CHAPTER ONE

Introduction

Synopsis

General issues associated with parameter estimation and inverse problems are introduced through the concepts of the forward problem and its inverse solution. Scaling and superposition properties that characterize linear systems are given, and common situations leading to linear and nonlinear mathematical models are discussed. Examples of discrete and continuous linear and nonlinear parameter estimation problems to be revisited in later chapters are shown. Mathematical demonstrations highlighting the key issues of solution existence, uniqueness, and instability are presented and discussed.

1.1. CLASSIFICATION OF PARAMETER ESTIMATION AND INVERSE PROBLEMS

Scientists and engineers frequently wish to relate physical parameters characterizing a **model**, m, to collected observations making up some set of **data**, d. We will commonly assume that the fundamental physics are adequately understood, so a function, G, may be specified relating m and d such that

$$G(m) = d. \tag{1.1}$$

In practice, d may be a function of time and/or space, or may be a collection of discrete observations. An important issue is that actual observations always contain some amount of noise. Two common ways that noise may arise are unmodeled influences on instrument readings and numerical round-off. We can thus envision data as generally consisting of noiseless observations from a "perfect" experiment, d_{true}, plus a noise component η,

$$d = G(m_{\text{true}}) + \eta \tag{1.2}$$

$$= d_{\text{true}} + \eta, \tag{1.3}$$

where d_{true} exactly satisfies (1.1) for m equal to the true model, m_{true}, and we assume that the forward modeling is exact. We will see that it is commonly mathematically possible, although practically undesirable, to also fit all or part of η by (1.1). It may seem remarkable, but it is often the case that a solution for m that is influenced by even a small

Parameter Estimation and Inverse Problems, Second Edition. DOI: 10.1016/B978-0-12-385048-5.00001-X

noise amplitude η can have little or no correspondence to m_{true}. Another key issue that may seem astounding at first is that commonly there are an infinite number of models aside from m_{true} which fit the perfect data, d_{true}.

When m and d are functions, we typically refer to G as an **operator**. G will be called a function when m and d are vectors. The operator G can take on many forms. In some cases G is an ordinary differential equation (ODE) or partial differential equation (PDE). In other cases, G is a linear or nonlinear system of algebraic equations.

Note that there is some inconsistency between mathematicians and other scientists in modeling terminology. Applied mathematicians usually refer to $G(m) = d$ as the "mathematical model" and to m as the "parameters." On the other hand, scientists often refer to G as the "forward operator" and to m as the "model." We will adopt the scientific parlance and refer to m as the "the model" while referring to the equation $G(m) = d$ as the "mathematical model."

The **forward problem** is to find d given m. Computing $G(m)$ might involve solving an ODE or PDE, evaluating an integral, or applying an algorithm for which there is no explicit analytical formula for $G(m)$. Our focus in this text is on the **inverse problem** of finding m given d. A third problem, not addressed here, is the **model identification problem** of determining G given examples of m and d.

In many cases, we will want to determine a finite number of parameters, n, to define a model. The parameters may define a physical entity directly (e.g., density, voltage, seismic velocity), or may be coefficients or other constants in a functional relationship that describes a physical process. In this case, we can express the model parameters as an n element vector $\mathbf{m}$. Similarly, if there are a finite number of data points then we can express the data as an m element vector $\mathbf{d}$. (Note that the use of the integer m here for the number of data points is easily distinguishable from the model m by its context.) Such problems are called **discrete inverse problems** or **parameter estimation problems**. A general parameter estimation problem can be written as a system of equations

$$G(\mathbf{m}) = \mathbf{d}. \tag{1.4}$$

In other cases, where the model and data are functions of continuous variables, such as time or space, the associated task of estimating m from d is called a **continuous inverse problem**. A central theme of this book is that continuous inverse problems can often be well-approximated by discrete inverse problems.

We will generally refer to problems with small numbers of parameters as "parameter estimation problems." Problems with a larger number of parameters, and which will often require the application of stabilizing constraints, will be referred to as "inverse problems." A key aspect of many inverse problems is that they are ill-conditioned in a sense that will be discussed later in this chapter. In both parameter estimation and inverse problems we solve for a set of parameters that characterize a model, and a key point of this text is that the treatment of all such problems can be sufficiently generalized so

that the distinction is largely irrelevant. In practice, what is important is the distinction between ill-conditioned and well-conditioned parameter estimation problems.

A type of mathematical model for which many useful results exist is the class of **linear systems.** Linear systems obey superposition

$$G(m_1 + m_2) = G(m_1) + G(m_2) \tag{1.5}$$

and scaling

$$G(\alpha m) = \alpha G(m). \tag{1.6}$$

In the case of a discrete linear inverse problem, (1.4) can always be written in the form of a linear system of algebraic equations (see Exercise 1.1).

$$G(\mathbf{m}) = \mathbf{Gm} = \mathbf{d}. \tag{1.7}$$

In a continuous linear inverse problem, G can often be expressed as a linear integral operator, where (1.1) has the form

$$\int_a^b g(x, \xi)m(\xi)d\xi = d(x) \tag{1.8}$$

and the function $g(x, \xi)$ is called the **kernel**. The linearity of (1.8) is easily seen because

$$\int_a^b g(x, \xi)(m_1(\xi) + m_2(\xi))d\xi = \int_a^b g(x, \xi)m_1(\xi)d\xi + \int_a^b g(x, \xi)m_2(x)d\xi \tag{1.9}$$

and

$$\int_a^b g(x, \xi)\alpha m(\xi)d\xi = \alpha \int_a^b g(x, \xi)m(\xi)d\xi. \tag{1.10}$$

Equations in the form of (1.8), where $m(x)$ is the unknown, are called **Fredholm integral equations of the first kind (IFK)**. IFKs arise in a surprisingly large number of inverse problems. A key characteristic of these equations is that they have mathematical properties which make it difficult to obtain useful solutions by straightforward methods.

In many cases the kernel in (1.8) can be written to depend explicitly on $x - \xi$, producing a **convolution equation**,

$$\int_{-\infty}^{\infty} g(x - \xi)m(\xi)d\xi = d(x). \tag{1.11}$$

Here we have written the interval of integration as extending from minus infinity to plus infinity, but other intervals can easily be accommodated by having $g(x-\xi)=0$ outside of the interval of interest. When a forward problem has the form of (1.11), determining $d(x)$ from $m(x)$ is called **convolution**, and the inverse problem of determining $m(x)$ from $d(x)$ is called **deconvolution**.

Another IFK arises in the problem of inverting a **Fourier transform**

$$\Phi(f)=\int_{-\infty}^{\infty} e^{-\imath 2\pi fx}\phi(x)dx \tag{1.12}$$

(where $\imath$ is $\sqrt{-1}$) to obtain $\phi(x)$. Although there are many tables and analytic methods of obtaining Fourier transforms and their inverses, numerical estimates of $\phi(x)$ may be of interest, such as where there is no analytic inverse, or where we wish to estimate $\phi(x)$ from spectral data collected at discrete frequencies.

It is an intriguing question as to why linearity appears in many interesting physical problems. A principal reason is that many such systems are associated with only small departures from equilibrium. One important geophysical example is seismic wave propagation, where the stresses associated with elastic fields are often very small relative to the elastic moduli of the medium. This situation leads to small strains and to a very nearly linear stress–strain relationship. Because of this, seismic wave field problems in many useful circumstances obey superposition and scaling. Other fields such as gravity and magnetism, at field strengths typically encountered in geophysics, also show effectively linear physics.

Because many important inverse problems are linear, and because linear theory is a key component in solving nonlinear problems, Chapters 2 through 8 of this book cover theory and methods for the solution of linear inverse problems. Nonlinear mathematical models arise when the parameters of interest have an inherently nonlinear relationship to the observables. This situation commonly occurs, for example, in electromagnetic field problems where we wish to relate geometric model parameters such as layer thicknesses to observed field properties. We discuss methods for nonlinear parameter estimation and inverse problems in Chapters 9 and 10, respectively.

1.2. EXAMPLES OF PARAMETER ESTIMATION PROBLEMS

Example 1.1

A canonical parameter estimation problem is the fitting of a function, defined by a collection of parameters, to a data set. In cases where this function fitting procedure can be cast as a linear inverse problem, the procedure is referred to as **linear regression**. An ancient example of linear regression is the characterization of a ballistic trajectory. In a basic take on this problem, the data, $\mathbf{y}$, are altitude observations of a ballistic body at a set of times $\mathbf{t}$

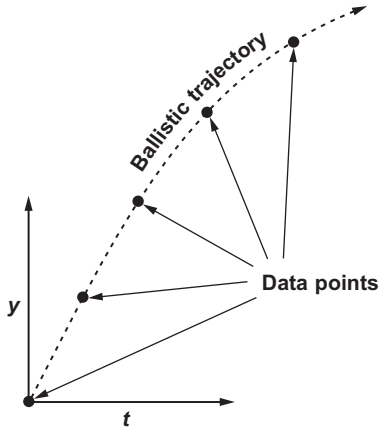

Figure 1.1 The parabolic trajectory problem.

(Figure 1.1). We wish to solve for a model, $\mathbf{m}$, that contains the initial altitude (m_1), initial vertical velocity (m_2), and effective gravitational acceleration (m_3) experienced by the body during its trajectory. This and related problems are naturally of practical interest in aeronautics and weaponry, but are also of fundamental geophysical interest, for example, in absolute gravity meters capable of estimating g from the acceleration of a falling object in a vacuum to accuracies on the order of one part in 10^9 [91].

The mathematical model is a quadratic function in the $(t,\ y)$ plane

$$y(t) = m_1 + m_2 t - (1/2) m_3 t^2 \tag{1.13}$$

that we expect to apply at all times along the trajectory, not just at the times t_i when we happen to have observations. The data will consist of m observations of the height of the body y_i at corresponding times t_i. Assuming that the t_i are measured precisely, and applying (1.13) to each observation, we obtain a system of equations with m rows and $n = 3$ columns that relates the data y_i to the model parameters, m_j

$$\begin{bmatrix} 1 & t_1 & -\frac{1}{2}t_1^2 \\ 1 & t_2 & -\frac{1}{2}t_2^2 \\ 1 & t_3 & -\frac{1}{2}t_3^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & -\frac{1}{2}t_m^2 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}. \tag{1.14}$$

Although the mathematical model of (1.13) is quadratic, the equations for the three parameters m_i in (1.14) are linear, so solving for $\mathbf{m} = [m_1,\ m_2,\ m_3]^T$ is a linear parameter estimation problem.

If there are more data points than model parameters in (1.14) ($m > n$), then the m constraint equations in (1.14) will likely be inconsistent, and it will be impossible to find

a model $\mathbf{m}$ that satisfies every equation exactly. The nonexistence of a model exactly satisfying the observations in such a case can be seen geometrically because no parabola will exist that goes through all of the (t_i, y_i) points (Exercise 1.2). Such a situation could arise in practice because of noise in the determinations of the y_i, and/or because the forward model of (1.13) is approximate. For example, we have neglected the physics of atmospheric drag, so a true trajectory will not be exactly parabolic and thus exactly modeled by (1.13).

In elementary linear algebra, where an exact solution is typically expected for a system of linear equations, we might throw up our hands at this point and simply state that no solution exists. However, useful solutions to such systems may be found in practice by solving for model parameters that satisfy the data in an approximate, or "best-fit," sense.

A reasonable approach to finding the "best" approximate solution to an inconsistent system of linear equations is to find an $\mathbf{m}$ that minimizes some misfit measure, calculated from the differences between the observations and the theoretical predictions of the forward problem, commonly called **residuals**. A traditional and very widely applied strategy is to find the model that minimizes the **2-norm** (Euclidean length) of the residual vector

$$\|\mathbf{y} - \mathbf{Gm}\|_2 = \sqrt{\sum_{i=1}^{m} (y_i - (\mathbf{Gm})_i)^2}. \tag{1.15}$$

However, (1.15) is not the only, or necessarily the best, misfit measure that can be applied to approximate solve systems of equations. An alternative misfit measure that is superior in many situations is the **1-norm**

$$\|\mathbf{y} - \mathbf{Gm}\|_1 = \sum_{i=1}^{m} |y_i - (\mathbf{Gm})_i|. \tag{1.16}$$

We shall see in Chapter 2 that a solution that minimizes (1.16) is less sensitive to data points that are wildly discordant with the mathematical model than one that minimizes (1.15). Solution techniques that are resistant to such data **outliers** are called **robust estimation procedures**.

Example 1.2

A classic nonlinear parameter estimation problem in geophysics is determining the space and time coordinates of an earthquake nucleation, the hypocenter, which is specified by

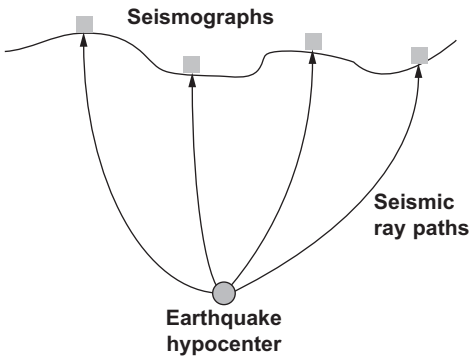

Figure 1.2 The earthquake location problem.

the four-vector

$$\mathbf{m} = \begin{bmatrix} \boldsymbol{\xi} \\ \tau \end{bmatrix} \tag{1.17}$$

where $\boldsymbol{\xi}$ is the three-dimensional earthquake location and the fourth element, τ, is the earthquake origin time (Figure 1.2). The hypocenter model we seek best fits a vector of seismic phase **arrival times**, $\mathbf{t}$, observed at an m-station seismic network. The mathematical model is

$$G(\mathbf{m}) = \mathbf{t} \tag{1.18}$$

where G models the physics of seismic wave propagation to map a hypocenter into a vector of predicted seismic arrival times at m stations. G depends on the seismic velocity structure, $v(\mathbf{x})$, which we assume here to be known.

The earthquake location problem is nonlinear even if $v(\mathbf{x})$ is a constant, c. In this case, all of the ray paths in Figure 1.2 would be straight, and the arrival time of the seismic signal at station i would be

$$t_i = \frac{\|\mathbf{S}_{\cdot,i} - \boldsymbol{\xi}\|_2}{c} + \tau \tag{1.19}$$

where the *ith* column of the matrix $\mathbf{S}$, $\mathbf{S}_{\cdot,i}$, specifies the coordinates for station i. Equation (1.19) is nonlinear with respect to the spatial parameters ξ_i in $\mathbf{m}$, and thus the problem cannot be expressed as a linear system of equations.

In a few special cases, a change of variables can convert a nonlinear problem to a linear one. More generally, nonlinear parameter estimation problems can often be solved by choosing a starting model and then iteratively improving it until a good solution is obtained. General methods for solving nonlinear parameter estimation problems are discussed in Chapter 9.

1.3. EXAMPLES OF INVERSE PROBLEMS

Example 1.3

In **vertical seismic profiling**, we wish to know the vertical seismic velocity of the material surrounding a borehole. A downward-propagating seismic wavefront is generated at the surface by a source, and seismic waves are sensed by a string of seismometers in the borehole (Figure 1.3).

The arrival times of the seismic wavefront at each instrument, which reflects the seismic velocity for vertically traveling waves as a function of depth, are measured from the recorded seismograms. The problem is nonlinear if it is expressed in terms of seismic velocities. However, we can linearize this problem via a simple change of variables where we parameterize the seismic structure in terms of **slowness**, $s(z)$, the reciprocal of the velocity $v(z)$. The observed travel time at depth z can then be expressed as the definite integral of the vertical slowness, s, from the surface to z:

$$t(z) = \int_0^z s(\xi)\, d\xi \tag{1.20}$$

$$= \int_0^\infty s(\xi) H(z-\xi)\, d\xi \tag{1.21}$$

where the kernel function H is the **Heaviside step function**, which is equal to one when its argument is nonnegative and zero when its argument is negative. The explicit dependence of the kernel on $z-\xi$ shows that (1.21) is a convolution.

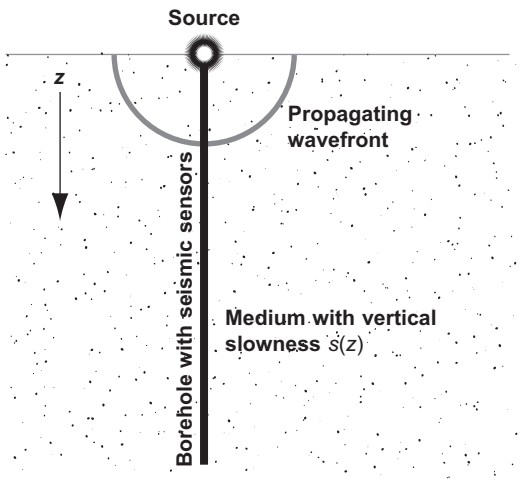

Figure 1.3 The vertical seismic profiling problem.

In theory, we can solve (1.21) quite easily because, by the fundamental theorem of calculus,

$$s(z) = \frac{dt(z)}{dz}. \tag{1.22}$$

In practice, simply differentiating the observations to obtain a solution may not be useful. This is because there will generally be noise present in the observed times $t(z)$, and applying (1.22) may result in a solution that includes unphysical values of $s(z)$ or other unrealistic model features.

Example 1.4

A further instructive linear inverse problem is the inversion of a vertical gravity anomaly, $d(x)$, observed at some height, h, to estimate an unknown buried line mass density distribution deviation from a background model, $m(x) = \Delta\rho(x)$ (Figure 1.4). The mathematical model for this problem can be written as an IFK, because the data are a superposition of the vertical gravity contributions from the differential elements comprising the line mass

$$d(x) = \Gamma \int_{-\infty}^{\infty} \frac{h}{\left((\xi - x)^2 + h^2\right)^{3/2}} m(\xi)\, d\xi \tag{1.23}$$

$$= \int_{-\infty}^{\infty} g(\xi - x) m(\xi)\, d\xi \tag{1.24}$$

where Γ is Newton's gravitational constant. Note that the kernel has the form $g(\xi - x)$, and (1.24) is thus a convolution. Because the kernel is a smooth function, $d(x)$ will be a smoothed and scaled transformation of $m(x)$. Conversely, solutions for $m(x)$ will be a

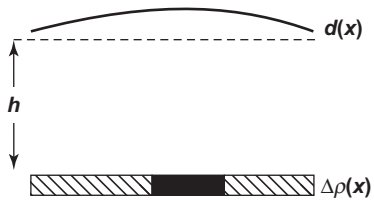

Figure 1.4 A linear inverse problem: determine the density deviation of a buried line mass, $\Delta\rho(x)$, relative to a background model, from gravity anomaly observations $d(x)$.

roughened transformation of $d(x)$. For this reason we again need to be wary of possibly severe deleterious effects of noise in the data.

Example 1.5

Consider a variation on Example 1.4, where the depth of the line density perturbation varies, rather than the density contrast. The gravity anomaly is now attributable to a variation in the burial depth, $m(x) = h(x)$, of an assumed known density perturbation, $\Delta\rho$ (Figure 1.5). The physics is the same as in Example 1.4, so the data are still given by the superposition of density perturbation contributions to the gravitational anomaly field, but the mathematical model now takes the form

$$d(x) = \Gamma \int_{-\infty}^{\infty} \frac{m(\xi)}{\left((\xi - x)^2 + m^2(\xi)\right)^{3/2}} \Delta\rho \, d\xi \,. \tag{1.25}$$

This problem is nonlinear in $m(x)$ because (1.25) does not follow the superposition and scaling rules (1.5) and (1.6).

Nonlinear inverse problems are generally significantly more difficult to solve than linear ones. In special cases, they may be solvable by coordinate transformations that globally linearize the problem or by other clever special–case methods. In other cases, the problem cannot be globally linearized, so nonlinear optimization techniques must be applied. The essential differences in the treatment of linear and nonlinear problems arise because, as we shall see, all linear problems can be generalized to be the "same" in an important sense, so that a single set of solution methods can be applied to all. In contrast, nonlinear problems tend to be nonlinear in mathematically different ways and

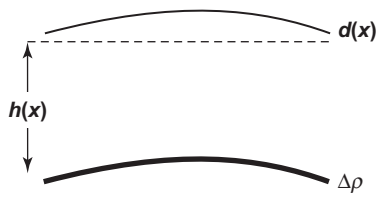

Figure 1.5 A nonlinear inverse problem: determine the depth to a buried line mass density anomaly $h(x)$ from observed gravity anomaly observations $d(x)$.

either require specific strategies or, more commonly, can by solved by iterative methods that may rely on local linearization.

Example 1.6

A classic pedagogical inverse problem is an experiment in which an angular distribution of illumination passes through a thin slit and produces a diffraction pattern, for which the intensity is observed (Figure 1.6; [141]).

The data, $d(s)$, are measurements of diffracted light intensity as a function of the outgoing angle $-\pi/2 \le s \le \pi/2$. Our goal is to find the intensity of the incident light on the slit, $m(\theta)$, as a function of the incoming angle $-\pi/2 \le \theta \le \pi/2$.

The forward problem relating d and m can be expressed as the linear mathematical model,

$$d(s) = \int_{-\pi/2}^{\pi/2} (\cos(s) + \cos(\theta))^2 \left(\frac{\sin(\pi(\sin(s) + \sin(\theta)))}{\pi(\sin(s) + \sin(\theta))} \right)^2 m(\theta) d\theta. \tag{1.26}$$

Example 1.7

Consider the problem of recovering the history of groundwater pollution at a source site from later measurements of the contamination at downstream wells to which the contaminant plume has been transported by advection and diffusion (Figure 1.7). This

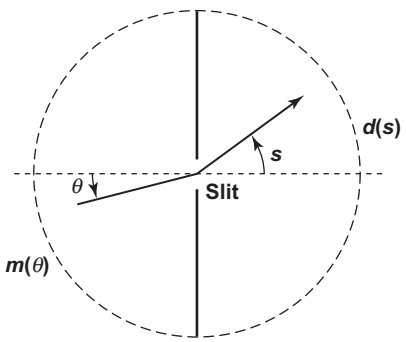

Figure 1.6 The Shaw diffraction intensity problem (1.26).

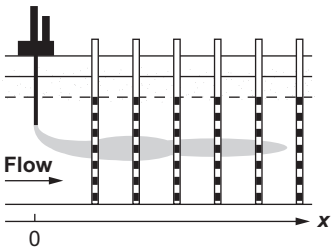

Figure 1.7 The contaminant plume source history reconstruction problem.

"source history reconstruction problem" has been considered by a number of authors [117, 144, 145].

The mathematical model for contaminant transport is an advection–diffusion equation

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - v\frac{\partial C}{\partial x} \tag{1.27}$$

$$C(0,\ t) = C_{in}(t)$$

$$C(x,\ t) \to 0 \ \text{ as } \ x \to \infty$$

where D is the diffusion coefficient and v is the velocity of groundwater flow. The solution to (1.27) at time T is the convolution

$$C(x,\ T) = \int_0^T C_{in}(t) f(x,\ T-t)\, dt, \tag{1.28}$$

where $C_{in}(t)$ is the time history of contaminant injection at $x = 0$, and the kernel is

$$f(x,\ T-t) = \frac{x}{2\sqrt{\pi D(T-t)^3}} \mathrm{e}^{-\frac{[x-v(T-t)]^2}{4D(T-t)}}. \tag{1.29}$$

Example 1.8

An important and instructive inverse problem is **tomography**, from the Greek roots *tomos*, "to section" or "to cut" (the ancient concept of an *atom* was that of an irreducible, uncuttable object) and *graphein*, "to write." Tomography is the general technique of determining models that are consistent with path-integrated properties such as attenuation (e.g., X-ray, radar, muon, seismic), travel time (e.g., electromagnetic, seismic, or

acoustic), or source intensity (e.g., positron emission). Here, we will use examples from seismic tomography. Although tomography problems originally involved determining models that were two-dimensional slices of three-dimensional objects, the term is now commonly used in situations where the model is two- or three-dimensional. Tomography has many applications in medicine, engineering, acoustics, and Earth science. One important geophysical example is cross-well tomography, where seismic sources are installed in a borehole, and the signals are received by sensors in another borehole. Another example is joint earthquake location/velocity structure inversion carried out on scales ranging from a fraction of a cubic kilometer to global [158, 159, 160].

The physical model for tomography in its most basic form (Figure 1.8) assumes that geometric ray theory (essentially the high-frequency limiting case of the wave equation) is valid, so that wave energy traveling between a source and receiver can be considered to be propagating along infinitesimally narrow ray paths. The density of ray path coverage in a tomographic problem may vary significantly throughout a section or volume, and provide much better constraints on physical properties in densely sampled regions than in sparsely sampled ones.

In seismic tomography, if the slowness at a point $\mathbf{x}$ is $s(\mathbf{x})$, and the ray path ℓ is known, then the travel time for seismic energy transiting along that ray path is given by the line integral along ℓ

$$t = \int_{\ell} s(\mathbf{x}(l))\, dl. \tag{1.30}$$

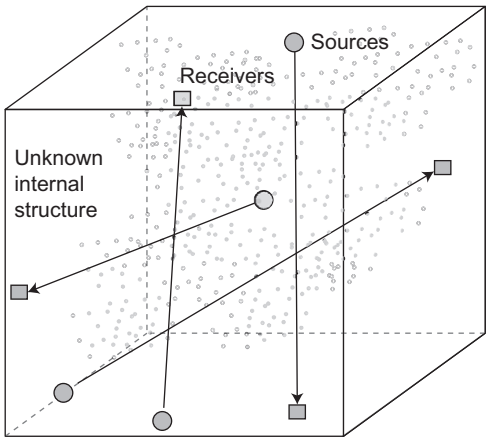

Figure 1.8 Conceptual depiction of ray path tomography. Sources and receivers may, in general, be either at the edges or within the volume, and ray paths may be either straight, as depicted, or curved.

Note that (1.30) is just a higher-dimensional generalization of (1.21), the forward problem for the vertical seismic profiling example. In general, seismic ray paths will be bent due to refraction and/or reflection. In cases where such effects are negligible, ray paths can be usefully approximated as straight lines (e.g., as depicted in Figure 1.8), and the forward and inverse problems can be cast in a linear form. However, if the ray paths are significantly bent by slowness variations, the resulting inverse problem will be nonlinear.

1.4. DISCRETIZING INTEGRAL EQUATIONS

Consider problems of the form

$$d(x) = \int_a^b g(x, \xi) m(\xi) d\xi. \tag{1.31}$$

Here $d(x)$ is a known function, typically representing observed data. The kernel $g(x, \xi)$ is considered to be given, and encodes the physics relating an unknown model $m(x)$ to observed data $d(s)$. The interval $[a, b]$ may be finite or infinite. The function $d(x)$ might in theory be known over an entire interval, but in practice we will only have measurements of $d(x)$ at a finite set of points.

We wish to solve for $m(x)$. This type of linear equation is called a Fredholm integral equation of the first kind, or IFK. A surprisingly large number of inverse problems, including all of the examples from the previous section, can be written as IFKs. Unfortunately, IFKs have properties that can make them very challenging to solve.

To obtain useful numerical solutions to IFKs, we will frequently discretize them into forms that are tractably solvable using the methods of linear algebra. We first assume that $d(x)$ is known at a finite number of points $x_1, x_2, \ldots, x_m$. We can then write the forward problem as

$$d_i = d(x_i) = \int_a^b g(x_i, \xi) m(\xi) d\xi \quad i = 1, 2, \ldots, m \tag{1.32}$$

or as

$$d_i = \int_a^b g_i(x) m(x) dx \quad i = 1, 2, \ldots, m, \tag{1.33}$$

where $g_i(x) = g(x_i, \xi)$. The functions $g_i(t)$ are referred to as **representers** or **data kernels**.

Here, we will apply a **quadrature rule** to approximate (1.33) numerically. Note that, although quadrature methods are applied in this section to linear integral equations, they also have utility in the discretization of nonlinear problems. The simplest quadrature approach is the **midpoint rule**, where we divide the interval $[a, b]$ into n subintervals, and pick points $x_1, x_2, \ldots, x_n$ in the middle of each interval. The points are given by

$$x_j = a + \frac{\Delta x}{2} + (j-1)\Delta x \tag{1.34}$$

where

$$\Delta x = \frac{b-a}{n}. \tag{1.35}$$

The integral (1.33) is then approximated by (Figure 1.9):

$$d_i = \int_a^b g_i(x)m(x)dx \approx \sum_{j=1}^{n} g_i(x_j)m(x_j)\Delta x, \qquad i = 1, 2, \ldots, m. \tag{1.36}$$

If we let

$$G_{i,j} = g_i(x_j)\Delta x \qquad \begin{pmatrix} i = 1, 2, \ldots, m \\ j = 1, 2, \ldots, n \end{pmatrix} \tag{1.37}$$

and

$$m_j = m(x_j) \quad j = 1, 2, \ldots, n, \tag{1.38}$$

then we obtain a linear system of equations $\mathbf{Gm} = \mathbf{d}$.

The approach of using the midpoint rule to approximate the integral is known as **simple collocation**. Of course, there are also more sophisticated quadrature rules for numerically approximating integrals (e.g., the trapezoidal rule, or Simpson's rule). In each case, we end up with an m by n linear system of equations, but the formulas for evaluating the elements of $\mathbf{G}$ will be different.

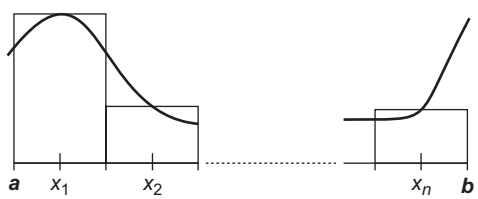

Figure 1.9 Grid for the midpoint rule.

Example 1.9

Consider vertical seismic profiling (Example 1.3), where we wish to estimate vertical seismic slowness using travel time measurements of downward propagating seismic waves (Figure 1.10). We discretize the forward problem (1.21) for m travel time data points, t_i, and n model depths, z_j, spaced at constant intervals of Δz. This discretization is depicted in Figure 1.10.

The discretized problem has

$$t_i = \sum_{j=1}^{n} H(y_i - z_j) s_j \Delta z, \tag{1.39}$$

where $n/m = \Delta y/\Delta z$ is an integer. The rows, $\mathbf{G}_{i,\cdot}$, of the matrix each consist of $i \cdot n/m$ elements equal to Δz on the left and $n - (i \cdot n/m)$ zeros on the right. For $n = m$, $\mathbf{G}$ is a lower triangular matrix with each nonzero entry equal to Δz.

Example 1.10

To discretize the Shaw problem (1.26), we apply the method of simple collocation with m and n intervals for the data and model functions, respectively. We additionally define

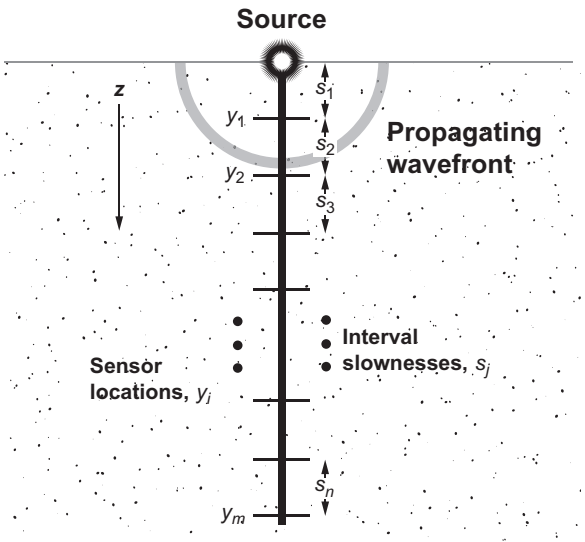

Figure 1.10 Discretization of the vertical seismic profiling problem ($n = 2m$) into uniform intervals.

the data and model points at m and n equally spaced respective angles given by

$$s_i = \frac{(i-0.5)\pi}{m} - \frac{\pi}{2} \quad i = 1,\ 2,\ \ldots,\ m \tag{1.40}$$

and

$$\theta_j = \frac{(j-0.5)\pi}{n} - \frac{\pi}{2} \quad j = 1,\ 2,\ \ldots,\ n. \tag{1.41}$$

Correspondingly, discretizing the data and model into m- and n-length vectors

$$d_i = d(s_i) \quad i = 1,\ 2,\ \ldots,\ m \tag{1.42}$$

and

$$m_j = m(\theta_j) \quad j = 1,\ 2,\ \ldots,\ n \tag{1.43}$$

leads to a discrete linear system $\mathbf{Gm} = \mathbf{d}$, where

$$G_{i,j} = (\cos(s_i) + \cos(\theta_j))^2 \left(\frac{\sin(\pi(\sin(s_i) + \sin(\theta_j)))}{\pi(\sin(s_i) + \sin(\theta_j))} \right)^2 \Delta\theta \tag{1.44}$$

and

$$\Delta\theta = \frac{\pi}{n}. \tag{1.45}$$

Example 1.11

We discretize the advection–diffusion problem (1.27), assuming that the parameters D and v in (1.29) are known. We wish to estimate $C_{in}(t)$ from simultaneous concentration observations at the locations x_i at some later time T. The convolution (1.28) for $C(x,\ T)$ is discretized as

$$\mathbf{d} = \mathbf{Gm} \tag{1.46}$$

where $\mathbf{d}$ is a vector of sampled concentrations at different well locations, $\mathbf{x}$, at time T, $\mathbf{m}$ is a vector of C_{in} values to be estimated, and

$$G_{i,j} = f(x_i,\ T - t_j)\Delta t \tag{1.47}$$

$$= \frac{x_i}{2\sqrt{\pi D(T-t_j)^3}} e^{-\frac{[x_i - v(T-t_j)]^2}{4D(T-t_j)}} \Delta t. \tag{1.48}$$

Example 1.12

A common way of discretizing the model in a tomographic problem is as uniform blocks (Figure 1.11). This approach is essentially applying the midpoint rule to the travel-time forward problem (1.30).

In this parameterization, the elements of **G** are just the lengths of the ray paths within corresponding blocks. Consider the example of Figure 1.11, where nine homogeneous blocks with sides of unit length and unknown slowness are crossed by eight ray paths. For straight ray paths, we map the two-dimensional slowness grid to a model vector using a row-by-row indexing convention. The constraint equations in the mathematical model are then

$$\mathbf{Gm} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ \sqrt{2} & 0 & 0 & 0 & \sqrt{2} & 0 & 0 & 0 & \sqrt{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} s_{1,1} \\ s_{1,2} \\ s_{1,3} \\ s_{2,1} \\ s_{2,2} \\ s_{2,3} \\ s_{3,1} \\ s_{3,2} \\ s_{3,3} \end{bmatrix} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \\ t_8 \end{bmatrix}. \tag{1.49}$$

Because there are nine unknown parameters $s_{i,j}$ in the model, but only eight constraints, the **G** matrix is clearly rank deficient. In fact, rank(**G**) is only seven. In addition,

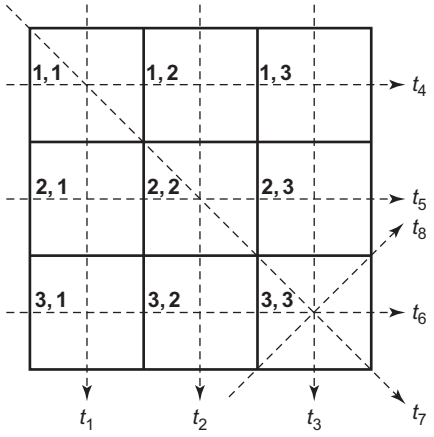

Figure 1.11 Discretization of a tomography problem into uniform blocks. Ray paths correspond to the constraint equations in (1.49).

there is clearly redundant information in (1.49), in that the slowness $s_{3,3}$ is completely determined by t_8, yet $s_{3,3}$ also influences the observations t_3, t_6, and t_7.

1.5. WHY INVERSE PROBLEMS ARE DIFFICULT

Scientists and engineers need to be concerned with far more than simply finding mathematically acceptable answers to parameter estimation and inverse problems. One reason is that there may be many models that adequately fit the data. It is essential to characterize just what solution has been obtained, how "good" it is in terms of physical plausibility, its ability to predict the data, and perhaps how consistent it is with other constraints. Essential issues that must be considered include **solution existence**, **solution uniqueness**, and **instability** of the solution process.

1. *Existence*. There may be no model that exactly fits the data. This can occur in practice because the mathematical model of the system's physics is approximate (or perhaps simply incorrect) or because the data contain noise.
2. *Uniqueness*. If exact solutions do exist, they may not be unique, even for an infinite number of exact data points. That is, there may be other solutions besides m_{true} that exactly satisfy $G(m) = d_{\text{true}}$. This situation commonly occurs in potential field problems. A classic example is the external gravitational field from a spherically symmetric mass distribution, which depends only on the total mass, and not on the radial density distribution.

   Nonuniqueness is a characteristic of rank deficient discrete linear inverse problems because the matrix $\mathbf{G}$ in this case has a nontrivial **null space**. In linear inverse problems, models ($\mathbf{m}_0$) that lie in the null space of $\mathbf{G}$ are solutions to $\mathbf{G}\mathbf{m}_0 = \mathbf{0}$. By superposition, any linear combination of these **null space models** can be added to a particular model that satisfies (1.7) and not change the fit to the data. There are thus an infinite number of mathematically acceptable models in such situations. In practical terms, suppose that there exists a nonzero model $\mathbf{m}_0$ that results in an instrument reading of zero. We cannot discriminate this situation from the situation where $\mathbf{m}_0$ is truly zero.

   An important and thorny issue with problems that have nonunique solutions is that an estimated model may be significantly smoothed or otherwise biased relative to the true situation. Characterizing such bias is essential to interpreting models in terms of their possible correspondence to reality. This issue falls under the general topic of **model resolution analysis**.
3. *Instability*. The process of computing an inverse solution can be, and often is, extremely unstable in that a small change in measurement (e.g., a small η in (1.3)) can lead to an enormous change in the estimated model. Inverse problems where this

situation arises are referred to as **ill-posed** in the case of continuous systems, or **ill-conditioned** in the case of discrete linear systems. A key point is that it is commonly possible to stabilize the inversion process by imposing additional constraints that bias the solution, a process that is generally referred to as **regularization**. Regularization is frequently essential to producing a usable solution to an otherwise intractable ill-posed or ill-conditioned inverse problem.

To examine existence, uniqueness, and instability issues, let us consider some simple mathematical examples using an IFK,

$$\int_0^1 g(x, \xi) m(\xi) d\xi = \gamma(x). \tag{1.50}$$

First, consider the trivial case where the kernel is a constant, such as

$$g(x, \xi) = 1, \tag{1.51}$$

which produces the integral equation

$$\int_0^1 m(\xi) d\xi = \gamma(x). \tag{1.52}$$

Because the left-hand side of (1.52) is independent of x, this system has no solution unless $\gamma(x)$ is a constant. Thus, there are an infinite number of mathematically conceivable data sets $\gamma(x)$ that are not constant and for which no exact solution exists. This is a simple illustration of a solution existence issue.

Where a solution to (1.52) does exist, the solution is nonunique because there are an infinite number of functions that, when integrated over the unit interval, produce the same constant and thus satisfy the IFK exactly. This demonstrates a uniqueness issue.

A more subtle example of nonuniqueness can be seen for

$$g(x, \xi) = x \cdot \sin(\pi \xi) \tag{1.53}$$

in (1.50), so that the IFK becomes

$$\int_0^1 x \cdot \sin(\pi \xi) m(\xi) d\xi = \gamma(x). \tag{1.54}$$

The functions $\sin(k\pi x)$ for integer values of k are orthogonal in the sense that

$$\int_0^1 \sin(k\pi x)\ \sin(l\pi x)dx = -\frac{1}{2}\int_0^1 \cos(\pi(k+l)x) - \cos(\pi(k-l)x)dx =$$

$$-\frac{1}{2\pi}\left(\frac{\sin(\pi(k+l))}{k+l} - \frac{\sin(\pi(k-l))}{k-l}\right) = 0 \quad (k \neq \pm l;\ k,\ l \neq 0). \tag{1.55}$$

Thus, in (1.54), for models of the form $m(x) = \sin(k\pi x)$, for $k = \pm 2, \pm 3, \ldots$, we have

$$\int_0^1 g(x,\ \xi)m(\xi)d\xi = \int_0^1 g(x,\ \xi)\ \sin(k\pi\xi)d\xi = 0. \tag{1.56}$$

Furthermore, because (1.54) is a linear system, we can add any function of the form

$$m_0(x) = \sum_{k=2}^{\infty} \alpha_k\ \sin(k\pi x) \tag{1.57}$$

to a solution, $m(x)$, and obtain a new model that fits the data equally well.

$$\int_0^1 x \cdot \sin(\pi\xi)(m(\xi) + m_0(\xi))d\xi = \int_0^1 x \cdot \sin(\pi\xi)m(\xi)d\xi + \int_0^1 s \cdot \sin(\pi\xi)m_0(\xi)d\xi$$

$$= \int_0^1 x \cdot \sin(\pi\xi)m(\xi)d\xi + 0. \tag{1.58}$$

There are thus an infinite number of very different solutions that fit the data equally well.

Finally, even if we do not encounter existence or uniqueness issues, instability is a fundamental feature of IFKs. It can be shown that, in the limit as k goes to infinity,

$$\lim_{k\to\infty} \int_{-\infty}^{\infty} g(x,\ \xi)\ \sin\ k\pi\xi\, d\xi = 0 \tag{1.59}$$

for *all* square-integrable functions $g(x,\ \xi)$. This result is known as the **Riemann-Lebesgue lemma** [134]. Examining (1.59) in more detail, we can better understand why this occurs. The oscillatory sine function is smoothed by integration with the kernel $g(\xi,\ x)$. For sufficiently large values of the sine frequency, k, the positive and negative excursions of the sine function will average out to zero. The inverse problem has this situation reversed, so that an inferred model can be extremely sensitive to small changes

in the data. If such small data changes are created by random noise that has nothing to do with the physical system that we are studying, then an inferred model from solving an inverse problem to fit this noise can easily have essentially no correspondence to the true model.

The unstable character of IFK solutions is similar to the situation encountered in solving linear systems of equations where the condition number of the matrix is very large, or equivalently, where the matrix is nearly singular. In both cases, the difficulty lies in the mathematical model itself, and not in the particular algorithm used to solve the problem. Fundamentally, most inverse problems are ill-posed.

1.6. EXERCISES

1. Consider a mathematical model of the form $G(\mathbf{m}) = \mathbf{d}$, where $\mathbf{m}$ is a vector of length n, and $\mathbf{d}$ is a vector of length m. Suppose that the model obeys the superposition and scaling laws and is thus linear. Show that $G(\mathbf{m})$ can be written in the form

$$G(\mathbf{m}) = \mathbf{\Gamma m} \tag{1.60}$$

where $\mathbf{\Gamma}$ is an m by n matrix. What are the elements of $\mathbf{\Gamma}$? Hint: Consider the standard basis, and write $\mathbf{m}$ as a linear combination of the vectors in the standard basis. Apply the superposition and scaling laws. Finally, recall the definition of matrix-vector multiplication.

2. Can (1.14) be inconsistent, even with only $m = 3$ data points? How about just $m = 2$ data points? If the system can be inconsistent, give an example. If not, explain why.

3. Consider the borehole vertical seismic profile problem of Examples 1.3 and 1.9 for $n = 100$ equally spaced seismic sensors located at depths of $z = 0.2,\ 0.4,\ \ldots,\ 20$ m, and for a model $\mathbf{m}$ describing n corresponding equal-length seismic slowness values for 0.2 m intervals having midpoints at $z - 0.1$ m.

a. Calculate the appropriate system matrix, $\mathbf{G}$, for discretizing the integral equation (1.21) using the midpoint rule.

b. For a seismic velocity model having a linear depth gradient specified by

$$v = v_0 + kz, \tag{1.61}$$

where the velocity at $z = 0$ is $v_0 = 1$ km/s and the gradient is $k = 40$ m/s per m, calculate the true slowness values, $\mathbf{m}_{\text{true}}$, at the midpoints of the n intervals. Integrate the corresponding slowness function for (1.61) using (1.21) to calculate a noiseless synthetic data vector, $\mathbf{d}$, of predicted seismic travel times at the sensor depths.

c. Solve for the slowness, $\mathbf{m}$, as a function of depth using your $\mathbf{G}$ matrix and analytically calculated noiseless travel times using the MATLAB backslash operator. Compare your result graphically with $\mathbf{m}_{\text{true}}$.

d. Generate a noisy travel time vector where independent normally distributed noise with a standard deviation of 0.05 ms is added to the elements of $\mathbf{d}$. Resolve the system for $\mathbf{m}$ and again compare your result graphically with $\mathbf{m}_{\text{true}}$. How has the model changed?

e. Repeat the problem, but for just $n = 4$ sensor depths and corresponding equal length slowness intervals. Is the recovery of the true model improved? Explain in terms of the condition numbers of your $\mathbf{G}$ matrices.

4. Find a journal article that discusses the solution of an inverse problem in a discipline of special interest to you. What are the data? Are the data discrete or continuous? Have the authors discussed possible sources of noise in the data? What is the model? Is the model continuous or discrete? What physical laws determine the forward operator G? Is G linear or nonlinear? Do the authors discuss any issues associated with existence, uniqueness, or instability of solutions?

1.7. NOTES AND FURTHER READING

Some important references on inverse problems in geophysics and remote sensing include [28, 60, 99, 126, 166]. Instructive examples of ill-posed problems and their solutions can be found in the book edited by Tikhonov and Goncharsky [162]. More mathematically oriented references on inverse problems include [8, 45, 59, 65, 67, 90, 94, 108, 112, 155, 161]. Tomography, particularly in medical imaging and seismology, is a very large field. Some general references on tomography are [70, 79, 84, 98, 100, 114, 121].