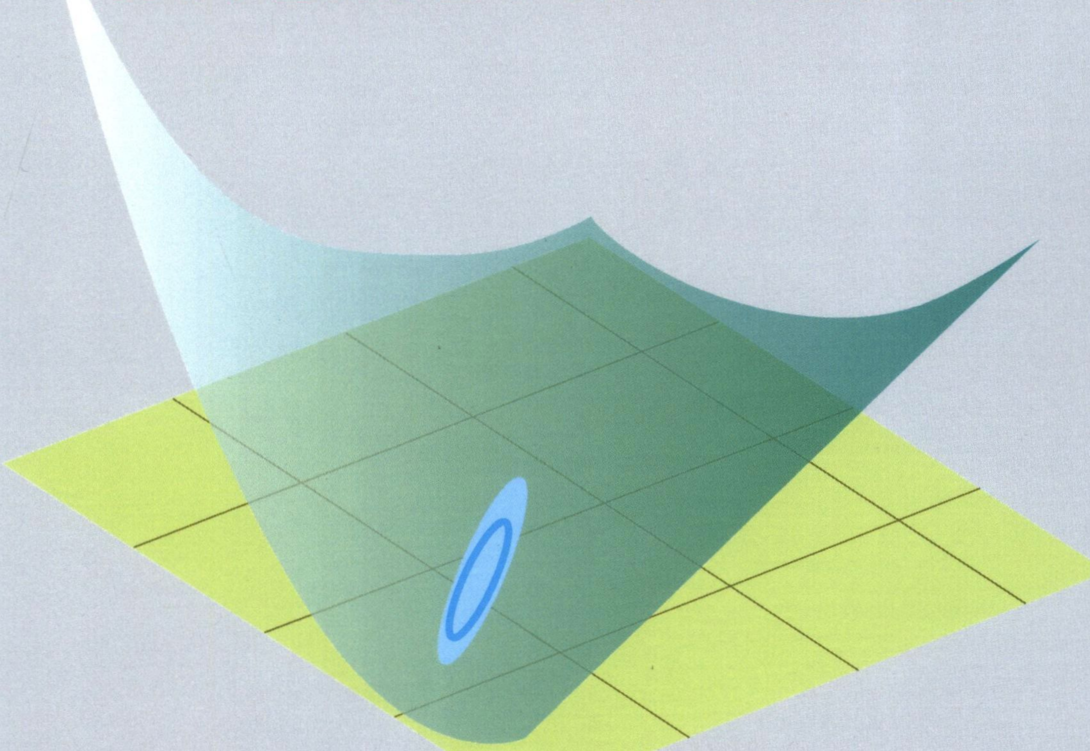
An Introduction to DATA ANALYSIS and UNCERTAINTY QUANTIFICATION for INVERSE PROBLEMS



LUIS TENORIO

• MATHEMATICS IN INDUSTRY •

Inverse problems are found in many applications, such as medical imaging, engineering, astronomy, and geophysics, among others. To solve an inverse problem is to recover an object from noisy, usually indirect observations. Solutions to inverse problems are subject to many potential sources of error introduced by approximate mathematical models, regularization methods, numerical approximations for efficient computations, noisy data, and limitations in the number of observations; thus it is important to include an assessment of the uncertainties as part of the solution. Such assessment is interdisciplinary by nature, as it requires, in addition to knowledge of the particular application, methods from applied mathematics, probability, and statistics.

This book bridges applied mathematics and statistics by providing a basic introduction to probability and statistics for uncertainty quantification in the context of inverse problems, as well as an introduction to statistical regularization of inverse problems. The author covers basic statistical inference, introduces the framework of ill-posed inverse problems, and explains statistical questions that arise in their applications.

An Introduction to Data Analysis and Uncertainty Quantification for Inverse Problems includes

- many examples that explain techniques which are useful to address general problems arising in uncertainty quantification,
- Bayesian and non-Bayesian statistical methods and discussions of their complementary roles, and
- analysis of a real data set to illustrate the methodology covered throughout the book.

This book is intended for senior undergraduates and beginning graduate students in mathematics, engineering and physical sciences. The material spans from undergraduate statistics and probability to data analysis for inverse problems and probability distributions on infinite-dimensional spaces. It is also intended for researchers working on inverse problems and uncertainty quantification in geophysics, astrophysics, physics, and engineering. Because the statistical and probability methods covered have applications beyond inverse problems, the book may also be of interest to those people working in data science or in other applications of uncertainty quantification.

Luis Tenorio is a faculty member in the Applied Mathematics and Statistics Department at the Colorado School of Mines. He obtained his PhD in mathematics at the University of California at Berkeley and worked on inverse problems in astrophysics as a member of George Smoot's astrophysics group in the Lawrence Berkeley National Laboratory. His main research interests are the statistical aspects of inverse problems with applications to astrophysics and geophysics.

For more information about SIAM books, journals, conferences, memberships, or activities, contact:

siam

Society for Industrial and Applied Mathematics
3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA
+1-215-382-9800 • Fax +1-215-386-7999
siam@siam.org • www.siam.org

MN03

ISBN 978-1-611974-91-1
90000
9781611974911

An Introduction to **DATA ANALYSIS** and **UNCERTAINTY QUANTIFICATION** for **INVERSE PROBLEMS**

LUIS TENORIO

siam.

MN03

An Introduction to DATA ANALYSIS and UNCERTAINTY QUANTIFICATION for INVERSE PROBLEMS

LUIS TENORIO
Colorado School of Mines
Golden, Colorado

siam.
Society for Industrial and Applied Mathematics
Philadelphia

10 9 8 7 6 5 4 3 2 1

Plot on front cover courtesy of Oscar Aguilar.

Publisher	David Marshall
Acquisitions Editor	Elizabeth Greenspan
Developmental Editor	Gina Rinelli Harris
Managing Editor	Kelly Thomas
Production Editor	David Riegelhaupt
Copy Editor	Samar Nour-El-Deen
Production Manager	Donna Witzleben
Production Coordinator	Cally Shrader
Compositor	Cheryl Hufnagle
Graphic Designer	Lois Sellers

Library of Congress Cataloging-in-Publication Data
Names: Tenorio, Luis.
Title: An introduction to data analysis and uncertainty quantification for inverse problems / Luis Tenorio, Colorado School of Mines, Golden, Colorado.
Description: Philadelphia : Society for Industrial and Applied Mathematics, [2017] | Series: Mathematics in industry ; 3 | Includes bibliographical references and index.
Identifiers: LCCN 2017012790 (print) | LCCN 2017016391 (ebook) | ISBN 9781611974928 (ebook) | ISBN 9781611974911 (print)
Subjects: LCSH: Inverse problems (Differential equations) | Parameter estimation.
Classification: LCC QA378.5 (ebook) | LCC QA378.5 .T46 2017 (print) | DDC 515/.357--dc23
LC record available at https://lccn.loc.gov/2017012790

An Introduction to DATA ANALYSIS and UNCERTAINTY QUANTIFICATION for INVERSE PROBLEMS

• MATHEMATICS IN INDUSTRY •

Series Volumes

Eldad Haber, *Computational Methods in Geophysical Electromagnetics*
Lyn Thomas, Jonathan Crook, David Edelman, *Credit Scoring and Its Applications, Second Edition*
Luis Tenorio, *An Introduction to Data Analysis and Uncertainty Quantification for Inverse Problems*

Preface

Roughly speaking, to solve an inverse problem is to recover an object (e.g., parameter or function) from noisy (typically indirect) observations. In most cases such recovery cannot be done exactly because the mathematical models that link data to the object are approximations, data are noisy, the number of observations is finite, and obtaining a solution may require further approximations for efficient numerical computations. The importance of assessing the reliability of solutions to inverse problems is evident given such potential sources of errors. This assessment step is part of what is now called uncertainty quantification (UQ). Uncertainty quantification for inverse problems and other problems in engineering requires familiarity with some basic methods from mathematics, probability, and statistics. But what I have observed during years of collaborations with scientists and applied mathematicians working on inverse problems is that they often do not feel as comfortable with their knowledge of probability or statistics as they do with their background in applied mathematics. The converse is also true: I have encountered statisticians interested in making contributions to inverse problems but who have not been exposed to the basic theory of inverse problems and the questions that arise in their applications. The objective of this book is therefore to serve as a bridge between the applied mathematics and statistics communities. I try to take advantage of the reader's mathematical background to provide a basic introduction to probability and statistics for UQ mainly in the context of inverse problems, a field with many important practical applications. In addition, the book provides a basic introduction to statistical regularization of inverse problems for those with a background in statistics. Since the reader is assumed to be comfortable with mathematical methods at the level of senior undergraduates and beginning graduate students in mathematics, engineering, and physical sciences, much ground can be covered: from undergraduate statistics and probability to probability distributions on infinite-dimensional spaces. For statisticians, the book uses classic linear regression and statistical inference to introduce the framework of ill-posed inverse problems and explain statistical questions that arise in their applications. A review of the mathematical analysis tools required for inverse problems is also included in the appendix. Since the statistical and probability methods covered have applications beyond inverse problems, the book may also be of interest to people working in data science or in other applications of UQ.

The selection of topics I cover has been strongly influenced by discussions I have had over the years with scientists, applied mathematicians, engineers, and students from a wide variety of fields. In particular, since advances in computational power have made the use of Bayesian methodology commonplace in many fields of application, I believe that the existence of different schools of inference to conduct UQ is a topic that deserves more attention. For example, I have encountered practitioners who were either not aware of the existence of non-Bayesian (e.g., frequentist) methods for inverse problems or could not tell (or care about) the difference. And among those who know about frequentist and Bayesian methods, there are many who are baffled by common heated discussions among statisticians regarding the merits or demerits of these two schools of inference. In practical applications we learn that Bayesian and frequentist methods (as well as likelihood methods) provide valuable tools for UQ and for statistical analysis in general. I therefore try to cover both frameworks and to explain their assumptions, corresponding interpretations, and their important complementary roles by means of examples. To keep the mathematics and probability theory accessible to a wide audience, I consider probability distributions mostly on finite-dimensional spaces but do provide some background and examples that serve as introduction to the infinite-dimensional case. However, even within the framework of finite-dimensional inverse problems, there is no single statistical methodology that will work in every application: UQ is highly problem dependent. I have chosen a particular framework that is widely used, has many practical applications, and provides basic tools for more complex problems.

We are all aware of how difficult it is to put to use new definitions and results as this requires techniques that are learned with experience. To help with this transition, each section includes examples with explicit calculations that introduce useful problem solving techniques relevant to the particular topic. Examples are also used to clarify theoretical concepts and to illustrate the type of applications for which the methods could be used. I include over 130 examples but choosing them has not been easy. I have tried to select simple illustrative examples that can be understood by a diverse audience. Although it may not be apparent, many of the examples are simplifications that capture the essence of more complex questions that arise in applications but which would require much background to explain fully. Some examples are in fact answers to questions I have received from students and collaborators over the years. Some sections also include more theoretical but important details to help warn the reader of subtle statistical/probabilistic issues that arise in applications of UQ and which could be easily overlooked.

The book is organized as follows. Chapter 1 provides an introduction to inverse problems and regularization. Chapters 2 and 4 cover probability and statistical methods whose applications to inverse problems are considered in Chapters 3, 4, and 5. Chapter 3 includes methods for data analysis, Chapter 4 focuses on Bayesian methods that are relevant to inverse problems, and Chapter 5 is dedicated to the data analysis of one particular set of experimental data. One of the goals of Chapter 5 is to illustrate the nuances that arise when we try to apply theory to the analysis of real data. The book includes two appendices: In order to make the book as self-contained as possible, and to establish the general terminology used throughout, Appendix A provides a summary of results from analysis that are used in different parts of the book. Given the importance of conditional probability for Bayesian inference, Appendix B provides a more careful discussion of conditional probability and conditional expectation, including the definition of regular conditional probability. Appendix B assumes some knowledge of measure theoretic probability but is not required for the understanding of the other chapters. It includes an introduction to an alternative approach to conditional probability based on disintegration which is not commonly taught. I believe this is a natural approach that may help some readers get a more intuitive understanding of conditional probability and expectation.

As explained above, the objective of this book is to provide a basic background in statistics and probability for UQ mainly in the framework of inverse problems. My hope is that this book can be used to complement other textbooks that focus on regularization, mathematical analysis or computational methods for inverse problems. Readers interested in learning more about regularization or the general theory of inverse problems may consult [93, 120, 150, 154], or [8, 126, 256] for more applied or computational introductions. The book [30] provides an edited collection of research papers and tutorials for UQ and large-scale inverse problems. For more material on Bayesian methods for finite-dimensional inverse problems see, for example, [49, 148, 246], and [241] for an introduction to Bayesian methods in the infinite-dimensional setting. The book should also help the reader learn the basic theory needed to study Markov chain Monte Carlo (MCMC) methods, which play a key role in Bayesian statistics but which I do not cover in this book. There are many good references dedicated to the theory or implementations of MCMC methods and Bayesian computation [42, 49, 113, 209, 234, 245]. The analysis of inverse problems also requires numerical optimization methods not discussed in this book. The reader may find introductions to optimization methods that are important for inverse problems in [8, 39, 192, 256]. Readers interested in learning more about general statistical methods, frequentist or Bayesian, may consult, for example, [22, 52, 55, 104, 163, 164, 224, 226]. The reader may also find [223] interesting as it provides a historical account of the role statistics has played in the twentieth century.

I would like to thank Fadil Santosa and the IMA for organizing and funding workshops where preliminary versions of this work have been used. I am also thankful to Roger Ghanem for giving me the opportunity to do the same at USC. I am very grateful to mentors, reviewers, colleagues, students, and friends who have either provided valuable feedback in the writing of this book, or who have played an important role in my understanding and appreciation of the subject. In particular, I am specially thankful to Oscar Aguilar, Vaughn Ball, Wolfgang Bangerth, Julianne and Tia Chung, Maarten de Hoop, Colin Fox, Mahadevan Ganesh, Eldad Haber, Alex Kalmikov, Paul Martin, Youssef Marzouk, Joyce McLaughlin, Bill Navidi, Aaron Porter, Juan Restrepo, John Scales, George Smoot, Alessio Spantini, and Philip Stark. Finally, I would like to thank Cheryl for her unfailing support and patience.

An apology regarding notation. Different areas of statistics, probability, mathematics, and physics have different notational conventions. For example, it is common in statistics to denote random variables with capital letters and their realizations with lower case (e.g., x is a value the random variable X takes). But in this book we need letters to denote sets, σ-algebras, random sets, scalars, vectors, matrices, random variables, random elements, functions, operators, measures, inner product spaces, normed spaces, etc. This makes it very difficult to follow any particular convention consistently. I hope the notation will be clear from the context.

Contents

Chapter 1

An introduction to inverse problems

In this chapter some classic examples are used to illustrate what an inverse problem is, what it means for the inverse problem to be ill-posed and what one can do to solve an ill-posed inverse problem. We will also see the role that statistics plays when some of the components of the mathematical model that defines the inverse problem are modeled as random.

1.1 ▪ What is an ill-posed inverse problem?

Suppose we have an input, θ, that is transformed by a physical system to produce an output, $\mathscr{K}(\theta)$. In this case $\mathscr{K}$ is called a *forward operator*, it represents the action of the physical system on the input θ. By *forward problem* we mean determining $\mathscr{K}(\theta)$ for a given θ. The *inverse problem* is the recovery of θ given the output $\mathscr{K}(\theta)$. That is, it is the recovery of an object from indirect observations. Inverse problems arise in many fields such as astronomy, physics, geophysics, engineering and medical imaging. The reader may find many examples in [27, 43, 198]. In this section we consider three examples of inverse problems that illustrate important characteristics of inverse problems in general.

The difficulties that arise in solving an inverse problem are different from those we face with the forward problem. For example, since the application of a physical system usually leads to information loss, it is to be expected that recovering the exact input is impossible without the use of complementary information that allows the recovery of what has been lost. Without this extra information one cannot choose a solution that is consistent with the data and is physically meaningful. This is one way the inverse problem is ill-posed. Informally, we may think of *regularization* as a way to include information for the recovery of the unknown input. But regularization can also used to stabilize the inversion and obtain solutions that are not dominated by noise.

Example 1.1 (Linear regression and least-squares. Unique but unstable solutions). We start with the classic linear regression and least-squares framework that is typically covered in elementary statistics courses. It is an important example that will be used throughout the book. A data vector, y, is modeled as

$$y = K\beta + \varepsilon, \tag{1.1.1}$$

where K is a known $n \times p$ matrix and ε is an error (noise) vector. The objective is to use the data to obtain an estimate of the parameter vector β. We make the usual linear regression assumptions that the columns of K are linearly independent and $n > p$. In particular, there are more observations than unknowns and the matrix $K^\top K$ is nonsingular. So, $Kx = 0$ only if $x = 0$. This is a particular type of inverse problem where the goal is to recover the input, β, from a vector, y, of indirect noisy observations linked to β through a linear forward operator. Since the vector y may not belong to the range of K because of the presence of noise, we look for an estimate, $\widehat{\beta}$, such that $K\widehat{\beta}$ is a best approximation of y, where best is defined in the sense of minimizing the Euclidean distance from y to the column-space of K. Hence, we need to find the minimizer of the function $F(x) = ||y - Kx||^2$. Since the columns of K are linearly independent, F defines a quadratic function with a unique minimizer that can be found by setting to zero the derivative of F:

$$\frac{\partial F(x)}{\partial x} = -2K^\top y + 2K^\top K x = 0.$$

Hence, the minimizer of F is

$$\widehat{\beta} = \arg\min_x ||y - Kx||^2 = (K^\top K)^{-1} K^\top y. \tag{1.1.2}$$

The vector $\widehat{\beta}$ is called the *least-squares estimate* of β.[1] The vector $K\widehat{\beta}$ is the linear combination of the columns of K that is closest to y in the Euclidean norm. (We will see that under some conditions on the noise, $\widehat{\beta}$ is also the maximum likelihood estimate of β.)

This simple linear regression problem of estimating β is similar to the inverse problems considered in this book except for one important difference: instead of a matrix, K, we often start with an operator, $\mathscr{K}$, defined on an infinite-dimensional vector space. Once the problem is discretized, we end up with a matrix K, as in linear regression, but this time the matrix $K^\top K$ will typically be *ill-conditioned*. That is, the ratio of its largest to smallest eigenvalue is very large. We will see in Example 1.5 that this makes the least-squares estimate very sensitive to noise. Thus, we need to find a way to obtain a solution that is not dominated by noise. One possibility is to use the penalized least-squares method described in Example 1.5. But this is certainly not the only way; one can include prior information in other ways such as inequality constraints and probability distributions. ■

In linear regression the noise is typically modeled as random while in the classic framework of inverse problems noise is often modeled as deterministic with an upper bound on its norm [250, 251]. In this book the noise will be modeled as random, an approach that seems to have been first considered in [242]. Some discussion on these two approaches can be found in [58, 85]. The assumption that the noise is random and zero-mean is not necessarily valid if, for example, the model is not specified properly, such as when there is a nonzero component introduced by an unmodeled physical process. Validation is an important step in the analysis of an inverse problem to check that the conclusions are not driven by unreasonable assumptions.

In Example 1.1 we considered inverse problems leading to a linear regression where $K^\top K$ is ill-conditioned. The inverse problem may also be ill-posed in the sense that the object to be recovered is unidentifiable given the information provided by the data.

---

[1] A "hat" on a parameter will denote an estimator of such parameter.

Example 1.2 (Nonuniqueness. Nonidentifiability). Consider again the model (1.1.1) but this time assume that K has a nontrivial nullspace (i.e., the columns of K are not linearly independent). Fix a nonzero vector v_0 in the nullspace of K. Then even though there is a true β that produced the data, any $\beta + \alpha v_0$ leads to the same data for any scalar α. There are infinitely many solutions that are consistent with the data; the data do not provide enough information to identify the true β. One can use a variety of methods, such as minimizing the quadratic data-misfit as in least-squares but with a penalty term (this is the method of *penalized least-squares* that will be discussed below), to choose a particular solution. The basic idea is to impose constraints to come up with an estimate that is reasonable. But to define reasonable constraints we need to have more information about the unknown β we hope to approximate. And once we have an estimate we still need to assess how good it is. In Chapter 3 we will consider methods to assess the quality of solutions to inverse problems and study the properties of penalized least-squares estimators. In section 2.6.2 we derive properties of the pseudoinverse least-squares estimate that can be used when K has linearly dependent columns. ■

In the next example we consider an inverse problem with a forward operator defined on an infinite-dimensional space. It is clear that such an operator most have a nullspace if it maps into a finite-dimensional vector space. Thus, it is the type of problem discussed in Example 1.2. However, we may choose to discretize the operator so that the resulting matrix is nonsingular but, as we will see later, we still run into problems caused by the ill-posed nature of the inverse problem because the reduced system is ill-conditioned as in Example 1.1.

Example 1.3 (Deblurring as an ill-posed inverse problem. Integral equations). We consider the classic problems of blurring and deblurring a signal described by a function f. This is a problem that arises in signal and image processing (e.g., in astronomy and seismic exploration) [27, 71, 212]. For simplicity we assume that f is defined on the interval $[0,1]$. The blurring can be modeled as a convolution of the signal with a kernel function, K, determined by the characteristics of the observing instrument. This convolution has the effect of smoothing the signal. In this case the forward operator is called a *convolution operator*. The forward problem consists of blurring the input signal f: For each x we have an operator $\mathcal{K}_x$ that maps f to an output $\mu(x)$:

$$\mu(x) = \mathcal{K}_x f = \int_0^1 K(x-t) f(t)\, dt. \tag{1.1.3}$$

For example, a Gaussian kernel is a function of the form $K(x) = e^{-x^2/\sigma^2}$, where σ controls the width of the kernel; the wider the kernel, the smoother the convolved signal. An equation of the form (1.1.3) is called a *Fredholm integral equation* of the first kind [154]. The inverse problem consists of deblurring $\mu(x)$ to recover f, thus undoing the effect of the forward operator. The effect of smoothing is the loss of high-frequency information; two very different signals may get mapped under $\mathcal{K}$ to very similar functions. For example, Figure 1.1 shows two very different functions, f and g, that under a Gaussian smoothing kernel are transformed to similar functions (here g is the smoothed version of f, i.e., a version of f without the high-frequency components). This behavior is often an indication that the problem is ill-posed (a more formal definition of ill-posed problem will be given in section 1.2). In addition to the loss of information caused by the smoothing, in practical applications we face the problem of

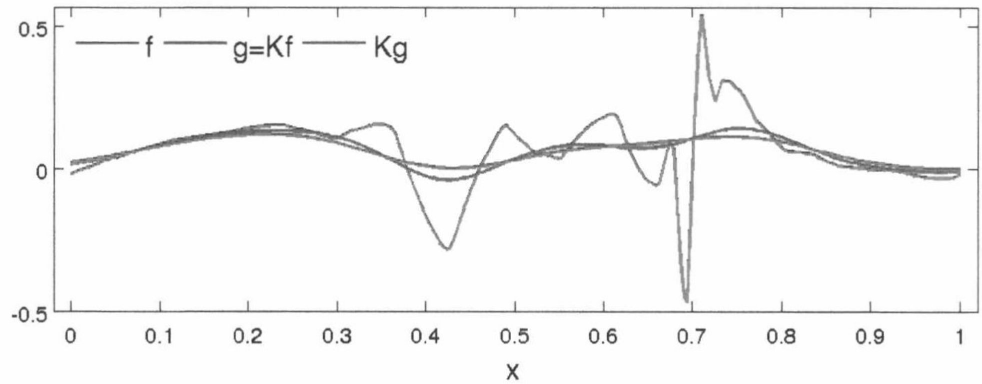


Figure 1.1. *The plot shows the effect that smoothing has on two functions, f and g. Here, g is already the smoothed version of f. See Example* 1.3. *Reprinted with permission from IOP Science* [247].

having only a finite amount of data. That is, we only have values of $\mu(x_i)$ at finitely many points $x_1, \dots, x_n$. Furthermore, the observations of $\mu(x_i)$ are usually contaminated by noise or other potential systematic errors. The noisy discrete data vector, $\boldsymbol{y}$, can be modeled as

$$\boldsymbol{y} = \mathcal{K} f + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y} = (y_1, \dots, y_n)^\top$, the vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top$ represents the noise, and $\mathcal{K}$ is the linear operator that maps f to $(\mathcal{K}_{x_1} f, \dots, \mathcal{K}_{x_n} f)^\top$. In particular, if f is modeled as an element of an infinite-dimensional linear space and $\mathcal{K}$ maps into $\mathbb{R}^n$, then $\mathcal{K}$ has a nontrivial nullspace, Null($\mathcal{K}$). So, if f is the true function and f_0 is a nonzero element in Null($\mathcal{K}$) (as in Example 1.2), then $f_\alpha = f + \alpha f_0$ generates exactly the same data for any scalar α. Such scalar can be chosen so as to make the difference $f - f_\alpha$ arbitrarily large. Therefore, even in the absence of noise the data for the deblurring problem do not provide enough information to choose any particular f_α. In section 1.2, we will show how to use regularization methods to reformulate the question and find a solution to the modified problem. In Chapters 3 and 5 we will discuss methods to check if the solution is meaningful by conducting validation and uncertainty analyses. ■

Operator equations of the form (1.1.3) arise in the study of boundary value problems (see, e.g., [154]). In fact, much of the work on integral equations has been motivated by the study of such problems. In the following example we use the Laplace transform to reduce an initial boundary value problem to an integral equation of the first kind.

Recall that the Laplace transform of a function, f, is defined as

$$\mathcal{L}[f](s) = \int_0^\infty f(t)e^{-st}\,dt.$$

We will use the following two well-known properties of the Laplace transform:

(i) If f is differentiable and $f(0) = 0$, then $\mathcal{L}[f'](s) = s\,\mathcal{L}[f](s)$.