods to determine failure probabilities. *Structural Safety* **4**(4) (1987), 293–309.
[232] Schwarz, G. Estimating the dimension of a model. *Annals of Statistics* **6**(2) (1978), 461–464.
[233] Shannon, C. E. and Weaver, W. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana and Chichago, IL, 1949.

www.EngineeringBooksPdf.com

[234] Shi, J. P. and Harrison, R. M. Regression modeling of hourly NO_x and NO_2 concentrations in urban air in London. *Atmospheric Environment* **31**(24) (1997), 4081–4094.
[235] Shi, S. and Shen, J. Study on ground motion attenuation relation in Shanghai and its adjacent region. *Earthquake Research in China* **18**(2) (2004), 105–113.
[236] Shi, T., Jones, N. P. and Ellis, J. H. Simultaneous estimation of system and input parameters from output measurements. *Journal of Engineering Mechanics (ASCE)* **126**(7) (2000), 746–753.
[237] Shinozuka, M. and Deodatis, G. Simulation of stochastic processes by spectral representation. *Applied Mechanics Review* **44**(4) (1991), 191–203.
[238] Shinozuka, M., Yun, C. B. and Imai, H. Identification of linear structural dynamic system. *Journal of Engineering Mechanics (ASCE)* **108**(6) (1982), 1371–1390.
[239] Shumway, R. H. and Stoffer, D. S. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis* **3**(4) (1982), 253–264.
[240] Silverman, B. W. *Density Estimators*. Chapman and Hall, New York, NY, 1986.
[241] Simiu, E. and Scanlan, R. H. *Wind Effects on Structures: Fundamentals and Applications to Design (3rd Edition)*. Wiley-Interscience, New York, NY, 2006.
[242] Sivia, D. S. *Data Analysis: A Bayesian Tutorial*. Oxford Science Publications, Oxford, UK, 1996.
[243] Smith, B. S. and Coull, A. *Tall Building Structures: Analysis and Design*. John Wiley & Sons, Ltd, Chichester, UK, 1991.
[244] Smyth, A. W., Masri, S. F., Caughey, T. K. and Hunter, N. F. Surveillance of intricate mechanical systems on the basis of vibration signature analysis. *Journal of Applied Mechanics (ASME)* **67**(3) (2000), 540–551.
[245] Soderstrom, T. and Stoica, P. *System Identification*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1989.
[246] Sohn, H. and Farrar, C. R. Damage diagnosis using time series analysis of vibration signals. *Smart Materials and Structures* **10**(3) (2001), 446–451.
[247] Sohn, H., Farrar, C. R., Hemez, F. M., Shunk, D. D., Stinemates, D. W. and Nadler, B. R. A review of structural health monitoring literature: 1996–2001. Technical Report LA-13976-MS, Los Alamos National Laboratory, Los Alamos, NM, 2003.
[248] Soong, T. T. *Active Structural Control: Theory and Practice*. Longman Scientific and Technical, Harlow, Essex, UK, 1990.
[249] Soong, T. T. and Grigoriu, M. *Random Vibration of Mechanical and Structural Systems*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1993.
[250] Sorenson, H. W. Least-squares estimation: from Gauss to Kalman. *IEEE Spectrum* **7**(2) (1970), 63–68.
[251] Soros, G. *The Alchemy of Finance*. John Wiley & Sons, Inc., New York, NY, 1987.
[252] Spencer, B. F. and Kaspari, D. C. Structural control design: a reliability-based approach. In *Proceedings of American Control Conference* (Baltimore, MD, 1994), pp. 1062–1066.
[253] Spencer, B. F., Kaspari, D. C. and Sain, M. K. Reliability-based optimal structural control. In *Proceedings of 5th U.S. National Conference on Earthquake Engineering* (EERI, Oakland, CA, 1994), Vol. I, pp. 703–712.
[254] Spiegelhalter, D. J., Abrams, K. R. and Myles, J. P. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley & Sons, Ltd, Chichester, UK, 2004.
[255] Stengel, R. F. and Ray, L. R. Stochastic robustness of linear time-invariant control systems. *IEEE Transactions on Automatic Control* **36**(1) (1991), 82–87.
[256] Streeter, V. L. and Wylie, E. B. *Fluid Mechanics*. McGraw-Hill Book Company, New York, NY, 1979.
[257] Taleb, N. N. *The Black Swan: The Impact of the Highly Improbable*. Random House, Inc., New York, NY, 2007.
[258] Tanaka, M., Matsumoto, T. and Yamamura, H. Application of BEM with extended Kalman filter to parameter identification of an elastic plate under dynamic loading. *Engineering Analysis with Boundary Elements* **28**(3) (2004), 213–219.
[259] Taroni, F., Aitken, C., Garbolino, P. and Biedermann, A. *Bayesian Networks and Probabilistic Inference in Forensic Science*. John Wiley & Sons, Ltd, Chichester, UK, 2006.
[260] Task Committee on Wind Forces, Committee on Loads and Stresses, Structural Division, ASCE. Wind forces on structures. *Transactions of ASCE* **126**(2) (1961), 1124–1198.
[261] Topole, K. G. and Stubbs, N. Non-destructive damage evaluation in complex structures from a minimum of modal parameters. *International Journal of Analytical and Experimental Modal Analysis* **10**(2) (1995), 95–103.
[262] Udwadia, F. E. Methodology for optimal sensor locations for parameters identification in dynamic systems. *Journal of Engineering Mechanics (ASCE)* **120**(2) (1994), 368–390.
[263] Ulrych, T. N. and Bishop, T. N. Maximum entropy spectral analysis and autoregressive decomposition. *Review of Geophysics and Space Physics* **13**(1) (1975), 183–200.

www.EngineeringBooksPdf.com

[264] Van der Auweraer, H. and Leuridan, J. Multiple input orthogonal polynomial parameter estimation. *Mechanical Systems and Signal Processing* **1**(3) (1987), 259–272.
[265] Van Rensburg, N. F. J. and van der Merwe, A. J. Natural frequencies and modes of a Tiomshenko beam. *Wave Motion* **44**(1) (2006), 58–69.
[266] Vandiver, J. K., Dunwoody, A. B., Campbell, R. B. and Cook, M. F. A mathematical basis for the random decrement vibration signature analysis technique. *Journal of Mechanical Design* **104**(2) (1982), 307–313.
[267] Vanik, M. W. A Bayesian probabilistic approach to structural health monitoring. Technical Report EERL 97-07, Earthquake Engineering Research Laboratory, California Institute of Technology, Pasadena, CA, 1997.
[268] Vanik, M. W., Beck, J. L. and Au, S. K. Bayesian probabilistic approach to structural health monitoring. *Journal of Engineering Mechanics (ASCE)* **126**(7) (2000), 738–745.
[269] Vong, C. M., Wong, P. K. and Li, Y. P. Prediction of automotive engine power and torque using least squares support vector machines and Bayesian inference. *International Journal of Engineering Application of Artificial Intelligence* **19**(3) (2006), 227–297.
[270] Wang, S., Yu, Y., Gao, A. and Yan, X. Development of attenuation relations for ground motion in China. *Earthquake Research in China* **16**(2) (2000), 99–106 (in Chinese).
[271] Wellstead, P. E., Edmunds, J. M., Prager, D. and Zanker, P. Self-tuning pole/zero assignment regulators. *International Journal of Control* **30**(1) (1979), 1–26.
[272] Wittrick, W. H. and Williams, F. W. A general algorithm for computing natural frequencies of elastic structures. *The Quarterly Journal of Mechanics and Applied Mathematics* **24**(3) (1971), 263–284.
[273] Wong, Y. L., Zhao, J. X. and Luo, Q. Attenuation characteristics of ground motions in Northern China. *Earthquake Engineering and Engineering Vibration* **1**(2) (2002), 161–166.
[274] Woodworth, G. G. *Biostatistics: A Bayesian Introduction*. John Wiley & Sons, Inc., Hoboken, NJ, 2004.
[275] Wright, J. R. Flutter test analysis in the time domain using a recursive system representation. *Journal of Aircraft (AIAA)* **11**(12) (1974), 774–777.
[276] Xu, Y. L., Chen, S. W. and Zhang, R. C. Modal identification of Di Wang building under typhoon York using the Hilbert–Huang transform method. *The Structural Design of Tall and Special Buildings* **12**(1) (2003), 21–47.
[277] Yaglom, A. M. *Correlation Theory of Stationary and Related Random Functions*. Springer-Verlag, Berlin, Germany, 1987.
[278] Yao, J. T. P. Special issue on structural health monitoring. *Computer-Aided Civil and Infrastructure Engineering* **16**(1) (2001).
[279] Yeh, H. H. and Mok, K. M. On turbulence in bores. *Physics of Fluids, Series A* **2**(5) (1990), 821–828.
[280] Young, P. C. An instrumental variable method for real-time identification of a noisy process. *Automatica* **6**(2) (1970), 271–287.
[281] Yuen, K.-V. Efficient model correction method with modal measurement. *Journal of Engineering Mechanics (ASCE),* **136**(1) (2010), 91–99.
[282] Yuen, K.-V., Au, S. K. and Beck, J. L. Two-stage structural health monitoring methodology and results for Phase I benchmark studies. *Journal of Engineering Mechanics (ASCE)* **130**(1) (2004), 16–33.
[283] Yuen, K.-V. and Beck, J. L. Reliability-based robust control for uncertain dynamical systems using feedback of incomplete noisy measurements. *Earthquake Engineering and Structural Dynamics* **32**(5) (2003), 751–770.
[284] Yuen, K.-V. and Beck, J. L. Updating properties of nonlinear dynamical systems with uncertain input. *Journal of Engineering Mechanics (ASCE)* **129**(1) (2003), 9–20.
[285] Yuen, K.-V., Beck, J. L. and Katafygiotis, L. S. Probabilistic approach for modal identification using non-stationary noisy response measurements only. *Earthquake Engineering and Structural Dynamics* **31**(4) (2002), 1007–1023.
[286] Yuen, K.-V., Beck, J. L. and Katafygiotis, L. S. Efficient model updating and monitoring methodology using incomplete modal data without mode matching. *Special Issue in Memory of Professor. T. K. Caughey in Structural Control and Health Monitoring* **13**(1) (2006), 91–107.
[287] Yuen, K.-V., Beck, J. L. and Katafygiotis, L. S. Unified probabilistic approach for model updating and damage detection. *Journal of Applied Mechanics (ASME)* **73**(4) (2006), 555–564.
[288] Yuen, K.-V., Hoi, K.-I. and Mok, K.-M. Selection of noise parameters for Kalman filter. *Earthquake Engineering and Engineering Vibration* **6**(1) (2007), 49–56.
[289] Yuen, K.-V. and Katafygiotis, L. Substructure identification and health monitoring using response measurement only. *Computer-Aided Civil and Infrastructure Engineering* **21**(4) (2006), 280–291.
[290] Yuen, K.-V. and Katafygiotis, L. S. Bayesian time-domain approach for modal updating using ambient data. *Probabilistic Engineering Mechanics* **16**(3) (2001), 219–231.

www.EngineeringBooksPdf.com

[291] Yuen, K.-V. and Katafygiotis, L. S. Bayesian modal updating using complete input and incomplete response noisy measurements. *Journal of Engineering Mechanics (ASCE)* **128**(3) (2002), 340–350.

[292] Yuen, K.-V. and Katafygiotis, L. S. Bayesian fast Fourier transform approach for modal updating using ambient data. *Advances in Structural Engineering – An International Journal* **6**(2) (2003), 81–95.

[293] Yuen, K.-V. and Katafygiotis, L. S. An efficient simulation method for reliability analysis of linear dynamical systems using simple additive rules of probability. *Probabilistic Engineering Mechanics* **20**(1) (2005), 109–114.

[294] Yuen, K.-V. and Katafygiotis, L. S. Model updating using response measurements without knowledge of the input spectrum. *Earthquake Engineering and Structural Dynamics* **34**(2) (2005), 167–187.

[295] Yuen, K.-V., Katafygiotis, L. S. and Beck, J. L. Spectral density estimation of stochastic vector processes. *Probabilistic Engineering Mechanics* **17**(3) (2002), 265–272.

[296] Yuen, K.-V., Katafygiotis, L. S., Papadimitriou, C. and Mickleborough, N. C. Optimal sensor placement methodology for identification with unmeasured excitation. *Journal of Dynamical Systems, Measurement and Control (ASME)* **123**(4) (2001), 677–686.

[297] Yuen, K.-V. and Lam, H. F. On the complexity of artificial neural networks for smart structures monitoring. *Engineering Structures* **28**(7) (2006), 977–984.

[298] Yuen, K.-V., Shi, Y.-F., Beck, J. L. and Lam, H.-F. Structural protection using MR dampers with clipped robust reliability-based control. *Structural and Multidisciplinary Optimization* **34**(5) (2007), 431–443.

[299] Zeldin, B. A. and Spanos, P. D. Spectral identification of nonlinear structural systems. *Journal of Engineering Mechanics (ASCE)* **124**(7) (1998), 728–733.

[300] Zellner, A. *An Introduction to Bayesian Inference in Econometrics*. John Wiley & Sons Ltd, Inc., New York, NY, 1996.

[301] Zhang, R. and Mahadevan, S. Bayesian methodology for reliability model acceptance. *Reliability Engineering and System Safety* **80**(1) (2003), 95–103.

[302] Zheng, S. and Wong, Y. L. Seismic ground motion relationships in Southern China based on stochastic finite-fault model. *Earthquake Engineering and Engineering Vibration* **3**(1) (2004), 11–22.

[303] Zolghadri, A. and Cazaurang, F. Adaptive nonlinear state-space modelling for the prediction of daily mean PM_{10} concentrations. *Environmental Modelling and Software* **21**(6) (2006), 885–894.

www.EngineeringBooksPdf.com

Index

adjoint, 111
AIC, 223, 236
aliasing, 117
almost everywhere (a.e.), 272, 276
ambient vibration survey (AVS), 61, 99
artificial neural network (ANN), 85
asymptotic expansion, 221
auto-regressive model, 80

Bélanger equation, 153
Bayesian fast Fourier transform approach, 190
Bayesian unified approach, 191
BIC, 224, 236
Boore-Joyner-Fumal seismic attenuation formula, 238
Bradley-Peterka curve, 155

candidate state, 50
Cauchy-Schwarz inequality, 44
causality, 1
central complex Wishart distribution, 113
Central Limit Theorem, 53
chain-like system, 5
characteristic equation, 265
Chi-square distribution, 27, 107, 113, 120
Cholesky Decomposition, 266
classical damping, 164
coefficient of determination, 88
coefficient of variation (COV), 53, 67, 122, 134, 175, 179, 183, 202, 236
conditional mean, 25, 270
conditional optimal, 46, 48, 64, 65
conditional PDF, 24, 25, 47, 48, 69, 70, 123, 126, 167, 168, 172, 176, 180, 186, 274
conditional probability, 1, 12
conditional variance, 25, 46, 270
conjugate prior, 21, 45, 231, 252
controlled Monte Carlo simulation method, 4
Coriolis effect, 139
correlation function, 41, 101, 103, 132, 154, 162, 164–166, 174
Crouse-McGuire model, 238
cumulant approximation, 132
cumulative distribution function (CDF), 23, 52, 120, 209

damped natural frequency, 155, 162, 166
damping ratio, 101, 104, 115, 123, 132, 139, 162, 167, 175
degree of believe, 2
degree of freedom (DOF), 68, 102, 164, 194, 252
Dirac delta function, 130, 251
discrete Fourier transform, 105, 111
domain decomposition method, 4
Dow Jones Industrial Average, 215
Duffing oscillator, 131

eigenvalue, 36, 102, 193, 194, 221
eigenvalue problem, 36, 193, 263
eigenvector, 36, 193, 194, 263, 265, 275
element stiffness matrix, 205
equivalent linear system, 132
error function, 23
Euclidean norm (2-norm), 34, 36, 42, 197, 198
Euler-Bernoulli beam, 62
evidence, 219–222, 225, 229, 231, 233, 234, 240, 250

Bayesian Methods for Structural Dynamics and Civil Engineering Ka-Veng Yuen
© 2010 John Wiley & Sons (Asia) Pte Ltd

www.EngineeringBooksPdf.com

exponential distribution, 21, 26, 31, 107, 109, 259
extended Kalman filter, 80

false negative rate, 15
false positive rate, 13
finite difference, 258
Fisher information matrix, 128
Fourier cosine function, 106
Fourier sine function, 106
Fourier transform, 166
frequency index set, 107, 115
frequentist, 1, 16
Froude number, 153
fundamental frequency, 5, 61–63, 66, 151, 168, 235

Gamma distribution, 21, 31, 252, 259
Gamma function, 21, 26
Gaussian, 23–25, 28, 31, 34, 36, 41, 44, 45, 47, 48, 52, 53, 56, 74, 81, 101, 102, 108–110, 112, 128, 163, 169, 172, 173, 176, 197, 198, 201, 206, 221, 225, 226, 231, 257, 263, 264, 269
Gaussian random process, 101, 162, 164–166
generalized coordinate, 69, 102
generalized eigenvalue problem, 102, 193
geophone spring-mass system, 141
globally model-identifiable, 35
goodness of fit, 34, 40, 44, 47, 53, 55, 194, 196, 230, 234, 239

half-or-double optimization algorithm, 75
Heaviside unit step function, 179, 182, 224
Hessian, 36, 45, 65, 108, 115, 128, 173, 201, 221, 225, 226, 231, 257, 259, 273
hidden layer, 86
hypercube, 15
hypocenter, 238

identity matrix, 5, 34, 37, 197
importance sampling, 3
improper prior, 21
impulse response function, 103, 165
independent and identically distributed (i.i.d.), 69, 80, 81, 83, 253
index of agreement, 88
information entropy, 22, 128, 129
inverse Gamma distribution, 25, 29, 45, 47, 231, 232
ISEE, 4

Jacobian matrix, 87
joint PDF, 13, 257

Kalman filter, 68, 80
Kermanshahi's rule of thumb, 89
kernel sampling density, 52, 228
Kronecker delta, 34, 111, 163, 166, 172

Lagrange function, 22
Lagrange multiplier, 22
law of total probability, 12, 14, 18, 219, 220, 254
leakage, 117
least squares, 41, 42, 109, 195, 221
Levenberg-Marquardt backpropagation algorithm, 86
likelihood function, 21, 24, 28, 37, 38, 44, 63, 70, 114, 163, 167, 172, 191, 197, 217, 230, 253
linear-quadratic-regulator controllers, 4
locally model-identifiable, 35
log-normal distribution, 17
Lyapunov equation, 57

magnetorheological (MR) damper, 4
marginal PDF, 13, 25, 26, 47, 136, 175, 180, 232
Markov chain, 50
Markov chain Monte Carlo (MCMC) simulation, 50
mathematical expectation, 12
matrix exponential, 69
maximum entropy distribution, 22
maximum likelihood solution, 21, 24
mean absolute percentage error, 88
mean upcrossing rate, 57
Meteorological and Geophysical Bureau, 78
Meteorological and Geophysical Bureau of Macao, 139
Metropolis-Hastings algorithm, 50
modal analysis, 102
modal coordinate, 102, 164, 165
modal forcing, 103
modal frequency, 54, 104, 167, 194, 195, 205
modal matrix, 102, 164
modal updating, 104, 167, 182, 235
mode matching, 194
mode shape, 102
model-identifiable, 35

www.EngineeringBooksPdf.com

model-unidentifiable, 35, 222
modeling error, 7, 214, 229
modulating function, 164, 182
moment magnitude (of an earthquake), 238
monsoon, 78
Monte Carlo simulation, 3, 49, 176
multilayer perceptrons (MLP), 85
multiplicity, 275

natural frequency, 101, 123, 132, 162, 175
neuron, 89
Nyquist frequency, 105, 117, 135

objective function, 24, 36, 45, 64, 71, 108, 109, 114, 173, 174, 198, 199, 201, 257–259
observation matrix, 34, 69, 81, 110, 128, 166, 197
occurrence rate, 30
Ockham factor, 221–223, 231, 241
Ockham's razor, 213
one-day-ahead prediction, 85
oscillating jump, 152
over-fitting, 89, 215, 231, 238

parameter-uncertainty ratio, 130
parametric identification, 2, 20, 34, 99, 213
peak ground acceleration (PGA), 238
Penrose generalized inverse, 195
plausibility, 2
Plexiglas, 155
Poisson distribution, 29
Poisson process, 56
prediction error, 34, 44, 45, 105, 110, 162, 166, 229
principal axes, 264
principle of model parsimony, 214
prior distribution, 21, 36, 45, 214, 251
probability density function (PDF), 12, 20
probability of detection, 94
probability of failure, 49
probability of false alarm, 97
process noise, 81
proposal PDF, 50

quantization error, 7

random decrement, 99
rank, 275
Rayleigh damping, 74, 235
regularizer, 36
relative entropy, 227
relative frequency, 31
reliability integral, 48
robust control, 4
robust failure probability, 56
robust reliability analysis, 3
roller generation-advection frequency, 153
root-mean-square error, 87

Saffir-Simpson hurricane scale, 138
screening test, 13
sensitivity, 216
separable, 231, 234
shear velocity, 243
sigmoid transfer function, 85
simulated annealing, 115
single-degree-of-freedom (SDOF) system, 58, 68, 71, 101, 104, 116, 154, 162, 179, 191
small-amplitude natural frequency, 133
soft constraint, 197
specificity, 14
spectral density function, 102, 103, 105, 155, 162, 164, 165
spectral intensity, 101–103, 105, 162
spectral set, 107
spherical subset simulation, 4
state vector, 69, 81, 84
stiffness matrix, 54, 69, 74, 102, 194–196
Strong Law of Large Numbers, 53
Strouhal number, 151
structural health monitoring (SHM), 61, 208
structural vibration control, 4
subset simulation method, 4
supercritical flow, 153
support, 16

tempering parameter, 228
time-dependent power spectral density, 166
Timoshenko beam, 61
torsional mode, 206
torsional motion, 206
training (of ANN), 86
transitional Markov chain Monte Carlo simulation, 228
tropical cyclone, 138
truncated Gaussian distribution, 23, 45
TVAR model, 80
TVAREX model, 82

www.EngineeringBooksPdf.com

Type I error, 13
Type II error, 15

undular jump, 152
uniform distribution, 23, 28, 44
updated PDF, 15, 18, 42, 64, 108, 114, 190, 221, 253
updated probability, 16

vortex shedding, 151

weak jump, 152
weighted sum of squared error, 41
white noise, 34, 56, 71, 74, 101, 102, 105, 110, 131, 154, 155, 162, 166, 179, 235
WHO, 77

www.EngineeringBooksPdf.com