Texts in Applied Mathematics

Volume 62

Editors-in-chief:

Stuart Antman, University of Maryland, College Park, USA

Leslie Greengard, New York University, New York City, USA

Philip Holmes, Princeton University, Princeton, USA

Series Editors:

John B. Bell, Lawrence Berkeley National Lab, Berkeley, USA

Joseph B. Keller, Stanford University, Stanford, USA

Robert Kohn, New York University, New York City, USA

Paul Newton, University of Southern California, Los Angeles, USA

Charles Peskin, New York University, New York City, USA

Robert Pego, Carnegie Mellon University, Pittsburgh, USA

Lenya Ryzhik, Stanford University, Stanford, USA

Amit Singer, Princeton University, Princeton, USA

Angela Stevens, Universität Münster, Münster, Germany

Andrew Stuart, University of Warwick, Coventry, UK

Thomas Witelski, Duke University, Durham, USA

Stephen Wright, University of Wisconsin-Madison, Madison, USA

Kody Law • Andrew Stuart • Konstantinos Zygalakis

Data Assimilation

A Mathematical Introduction

Kody Law
Oak Ridge National Laboratory
Computer Science and Mathematics Division
Oak Ridge, USA

Andrew Stuart
Mathematics Institute
University of Warwick
Coventry, UK

Konstantinos Zygalakis
Division of Mathematical Sciences
University of Southampton
Southampton, UK

ISSN 0939-2475 ISSN 2196-9949 (electronic)
Texts in Applied Mathematics
ISBN 978-3-319-20324-9 ISBN 978-3-319-20325-6 (eBook)
DOI 10.1007/978-3-319-20325-6

Library of Congress Control Number: 2015945224

Mathematics Subject Classification (2010): 34, 35, 37, 60, 62, 65

Springer Cham Heidelberg New York Dordrecht London

© Springer International Publishing Switzerland 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

I dedicate this book to my wife, Natasha, to my parents, John and Julie, and to my sister, Amity.

KJHL.

For my step-children: Elliot, Gregory, and Zoe.

AMS.

Για την Γεωργία και το μικρό μας καραμελάκι, Χριστίνα.

Κώστας.

Contents

Preface	xi
List of Symbols	xvi
1 Mathematical Background	1
1.1 Probability	1
1.1.1 Random Variables on $\mathbb{R}^\ell$	1
1.1.2 Gaussian Random Variables	3
1.1.3 Conditional and Marginal Distributions	6
1.1.4 Bayes's Formula	7
1.1.5 Independence	8
1.2 Dynamical Systems	9
1.2.1 Iterated Maps	9
1.2.2 Differential Equations	10
1.2.3 Long-Time Behavior	12
1.2.4 Controlled Dynamical Systems	14
1.3 Probability Metrics	15
1.3.1 Metric Properties	15
1.3.2 Metrics on Spaces of Probability Measures	15
1.4 Probabilistic View of Dynamical Systems	18
1.4.1 Markov Kernel	19
1.4.2 Ergodicity	20
1.4.3 Bayes's Formula as a Map	21
1.5 Bibliography	22
1.6 Exercises	22
2 Discrete Time: Formulation	25
2.1 Setup	25
2.2 Guiding Examples	26
2.3 Smoothing Problem	33
2.3.1 Probabilistic Formulation of Data Assimilation	33
2.3.2 Stochastic Dynamics	34
2.3.3 Reformulation of Stochastic Dynamics	36
2.3.4 Deterministic Dynamics	37
2.4 Filtering Problem	38
2.5 Filtering and Smoothing are Related	39
2.6 Well-Posedness	40
2.7 Assessing the Quality of Data-Assimilation Algorithms	43

2.8	Illustrations	45
2.9	Bibliographic Notes	49
2.10	Exercises	52
3	Discrete Time: Smoothing Algorithms	53
3.1	Linear Gaussian Problems: The Kalman Smoother	54
3.2	Markov Chain–Monte Carlo Methods	57
3.2.1	The MCMC Methodology	57
3.2.2	Metropolis–Hastings Methods	58
3.2.3	Deterministic Dynamics	60
3.2.4	Stochastic Dynamics	61
3.3	Variational Methods	65
3.4	Illustrations	68
3.5	Bibliographic Notes	74
3.6	Exercises	76
4	Discrete Time: Filtering Algorithms	79
4.1	Linear Gaussian Problems: The Kalman Filter	80
4.2	Approximate Gaussian Filters	82
4.2.1	3DVAR	84
4.2.2	Extended Kalman Filter	84
4.2.3	Ensemble Kalman Filter	84
4.2.4	Ensemble Square-Root Kalman Filter	86
4.3	The Particle Filter	87
4.3.1	The Basic Approximation Scheme	87
4.3.2	Sequential Importance Resampling	88
4.3.3	Improved Proposals	94
4.4	Large-Time Behavior of Filters	95
4.4.1	The Kalman Filter in One Dimension	95
4.4.2	The 3DVAR Filter	99
4.4.3	The Synchronization Filter	101
4.5	Illustrations	102
4.6	Bibliographic Notes	105
4.7	Exercises	113
5	Discrete Time: MATLAB Programs	115
5.1	Chapter 2 Programs	115
5.1.1	p1.m	116
5.1.2	p2.m	117
5.2	Chapter 3 Programs	119
5.2.1	p3.m	119
5.2.2	p4.m	122
5.2.3	p5.m	124
5.2.4	p6.m	126
5.2.5	p7.m	128
5.3	Chapter 4 Programs	130
5.3.1	p8.m	130
5.3.2	p9.m	133
5.3.3	p10.m	135
5.3.4	p11.m	137

5.3.5	p12.m	139
5.3.6	p13.m	141
5.3.7	p14.m	143
5.3.8	p15.m	145
5.4	ODE Programs	147
5.4.1	p16.m	147
5.4.2	p17.m	147
6	Continuous Time: Formulation	151
6.1	Setup	151
6.1.1	Derivation of the Continuous-Time Model	151
6.1.2	Stochastic Integration	153
6.2	Guiding Examples	156
6.3	Smoothing Problem	161
6.3.1	Probabilistic Formulation of Data Assimilation	161
6.3.2	Girsanov Formula	161
6.3.3	Conditional Probabilities	162
6.3.4	Stochastic Dynamics	163
6.3.5	Deterministic Dynamics	165
6.4	Filtering Problem	166
6.5	Illustrations	168
6.6	Bibliographic Notes	171
6.7	Exercises	173
7	Continuous Time: Smoothing Algorithms	175
7.1	Linear Gaussian Problems: The Kalman–Bucy Smoother	175
7.2	Markov Chain–Monte Carlo Methods	177
7.2.1	Deterministic Dynamics	177
7.2.2	Stochastic Dynamics	177
7.3	Variational Methods	179
7.4	Illustrations	180
7.5	Bibliographic Notes	184
7.6	Exercises	185
8	Continuous Time: Filtering Algorithms	187
8.1	Linear Gaussian Problems: The Kalman–Bucy Filter	188
8.2	Approximate Gaussian Filters	189
8.2.1	3DVAR	189
8.2.2	Extended Kalman Filter	190
8.2.3	Ensemble Kalman Filter	191
8.3	The Particle Filter	194
8.4	Large-Time Behavior of Filters	194
8.4.1	The Kalman–Bucy Filter in One Dimension	194
8.4.2	The 3DVAR Filter	195
8.5	Illustrations	197
8.6	Bibliographic Notes	198
8.7	Exercises	201

9	Continuous Time: MATLAB Programs	207
9.1	Chapter 6 Programs	207
9.1.1	p1c.m	207
9.1.2	p2c.m	208
9.1.3	p3c.m	209
9.2	Chapter 7 Programs	210
9.2.1	p4c.m	211
9.2.2	p5c.m	213
9.2.3	p6c.m	215
9.2.4	p7c.m	217
9.3	Chapter 8 Programs	219
9.3.1	p8c.m	219
9.3.2	p10c.m	221
9.3.3	p11c.m	222
9.3.4	p12c.m	222
9.3.5	p13c.m	225
9.3.6	p14c.m	225
9.3.7	p15c.m	225
9.3.8	p16c.m	229
9.3.9	p17c.m	229
	References	233
	Index	239

Preface

A central research challenge for the mathematical sciences in the twenty-first century is the development of principled methodologies for the seamless integration of (often vast) data sets with sophisticated mathematical models. Such data sets are becoming routinely available in almost all areas of engineering, science, and technology, while mathematical models describing phenomena of interest are often built on decades, or even centuries, of human knowledge. Ignoring either the data or the models is clearly unwise, and so the issue of combining them is of paramount importance. When the underlying mathematical model is a (possibly stochastic) dynamical system and the data may be time-ordered, combining model and data is referred to as **data assimilation**.

The research area of data assimilation has been driven, to a large extent, by practitioners working in the atmospheric and oceanographic sciences and in other areas of the geosciences, such as oil recovery. The resulting research has led to a host of algorithmic approaches and a number of significant algorithmic innovations. However, there has been no systematic treatment of the mathematical underpinnings of the subject. The goal of this book is to provide such a treatment. Specifically, we develop a unified mathematical framework in which a Bayesian formulation of the problem provides the bedrock for the derivation and development of algorithms; furthermore, the examples used in the text, together with the algorithms that are introduced and discussed, are all illustrated by MATLAB software detailed in the book and freely available online via the Springer website, the authors' personal web pages, and at the following link:

<http://tiny.cc/damat> .

It is important to appreciate that this book develops the subject of data assimilation in a manner that differs significantly from both its historical development and its typical presentation in other books. We begin with a “gold-standard” Bayesian formulation of the problem, which while out of reach for current online geophysical applications such as weather forecasting, provides a clearly defined mathematical problem whose solution would provide the ideal combination of mathematical model and available data, given known statistical uncertainties in both. We then describe various algorithmic approaches within the context of the underpinning Bayesian formulation. The reader interested in understanding the historical development of data assimilation in an applied context, or in more practically oriented mathematical and computational treatments of the field, has a number of options for additional reading, which we now describe. First, we mention that the probabilistic “smoothing” and “filtering” approaches that we describe in this book grew out of the calculus of variations and the control-theoretic viewpoints on data assimilation, resulting, respectively, from the papers of Talagrand and Courtier [135, 37] and Kalman [79]. The current practice of data assimilation in the context of the atmospheric sciences, and weather prediction in particular, is covered in the book by Kalnay [81]. An introduction in which the presentation is motivated by applications in oceanography is Bennett’s book [12]. The book by Evensen [50] provides

a good overview of many computational aspects of the subject, reflecting the author's experience in geophysical applications including oceanography and oil recovery; application of data assimilation in oil recovery is the primary focus of the book [115] by Oliver, Reynolds, and Li. The recent book by Abarbanel provides a physics and dynamical systems perspective on data assimilation [1], with motivation coming not only from weather forecasting, but also from neuroscience. The reader interested in state-of-the-art novel online algorithms based on mathematical insight and physical reasoning may consult the book by Majda and Harlim [100]; there the focus is on complex dynamical systems exhibiting turbulent behavior. And finally, the book by Cotter and Reich [31] provides a novel numerical-analysis-based perspective on the field, with primary motivation coming from geophysical applications. Our book provides a mathematical perspective on the subject via which all of these existing books can be interpreted.

The book is organized into nine chapters: the first contains a brief introduction to the mathematical tools around which the material is organized; the next four are concerned with discrete-time dynamical systems and discrete-time data, while the last four are concerned with continuous-time dynamical systems and continuous-time data; continuous-time dynamical systems together with discrete-time data can be reformulated as discrete-time dynamical systems with discrete-time data, and so this problem is not covered explicitly in the book. The four chapters on discrete and continuous time are organized identically: in the first, we frame the problem, the second and third are devoted to smoothing algorithms and filtering algorithms respectively, and the fourth contains MATLAB programs and discussion of them. Both underlying stochastic and deterministic dynamics are studied.

Chapter 1 is organized around four themes: probability, dynamical systems, probability metrics, and dynamical systems for probability measures. These subjects are given a terse overview that cannot do justice to the richness of these areas of mathematics. However, the introductory style of the material serves to orient the reader toward the form of mathematical thinking that underpins the approach taken to data assimilation in this book. Chapter 2 sets up the problem of data assimilation, based on an underlying stochastic dynamical system in which the noise appears additively; the degenerate case in which the noise disappears, deterministic dynamics, is also considered. Nonadditive noise can be studied similarly to our development, but it does lead to additional complications in some scenarios, in particular when the transition density of the underlying Markov chain is not known explicitly; for pedagogical reasons, we hence focus on additive noise. Similarly, the data is assumed to be found from a nonlinear function of the output of the dynamical system, at each discrete time, also with additive noise. Furthermore, both the model and data noise are assumed to be Gaussian in order to simplify our exposition; this may easily be relaxed to include situations in which the noise has a known density with respect to Lebesgue measure. Examples of illustrative maps are given, including a pair of examples arising from quadratic differential equations, both due to Lorenz, of dissipative character and motivated by geophysical applications. The central conceptual idea underlying the chapter is to frame the data-assimilation problem as that of determining the probability distribution on the output of the dynamical system when conditioned on the observed noisy data—the fully Bayesian formulation of the problem. There are two key variants on this perspective: smoothing, in which data in the future may condition knowledge of the output of the dynamical system in its past; and filtering, in which data in the future is not used to estimate the conditional probability distribution at a given time. We define these two probabilistic problems, demonstrate that they are related to one another, and describe a well-posedness theory, showing that the desired conditional probability distributions are Lipschitz with respect to small changes in the data.

The two probabilistic problems of smoothing and filtering lead to two different classes of algorithms, and these are described, respectively, in Chapters 3 and 4. Smoothing algorithms are

typically more computationally intensive than filtering algorithms. This is because smoothing algorithms involve the incorporation of data that is distributed over a time interval, while filtering algorithms incorporate the data sequentially, at each time point, as it arrives. Adapting the output of a dynamical system to fit a stream of data all at once, rather than sequentially, is typically a more demanding task; however, it does have potential rewards, and at least currently, weather forecasts based on the former approach give better predictions than those based on sequential incorporation of data. Chapter 3 starts with a fully Bayesian algorithmic approach, namely the use of Monte Carlo–Markov Chain (MCMC) methods to sample the distribution on the output of the dynamical system, given the data. These methods are introduced in a general framework for an arbitrary target density defined via its Lebesgue density, and an arbitrary proposal distribution; then a variety of specific methods are described, exploiting the specific structure of the target distribution arising in the data-assimilation problem to construct the proposal. The chapter concludes by linking the Bayesian approach to variational methods; specifically, it is shown how the well-known 4DVAR (for deterministic dynamics) and weak constraint 4DVAR (for stochastic dynamics, denoted by w4DVAR in this book) methods correspond to maximizing the desired probability distribution on signal given data. In Chapter 4, filtering algorithms are described. For the derivation of many of the algorithms, we adopt a minimization approach; this unifies all the methods described, with the exception of the particle filter. The chapter begins by studying linear Gaussian problems, where the Kalman filter may be used to characterize the resulting Gaussian probability distribution—the filtering distribution—exactly. The propagation of the mean is shown to be derived from a sequential minimization principle in which a quadratic functional, representing a compromise between fitting the model and the data, is minimized as each data point is acquired in time. This sequential minimization approach is then used as the basis from which to derive various approximate algorithms such as 3DVAR, the extended Kalman filter, and the ensemble Kalman filter. Although widely used in practice, these approximate algorithms cannot in general accurately reproduce the desired conditional probability distribution, except for linear Gaussian problems. Our final algorithm in this chapter is the particle filter, which, although known to behave poorly in high dimensions, can in principle reproduce the true filtering distribution as the number of particles tends to infinity. Neither the MCMC method for the smoothing problem nor the particle filter for the filtering problem is currently practical for high-dimensional applications such as those arising in the geophysical sciences. However, they play an important role in the subject, since they provide benchmarks against which the more practical, yet less well founded, algorithms may be compared. Furthermore, their formulation provides guiding principles that can be used in the development of new algorithms targeted at the high-dimensional setting.

In Chapter 2, the guiding dynamical system examples used throughout the text are illustrated with the output of MATLAB programs, while Chapters 3 and 4 include the output of MATLAB programs to illustrate the algorithms we introduce for smoothing and filtering. In Chapter 5, we provide and discuss the MATLAB programs used to produce all of this material. These programs are provided for two reasons: firstly, for some readers, code can provide a very direct way to understand both the theory and algorithms presented within the text; and secondly, the codes can form the basis of more sophisticated algorithms that the reader can develop by building upon them. We have not aimed for the most succinct or speediest implementation of the algorithms, but rather have tried to maintain a uniform presentation in which connections to the theoretical developments in the text are fairly transparent.

Chapter 6 introduces data assimilation in continuous time. Our perspective in this chapter, and in the two subsequent algorithmic chapters concerning continuous time, is to motivate much of the mathematics by taking formal asymptotic limits of the discrete-time setting, in which the underlying dynamical system behaves like a stochastic differential equation

(SDE), as does the observation model. Continuous-time data assimilation is in principle a very technical area, primarily because it involves probability on infinite-dimensional spaces. Our approach is to build intuition about the problems using the discrete setting as a stepping-stone, deriving continuous-time limits from increasingly frequent discrete-time observations of an underlying continuous process. Although we must forgo complete rigorous proofs in order to adopt this approach within a short book, we believe that the resulting intuition will help the reader access the more technical published literature in this area, and that the lack of rigorous proofs is compensated for by significant insight. Various important results from the theory of stochastic differential equations are stated and then used throughout the chapter, including the Itô and the Girsanov formulae, together with basic existence and uniqueness theorems for equations with additive Brownian noise and equipped with moment inequalities; such inequalities arise naturally for the dissipative quadratic dynamical systems introduced in Chapter 2. As in discrete time, the smoothing and filtering distributions are both introduced. Furthermore, the Zakai and Kushner–Stratonovich stochastic partial differential equations (SPDEs) for the evolution of the density of the filtering distribution are also derived.

In Chapter 7, we describe MCMC methods to sample the posterior distribution for the smoothing problem in continuous time. We include both stochastic dynamics, for which the probability distribution of interest is on an infinite-dimensional space (the pathspace of the solution on a time interval $[0, T]$), and deterministic dynamics, for which it is on a finite-dimensional space (where the initial condition lies). As in the discrete-time setting, we also discuss variational methods, which lead to problems in the calculus of variations. Chapter 8 is devoted to filtering algorithms in continuous time. The Kalman–Bucy filter is derived by taking the small interobservation time limit of a discretely observed continuous process. The same methodology is then applied to the 3DVAR method, to the extended and ensemble Kalman filters, and to the particle filter. In Chapter 9, we provide and discuss the MATLAB code used to produce the figures in the preceding three chapters; our motivation for including this code is the same as in Chapter 5, where we consider discrete-time data assimilation.

Warning We reiterate an important issue relating to the perspective we adopt in this book. Our aim is to provide clearly defined mathematical foundations for data assimilation, and this leads to the Bayesian problem of finding a probability distribution on the signal, given the data. In the very-high-dimensional problems arising in many applications, especially those of geophysical origin referred to above, it is currently beyond reach to expect to be able to compute the resulting probability distributions accurately in high dimensions, in either the filtering or the smoothing context. Thus many of the algorithms we describe are currently impractical for such applications, especially in online scenarios. For this reason, of all the algorithms we describe here, only the 3DVAR, 4DVAR, and ensemble Kalman filter are routinely used in real geophysical applications. The use of MCMC methods for smoothing problems, and the extended Kalman filter and the particle filter, is currently impractical for real online applications arising for large systems such as those in geophysical applications. However, this fact should not diminish their importance. Our book provides a clear and unequivocal statement and analysis of the desired problem to be solved in many such applications, and methods that can be used to accurately solve the problem in various simplified scenarios. As such, we believe that it provides an important cornerstone in the field that can help guide applied researchers.

Target Audience/Prerequisites The book is aimed at mathematical researchers interested in a systematic development of this interdisciplinary field and at researchers from the geosciences and a variety of other scientific fields who use tools from data assimilation to combine data with time-dependent models. As well as being suitable for self-study by such researchers,

the book can also be used as a text, and the examples, exercises, and MATLAB programs help to make it usable in this context. Although we review some basic mathematical prerequisites in the first chapter, a first course in each of differential equations, dynamical systems, and probability is assumed of the reader. It is thus suitable for a master's- or PhD-level lecture course.

Notation Throughout, $(\langle \cdot, \cdot \rangle, |\cdot|)$ denotes the Euclidean inner product and norm on $\mathbb{R}^\ell$, for an integer ℓ ; and $|\cdot|$ will also denote the induced operator norm. In the sections on continuous time, the same notation will also be used for the Hilbert space structure on the space $L^2([0, T]; \mathbb{R}^\ell)$. For a positive definite symmetric matrix $A \in \mathbb{R}^{\ell \times \ell}$, we introduce the inner product $\langle \cdot, \cdot \rangle_A = \langle A^{-\frac{1}{2}} \cdot, A^{-\frac{1}{2}} \cdot \rangle$ and the resulting norm $|\cdot|_A = |A^{-\frac{1}{2}} \cdot|$; this may also be generalized to the Hilbert space setting of $L^2([0, T]; \mathbb{R}^\ell)$. We let $\|\cdot\|_F$ denote the Frobenius norm of a matrix, found as the square root of the sum of the squares of the entries. The outer product of the ordered pair of vectors a, b in $\mathbb{R}^\ell$ is the linear operator $a \otimes b$ with property $(a \otimes b)c = \langle b, c \rangle a$. The symbol $\wedge$ is used to denote the (symmetric) operation on a pair of real numbers that delivers the minimum of that pair: $a \wedge b = a$ if $a \leq b$.

The symbol $\mathbb{N} = \{1, 2, \dots\}$ denotes the positive integers, and $\mathbb{Z}^+ := \mathbb{N} \cup \{0\} = \{0, 1, 2, \dots\}$. We write $\mathbb{R}$ for $(-\infty, \infty)$ and $\mathbb{R}^+$ for $[0, \infty)$. We use > 0 (respectively ≥ 0) to denote positive definite (respectively positive semidefinite) for real symmetric matrices. The notation $\mathbb{R}_{\text{sym}}^{\ell \times \ell}$ is used to denote symmetric $\ell \times \ell$ matrices. The symbols $\mathbb{P}$ and $\mathbb{E}$ are used to denote probability (measure and density function) and expectation.

A Gaussian probability distribution on $\mathbb{R}^\ell$, or on the Hilbert space $L^2([0, T]; \mathbb{R}^\ell)$, will be denoted by $N(m, C)$, with m the *mean* and *covariance* C . The inverse of the covariance, the *precision*, is denoted by the symbol L . The symbols σ and γ are typically used to denote standard deviations. The symbol *cf* denotes a characteristic function, and *SI* a stochastic integral. The symbol *z* denotes an i.i.d. (independent and identically distributed) sequence of uniform random variables on $[0, 1]$, and *v* an i.i.d. sequence of Gaussian random variables, both arising in the definition of MCMC methods.

We use v to denote the (unknown) signal, and y the data, both indexed by discrete time j . The symbol Y_j denotes the accumulated data to time j . In continuous time, the unknown signal is again v , now indexed by continuous time t , and z denotes the data, also indexed by t . The maps Ψ and f on $\mathbb{R}^n$ define the systematic part of the dynamics model in discrete and continuous time respectively, while h denotes the observation operator in both discrete and continuous time. The symbols ξ and η denote the noise sequences entering the signal and data models, respectively, in discrete time; when these random variables are Gaussian, their covariances are Σ and Γ respectively. In continuous time, noises are typically Brownian, denoted by B and W (possibly with suffixes) and variances Σ_0 and Γ_0 . The symbols *v* and *q* denote variables determined from v and h respectively.

The symbols $\mu, \mu_0, \mu', \mu_j, \vartheta, \vartheta_0, \nu$, some possibly with suffixes, denote probability measures. The symbols $\rho, \rho_0, \rho', \rho_j, \pi, \pi_0, \varrho$, again possibly with suffixes, denote probability density functions (pdfs). We use μ_∞ (respectively ρ_∞) to denote (typically ergodic) invariant measures (respectively densities). Furthermore, p denotes a Markov kernel. The symbol P is used to denote the linear map on measures (or their densities) implied by a Markov kernel, and L_j the nonlinear map on measures (or their densities) implied by application of Bayes's formula to take a prior into a posterior.

The symbols *I* and *J*, possibly with suffixes *det* or *r*, are used to denote nonlinear real-valued functionals arising in variational methods, and the functionals Φ (in discrete time) and Ξ_1, Ξ_2 , and Ξ_3 (in continuous time) appear in their definition.

Acknowledgments Some of the research described in the book has emerged from collaborations with others, and the authors are indebted to all of their coauthors on collaborative work cited in the bibliography sections. The authors are very grateful to Sergios Agapiou, Daniel Sanz-Alonso, and Yuan-Xiang Zhang for help in typesetting parts of this book, and for input regarding the presentation; furthermore, Ian Hawke and Gregory Ashton provided help with the preparation of the figures, while Daniel Sanz-Alonso road-tested many of the exercises. The authors also wish to thank Mel Ades and Håkon Hoel for useful feedback, which helped to improve the presentation. The anonymous reviewers chosen by Springer and editors at Springer also gave very useful feedback, and Achi Dosanjh from Springer provided very helpful overall guidance and support throughout the publication process. And finally, the authors thank the student audiences at Peking University (2012), KAUST (2014), the University of Warwick (2014), and Fudan University (2015), whose interest helped to shape the notes. AMS is grateful to EPSRC, ERC, ESA, and ONR for financial support for research whose development suggested the need for a mathematization of the subject of data assimilation. KJHL was supported as a member of the SRI-UQ Center at KAUST while much of the writing of this book was undertaken.

It is hard to perfect a book. The authors would be grateful to receive any comments and suggestions for improvement, by email:

`a.m.stuart@warwick.ac.uk` .

Please note that in making a bibliography for this book, we have not tried to be comprehensive, but rather to ensure that we have included key references relating to the mathematical underpinnings of data assimilation that are our focus here, together with texts and other references that support this mathematical perspective.

Oak Ridge, USA
Coventry, UK
Southampton, UK
April 2015

K.J.H. Law
A.M. Stuart
K.C. Zygalakis

List of symbols

Symbol	Description
$(\langle \cdot, \cdot \rangle, \cdot)$	Euclidean inner product and norm on $\mathbb{R}^\ell$, for any integer ℓ ; occasionally extended to the Hilbert space $L^2([0, T]; \mathbb{R}^\ell)$
$\langle \cdot, \cdot \rangle_A = \langle A^{-\frac{1}{2}} \cdot, A^{-\frac{1}{2}} \cdot \rangle$	weighted inner product (A positive definite symmetric)
$ \cdot _A = A^{-\frac{1}{2}} \cdot $	A induced norm (A positive definite symmetric)
$\ \cdot\ _F$	Frobenius norm of a matrix F
$\mathbb{N}$	$\{1, 2, \dots\}$
$\mathbb{Z}^+$	$\{0, 1, 2, \dots\}$
$a \otimes b$	outer product of the ordered pair of vectors a, b in $\mathbb{R}^\ell$ (with property $(a \otimes b)c = \langle b, c \rangle a$)
$\mathbb{R}$	$(-\infty, \infty)$
$\mathbb{R}^+$	$[0, \infty)$
$A > 0$	A positive definite real symmetric matrix
$A \geq 0$	A positive semi-definite real symmetric matrix
$\mathbb{R}_{\text{sym}}^{\ell \times \ell}$	set of symmetric $\ell \times \ell$ matrices
$\mathbb{P}, \mathbb{E}$	probability (measure or density) and expectation
$N(m, C)$	Gaussian probability distribution with mean m and covariance C on $\mathbb{R}^\ell$ or on the Hilbert space $L^2([0, T]; \mathbb{R}^\ell)$
L	inverse of the covariance matrix or operator (also called precision)
σ, γ	standard deviations
cf	characteristic function
St	stochastic integral
r	i.i.d. sequence of uniform random variables on $[0, 1]$ arising in the definition of MCMC methods
ι	i.i.d. sequence of Gaussian random variables arising in the definition of MCMC methods
v	unknown signal (indexed by discrete time j)
y	data (indexed by discrete time j)
Y_j	accumulated data up to time j : $\{y_\ell\}_{\ell=1}^j$
v	unknown signal (indexed by continuous time t)
z	data (indexed by continuous time t)
Ψ	systematic part of dynamics (discrete time)
f	systematic part of dynamics (continuous time)
ξ	noise sequence entering the signal model (discrete time)
η	noise sequence entering the data model (discrete model)
Σ, Γ	covariance of ξ and η
B	Brownian motion entering the signal model (continuous time)
W	Brownian motion entering the data model (continuous time)
Σ_0, Γ_0	covariance of B and W respectively.
v	variable determined from v (used to prove well-posedness)
q	variable determined from h (used in derivation of Zakai equation)
$\mu, \mu_0, \mu', \mu_j, \vartheta, \vartheta_0, \nu$	probability measures
$\rho, \rho_0, \rho', \rho_j, \pi, \pi_0, \varrho$	probability density functions (associated with those measures)
μ_∞	ergodic invariant (probability) measure
ρ_∞	ergodic invariant (probability) density function.

p	Markov kernel
P	linear map on measures (or densities) implied by a Markov kernel
L_j	nonlinear map on measures (or their densities) implied by application of Bayes's formula to take a prior into a posterior
I and J	real-valued functionals arising in variational methods; possibly with suffices denoting deterministic or reformulated problems
Φ, Ξ_1, Ξ_2 and Ξ_3	appear in definition of these functionals arising in variational methods