Large-Scale Inverse Problems and Quantification of Uncertainty

Edited by

Lorenz Biegler
Carnegie Mellon University, USA

George Biros
Georgia Institute of Technology, USA

Omar Ghattas
University of Texas at Austin, USA

Matthias Heinkenschloss
Rice University, USA

David Keyes
KAUST and Columbia University, USA

Bani Mallick,
Texas A&M University, USA

Youssef Marzouk
Massachusetts Institute of Technology, USA

Luis Tenorio
Colorado School of Mines, USA

Bart van Bloemen Waanders
Sandia National Laboratories, USA

Karen Willcox
Massachusetts Institute of Technology, USA

A John Wiley and Sons, Ltd., Publication

Large-Scale Inverse Problems and Quantification of Uncertainty

Wiley Series in Computational Statistics

Consulting Editors:

Wiley Series in Computational Statistics is comprised of practical guides and cutting edge research books on new developments in computational statistics. It features quality authors with a strong applications focus. The texts in the series provide detailed coverage of statistical concepts, methods and case studies in areas at the interface of statistics, computing, and numerics.

With sound motivation and a wealth of practical examples, the books show in concrete terms how to select and to use appropriate ranges of statistical computing techniques in particular fields of study. Readers are assumed to have a basic understanding of introductory terminology.

The series concentrates on applications of computational methods in statistics to fields of bioinformatics, genomics, epidemiology, business, engineering, finance and applied statistics.

Titles in the Series

Large-Scale Inverse Problems and Quantification of Uncertainty

Edited by

Lorenz Biegler
Carnegie Mellon University, USA

George Biros
Georgia Institute of Technology, USA

Omar Ghattas
University of Texas at Austin, USA

Matthias Heinkenschloss
Rice University, USA

David Keyes
KAUST and Columbia University, USA

Bani Mallick,
Texas A&M University, USA

Youssef Marzouk
Massachusetts Institute of Technology, USA

Luis Tenorio
Colorado School of Mines, USA

Bart van Bloemen Waanders
Sandia National Laboratories, USA

Karen Willcox
Massachusetts Institute of Technology, USA

A John Wiley and Sons, Ltd., Publication

This edition first published 2011

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

Library of Congress Cataloging-in-Publication Data

Large-scale inverse problems and quantification of uncertainty / edited by Lorenz Biegler ... [et al.].
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-69743-6 (cloth)
1. Bayesian statistical decision theory. 2. Inverse problems (Differential equations) 3. Mathematical optimization. I. Biegler, Lorenz T.
QA279.5.L37 2010
515′.357 – dc22

2010026297

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

Sandia National Laboratories is a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Print ISBN: 978-0-470-69743-6
ePDF ISBN: 978-0-470-68586-0
oBook ISBN: 978-0-470-68585-3

Typeset in 10/12pt Computer Modern font by Laserwords Private Limited, Chennai, India

Contents

List of Contributors

D. Calvetti
Department of Mathematics
Case Western Reserve University
Cleveland, OH, USA

A. Datta-Gupta
Department of Petroleum Engineering
Texas A&M University
College Station, TX, USA

F. Delbos
Institut Francais du petrole
Rueil-Malmaison, France

C. Duffet
Institut Francais du petrole
Rueil-Malmaison, France

Y. Efendiev
Department of Mathematics
Texas A&M University
College Station, TX, USA

G. Evensen
Statoil Research Centre
Bergen, Norway

M. Frangos
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, MA, USA

E. Haber
Department of Mathematics
University of British Columbia
Vancouver, Canada

S. Habib
Nuclear & Particle Physics, Astrophysics, and Cosmology Group
Los Alamos National Laboratory
Los Alamos, NM, USA

K. Heitmann
International, Space and Response Division
Los Alamos National Laboratory
Los Alamos, NM, USA

D. Higdon
Statistical Sciences
Los Alamos National Laboratory
Los Alamos, NM, USA

L. Horesh
IBM T. J. Watson Research Center
Yorktown Heights, NY, USA

J. Hove
Statoil Research Centre
Bergen, Norway

D.B.P. Huynh
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA, USA

K. Hwang
Department of Statistics
Texas A&M University
College Station, TX, USA

P.K. Kitanidis
Department of Civil and Environmental Engineering
Stanford University
Stanford, CA, USA

E. Lawrence
Statistical Sciences
Los Alamos National Laboratory
Los Alamos, NM, USA

X. Ma
Department of Petroleum Engineering
Texas A&M University
College Station, TX, USA

J. McFarland
Southwest Research Institute
San Antonio, TX, USA

B. Mallick
Department of Statistics
Texas A&M University
College Station, TX, USA

Y. Marzouk
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, MA, USA

I. Myrseth
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim, Norway

N.C. Nguyen
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA, USA

H. Omre
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim, Norway

A.T. Patera
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA, USA

G. Rozza
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA, USA

A. Sandu
Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, VA, USA

A. Seiler
Statoil Research Centre
Bergen, Norway

D. Sinoquet
Institut Francais du petrole
Rueil-Malmaison, France

J.-A. Skjervheim
Statoil Research Centre
Bergen, Norway

E. Somersalo
Department of Mathematics
Case Western Reserve University
Cleveland, OH, USA

P.B. Stark
Department of Statistics
University of California
Berkeley, CA, USA

L. Swiler
Optimization and Uncertainty Quantification
Sandia National Laboratories
Albuquerque, NM, USA

L. Tenorio
Mathematical and Computer Sciences
Colorado School of Mines
Golden, CO, USA

J.G. Vabø
Statoil Research Centre
Bergen, Norway

B. van BloemenWaanders
Applied Mathematics and Applications
Sandia National Laboratories
Albuquerque, NM, USA

R.D. Wilkinson
School of Mathematical Sciences
University of Nottingham
Nottingham, United Kingdom

K. Willcox
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, MA, USA

N. Zabaras
Sibley School of Mechanical and Aerospace Engineering
Cornell University
Ithaca, NY, USA

Chapter 1

Introduction

1.1 Introduction

To solve an inverse problem means to estimate unknown objects (e.g. parameters and functions) from indirect noisy observations. A classic example is the mapping of the Earth's subsurface using seismic waves. Inverse problems are often ill-posed, in the sense that small perturbations in the data may lead to large errors in the inversion estimates. Furthermore, many practical and important inverse problems are large-scale; they involve large amounts of data and high-dimensional parameter spaces. For example, in the case of seismic inversion, typically millions of parameters are needed to describe the material properties of the Earth's subsurface. Large-scale, ill-posed inverse problems are ubiquitous in science and engineering, and are important precursors to the quantification of uncertainties underpinning prediction and decision-making.

In the absence of measurement noise, the connection between the parameters (input) and the data (output) defines the forward operator. In the inverse problem we determine an input corresponding to given noisy data. It is often the case that there is no explicit formula for the inversion estimate that maps output to input. We thus often rely on forward calculations to compare the output of different plausible input models to choose one model, or a distribution of models, that is consistent with the data. A further complication is that the forward operator itself may not be perfectly known, as it may depend on unknown tuning parameters. For example, the forward operator in seismic inversion may depend on a velocity model that is only partly known.

It is then clear that there is a need in inverse problems for a framework that includes computationally efficient methods capable of incorporating

Large-Scale Inverse Problems and Quantification of Uncertainty Edited by L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, Y. Marzouk, L. Tenorio, B. van Bloemen Waanders and K. Willcox © 2011 John Wiley & Sons, Ltd

prior information and accounting for uncertainties at different stages of the modeling procedure, as well as methods that can be used to provide measures of the reliability of the final estimates. However, efficient modeling of the uncertainties in the inputs for high-dimensional parameter spaces and expensive forward simulations remain a tremendous challenge for many problems today – there is a crucial unmet need for the development of scalable numerical algorithms for the solution of large-scale inverse problems. While complete quantification of uncertainty in inverse problems for very large-scale nonlinear systems has often been intractable, several recent developments are making it viable: (1) the maturing state of algorithms and software for forward simulation for many classes of problems; (2) the arrival of the petascale computing era; and (3) the explosion of available observational data in many scientific areas.

This book is focused on computational methods for large-scale inverse problems. It includes methods for uncertainty quantification in input and output parameters and for efficient forward calculations, as well as methodologies to incorporate different types of prior information to improve the inversion estimates. The aim is to cross-fertilize the perspectives of researchers in the areas of data assimilation, statistics, large-scale optimization, applied and computational mathematics, high performance computing, and cutting-edge applications.

Given the different types of uncertainties and prior information that have to be consolidated, it is not surprising that many of the methods in the following chapters are defined in a Bayesian framework with solution techniques ranging from deterministic optimization-based approaches to variants of Markov chain Monte Carlo (MCMC) solvers. The choice of algorithms to solve a large-scale problem depends on the problem formulation and a balance of computational resources and a complete statistical characterization of the inversion estimates. If time to solution is the priority, deterministic optimization methods offer a computationally efficient strategy but at the cost of statistical inflexibility. If a complete statistical characterization is required, large computational resources must accommodate Monte Carlo type solvers. It is often necessary to consider surrogate modeling to reduce the cost of forward solves and/or the dimension of the state and inversion spaces.

The book is organized in four general categories: (1) an introduction to statistical Bayesian and frequentist methodologies, (2) approximation methods, (3) Kalman filtering methods, and (4) deterministic optimization based approaches.

1.2 Statistical Methods

Inverse problems require statistical characterizations because uncertainties and/or prior information are modeled as random. Such stochastic structure helps us deal with the complex nature of the uncertainties that plague

many aspects of inverse problems including the underlying simulation model, data measurements and the prior information. A variety of methods must be considered to achieve an acceptable statistical description at a practical computational cost. Several chapters discuss different approaches in an attempt to reduce the computational requirements. Most of these methods are centered around a Bayesian framework in which uncertainty quantification is achieved by exploring a posterior probability distribution. The first three chapters introduce key concepts of the frequentist and Bayesian framework through algorithmic explanations and simple numerical examples. Along the way they also explain how the two different frameworks are used in geostatistical applications and regularization of inverse problems. The subsequent two chapters propose new methods to further build on the Bayesian framework.

Brief summaries of these chapters follow:

- Stark and Tenorio present frequentist and Bayesian methods for inverse problems. They discuss the different ways prior information is used by each school and explain basic statistical procedures such as estimators, confidence intervals and credible regions. They also show how decision theory is used to compare statistical procedures. For example, a frequentist estimator can be compared to a Bayesian one by computing the frequentist mean squared error of each. Credible regions can be compared to frequentist regions via their frequentist coverage. Stark and Tenorio provide illustrative examples of these and other types of comparisons.

- Calvetti and Somersalo clarify where the subjectivity in the Bayesian approach lies and what it really amounts to. The focus is on the interpretation of the probability and on its role in setting up likelihoods and priors. They show how to use hierarchical Bayesian methods to incorporate prior information and uncertainty at different levels of the mathematical model. Algorithms to compute the maximum a-posteriori estimate and sampling methods to explore the posterior distribution are discussed. Dynamic updating and the classic Kalman filter algorithm are introduced as a prelude to the chapters on Kalman and Bayesian filtering.

- Kitanidis presents the Bayesian framework as the appropriate methodology to solve inverse problems. He explains how the Bayesian approach differs from the frequentist approach, both in terms of methodology and in terms of the meaning of the results one obtains, and discusses some disagreements between Bayesians and non-Bayesians in the selection of prior distributions. Bayesian methods for geostatiscal analysis are also discussed.

- Higdon *et al.* consider the problem of making predictions based on computer simulations. They present a Bayesian framework to combine

available data to estimate unknown parameters for the computer model and assess prediction uncertainties. In addition, their methodology can account for uncertainties due to limitations on the number of simulations. This chapter also serves as an introduction to Gaussian processes and Markov chain Monte Carlo (MCMC) methods.

- Efendiev *et al.* present a strategy to efficiently sample from a surrogate model obtained from a Bayesian Partition Model (BPM) which uses Voronoi Tessellations to decompose the entire parameter space into homogeneous regions and use the same probability distribution within each region. The technique is demonstrated on an inversion of permeability fields and fractional flow simulation from the Darcy and continuity equations. The high dimensional permeability field is approximated by a Karhunen-Loeve expansion and then combined using regression techniques on different BPM regions. A two-stage MCMC method has been employed, where at the first stage the BPM approximation has been used thereby creating a more efficient MCMC method.

1.3 Approximation Methods

The solution of large-scale inverse problems critically depends on methods to reduce computational cost. Solving the inverse problem typically requires the evaluation of many thousands of plausible input models through the forward problem; thus, finding computationally efficient methods to solve the forward problem is one important component of achieving the desired cost reductions. Advances in linear solver and preconditioning techniques, in addition to parallelization, can provide significant efficiency gains; however, in many cases these gains are not sufficient to meet all the computational needs of large-scale inversion problems. We must therefore appeal to approximation techniques that seek to replace the forward model with an inexpensive surrogate. In addition to yielding a dramatic decrease in forward problem solution time, approximations can reduce the dimension of the input space and entail more efficient sampling strategies, targeting a reduction in the number of forward solves required to find solutions and assess uncertainties. Below we describe a number of chapters that employ combinations of these approximation approaches in a Bayesian inference framework.

- Frangos *et al.* summarize the state of the art in methods to reduce the computational cost of solving statistical inverse problems. A literature survey is provided for three classes of approaches – approximating the forward model, reducing the size of the input space, and reducing the number of samples required to compute statistics of the posterior. A simple example demonstrates the relative advantages of polynomial chaos-based surrogates and projection-based model reduction of the forward simulator.

- Nguyen *et al.* present a reduced basis approximation approach to solve a real time Bayesian parameter estimation problem. The approach uses Galerkin projection of the nonlinear partial differential equations onto a low-dimensional space that is identified through adaptive sampling. Decomposition into 'Offline' and 'Online' computational tasks achieves solution of the Bayesian estimation problem in real time. A posteriori error estimation for linear functionals yields rigorous bounds on the results computed using the reduced basis approximation.

- Swiler *et al.* present a Bayesian solution strategy to solve a model calibration in which the underlying simulation is expensive and observational data contains errors or uncertainty. They demonstrate the use of Gaussian surrogates to reduce the computational expense on a complex thermal simulation of decomposing foam dataset.

- Wilkinson discusses emulation techniques to manage multivariate output from expensive models in the context of calibration using observational data. A focus of this work is on calibration of models with long run time. Consequently an ensemble comprising of only a limited number of forward runs can be considered. A strategy is presented for selecting the design points that are used to define the ensemble. When the simulation model is computationally expensive, emulation is required and here a Bayesian approach is used.

1.4 Kalman Filtering

The next two chapters discuss filtering methods to solve large-scale statistical inverse problems. In particular, they focus on the ensemble Kalman filter, which searches for a solution in the space spanned by a collection of ensembles. Analogous to reducing the order of a high dimensional parameter space using a stochastic spectral approximation or a projection-based reduced-order model, the ensemble Kalman filter assumes that the variability of the parameters can be well approximated by a small number of modes. As such, large numbers of inversion parameters can be accommodated in combination with complex and large-scale dynamics. This comes however, at the cost of less statistical flexibility, since approximate solutions are needed for nonlinear non-Gaussian problems.

- Myrseth *et al.* provide an overview of the ensemble Kalman filter in addition to various other filters. Under very specific assumptions about linearity and Gaussianity, exact analytical solutions can be determined for the Bayesian inversion, but for any deviation from these assumptions, one has to rely on approximations. The filter relies on simulation based inference and utilizes a linearization in the data conditioning. These approximations make the ensemble Kalman filter computationally efficient and well suited for high-dimensional hidden Markov models.

- Seiler *et al.* discuss the use of the ensemble Kalman filter to solve a large inverse problem. This approach uses a Monte Carlo process for calculating the joint probability density function for the model and state parameters, and it computes the recursive update steps by approximating the first and second moments of the predicted PDF. The recursive Bayesian formulation can be solved using the ensemble Kalman filter under the assumption that predicted error statistics are nearly Gaussian. Instead of working with the high-dimensional parameter space, the inverse problem is reduced to the number of realizations included in the ensemble. The approach is demonstrated on a petroleum reservoir simulation dataset in which a history matching problem is solved.

1.5 Optimization

An alternative strategy to a Bayesian formulation is to pose the statistical inverse problem as an optimization problem. Computational frameworks that use this approach build upon state-of-the-art methods for simulation of the forward problem, as well as the machinery for large-scale deterministic optimization. Typically, Gaussian assumptions are made regarding various components of the data and models. This reduces the statistical flexibility and perhaps compromises the quality of the solution. However as some of these chapters will demonstrate, these methods are capable of addressing very large inversion spaces while still providing statistical descriptions of the solution.

- Horesh *et al.* present optimal experimental design strategies for large-scale nonlinear ill-posed inverse problems. In particular, strategies for a nonlinear impedance imaging problem are presented. Optimal selection of source and receiver locations is achieved by solving an optimization problem that controls the performance of the inversion estimate subject to sparsity constraints.

- Zabaras outlines a methodology to perform estimation under multiple sources of uncertainties. By relying on the use of the deterministic simulator, the solution of the stochastic problem is constructed using sparse grid collocation. Furthermore, the stochastic solution is converted to a deterministic optimization problem in a higher dimensional space. Stochastic sensitivities are calculated using deterministic calculations. The evolution of a PDF requires the solution of a billion DOFs at each stochastic optimization iteration. The technique is demonstrated on a heat flux problem.

- Frederic *et al.* present an uncertainty analysis approach for seismic inversion and discuss tomography and ray tracing, which are efficient methods to predict travel time. The deterministic inversion approach makes the connection between the Hessian and the covariance. The

prior and posterior PDFs are assumed to be Gaussian. By estimating just the diagonal terms of the covariance matrix for the prior and therefore assuming the errors are uncorrelated, the computational expense is reduced considerably. By assuming Gaussianity and calculating the diagonal terms, the inversion formulation has a specific form that involves the Hessian. Two strategies are presented, one where the full Hessian is inverted to give the covariance and the other where a multi-parameter approach is used to reduce the computational expense of the Hessian inversion.

- Sandu discusses the use of adjoint methods, which are at the core of many optimization strategies for large-scale inverse problems. This chapter presents an analysis of the properties of Runge-Kutta and linear multistep methods in the context of solving ODEs that arise in adjoint equations. An example shows the use of discrete adjoint methods in the solution of large-scale data assimilation problems for air quality modeling.

Chapter 2

A Primer of Frequentist and Bayesian Inference in Inverse Problems

P.B. Stark[1] and L. Tenorio[2]
[1] *University of California at Berkeley, USA*
[2] *Colorado School of Mines, USA*

2.1 Introduction

Inverse problems seek to learn about the world from indirect, noisy data. They can be cast as statistical estimation problems and studied using statistical decision theory, a framework that subsumes Bayesian and frequentist methods. Both Bayesian and frequentist methods require a stochastic model for the data, and both can incorporate constraints on the possible states of the world. Bayesian methods require that constraints be re-formulated as a *prior probability distribution*, a stochastic model for the unknown state of the world. If the state of the world is a random variable with a known distribution, Bayesian methods are in some sense optimal. Frequentist methods are easier to justify when the state of the world is unknown but not necessarily random, or random but with an unknown distribution; some frequentist methods are then optimal – in other senses.

Parameters are numerical properties of the state of the world. *Estimators* are quantities that can be computed from the data without knowing the state of the world. Estimators can be compared using *risk functions*, which quantify

Large-Scale Inverse Problems and Quantification of Uncertainty Edited by L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, Y. Marzouk, L. Tenorio, B. van Bloemen Waanders and K. Willcox

the expected 'cost' of using a particular estimator when the world is in a given state. (The definition of cost should be dictated by the scientific context, but some definitions, such as mean squared error, are used because they lead to tractable computations.)

Generally, no estimator has the smallest risk for all possible states of the world: there are tradeoffs, which Bayesian and frequentist methods address differently. Bayesian methods seek to minimize the expected risk when the state of the world is drawn at random according to the prior probability distribution. One frequentist approach – *minimax estimation* – seeks to minimize the maximum risk over all states of the world that satisfy the constraints.

The performance of frequentist and Bayesian estimators can be compared both with and without a prior probability distribution for the state of the world. Bayesian estimators can be evaluated from a frequentist perspective, and vice versa. Comparing the minimax risk with the Bayes risk measures how much information the prior probability distribution adds that is not present in the constraints or the data.

This chapter sketches frequentist and Bayesian approaches to estimation and inference, including some differences and connections. The treatment is expository, not rigorous. We illustrate the approaches with two examples: a concrete one-dimensional problem (estimating the mean of a Normal distribution when that mean is known to lie in the interval $[-\tau, \tau]$), and an abstract linear inverse problem.

For a more philosophical perspective on the frequentist and Bayesian interpretations of probability and models, see Freedman and Stark (2003); Freedman (1995). For more careful treatments of the technical aspects, see Berger (1985); Evans and Stark (2002); Le Cam (1986).

2.2 Prior Information and Parameters: What Do You Know, and What Do You Want to Know?

This section lays out some of the basic terms, most of which are shared by frequentist and Bayesian methods. Both schools seek to learn about the state of the world – to estimate parameters – from noisy data. Both consider constraints on the possible states of the world, and both require a stochastic measurement model for the data.

2.2.1 The State of the World, Measurement Model, Parameters and Likelihoods

The term 'model' is used in many ways in different communities. In the interest of clarity, we distinguish among three things sometimes called 'models,' namely, the state of the world, the measurement model, and parameters.

The state of the world, denoted by θ, is a mathematical representation of the physical system of interest, for example, a parametrized representation of seismic velocity as a function of position in the Earth, of the angular velocity of material in the Sun, or of the temperature of the cosmic microwave background radiation as a function of direction. Often in physical problems, some states of the world can be ruled out by physical theory or prior experiment. For instance, mass densities and energies are necessarily nonnegative. Transmission coefficients are between 0 and 1. Particle velocities are less than the speed of light. The rest mass of the energy stored in Earth's magnetic field is less than the mass of the Earth (Backus 1989). The set Θ will represent the possible states of the world (values of θ) that satisfy the constraints. That is, we know a priori that $\theta \in \Theta$.[1]

The observations Y are related to the particular state of the world θ through a *measurement model* that relates the probability distribution of the observations to θ. The set of possible observations is denoted $\mathcal{Y}$, called the *sample space*. Typically, Y is an n-dimensional vector of real numbers; then, $\mathcal{Y}$ is $\mathbb{R}^n$. Depending on θ, Y is more likely to take some values in $\mathcal{Y}$ than others. The probability distribution of the data Y when the state of the world is η (i.e., when $\theta = \eta$) is denoted $\mathbb{P}_\eta$; we write $Y \sim \mathbb{P}_\eta$. (The 'true' state of the world is some particular – but unknown – $\theta \in \Theta$; η is a generic element of Θ that might or might not be equal to θ.) We shall assume that the set of distributions $\mathcal{P} \equiv \{\mathbb{P}_\eta : \eta \in \Theta\}$ is dominated by a common measure μ.[2] (In the special case that μ is Lebesgue measure, that just means that all the probability distributions $\mathcal{P} \equiv \{\mathbb{P}_\eta : \eta \in \Theta\}$ have densities in the ordinary sense.) The *density of* $\mathbb{P}_\eta$ *(with respect to* μ*) at* y is

$$p_\eta(y) \equiv d\mathbb{P}_\eta/d\mu|_y. \tag{2.1}$$

The *likelihood of* η *given* $Y = y$ is $p_\eta(y)$ viewed as a function of η, with y fixed.

For example, suppose that for the purposes of our experiment, the state of the world can be described by a single number $\theta \in \mathbb{R}$ that is known to be in the interval $[-\tau, \tau]$, and that our experiment measures θ with additive Gaussian noise that has mean zero and variance 1. Then the measurement model is

$$Y = \theta + Z, \tag{2.2}$$

where Z is a standard Gaussian random variable (we write $Z \sim N(0,1)$). The set $\Theta = [-\tau, \tau]$. Equivalently, we may write $Y \sim N(\theta, 1)$ with $\theta \in [-\tau, \tau]$.

[1] There is a difference between the physical state of the world and a numerical discretization or approximation of the state of the world for computational convenience. The numerical approximation to the underlying physical process contributes uncertainty that is generally ignored. For discussion, see, e.g., Stark (1992b).

[2] By defining μ suitably, this can allow us to work with 'densities' even if the family $\mathcal{P}$ contains measures that assign positive probability to individual points. Assuming that there is a dominating measure is a technical convenience that permits a general definition of likelihoods.

(The symbol $\sim$ means 'has the probability distribution' or 'has the same probability distribution as.') Thus $\mathbb{P}_\eta$ is the Gaussian distribution with mean η and variance 1. This is called the *bounded normal mean* (BNM) problem. The BNM problem is of theoretical interest, and is a building block for the study of more complicated problems in higher dimensions; see, for example, Donoho (1994). The dominating measure μ in this problem can be taken to be Lebesgue measure. Then the density of $\mathbb{P}_\theta$ at y is

$$\varphi_\theta(y) \equiv \frac{1}{\sqrt{2\pi}} e^{-(y-\theta)^2/2}, \tag{2.3}$$

and the likelihood of η given $Y = y$ is $\varphi_\eta(y)$ viewed as a function of η with y fixed.

As a more general example, consider the following canonical linear inverse problem. The set Θ is a ball in a norm or semi-norm in a separable, infinite-dimensional Banach space (for example, Θ might be a set of functions whose integrated squared second derivative is less than some constant $C < \infty$, or a set of nonnegative functions that are continuous and bounded).[3] The data Y are related to the state of the world θ through the action of a linear operator K from Θ to $\mathcal{Y} = \mathbb{R}^n$, with additive noise:

$$Y = K\theta + \varepsilon, \quad \theta \in \Theta, \tag{2.4}$$

where the probability distribution of the noise vector ε is known. We assume that $K\eta = (K_1\eta, K_2\eta, \ldots, K_n\eta)$ for $\eta \in \Theta$, where $\{K_j\}_{j=1}^n$ are linearly independent bounded linear functionals on Θ. Let $f(y)$ denote the density of ε with respect to a dominating measure μ. Then $p_\eta(y) = f(y - K\eta)$. The BNM problem is an example of a linear inverse problem with K the identity operator, $\mathcal{Y} = \mathbb{R}$, $\Theta \equiv [-\tau, \tau]$, $\varepsilon \sim N(0,1)$, $f(y) = \varphi_0(y)$.

A *parameter* $\lambda = \lambda[\theta]$ is a property of the state of the world. The entire description of the state of the world, θ, could be considered to be a parameter; then λ is the identity operator. Alternatively, we might be interested in a simpler property of θ. For example, in gravimetry, the state of the world θ might be mass density as a function of position in Earth's interior, and the parameter of interest, $\lambda[\theta]$, might be the average mass density in some region below the surface. In that case, the rest of θ is a *nuisance parameter*: it can affect the (probability distribution of the) measurements, but it is not of primary interest.

Our lead examples in this paper are the 'bounded normal mean' problem and the canonical linear inverse problem just described.

2.2.2 Prior and Posterior Probability Distributions

In the present framework, there is prior information about the state of the world θ expressed as a constraint $\theta \in \Theta$. Frequentists use such constraints

[3]Separable Banach spaces are measurable with respect to the σ-algebra induced by the norm topology, a fact that ensures that prior probability distributions – required by Bayesian methods – can be defined.