Finding Connections Between Data and Theory: Applications in Geophysical Sciences [1]

Douglas R. MacAyeal and Victor Barcilon

Department of Geophysical Sciences
University of Chicago
Chicago, Illinois

October 7, 1998

[1]This manuscript represents a rough draft of a book developed from the class notes of Geophysical Sciences 235 taught by Doug MacAyeal at the University of Chicago. The current draft is intended for inspection only. It is anticipated that approximately 60% of the subjects intended for this book have yet to be written.

Contents

I Methodology **2**

1 Radiometric Dating and Least-Squares Line Fitting **3**

1.1 Introduction 3

1.2 Debate Over the Age of the Earth 4

1.3 Radioactive Decay 6

1.4 Radiometric Dating 7

1.5 Isochron Method 8

1.6 Example: The Age of a Lunar Basalt 11

1.7 Linear Algebra of Line Fitting 11

1.7.1 LU Decomposition 13

1.8 Least-Squares Inverse 16

1.8.1 Using Calculus to Find the Minimum of J 16

1.8.2 A Problem Ahead 20

1.9 Summary 20

1.10 Bibliography 21

1.11 Appendix: Newton-Raphson Method for Finding Roots 21

1.11.1 Example 23

1.12 Lab 1 - The Radiometric Determination of the Age of the Earth 26

1.12.1 Look at the Data 26

1.12.2 The Forward Problem 27

1.12.3 The Least-Squares Inverse 28

1.12.4 Inverting an Intransitive Relationship 29

2 Underdetermined Inverse Problems: Minimum-Norm Line Fitting 30

2.1 Introduction 30

2.2 An Absurd Inverse Problem: Fitting an Isochron Through One Point 31

2.2.1 Lagrange Undetermined Multiplier 32

3 Dealing with Uncertainty 35

3.1 Introduction 35

3.2 Expectation Operators and Covarience Matrices 36

3.2.1 Expectation Operator 37

3.3 Two Questions 37

3.3.1 Least-Squares Inverse with Data Uncertainty 38

3.3.2 Model Covarience 38
3.3.3 Example: Lunar Basalt Isochron Uncertainty 40
3.4 Data Independence 42
3.4.1 Example: Data Independence Matrix for the Lunar Basalt Problem 44
3.5 Uncertainty in Underdetermined Inverse Problems 44
3.6 Model Resolution 45
3.7 Bibliography 47

4 Singular Value Decomposition: Geometric Interpretation 48

4.1 Introduction 48
4.2 Geometrical Interpretations of Linear Operators 49
4.2.1 Mappings of the Unit Sphere 49
4.3 Eigenvalues and Eigenvectors of a Symmetric Matrix 54
4.4 SVD of General, Nonsymmetric, Square Matrices 58
4.5 SVD of Rectangular Matrices 61
4.6 The Moore-Penrose Inverse 63
4.7 Solving Overdetermined Linear Problems with SVD 65
4.8 Data Independence and Model Resolution 68
4.9 Model Covarience 69
4.10 Testing Data Sufficiency 70

4.11 Summary 70

4.12 Bibliography 71

4.13 Laboratory Exercises 72

5 Idiosyncratic Line-Fitting Algorithms: Control Method and Simulated Annealing 73

5.1 Introduction 73

5.2 Radiometric Dating Redux 75

5.2.1 Control Algorithm 76

5.3 Example: Lunar Basalt Isochron 78

5.4 Error Analysis Associated with the Control Method 81

5.5 Radiometric Dating Redux^2: Simulated Annealing 85

5.6 Bibliography 92

II Applications 93

6 Sea-Floor Spreading Models and Ocean Bathymetry 94

6.1 Overview 94

6.2 Sea-Floor Spreading and Continental Drift 95

6.2.1 Kelvin's Solution for the Cooling of the Earth 97

6.2.2 Laplace Transform 98

6.2.3 Solution of the Subsidiary Equation 99

6.2.4 Inverse Laplace Transform 100

6.2.5 Integrating the Bromwich Integral 100

6.2.6 Geothermal Gradient 105

6.2.7 Kelvin's Edinburgh Calculation 106

6.3 Theory of Sea-Floor Subsidence 108

6.4 A Plate Model of Oceanic Crust 110

6.5 Bibliography 117

6.6 Lab: Fitting an Oceanic Crust Model to Oceanic Depth and Heat Flow Data 118

6.6.1 Cooling Half Space Model 118

6.6.2 Cooling Plate Model 119

7 Borehole Paleothermometry **122**

7.1 Overview 122

7.2 An Ideal Borehole Paleothermometry Problem 123

7.3 Solution of the Forward Problem 125

7.3.1 Green's Function Approach 126

7.3.2 Determination of $G_z(0,t;\xi)$ using Duhammel's Theorem 127

7.4 A Family of Inverse Probems 130

7.4.1 Inverse Problem 1. 130

7.4.2 Inverse Problem 2. 133

7.4.3 Inverse Problem 3. 133

7.5 Least Squares Solution: Continuous Case 137

7.6 Discrete Version of the Forward Problem 139

7.7 Discrete Version of the Least-Squares Solution 142

7.8 Least-Squares Solution: Discrete Case 144

7.9 Limits to Resolution of Past Thermal History 145

7.10 Uncertainty 147

7.11 Paleothermometry by Control Methods 148

7.11.1 Continuous Version of a Control Method 148

7.11.2 Discrete Version of a Control Method 151

7.12 Bibliography 155

7.13 Laboratory Exercise: Warming Anomaly in Permafrost of Northern Alaska 157

8 Kalman Smoother: Estimation of Atmospheric Trace-Gas Emissions 158

8.1 Overview 158

8.2 An Idealized Atmospheric-Chemistry Model 159

8.3 A Simple Inverse Problem 160

8.4 Green's Function Approach to the Forward Problem 161

8.4.1 Legendre Equation 162

8.4.2 Legendre Polynomials 163

8.4.3 Time Dependence 164
8.4.4 General Solution to Homogeneous Forward Problem . . 165
8.4.5 The Inhomogeneous Forward Problem 165
8.4.6 Spectral Form of the Solution 167
8.5 Inverse Problem Restated 168
8.6 Continuous Version of the Kalman Filter 168
8.7 Discrete Version of the Kalman Filter 172
8.7.1 Finite-Difference version of the Forward Problem . . . 172
8.7.2 Inverse Problem in Discrete Form 174
8.7.3 Matrix Form of the Kalman Filter 175
8.7.4 Estimate Covarience 176

9 Backus-Gilbert Method: Free Oscillations of Lake Michigan 179

9.1 Introduction 179
9.2 Free-Oscillations of a Long, Narrow Lake 180
9.3 Eigenfunctions and Eigenfrequencies of a Flat-Bottomed Lake 184
9.3.1 A Discretized Analysis of the Flat-Bottomed Lake . . 185
9.4 An Inverse Problem 190
9.5 Linearization of the Inverse Problem 190
9.6 A Fourier-Series Approach 193
9.7 A Minimum-Norm Solution 195

9.8 Example: Minimum-Norm Solution with Lake Michigan Data 196

9.9 Dirichlet Spread of the Model-Resolution Matrix 205

9.10 The Backus-Gilbert Spread of the Model-Resolution Matrix . 206

9.11 Derivation of the Backus-Gilbert Inverse 207

9.12 Example: Backus-Gilbert Solution with Lake Michigan Data . 210

9.13 An Alternative Definition of the Backus-Gilbert Spread 215

9.14 Conclusion 218

9.15 Bibliography 220

Part I

Methodology

Chapter 1

Radiometric Dating and Least-Squares Line Fitting

1.1 Introduction

How do you fit a line through a scatter of data points on a piece of graph paper? If you have just two data points and have a sharp pencil, then the job is easy: just connect the two points with a stiff ruler and draw the line. Suppose you have three or more data points which you believe *should* lie on a straight line, but actually don't. In this situation it is likely that you will attempt to *fit* a line through the data points by minimizing a *least-square* measure of the misfit between the data points and the line.

This kind of problem arises in all branches of science and engineering. We shall learn about linear least-squares inverse methods using a very simple example derived from geochemistry (radiometric dating). This example has historical significance because it involves the theory and data which were used to determine the age of the Earth and resolve a controversy which lasted throughout the ninteenth century.

1.2 Debate Over the Age of the Earth

One of the turning points in the history of science was the development of an absolute time scale for geologic history. Before this absolute time scale was developed, the age of the Earth was the subject of great debates involving scientists from many disciplines. The discovery of radioactive elements ended this debate and gave the scientific viewpoint that the earth is 4.5 billion years old. A good review of this debate may be found in van Andel [1985].

Before the late 1700's, scientists and philosophers of Western cultures generally looked towards the Old Testament of the Bible for data needed to determine the age of the earth. Isaac Newton (1642-1627), the father of classical mechanics, for example, spent much of his time studying the biblical scriptures to estimate the time at which geologic history began. By counting the generations listed in the Book of Genesis, Newton and other scholars estimated that the earth was roughly 6000 years old.

In the late 1700's, a Scottish gentleman named James Hutton (1726-1797) noticed that certain predictions made by the Bible about the appearance and location of rocks were falsified by direct observations (*e.g.,* sandstones could be found within granites, rock strata could be seen resting on the eroded remnants of former mountain ranges, etc.). Hutton proposed that the biblical account of earth geology was misleading and that the age of the earth was much longer than the 6000-year figure proposed earlier. Hutton believed, in fact, that the earth's age was so great that it would defy quantitative determination by scientific means. This idea lasted through the early 1800's, and gave rise to a number of related ideas including Darwin's (1809-1882) notion of *natural selection* and the evolution of life.

During the latter half of the 19th century, Lord Kelvin (known as William Thompson, 1824-1907, before his elevation to peerage by Queen Victoria) performed a mathematical calculation that stunned the scientific world and led to the replacement of Hutton's point of view concerning the determinability of the earth's age. Kelvin, an important physicist of his day responsible for many theoretical developments in thermodynamics, realized that the Earth, like any other object in the physical universe, must obey the laws of ther-

modynamics. In particular, the earth and it's internal heat must exemplify the conservation of heat energy. With this realization, Kelvin suggested that the Earth's age was not "indeterminant" as Hutton suggested, but could be measured by considering the physics which governed the way the Earth cooled from its initial, molten state.

Kelvin's calculation (which he reported in a scientific paper published in 1864) involved the temperature profiles measured in deep mine shafts and wells scattered throughout England and Scotland. These temperature profiles showed then, as they do today, that the ambient rock temperature increases by several degrees with every 100 m increase in depth below the earth's surface. If the deep earth is so warm, Kelvin reasoned, then it couldn't have been cooling down from some original molten state for very long. Using the mathematics which we will review in Chapter (6), and data from a deep well in Edinburgh, Scotland, Kelvin estimated the Earth's age to be 20-100 million years. Kelvin's theory created a crisis in geology because many other theories (such as Darwin's theory of natural selection) depended on earth being much older than Kelvin's figure.

For 50 years, this crisis went on unresolved. It was only after the discovery of radioactive elements at the beginning of the 20th century that the impasse between Kelvin's relatively young Earth and the prevailing view of geologists familiar with other indicators of geologic time was resolved. Once it was determined that heat was liberated from radioactive radium salts during the process of radioactive decay, the assumptions Kelvin used to estimate the age of the Earth were found to be incorrect.

Kelvin assumed that the heat now flowing out of the earth's interior was entirely *primeval*, *i.e.,* was left over from the assembly of the planet from the hot gasses and dust of the early solar nebula. He assumed that no new heat could be added to the earth to replace that which seeped out by conduction through the crust. In 1896, Becquerel (1852-1908) discovered an element called radium (Ra). Soon thereafter, Mdm. Marie Sklodowska Curie (1867-1934) determined the significance of this element and others commonly found within the earth which are subject to a newly discovered phenomena called radioactive decay. Heat is emitted when radium, uranium, thorium and other elements undergo nuclear decay. The minor abundance

of radioactive elements within the earth's interior is sufficient account for the heat flow Kelvin observed in the mine shafts without requiring that the age of the earth be as short as Kelvin had estimated. Kelvin was forced to retract his estimate of the earth's age in 1904, and admit that the geologists who held-out for a much older Earth were correct.

The discovery of radioactive decay became significant in a more crucial way in the middle of the twentieth century when when efforts were begun to determine an absolute geologic time scale. Much of this work was done at the University of Chicago by scientists who were formerly employed by the Manhattan Project (a government sponsored project during World War II designed to create the atom bomb). Harold Urey (1893-19), his students, and colleagues determined a way to recognize the existence of natural radiometric clocks found within natural rock samples. In 1956, a former University of Chicago student, Clair Patterson, determined that the earth and all of the meteorites that have fallen on the earth formed (differentiated from a well-mixed solar nebula) at the same time. This age, 4.5 billion years, today stands as the best estimate of the span of geologic history. We will next learn more about Patterson's methodology and repeat in the lab exercises associated with this Chapter his data analysis as an example of a least-squares inverse problem.

1.3 Radioactive Decay

In the examples of radiometric dating discussed below, we will consider three distinct radioactive decay series: rubidium to stronsium decay ($^{87}\mathrm{Rb} \rightarrow ^{87}\mathrm{Sr}$), and two uranium to lead decays ($^{238}\mathrm{U} \rightarrow ^{206}\mathrm{Pb}$ and $^{235}\mathrm{U} \rightarrow ^{207}\mathrm{Pb}$). Each of these decay series can be described by the same mathmatical law. Consider a rock which contains one or more radioactive elements. Denoting the concentration of a radioactive *parent* element by P and the concentration of the stable *daughter* element into which the parent element decays by D, we have

$$P(t) = P_o \exp(-\lambda t) \tag{1.1}$$

$$D(t) = D_o + (P_o - P(t)) \tag{1.2}$$

where t is time, P_o and D_o are initial concentrations of P and D at $t = 0$, and λ is a decay constant that is inversly related to the *half-life*. In the exercises associated with this chapter, you will refer to the decay constants for ^{87}Rb$\rightarrow^{87}$Sr ($\lambda = 1.42 \times 10^{-11}$ per year), ^{238}U$\rightarrow^{206}$Pb ($\lambda = 1.55 \times 10^{-10}$ per year), and ^{235}U$\rightarrow^{207}$Pb ($\lambda = 9.85 \times 10^{-10}$ per year) [Faure, 1986].

1.4 Radiometric Dating

On first glance, (1.1) might seem to provide an easy way to date rocks to determine the time since their formation. If you knew P_o, and could measure P, then you could simply evaluate (1.1) to determine the age of the rock, T:

$$T = \frac{-\ln \frac{P}{P_o}}{\lambda} \tag{1.3}$$

This method has no promise in practical applications for the very simple reason that P_o is never known.

^{40}K $\rightarrow$ ^{40}Ar dating

In some circumstances, a radiometric age can be determined by considering the differences in chemical behavior between the parent and the daughter elements. In the ^{40}K $\rightarrow$ ^{40}Ar decay series, the daughter element is an inert gas which easily effuses from molten rocks. The age of igneous rocks (rocks which form from a melt) can be determined by measuring the amount of the inert daughter gas which builds up within the the rock after gas effusion is effectively shut off by solidification. In ^{40}K$\rightarrow^{40}$Ar decay ($\lambda = 5.54 \times 10^{-10}$ per year) for example, Ar is a noble gas which has a zero concentration while the rock is molten. Once the rock becomes solid, ^{40}Ar gas that is created by ^{40}K decay is locked in.

In this circumstance, it is easy to manipulate Eqn. (1.1) and Eqn. (1.2) to yield

$$\frac{D}{P} = \exp(\lambda T) - 1 \tag{1.4}$$

where T is now interpreted as the age of the rock since it solidified. This age is readily determined by measuring the ratio of the concentration of daughter to parent (^{40}A to ^{40}K):

$$T = \frac{\ln(1 + \frac{D}{P})}{\lambda} \tag{1.5}$$

Potasium-argon dating is used in many applications to date igneous rocks that are associated with volcanic lava flows. By correlating lava flow ages with the magnetic-reversal and biostratigraphic chronologies, most sedimentary rocks currently found on the earth can be dated. The trouble with this form of dating is that it doesn't give the age of the earth's formation as a planet, only the ages of the oldest lava flows. (This method also suffers from the fact that Ar gas can escape from even the most solid rocks, thus ages determined by this method tend to under estimate the true age of the igneous rock.)

1.5 Isochron Method

Another way to overcome the problem of not knowing the initial concentration of the parent element is known as the *isochron* method [Faure, 1986]. To see how this works, we divide (1.2) by the equation which governs the time-evolution of the concentration of a *stable* isotope of the daughter element:

$$D_s(t) = D_{so} \tag{1.6}$$

and use (1.1) to get

$$\frac{D}{D_s} = \frac{D_o}{D_s} + \frac{P}{D_s}(e^{\lambda T} - 1) \tag{1.7}$$

Two advantages are gained in deriving the above expression which involves isotopic ratios. First is that only one initial condition needs to be known to solve for T (*i.e.,* D_o/D_s). The second is that the laboratory practices needed to measure isotopic ratios of the daughter element are generally less stringent than for the measurement of absolute concentrations of the isotopes. (In other words, the chemist can spill half of a sample on the floor and still measure D/D_s to the same precision.)

Notice that the equation for $\frac{D}{D_s}$ as a function of $\frac{P}{D_s}$ in (1.7) is that of a line. The intercept of this line is ($\frac{D_o}{D}$, and $(e^{\lambda T} - 1)$ is the slope. This line is called the *isochron*. If $\frac{D}{D_s}$ points for various minerals within a given rock were to be plotted as a function of $\frac{P}{D_s}$ on a graph, the data points would lie on a line. (To make this point as clearly as possible, we assume that measurement error and other disturbances to the rock samples do not distort the location of the data points. We consider the effects of measurement uncertainty in the next chapter.) The slope of the isochron, α, is related to T, the age of the rock since an initial time when $D = D_o$ (I will explain the physical significance of this so-designated initial time below):

$$\alpha = (e^{\lambda T} - 1) \tag{1.8}$$

and

$$T = \frac{1}{\lambda} \ln(\alpha + 1) \tag{1.9}$$

Notice that T can be determined without ever knowing $\frac{D_o}{D_s}$. Only the slope of the line is important in determining the age of the rock. It is easy, however, to determine $\frac{D_o}{D_s}$ as well simply by locating the intercept of the isochron on the $\frac{P}{D_s}$-axis of the plotted data. A diagram displaying the concept behind the isochron method of radiometric dating is provided in Fig. (1.1).

Given that $\frac{D_o}{D_s}$ does not change with time, and that $\frac{D}{D_s}$ and $\frac{P}{D_s}$ are both observable by laboratory analysis on samples collected today, one can determine the age of a rock by finding the slope of the isochron. This task will provide the *entré* to linear least-squares inverse methods in the discussion which follows.

Before considering an example of the isochron method of radiometric dating, we will need to add one more twist to the technique for determining the age of rock using isotopic measurements. This additional twist was used by Patterson [1956] to determine the 4.5-billion year age of the earth using two independent uranium to lead decay series as measured in meteorites. The motivation for Patterson's improvement of the isochron method is that it is very difficult to measure a P/D_s ratio compared to a D/D_s ratio. For example, many elements that are rich in lead isotopes have very little affinity for uranium, thus, to get the age of an entire planet (the earth) it would be

best to devise a dating strategy that depended only on the measurement of lead isotopes. Fortunately, this is possible due to the fact that there are *two* different isotopes of uranium which decay independently at two different rates to two different lead isotopes.

Consider the $^{238}\text{U}\rightarrow^{206}\text{Pb}$ ($\lambda = 9.85 \times 10^{-10}$ per year) , and $^{235}\text{U}\rightarrow^{207}\text{Pb}$ ($\lambda = 1.55 \times 10^{-10}$ per year)decay series. Each seperately satisfies (1.7), thus

$$\frac{^1D}{D_s} = \frac{^1D_o}{D_s} + \frac{^1P}{D_s}(e^{\lambda_1 T} - 1) \tag{1.10}$$

$$\frac{^2D}{D_s} = \frac{^2D_o}{D_s} + \frac{^2P}{D_s}(e^{\lambda_2 T} - 1) \tag{1.11}$$

Here, we take D_s to be the concentration of a stable lead isotope, say ^{204}Pb. If we subtract the ratio of the initial concentration to the stable concentration from both sides of (1.10) and (1.11) and then divide (1.10) by (1.11), we obtain

$$\left(\frac{^1D}{D_s} - \frac{^1D_o}{D_s}\right)\left(\frac{^2D}{D_s} - \frac{^2D_o}{D_s}\right)^{-1} = \left(\frac{^1P}{^2P}\right)\left(\frac{e^{\lambda_1 T} - 1}{e^{\lambda_2 T} - 1}\right) \tag{1.12}$$

following the same arguement as before, we see that the ratio $\frac{^1D}{D_s}$ is a linear function of $\frac{^2D}{D_s}$, and that the slope of the line povides a means to estimate the age of the rock:

$$\alpha = \left(\frac{^1P}{^2P}\right)\left(\frac{e^{\lambda_1 T} - 1}{e^{\lambda_2 T} - 1}\right) \tag{1.13}$$

Notice that the expression in (1.13) involves the ratio of $\frac{^1P}{^2P}$. This ratio happens to be the same in all meteorites: $\frac{1}{137.8}$.

Unfortunately, it is not possible to invert (1.13) to get a tidy expression for T. We will discuss the inversion of (1.13) for T once α is known after first describing the least-squares line fitting procedure that provides us with the estimate of α when there are more than two points on the plot of $^1D/D_s$ vs. $^2D/D_s$. The object of the lab exercises associated with this chapter will be to use (1.13) to reproduce Patterson's famous calculation of the age of the earth.

1.6 Example: The Age of a Lunar Basalt

As an example of the isochron method, we determine the age of a lunar basalt collected by the Apollo 12 astronauts using ^{87}Rb$\rightarrow^{87}$Sr decay with ^{86}Sr serving as the stable daughter isotope [Nyquist *et al*, 1979]. The data we will use is presented in Table (3) of Nyquist *et al.*'s paper, and a plot of the data points is displayed in Fig. (1.2). The data represent the various measured isotopic ratios of several mineral separates of the whole rock sample that were extracted for the radiometric dating analysis: plagioclase (Plag), pyroxene (Px), ilmenite (Ilm), and the bulk rock itself (WR). Close inspection of the plot of data in Fig. (1.2) suggests that the four data points do not exacly lie on the same line. This is due to measurement error. (We sall postpone the discussion of measurement error to the next chapter.) Casual inspection of these data and an eye-ball guess suggest a slope for the isochron on the order of 0.04. This rough estimate of the slope corresponds to an age of 2.7-billion years using Eqn. (1.9). (The date determined by Nyquist and colleagues is 3.29 ± 0.11 billion years.) Our goal, however, is to determine these parameters as systematically and as accurately as possible. We shall do so by solving a linear inverse problem.

1.7 Linear Algebra of Line Fitting

Let's define the vector $\mathbf{d}$ to be the column vector of ^{87}Sr/^{86}Sr values obtained from the mineral separates and the whole rock. Using the data in Table (3) of Nyquist *et al.*, we have:

$$\mathbf{d} = \begin{bmatrix} 0.70096 \\ 0.69989 \\ 0.70200 \\ 0.70490 \end{bmatrix} \tag{1.14}$$

The vector $\mathbf{m}$ is defined to be the column vector which contains the estimate of the isochron's slope α and intercept β:

$$\mathbf{m} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{1.15}$$

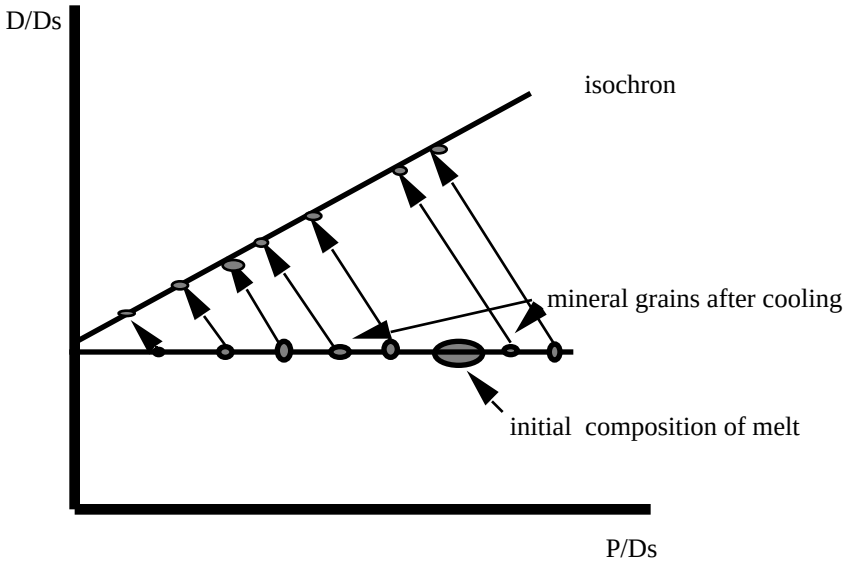

Figure 1.1: Change in slope of an isochron as a function of time.

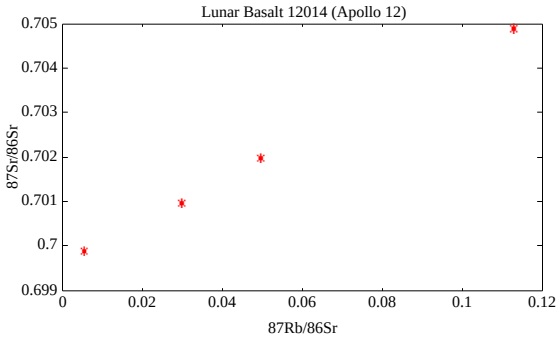

Figure 1.2: Lunar-basalt isotopic-ratio data of Nyquist *et al.* [1979].

The relation between the model vector $\mathbf{m}$ and the data vector $\mathbf{d}$ is

$$\mathbf{Am} = \mathbf{d} \tag{1.16}$$

where,

$$\mathbf{A} = \begin{pmatrix} R_{wr} & 1 \\ R_{Plag} & 1 \\ R_{Px} & 1 \\ R_{Ilm} & 1 \end{pmatrix} = \begin{pmatrix} 0.0296 & 1 \\ 0.00537 & 1 \\ 0.0492 & 1 \\ 0.1127 & 1 \end{pmatrix} \tag{1.17}$$

where R_{xx} denotes the $^{87}\mathrm{Rb}/^{86}\mathrm{Sr}$ value of the xxth mineral separate.

Notice that the matrix $\mathbf{A}$ to be inverted for $\mathbf{m}$ in Eqn. (1.16) is not square. It has four rows but only two columns. this reflects the fact that there are more data points than model parameters. Thus the problem is said to be *overdetermined*. In circumstances where measurement error scatters the data points so that they are not colinear, Eqn. (1.16) cannot be solved exactly.

1.7.1 LU Decomposition

Suppose that we wish to determine the slope and intercept of the lunar basalt isochron using only two data points corresponding to the ilmenite and plagioclase mineral separates (the two end points in Fig. 1.2). In this circumstance, $\mathbf{d}$ and $\mathbf{m}$ have the same dimension (2), and $\mathbf{A}$ is a square, 2×2 matrix. One of the classic techniques for solving a linear system of equations is called the *LU-decomposition*. (This method also goes by the name of Gaussian elimination and back substitution.)

The main idea behind the LU-decomposition can be seen by considering a square matrix $\mathbf{U}$ which contains only zeros below the diagonal:

$$\mathbf{U} = \begin{pmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{pmatrix} \tag{1.18}$$

Suppose we wish to solve the following linear equation involving the matrix $\mathbf{U}$:

$$\mathbf{Ux} = \mathbf{y} \tag{1.19}$$

or

$$\begin{pmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \tag{1.20}$$

where $\mathbf{x}$ and $\mathbf{y}$ are arbitrary column vectors, $\mathbf{y}$ has known values and $\mathbf{x}$ is the desired vector to be solved for. Due to the upper-triangular structure of $\mathbf{U}$, we can immediately perform a back substitution to give:

$$x_2 = \frac{y_2}{u_{22}} \tag{1.21}$$

and

$$x_1 = \frac{y_1}{u_{11}} - \frac{x_2 u_{12}}{u_{11}} \tag{1.22}$$

The reason the solution is so easily found is that we only encounter one unknown at a time during each back substitution step. If we had a lower-triangular matrix $\mathbf{L}$, we could solve a similar problem by performing forward substitution.

The simplicity offered by the structure of upper- and lower-triangular matrices can be exploited to solve general square matrices that have non zero determinants. This is because for any matrix $\mathbf{A}$ it is possible to construct an upper- and a lower-triangular matrix, $\mathbf{U}$ and $\mathbf{L}$ respectively such that

$$\mathbf{A} = \mathbf{LU} \tag{1.23}$$

The triangular structure of $\mathbf{L}$ and $\mathbf{U}$ is exploited to solve for the slope and intercept of the isochron in the following way:

$$\mathbf{Am} = \mathbf{LUm} = \mathbf{d} \tag{1.24}$$

First, define a temporary vector $\mathbf{y}$ such that

$$\mathbf{Ly} = \mathbf{d} \tag{1.25}$$

We use forward substitution to solve for $\mathbf{y}$:

$$y_1 = \frac{d_1}{l_{11}} \tag{1.26}$$

$$y_2 = \frac{d_2}{l_{22}} - \frac{y_1 l_{21}}{l_{22}} \tag{1.27}$$

Next, we use back substitution to sove for $\mathbf{m}$ in terms of the now-known $\mathbf{y}$:

$$m_2 = \frac{y_2}{u_{22}} \tag{1.28}$$

$$m_1 = \frac{y_1}{u_{11}} - \frac{m_2 u_{12}}{u_{11}} \tag{1.29}$$

The LU-decomposition of the 2×2 matrix $\mathbf{A}$ obtained when we use only two data points to determine the slope and intercept of the lunar basalt isochron is found by judiciously multiplying rows of $\mathbf{A}$ by constants and subtracting various rows from various other rows. First add $\frac{-a_{21}}{a_{11}}$ times the first row to the second row. This gives:

$$\begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} - \frac{a_{21}a_{12}}{a_{11}} \end{pmatrix} \tag{1.30}$$

This is the matrix $\mathbf{U}$ that we seek. When the first row (multiplied by a factor) was added to the second, the corresponding elements of $\mathbf{d}$ should also be added to give:

$$\begin{pmatrix} d_1 \\ d_2 - \frac{-a_{21}}{a_{11}} d_1 \end{pmatrix} \tag{1.31}$$

Thus,

$$\mathbf{L} = \begin{pmatrix} 1 & 0 \\ \frac{a_{21}}{a_{11}} & 1 \end{pmatrix} \tag{1.32}$$

One can readily see that

$$\mathbf{LU} = \mathbf{A} \tag{1.33}$$

It is easy to see why the construction of the LU-decomposition depends on the determinant of $\mathbf{A}$ being nonzero. Since the $\mathbf{L}$ we constructed has only ones on its diagonal, failure of the LU-decomposition depends on there being a zero diagonal element of $\mathbf{U}$. When this happens, the back substitution step fails due to an attempt to divide by zero. It is easily shown that

$$\det(\mathbf{A}) = \prod_{i=1}^{2} u_{ii} \tag{1.34}$$

Thus a non-zero determinant assures us that none of the diagonal elements of $\mathbf{U}$ will be zero.

1.8 Least-Squares Inverse

One way to overcome the problem that there are more than two data points in the lunar basalt problem is to perform a least-squares determination of $\mathbf{m}$ in Eqn. (1.16). To do this, we define a performance index J which measures the *distance* between the data vector $\mathbf{d}$ and the predicted data vector $\mathbf{d}_p$ associated with a given estimate of $\mathbf{m}$

$$J = \frac{1}{2}(\mathbf{d}_p - \mathbf{d})' \cdot (\mathbf{d}_p - \mathbf{d}) \tag{1.35}$$

where the prime denotes the transpose. Notice that Eqn. (1.18) represents the dot-product between two difference vectors. This dot-product yeilds the classic sum of squares that the least-squares method is named for. A map of J is displayed in Fig. (1.3). The goal of this section is to determine an efficient and automatic algorithm for finding the $\mathbf{m}$ which minimizes J.

1.8.1 Using Calculus to Find the Minimum of J

To minimize J, we must find an $\mathbf{m}$ which satisfies

$$\begin{aligned} \tfrac{\partial J}{\partial m_i} &= (\mathbf{Am} - \mathbf{d})_j a_{ji} \\ &= a_{ji}a_{jk}m_k - a_{ji}d_j \\ &= 0 \end{aligned} \tag{1.36}$$

Observe that the summation convention is used; thus indicies j and k are summed over their respective ranges (in this case from 1 to 4 for j, and 1 to 2 for k).

The index notation used above can be written in a more illuminating fashion by using matrix-vector multiplications. The km'th component of the product of two matrices $\mathbf{A}$ and $\mathbf{B}$ can be expressed as:

$$(\mathbf{AB})_{km} = a_{ki}b_{im} \tag{1.37}$$

Also, the ij'th component of the transpose $\mathbf{A}'$ of $\mathbf{A}$ can be written:

$$a'_{ij} = a_{ji} \tag{1.38}$$

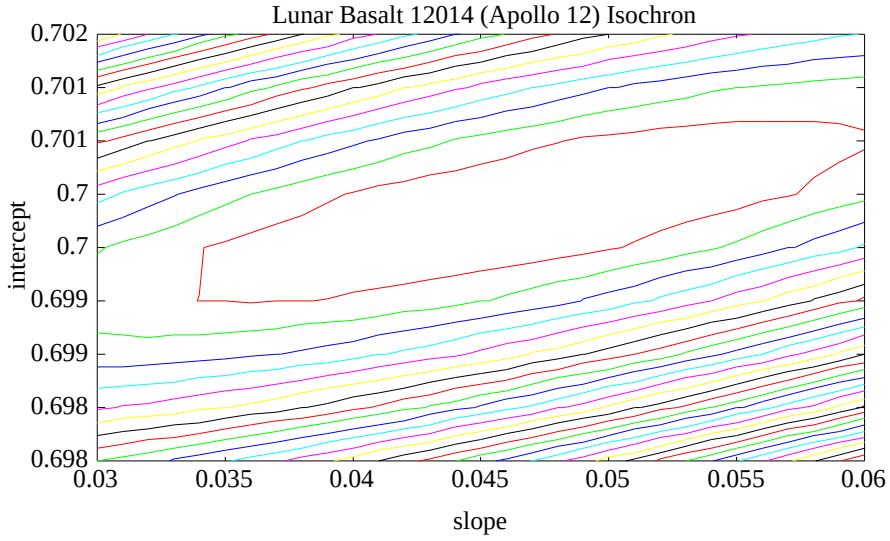

Figure 1.3: Contours of the least-squares performance index J as a function of the slope and intercept of the isochron.

Thus the jk'th component of the product of $\mathbf{A}'$ and $\mathbf{A}$ can be written

$$(\mathbf{A}'\mathbf{A})_{jk} = a_{ji}a_{jk} \tag{1.39}$$

Using Eqn. (1.39), it is now easy to see that Eqn. (1.36) can be written

$$\mathbf{A}'\mathbf{A}\mathbf{m} - \mathbf{A}'\mathbf{d} = \mathbf{0} \tag{1.40}$$

Thus,

$$\mathbf{m} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{d} \tag{1.41}$$

The matrix product $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}$ is refered to as the *least-squares inverse* of the rectangular matrix $\mathbf{A}$. Notice that $\mathbf{A}'\mathbf{A}$ is a square matrix that is also symmetric. This helps to assure us that $\mathbf{A}'\mathbf{A}$ can be inverted (we must still watch out for the possibility that $\mathbf{A}'\mathbf{A}$ is defective, or ill-conditioned in some sense, and that this may prevent us from finding its inverse).

We can now solve for the isochron of the lunar basalt using all four data points. As will be shown in the laboratory exercise, the least-squares inversion of the rectangular matrix is easily performed with the aid of MATLAB®:

```
   >>       m=(A'*A)/A'*d
m =
0.0469
0.6996
```

A plot of the Lunar basalt isotopic ratio data and the least-squares line which runs through the data is provided in Fig. (1.4).

Having found the least-squares solution for the isochron, it is now possible to substitute the numerical value of the slope into Eqn. (1.9) to get

$$T = \frac{1}{1.42 \times 10^{-11}} \ln(0.0469 + 1) = 3.23 \times 10^9 \quad \text{years} \tag{1.42}$$

According to Nyquist *et al.* [1979], the estimated uncertainty of T is 0.11×10^9 years. (Our estimate of this uncertainty will be approximately twice as large,

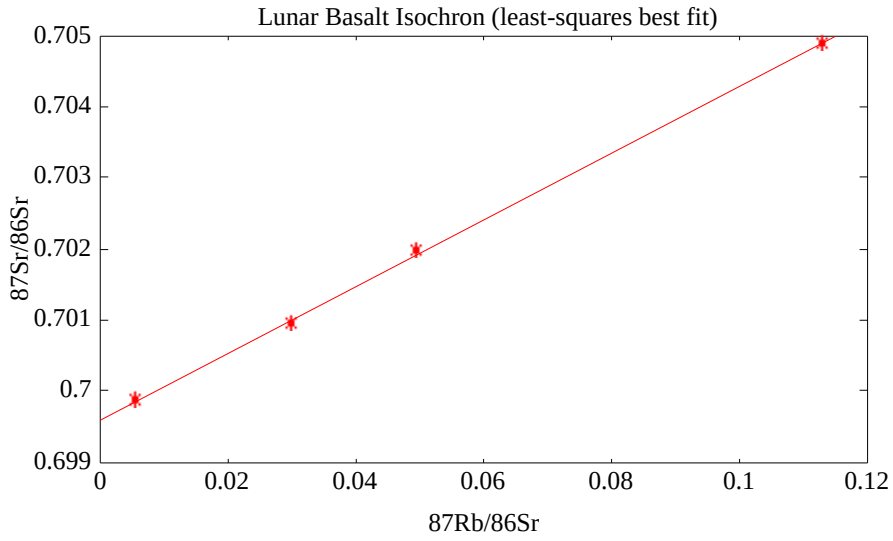

Figure 1.4: The lunar basalt isochron determined by solving the least-squares inverse problem.

as demonstrated in the next chapter.) This age corresponds to the time when great floods of molten basalt created the lunar mare; and is significant because it suggests that the majority of lunar geologic history corresponds to the very early history of the Earth.

1.8.2 A Problem Ahead

Before leaving the subject of radiometric dating it is important to point out that the method for estimating the slope and intercept of an isochron developed in this chapter has not allowed for the fact that measured data appear both in the operator **A** and in the data vector **d**. If we were to arbitrarily switch the data vector in Eqn. (1.14) with the first column of the operator in Eqn. (1.17), the resulting isochron found from inverting Eqn. (1.16) could yield a different age when the *inverse* of its slope is substituted into Eqn. (1.9). In the Lunar-basalt example, the age of the isochron was the same for either case; but, as you will see in Lab 1, the age of the lead-iosotope isochron will differ significantly depending on what data is partitioned into the linear operator **A** and what is partitioned into **d**. Perhaps of greatest concern is the fact that the *uncertainty analysis* to be taken up in Chapter 2 will suffer from the fact that measurement errors associated with the first column of **A** cannot be treated in a systematic manner.

1.9 Summary

In this chapter, we have accomplished two things. First, we have reviewed the history and techniques involved in radiometric dating. Second, we have learned how to solve linear algebra problems involving a rectangular matrix which maps a small number of parameters (the slope and intercept of an isochron, in the example discussed here) to a large number of data points. The key to the solution of such overdetermined linear inverse problems is the fact that we recognize that an exact inversion of the rectangular matrix is *impossible* from the start. We proceed by formulating an inexact problem, the least-squares problem, which we solve readily using the linear algebra of

rectangular matrices.

1.10 Bibliography

Faure, G., 1986. *Principles of Isotope Geology.* (John Wiley and Sons, New York)

Nyquist, L. E., C.-Y. Shih, J. L. Wooden, B. M. Bansal, and H. Wisemann, 1979. The Sr and Nd isotopic record of Apollo 12 basalts: Implications for lunar geochemical evolution. *Proc. 10th Lunar Planet. Sci. Conf.*, 77-114.

Patterson, C., 1956. Age of meteorites and the earth. *Geochimica et Cosmochimica Acta*, **10**, 230-237.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1989. *Numerical Recipes.* (Cambridge University Press, Cambridge, U.K., 702 pp.)

van Andel, T. H., 1985. *New Views on an Old Planet, Continental Drift and the History of Earth.* (Cambridge University Press, Cambridge, U.K., 324 pp.)

1.11 Appendix: Newton-Raphson Method for Finding Roots

The expression relating the slope of the isochron and the age of the isocrhon given by Eqn. (1.13) is transendental. While $\alpha(T)$ is expressed as an analytic function, $T(\alpha)$ cannot be expressed analytically. This is due to the fact that T appears in the exponential functions of both the numerator and denominator on the right-hand side of Eqn. (1.13). These expressions cannot be inverted by taking a logarithm to form a simple, analytic expression. This presents us, as it did Clair Patterson, with a problem. We can determine α; but once determined, how do we then use α to determine T using Eqn. (1.13)?

We will use the Newton-Raphson technique to determine T from a known α. This technique is just one of many that are recommended for such problems. A good review of these techniques can be found in Press *et al.* [1989]. The Newton-Raphson method is an algorithm that is good for finding roots of a function $f(x)$. A root is a special value of x, say $\tilde{x}$, where $f(\tilde{x}) = 0$. The Newton-Raphson algorithm finds the root by an iterative procedure whereby an initial guess of $\tilde{x}$, say x_o, is corrected through evaluation of f and its derivative f' at $x = x_o$. This algorithm is derived by considering a truncated Taylor series which expresses $f(x)$ in the neighborhood of the initial guess x_o:

$$f(x_o + \delta) = f(x_o) + \delta f'(x_o) \qquad (A1)$$

We know that this expression is not valid when δ is large because other terms in the Taylor series expression that have been truncated from (A1) are significant. Nevertheless, we shall assume that Eqn. (A1) is accurate and "hope for the best". If we define our intial guess, x_o to be separated from the desired root by δ, $x_o = \tilde{x} - \delta$, then

$$\delta = -\frac{f(x_o)}{f'(x_o)} \qquad (A2)$$

We can improve our guess of $\tilde{x}$ by making use of the estimate for δ given by Eqn. (A2):

$$x_1 = x_o + \delta = x_o - \frac{f(x_o)}{f'(x_o)} \qquad (A3)$$

We can continue this procedure *ad infinitum* to eventually reach the root. Each successive guess at the root will be related to the previous guess by the following recursion rule:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \qquad (A4)$$

Eventually, we might wish to stop the iterative process when x_{i+1} and x_i no longer differ by a significant amount. This stopping point is a matter of judgement, and often is related to the level of observational error found in the observations which go into defining $f(x)$, or the accuracy of the computer being used to perform the arithmetic. (For example, on a Macintosh, using single-precision arithmetic, two floating point numbers are the same if their mantissas are the same to the 7'th decimal place.)

To solve Eqn. (1.13) for T when α is known, we can define a function

$$f(T) = \alpha - \left(\frac{^1P}{^2P}\right)\left(\frac{e^{\lambda_1 T} - 1}{e^{\lambda_2 T} - 1}\right) \qquad (A5)$$

Clearly, when $f(T) = 0$, Eqn. (1.13) is satisfied.

1.11.1 Example

Let's use the Newton-Raphson technique to determine the age of the lead-isotope isochron considered by Patterson [1956], and which we will assume has a slope of $\alpha = 0.6$. To impliment the algorithm expressed by Eqn. (A4), we make use of the derivative of $f(T)$, which can easily be determined analytically using the chain rule:

$$f'(T) = \left(\frac{^1P}{^2p}\right)\left\{\frac{-\lambda_1 e^{\lambda_1 T}}{e^{\lambda_2 T} - 1} + \frac{\lambda_2 e^{\lambda_2 T}(e^{\lambda_1 T} - 1)}{(e^{\lambda_2 T} - 1)^2}\right\} \qquad (A6)$$

Values of the constants in Eqns. (A5) and (A6) are: $\frac{^1P}{^2P} = \frac{1}{138.7}$, $\lambda_1 = 9.72 \times 10^{-10}$ yr^{-1}, and $\lambda_2 = 1.537 \times 10^{-10}$ yr^{-1}. First, let's define the function and its derivative using MATLAB® script-files:

```
% ************************
% Function isochron
% *************************
% When this function is zero, T must be the age of the
% lead-isotope isochron of slope alpha:
%
function f = isochron(T);
f=alpha-(1/138.7)*(exp(lambda1*T)-1)/(exp(lambda2*T)-1);
%
% End of Function.


% ************************
% Function isochronprime
```

```
% *************************
% This is the derivative of the isochron-function with respect to
T:
%
%
function fp = isochronprime(T);
fp=-(1/138.7)*(lambda1*exp(lambda1*T)/(exp(lambda2*T)-1)+ ...
lambda2*exp(lambda2*T)*(exp(lambda1*T)-1)/(exp(lambda2*T)-1)%
% End of Function
```

Next, let's write a script file that runs the Newton-Raphson algorithm:

```
% ***********************
% Script which runs Newton-Raphson Algorithm
% *************************
% This routine performs the Newton-Raphson algorithm to seek
% the root (T) of the function "isochron()" which is stored as
% a MatLab function, using the derivative of the function
% "isochronprime()":
%
% Initial guess:
%
Tn=1.5e9;
test=1.e12;
%
% Do the NR-loop until convergence to limit 1.0e4 (10,000 years):
%
counter=0;
while (abs(test)> 1.0e4)
counter=counter+1;
Tnp1=Tn-isochron(Tn)/isochronprime(Tn);
test=Tnp1-Tn;
Tn=Tnp1;
end
%
% Write the result:
%
```

```
counter
%
Tn
%
isochron(Tn)
% End of script.
```

Now, let's use these functions and script in a MATLAB® session:

```
>> alpha=0.6;
>> global alpha;
>> lambda1=9.72e-10;
>> lambda2=1.537e-10;
>> global lambda1 lambda2;
>> NR
counter =

15

Tn =

4.5843e+09

ans =

3.0532e-06
```

Let's check the answer by using the MATLAB® -native routine `fzero()`:

```
>>fzero('isochron',2e9)

ans =

4.5843e+09
```

1.12 Lab 1 - The Radiometric Determination of the Age of the Earth

To test your understanding of the concepts of radiometric dating and linear least-square inverse methods, try to repeat the famous lead-isotope chronometry performed by Clair Patterson [1956] to determine the 4.55-billion year age of the earth (and meteorites). Patterson's paper will be handed to you in class, so you can get the data you need from his Table (1). The method I ask you to use in this lab is slightly different than his. (Notice that his Eqn. (1) does not address the overdetermined nature of his data-analysis task. He has five meteorite lead-isotope measurements to deal with; yet his Eqn. (1) is only capable of dealing with two at a time.) Please use the linear least-squares inverse methods discussed in Chapter 1.

1.12.1 Look at the Data

Before setting up the linear-algebra problem represented by Eqn. (1.16), plot your lead-isotope data and estimate the slope and intercept of the meteorite isochron.

Problem 1. Use the MATLAB® plotting routines to construct a graph of the $\frac{^{207}\mathrm{Pb}}{^{204}\mathrm{Pb}}$ *vs.* $\frac{^{206}\mathrm{Pb}}{^{204}\mathrm{Pb}}$ values for the 5 meteorites listed in Table (1) of Patterson's [1956] paper. Use the text labeling routines of MATLAB® to identify each point.

Problem 2. Estimate the slope of the isochron, and compare your estimate with the slope estimated by Patterson. To determine Patterson's estimated slope, plug his determined age of the isochron (4.55×10^9 years) into Eqn. (1.13).

1.12.2 The Forward Problem

Before confirming your estimate of the slope of the isochron using least-squares inverse methods, let's see what the time-evolution of the data should have been for an earth 4.55-billion years old.

Problem 3. Two of the five meteorites analyzed by Patterson were of the iron variety (the other three were stony). These iron meteorites (*Henbury, Australia* and *Canyon Diablo, Arizona*) came from the iron core of a planetary body that was demolished early in the history of the solar system. The two iron meteorites are fragments of the original iron core. The three stony meteorites are probably fragments of the stony part of the original planetary body.

An important chemistry fact is that ^{238}U and ^{235}U (the parent elements of ^{206}Pb and ^{207}Pb) are not *siderophilic*. (A siderophile is an element that likes to chemically join with the metalic elements of the Fe and Ni core of planetary bodies.) This means that the lead-isotope ratios in the iron meteorites are probably primeval. In other words, uranium probably did not get into the cores of the planetary bodies at the time they were formed. Thus, the lead-isotope ratios represented by the two iron meteorites are probably the *initial* ratios for these isotopes for all material in the solar system (even the parts of planets, such as the earth's crust and mantle, which received a large dose of uranium during the differentiation process).

For this problem, I ask you to devise a series of 5 graphs like the one you plotted in Problem 1 which show the time-evolution of the isotopic concentrations of the five meteorites. The 5 graphs should plot the data that would have been measured at 1, 2, 3, 4 and 4.5 billion year intervals after the formation of the planets. To do this, you will have to devise an expression which indicates what the ratios $\frac{^{235}\mathrm{U}}{^{204}\mathrm{Pb}}$ and $\frac{^{238}\mathrm{U}}{^{204}\mathrm{Pb}}$ were at the time of planetary differentiation. This should not be hard to do, just use Eqns. (1.10), (1.11), and assume that the age of the earth is indeed what Patterson measured (4.55-billion years). Remember to put labels and titles on your 5 graphs to indicate which data points correspond to which meteorites and for what time the data is being plotted.

1.12.3 The Least-Squares Inverse

Now that you have pretty much solved the problem by the normal trial-and-error technique, let's do it the easy and sophisticated way using the linear algebra of overdetermined linear systems.

Problem 4. Construct the matrix $\mathbf{A}$ and the vector $\mathbf{d}$ for the linear problem (1.5.3) applied to Patterson's data. (In this circumstance the definitions of the data and the entries in the matrix are different from what was done in Eqns. (1.5.1) and (1.5.4) due to the fact that only lead-isotopes are being fit to a line.) Enter your matrix and data vector as MATLAB® variables in an interactive MATLAB® session. (Try printing a portion of the screen to serve as a record of your data entry.)

Next, construct the matrix $\mathbf{A}'\mathbf{A}$ and the modified data vector $\mathbf{A}'\mathbf{d}$ using the MATLAB® commands. Agian, show your results. Determine the LU-decomposition of $\mathbf{A}'\mathbf{A}$ (use the MATLAB® -native LU-decomposition routine), and apply this decomposition to the modified data vector $\mathbf{A}'\mathbf{d}$ to obtain the least-squares estimate of the slope and intercept of the isochron for the meteorites (in other words, the vector $\mathbf{m}$). (Note, you will have to devise your own forward substitution and back substitution algorithms. This will give you practice in setting up 'for'-loops and MATLAB® script files.)

Finally, do the above steps the easy way by using the backslash notation of MATLAB® to eliminate the need to explicitly work out the LU-decomposition and perform the forward and backward substitutions.

Problem 5. Plot the lead-isotope isochron you found in Problem 4 on the graph of the lead-isotope data determined in Problem 1.

Problem 6. Repeat your determination of the slope of the lead-isotope isochron, but this time swap the data contained in the first column of the $\mathbf{A}$ with the data contained in $\mathbf{d}$. How does the inverse of this slope ($\frac{1}{\alpha}$) compare with the slope you determined in Problem 4. Will the age of the isochron determined in this problem be different from that of the isochron determined in problem 4.

1.12.4 Inverting an Intransitive Relationship

One of the tough parts of the analysis is the fact that the relation between the slope of the isochron and the age of the isochron cannot be analytically inverted. While you can write α as a function of T, you can't write T as a function of α. To determine the age of the isochron who's slope was found above, you will have to use some approximation strategy to invert the formula in Eqn. (1.13).

Problem 7. Determine T, the age of meteorites, for the isochrons determined in Problems 4 and 6 by using the *Newton-Raphson* method. This method involves finding the root (zero) of the following function:

$$f(T) = \alpha - \frac{1}{137.8}\left(\frac{e^{\lambda_1 T} - 1}{e^{\lambda_2 T} - 1}\right) \qquad (L4.0.1)$$

where λ_1 and λ_2 are the decay constants for ^{238}U and ^{235}U, respectively. Devise a MATLAB® script which performs the Newton-Raphson algorithm for an arbitrary MATLAB® function; you may wish to use this algorithm later. Check your answer by using the MATLAB® -native routine `fzero()`.

Discuss the significance of your solution to the age of the earth. Suggest reasons why your result is different than that determined by Patterson [1956].

Chapter 2

Underdetermined Inverse Problems: Minimum-Norm Line Fitting

2.1 Introduction

In Chapter (**??**), we learned how to solve an inverse problem in which the number of observations exceeds the number of unknown model parameters. Here we consider the opposite situation, *i.e.,* an inverse problem in which the number of undetermined model parameters exceeds the number of observations. Problems of this nature do not have unique solutions; thus, additional constraints must be imposed artificially on the undetermined model parameters to select a single solution from the multitude of solutions which satisfy the constraints imposed by the observations. The technique we shall derive here is referred to as the *minimum norm* inverse of underdetermined problems. The term, minimum norm, refers to an artificial constraint that the solution be simple in some sense, such as having minimum norm in the vector space which contains it. Taken together, the minimum norm inverse and the *least squares* inverse described in Chapter (**??**) provide a means to solve most inverse problems encountered in the geophysical sciences.

2.2 An Absurd Inverse Problem: Fitting an Isochron Through One Point

For continuity, we shall consider the lunar-basalt isochron problem developed in Chapter (**??**), but with one rather strange twist. We shall assume that only one measurement of the $^{87}Sr/^{86}Sr$ and $^{87}Rb/^{86}Sr$ values exist, say the data pair for the whole-rock sample. Obviously, with only one single data point, an infinite number of possible isochrons fit the data. The question of dating the lunar basalt under such circumstances becomes absurd. There is no unique date.

While the above problem may seem absurd, it serves to illustrate both the nature of all underdetermined inverse problems (which often appear to be very reasonable despite the non-uniqueness of their solution) and the solution method. We thus proceed with the problem of determining $\mathbf{m} = [\alpha\ \beta]' \in \mathcal{R}^2$ from a single data point $\mathbf{d} \in \mathcal{R}^1$ subject to the constraint

$$\mathbf{A}\mathbf{m} = \mathbf{d} \tag{2.1}$$

where, following § (1.7)

$$\mathbf{d} = [0.70096] \tag{2.2}$$

and

$$\mathbf{A} = [\, R_{wr} \quad 1\,] = [\, 0.0296 \quad 1\,] \tag{2.3}$$

The 1×2 matrix $\mathbf{A} : \mathcal{R}^2 \rightarrow \mathcal{R}^1$ is rectangular and thus cannot be inverted.

To make headway, suppose that we have a hunch that the lunar basalt is approximately 3.5 $\times 10^9$ years old (corresponding to an isochron with a slope of 0.0497) and that the initial $^{87}Sr/^{86}Sr$ ratio at the time the basalt formed was 0.7000. This hunch might lead us to seek the $\mathbf{m}$ which satisfies Eqn. (2.1) and at the same time minimizes the difference between $\mathbf{m}$ and $\mathbf{m}_h = [0.0497\ 0.7000]'$, where $\mathbf{m}_h$ represents the value of $\alpha = \alpha_h$ and $\beta = \beta_h$ (the slope and intercept of the isochron, respectively) associated with the "hunch". In mathematical terms, we wish to minimize the following performance index

$$\begin{aligned} J = &\quad [\mathbf{m} - \mathbf{m}_h]'\,[\mathbf{m} - \mathbf{m}_h] \\ = &\quad (\alpha - \alpha_h)^2 + (\beta - \beta_h)^2 \end{aligned} \tag{2.4}$$

subject to Eqn. (2.1) as a constraint. The scalar quantity J can be referred to as a *norm*, or way of measureing the length of vectors in $\mathcal{R}^2$. Thus, the inverse problem we wish to solve is to minimize the norm of $\mathbf{m}$ subject to the constraint that it satisfies the data expressed by Eqn. (2.1).

The minimization of J subject to Eqn. (2.1) is relatively straightforward. First, we write $m_1 = \alpha$ as a function of $m_2 = \beta$ and $\mathbf{d}$:

$$m_1 = \alpha = \frac{d_1 - A_{12}m_2}{A_{11}} = \frac{d_1 - A_{12}\beta}{A_{11}} \tag{2.5}$$

Substitution into Eqn. (2.4) gives,

$$J = \left(\frac{d_1 - A_{12}\beta}{A_{11}} - \alpha_h\right)^2 + (\beta - \beta_h)^2 \tag{2.6}$$

We can now appeal to calculus to find the constraints which minimize J:

$$\frac{dJ}{d\beta} = 0 = -2\frac{A_{12}}{A_{11}}\left(\frac{d_1 - A_{12}\beta}{A_{11}} - \alpha_h\right) + 2(\beta - \beta_h) \tag{2.7}$$

The above equation gives

$$\beta = \frac{\frac{A_{12}d_1}{A_{11}^2} + \frac{A_{12}\alpha_h}{A_{11}^2} + \beta_h}{\left(\frac{A_{12}}{A_{11}}\right)^2 + 1} \tag{2.8}$$

which may be substituted into Eqn. (2.5) to obtain the corresponding α.

2.2.1 Lagrange Undetermined Multiplier

Another way to satisfy the constraint is to augment the performance index J with the addition of a *Lagrange multiplier* term. We define this augmented performance index H as follows

$$H = [\mathbf{m} - \mathbf{m}_h]'[\mathbf{m} - \mathbf{m}_h] + 2\underline{\lambda}'(\mathbf{Am} - \mathbf{d}) \tag{2.9}$$

where $\underline{\lambda} \in \mathcal{R}^1$ is the Lagrange undetermined multiplier vector. We now minimize H with respect to *two* unknowns: $\mathbf{m}$ and $\underline{\lambda}$. From our understanding

of calculus, we seek the $\mathbf{m}$ and $\underline{\lambda}$ which make the partial derivatives of H with respect to the components of $\mathbf{m}$ and $\underline{\lambda}$ equal to zero:

$$\frac{\partial H}{\partial \mathbf{m}} = 2\mathbf{m}' - 2\mathbf{m}'_h + 2\underline{\lambda}'\mathbf{A} = \mathbf{0} \tag{2.10}$$

$$\frac{\partial H}{\partial \underline{\lambda}} = 2\left(\mathbf{m}'\mathbf{A}' - \mathbf{d}'\right) = \mathbf{0} \tag{2.11}$$

The two *Euler-Lagrange* conditions implied by Eqns. (2.10) and (2.11) are

$$\mathbf{m} = \mathbf{m}_h - \mathbf{A}'\underline{\lambda} \tag{2.12}$$

$$\mathbf{A}\mathbf{m} - \mathbf{d} = 0 \tag{2.13}$$

Substitution the expression for $\mathbf{m}$ given in Eqn. (2.12) into Eqn (2.13) , we obtain an expression for $\underline{\lambda}$

$$\underline{\lambda} = \left[\mathbf{A}\mathbf{A}'\right]^{-1}\mathbf{A}\mathbf{m}_h - \left[\mathbf{A}\mathbf{A}'\right]^{-1}\mathbf{d} \tag{2.14}$$

Observe that the matrix $[\mathbf{A}\mathbf{A}']$ is a square, 1×1 matrix which has an inverse. (Do not be alarmed by the formal rigor of retaining matrix algebra notation despite the matrix having only one row and one column. The results will generalize to problems involving higher dimensional vector spaces readily.) Substituting the expression for $\underline{\lambda}$ given in Eqn. (2.14) back into Eqn. (2.12) gives the final solution:

$$\begin{aligned}\mathbf{m} &= \mathbf{m}_h - \mathbf{A}'\left[\mathbf{A}\mathbf{A}'\right]^{-1}\mathbf{A}\mathbf{m}_h + \mathbf{A}'\left[\mathbf{A}\mathbf{A}'\right]^{-1}\mathbf{d} \\ &= \left[\mathbf{I} - \mathbf{A}'\left[\mathbf{A}\mathbf{A}'\right]^{-1}\mathbf{A}\right]\mathbf{m}_h + \mathbf{A}'\left[\mathbf{A}\mathbf{A}'\right]^{-1}\mathbf{d}\end{aligned} \tag{2.15}$$

Defintion: A General Minimum-Norm Inverse

The expression given by Eqn. (2.15) when $\mathbf{m}_h = \mathbf{0}$ is often referred to as the *minimum-norm* inverse of the following problem

$$\mathbf{A}\mathbf{m} = \mathbf{d} \tag{2.16}$$

when $\mathbf{m} \in \mathcal{R}^N$, $\mathbf{d} \in \mathcal{R}^M$, and $\mathbf{A} : \mathcal{R}^N \rightarrow \mathcal{R}^M$ is a rectangular $M \times N$ matrix with $N > M$. In other words, the minimum-norm inverse is

$$\mathbf{m} = \mathbf{A}'\left[\mathbf{A}\mathbf{A}'\right]^{-1}\mathbf{d} \tag{2.17}$$

■

Example

Using MATLAB® , the solution given by Eqn. (2.15) was evaluated using the data given in Eqns. (2.2) and (2.3) with $\mathbf{m}_h = [0.0497\ 0.7000]'$:

```
>>m=[eye(2,2) - A'*inv(A*A')*A]*mh + A'*inv(A*A')*d

m =

0.0497
0.6995
```

Clearly, this solution is far from satisfactory because it represents an isochron that is far older than the known age of the lunar basalt derived in Chapter (1). This inaccuracy serves to illustrate the pitfalls inherent in underdetermined inverse problems. In the current circumstance, the "hunch" $\mathbf{m}_h$ used to select one solution from the multitude of possible solutions was inaccurate. Unfortunately for most underdetermined inverse problems, there is no *a priori* way of evaluating the accuracy of the minimum norm solution.

Chapter 3

Dealing with Uncertainty

3.1 Introduction

In Chapter (**??**) we developed a method for fitting a line through a scatter of data points. The pay off was a means to estimate the age of the earth by determining the slope of the isochron which passed through isotopic data measured in five meteorites collected on the earth's surface. What made this problem hard was the fact that the five meteorite samples *overdetermine* the two unknowns necessary to describe the isochron.

What we didn't consider in Chapter (**??**) was the fact that the isotopic measurements are subject to uncertainty introduced by instrumental errors, sample handling, and other causes. In this chapter, we will account for this uncertainty in the methodology for estimating the slope and intercept of the isochron. In addition, we will examine how uncertainty in the isotopic measurements propagates through the least-squares line fitting method to ultimately yield a quantifyable uncertainty in the age of the earth.

3.2 Expectation Operators and Covarience Matrices

Consider the overdetermined problem defined in §(1.7). The vector $\mathbf{m} \in \mathcal{R}^N$ with $N = 2$ represents the slope and intercept of the isochron, the data vector $\mathbf{d} \in \mathcal{R}^M$ represents the $M > N$ measured isotopic ratios, and the rectangular $M \times N$ matrix $\mathbf{A}$ represents the mapping which converts the slope and intercept of the isochron into predicted values of the isotopic ratios. Restating the problem, our assumption is that $\mathbf{m}$ and $\mathbf{d}$ are related linearly, *i.e.,*

$$\mathbf{Am} = \mathbf{d} \tag{3.1}$$

and that $\mathbf{d}$ is known to us, but $\mathbf{m}$ is not. The purpose of the previous chapter was to devise a scheme to determine $\mathbf{m}$ in the face of the fact that $\mathbf{A}$ is a rectangular matrix. Here, we shall consider an additional difficulty in solving Eqn. (3.1), namely the difficulty which arises when the data $\mathbf{d}$ is subject to error.

We assume that the data actually observed $\mathbf{d}^o$ represents the sum of the "actual" state of the vector $\mathbf{d}$ and a random error vector ϵ:

$$\mathbf{d}^o = \mathbf{d} + \underline{\epsilon} \tag{3.2}$$

We shall assume that we know in advance certain statistical properties of the errors, namely,

$$\langle \underline{\epsilon} \rangle = \mathbf{0} \tag{3.3}$$

$$\langle \underline{\epsilon}\epsilon' \rangle = \langle (\underline{\epsilon} - \langle \underline{\epsilon} \rangle)(\underline{\epsilon} - \langle \epsilon \rangle)' \rangle = \mathbf{Q} \tag{3.4}$$

where we use the notation $\langle \cdot \rangle$ to denote the expectation value of the variable enclosed by the angle brackets. The $M \times M$ matrix $\mathbf{Q}$ is called the covarience matrix. It's diagonal elements represent the standard deviations of the individual components of ϵ, and it's off-diagonal elements represent the correlations between different components of ϵ. In circumstances where ϵ arises because of random measurement error associated with instrumentation (*i.e.,* not because of inadequacy of our assumed linear relationship expressed in Eqn. (3.1)), we expect $\mathbf{Q}$ to be diagonal.

The expectation value, $\langle\cdot\rangle$, is formally defined using the notion of probability:

3.2.1 Expectation Operator

The probability that a given random variable $\underline{\epsilon}$ will lie within $d\underline{\epsilon}$ of a given value $\underline{\epsilon}_o$ is defined to be $P(\underline{\epsilon}_o)d\underline{\epsilon}_o$. The mean (expectation value) and covariance (second moment) of the random variable are defined to be

$$\langle\underline{\epsilon}\rangle = \int_{\mathcal{R}^M} \underline{\epsilon}\ P(\underline{\epsilon})\ d\underline{\epsilon} \tag{3.5}$$

$$Q_{ij} = \int_{\mathcal{R}^M} (\underline{\epsilon} - \langle\underline{\epsilon}\rangle)_i\ (\underline{\epsilon} - \langle\underline{\epsilon}\rangle)_j\ P(\underline{\epsilon})\ d\underline{\epsilon} \tag{3.6}$$

where the integrations are over the entire vector space $\mathcal{R}^M$ which contains $\underline{\epsilon}$. A good explanation of the above definitions is provided in chapter 2 of Menke (1989).

3.3 Two Questions

Given the existence of ϵ, we are confronted with two basic questions. First, should we modify the least-squares inverse derived in §(1.8) to account for the fact that some components of $\mathbf{d}^o$ are subject to greater errors (presumably) than others? Second, once the first question is answered and a solution

$$\hat{\mathbf{m}} = \mathbf{m} + \underline{\zeta} \tag{3.7}$$

is found to the equation

$$\mathbf{A}\hat{\mathbf{m}} = \mathbf{d}^o \tag{3.8}$$

determined via a method which satisfactorily addresses the first question, what will be the statistical description of its error $\underline{\zeta}$?

An appreciation for the first question can be derived by considering Fig. (3.1). Three data points with error bars (denoting expected standard deviation of the error in the ordinate value of the data) are displayed along with two possible lines which are "fit" in some sense to the data. The dashed line represents the least-squares fit which simply represents the minimization of J defined in §(1.8). It clearly does not account for the fact that the error bar on the right-most data point is much larger than the error bars on the two other data points. The solid line represents what is considered to be a superior fit to the data (*i.e.,* its slope is positive, which is physically required of all isochrons in radiometric dating). To achieve the superior fit, it is necessary to weight the misfit between the line and the data less for data points in which the uncertainty is high, and more for data points in which the uncertainty is low. The following modification of the least-squares inverse derived in §(1.8) will do the trick.

3.3.1 Least-Squares Inverse with Data Uncertainty

Consider the following least-squares performance index J:

$$J = (\mathbf{Am} - \mathbf{d}^o)' \mathbf{Q}^{-1} (\mathbf{Am} - \mathbf{d}^o) \tag{3.9}$$

The inverse of $\mathbf{Q}$ appears on the right-hand side of the above definition as a means to preferentially weight the misfit between the predicted and observed data components. The solution $\hat{\mathbf{m}}$ which minimizes J is

$$\hat{\mathbf{m}} = \left[\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\right]^{-1} \mathbf{A}'\mathbf{Q}^{-1}\mathbf{d}^o \tag{3.10}$$

This solution of the inverse problem is preferable to that derived in the previous chapter because, in situations such as that depicted in Fig. (3.1), the line chosen to fit the data will account properly for non-uniform error bars.

3.3.2 Model Covarience

Having defined the least-sqares solution to Eqn. (3.8) in a manner consistent with the presence of data uncertainty (represented by the matrix $\mathbf{Q}$), we next

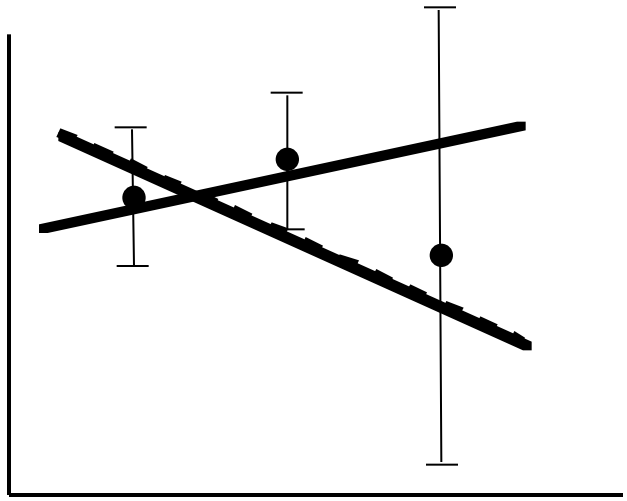

Figure 3.1: Two possible lines fit to three data points. Vertical bars indicate expected errors in data. The solid line represents a solution to the line-fitting problem which properly accounts for the uncertainty of the data. The dashed line represents a solution which does not.

consider how to describe the uncertainty of $\hat{\mathbf{m}}$. We define the covarience matrix $\mathbf{E}$ to be the covarience of $\underline{\zeta}$ the random errors in $\hat{\mathbf{m}} = \mathbf{m} + \underline{\zeta}$:

$$\mathbf{E} = \left\langle \underline{\zeta}\underline{\zeta}' \right\rangle \tag{3.11}$$

Our goal is to express $\mathbf{E}$ in terms of $\mathbf{Q}$. To achieve this goal, we note the definition

$$\underline{\zeta} = \hat{\mathbf{m}} - \mathbf{m} \tag{3.12}$$

and make use of the assumption that $\mathbf{Am} = \mathbf{d}$ (*i.e.,* misfit between the line and the measured data is due to errors $\underline{\epsilon}$ in the data only) to write

$$\begin{aligned} \mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\underline{\zeta} &= \mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\hat{\mathbf{m}} - \mathbf{A}'\mathbf{Q}^{-1}\mathbf{Am} && (3.13)\\ &= \mathbf{A}'\mathbf{Q}^{-1}\mathbf{d}^o - \mathbf{A}'\mathbf{Q}^{-1}\mathbf{d} && (3.14)\\ &= \mathbf{A}'\mathbf{Q}^{-1}\underline{\epsilon} && (3.15) \end{aligned}$$

thus,

$$\underline{\zeta} = \left[\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}'\mathbf{Q}^{-1}\underline{\epsilon} \tag{3.16}$$

Substitution of the above expression into Eqn. (3.11) gives,

$$\begin{aligned} \left\langle \underline{\zeta}\underline{\zeta}' \right\rangle = \mathbf{E} &= \left[\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}'\mathbf{Q}^{-1}\left\langle \underline{\epsilon}\underline{\epsilon}' \right\rangle \left[\mathbf{Q}^{-1}\right]'\mathbf{A}\left[\left[\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\right]^{-1}\right]' && (3.17)\\ &= \left(\left[\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}'\left[\mathbf{Q}^{-1}\right]'\mathbf{A}\right)\left[\left[\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\right]^{-1}\right]' && (3.18)\\ &= \left[\left[\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\right]^{-1}\right]' = \left[\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\right]^{-1} && (3.19) \end{aligned}$$

In simplifying the above expression, we have made use of the fact that $\mathbf{E}$, $\mathbf{Q}$ and its inverse are symmetric matrices (*i.e.,* $\mathbf{Q}' = \mathbf{Q}$). The bottom line is that we have a precise description of the uncertainty in the derived quantity $\hat{\mathbf{m}}$ that is a simple linear function of $\mathbf{Q}$, the uncertainty of the data.

3.3.3 Example: Lunar Basalt Isochron Uncertainty

As an example of the above error-analysis works, we reconsider the problem discussed in §(1.8) in which the age of a lunar basalt is determined using

$^{87}\mathrm{Rb} \rightarrow {}^{87}\mathrm{Sr}$ dating. The covariance matrix $\mathbf{Q}$ associated with the $^{87}\mathrm{Sr}/^{86}\mathrm{Sr}$ data is [Nyquist *et al.*, 1979]

$$\mathbf{Q} = \begin{bmatrix} (3.5 \times 10^{-5})^2 & 0 & 0 & 0 \\ 0 & (4.5 \times 10^{-5})^2 & 0 & 0 \\ 0 & 0 & (2.5 \times 10^{-5})^2 & 0 \\ 0 & 0 & 0 & (3.0 \times 10^{-5})^2 \end{bmatrix} \tag{3.20}$$

Two least-squares solutions to the problem $\mathbf{A}\tilde{\mathbf{m}} = \mathbf{d}^o$ posed in §(1.8) can be constructed. The first is the solution developed in the previous chapter which takes no account of data uncertainty,

$$\tilde{\mathbf{m}} = [\mathbf{A}'\mathbf{A}]^{-1}\mathbf{A}'\mathbf{d}^o \tag{3.21}$$

The second is that derived above, *i.e.,*

$$\hat{\mathbf{m}} = \left[\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}'\mathbf{Q}^{-1}\mathbf{d}^o \tag{3.22}$$

(To be rigorous, a third least-squares solution will be discussed in a future chapter, and arises due to the fact that the matrix $\mathbf{A}$ contains measured $^{87}\mathrm{Rb}/^{86}\mathrm{Sr}$ ratios which are also subject to uncertainty which is not accounted for here.) Using the Nyquist *et al.* [1979] data for $\mathbf{A}$ and $\mathbf{d}$, and the above estimate of $\mathbf{Q}$ (this estimate was made using data presented in Table 3 of Nyquist *et al.*, and the assumption that one standard deviation of the data reported by Nyquist *et al.* should be taken to be the diagonal components of $\mathbf{Q}$), the values of $\mathbf{m}$ and $\hat{\mathbf{m}}$ were computed, and found to differ only in the fifth significant digit, *i.e.*,

$$\hat{\mathbf{m}} - \mathbf{m} = \begin{bmatrix} -8.9 \\ 1.5 \end{bmatrix} \times 10^{-5} \tag{3.23}$$

Proper accounting of data uncertainty thus does not lead to much difference in the least-squares determination of the slope of the lunar-basalt isochron.

What is more important, however, is the determination of $\mathbf{E}$, the covarience matrix for the least-squares estimate of $\hat{\mathbf{m}}$. Using Eqn. (3.19), we find

$$\mathbf{E} = \begin{bmatrix} 0.1827 & -0.0105 \\ -0.0105 & 0.0009 \end{bmatrix} \times 10^{-6} \tag{3.24}$$

One standard deviation in the uncertainty of the slope $\alpha = m_1$ is thus 0.43×10^{-3}, or about 0.9%. To translate this uncertainty into the uncertainty of the age of the lunar basalt, we make use of Eqn. (**??**):

$$T = \frac{1}{\lambda} \ln(\alpha + 1) \tag{3.25}$$

Taking the derivative of T with respect to α leads to the expression

$$dT = \frac{1}{\lambda} \frac{1}{\alpha + 1} d\alpha \tag{3.26}$$

Using $d\alpha = 0.43 \times 10^{-3}$ we find that $dT = 2.9 \times 10^7$ years, which is about 0.9% of the estimated 3.23×10^9 year age of the lunar basalt.

We remark that the estimated uncertainty (0.06×10^9 years at the 2σ confidence level) derived here is substantially smaller than the 0.11×10^9 years estimated by Nyquist *et al.* [1979]. This discrepancy stems from the fact that Nyquist *et al.* used a more sophisticated method to derive the isochron which accounted for uncertainty in the $^{87}Rb/^{86}Sr$ data which appear in the matrix **A**. To this point, we have disregarded the effects of data uncertainty in **A** for the purpose of illustrating the most simple linear least-squares method. We shall investigate the effects of data uncertainty in **A** in Chapter (5).

3.4 Data Independence

Another question associated with data uncertainty in over-determined least-squares problems concerns the subject of data independence. Suppose the researchers who measured the five elements of $\mathbf{d}^o$ were given funds to make a re-measurement of only one of the components of $\mathbf{d}^o$. Which one(s) would they remeasure? Logically, the researchers would like to remeasure the component of $\mathbf{d}^o$ that is most influential in determining $\hat{\mathbf{m}}$. Our goal here is to develop a means to make this determination, *i.e.*, to identify the data which is most faithfully represented by $\hat{\mathbf{m}}$.

We proceed by determining a matrix **D** referred to as the data independence matrix. Suppose we invert a particular set of data $\mathbf{d}^o$ for $\hat{\mathbf{m}}$, and then

operate on $\hat{\mathbf{m}}$ with the matrix $\mathbf{A}$ to form a "retrodiction"of the data $\mathbf{d}^r$:

$$\mathbf{d}^r = \mathbf{A}\hat{\mathbf{m}} = \mathbf{A}\left[\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}'\mathbf{Q}^{-1}\mathbf{d}^o = \mathbf{D}\mathbf{d}^o \tag{3.27}$$

where

$$\mathbf{D} = \mathbf{A}\left[\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}'\mathbf{Q}^{-1} \tag{3.28}$$

In a perfect world, where $\mathbf{d}^r$ closely resembled $\mathbf{d}^o$, the matrix $\mathbf{D}$ would closely resemble the identity matrix. In most circumstances, however $\mathbf{D}$ will have numerous off-diagonal elements that are different from zero. Rows of $\mathbf{D}$ which have strongest diagonal dominance suggest that the corresponding element of the data vector will be most faithfully preserved through the inversion process. Elements of the data vector which correspond to the most diagonally dominant rows of $\mathbf{D}$ are therefore most influential in determining $\hat{\mathbf{m}}$, and are thus prime candidates for remeasurement. Conversely, elements of the data vector which correspond to the least diagonally dominant rows of $\mathbf{D}$ have little significance in determining $\hat{\mathbf{m}}$. It would not be desireable to remeasure these components.

One of the crucial advantages of the data-independence matrix is that it depends only on the matrices $\mathbf{A}$ and $\mathbf{Q}$. Thus, it can be calculated in advance of making the initial measurements of $\mathbf{d}^o$. This point should be kept in mind for circumstances where you have to defend a particular measurement strategy in advance of actually making the measurements. If you know how you intend to process your data in advance, *i.e.,* you know the matrices $\mathbf{A}$ and $\mathbf{Q}$ in advance, then you can compute $\mathbf{D}$ and use it to argue for efforts to concentrate on making better measurements of the most influential elements of the data.

3.4.1 Example: Data Independence Matrix for the Lunar Basalt Problem

The data-independence matrix for the lunar basalt isochron described in §(3.3.3) can be determined readily using Eqn. (3.28):

$$\mathbf{D} = \begin{bmatrix} 0.3212 & .03204 & 0.3469 & 0.0116 \\ 0.4119 & 0.4613 & 0.3714 & -0.2447 \\ 0.2478 & 0.2063 & 0.3270 & 0.2189 \\ 0.0099 & -0.1631 & 0.2627 & 0.8905 \end{bmatrix} \tag{3.29}$$

The second and fourth rows are the most diagonally dominant rows of $\mathbf{D}$; thus, the data corresponding to the second and fourth components of $\mathbf{d}^o$ are most influential in determining the least-squares solution $\hat{\mathbf{m}}$. Reference to Eqn. (**??**) and Fig. (4) of the previous chapter indicates that the second and fourth components of $\mathbf{d}^o$ are the *outlying* points through which the isochron must pass. It makes intuitive sense that this should be true. Data points clustered in the middle of the overall spread of data points do little to determine the slope of the line (as is suggested by the lack of diagonal dominance in the first and third rows of $\mathbf{D}$). The data points clustered near the two extremes of the data range have more influence in determining the slope, and are thus of greater interest in measurement efforts. If Nyquist *et al.* were to wish to remeasure one or two of the mineral-separate $^{87}Sr/^{86}Sr$ values which comprise the components of $\mathbf{d}^o$ in an effort to improve overall accuracy of the resulting age of the lunar basalt, the data-independence matrix $\mathbf{D}$ computed above suggests that the plagioclase (second) and ilmenite (fourth) mineral grains would be the best candidates for remeasurement.

3.5 Uncertainty in Underdetermined Inverse Problems

In the previous sections of this chapter, we have focussed on the description of uncertainty in *overdetermined* inverse problems. Here, we turn our attention to the question of uncertainty in *underdetermined* inverse problems,

such as that discussed in Chapter (**??**). Uncertainty in the solution of underdetermined problems arises from two sources: error in $\mathbf{d}$, and uncertainty in the additional constraints imposed to select the solution from the multitude of possible solutions which satisfy the data. In other words, for the problem

$$\mathbf{Am} = \mathbf{d} \tag{3.30}$$

in which $\mathbf{m} \in \mathcal{R}^N$, $\mathbf{d} \in \mathcal{R}^M$, $\mathbf{A} : \mathcal{R}^N \rightarrow \mathcal{R}^M$, $N > M$, and where a "hunch" $\mathbf{m}_h$ is used to define the norm of $\mathbf{m}$ to be minimized, error can arise from uncertainty in both $\mathbf{d}$ and $\mathbf{m}_h$. Here we assume that the covarience of errors in the data, $\mathbf{Q}$, and of the "hunch"$\mathbf{m}_h$, $\mathbf{S}$, are known in advance.

In the circumstance when both $\mathbf{d}$ and $\mathbf{m}_h$ are known to be uncertain, it is not appropriate to apply the minimum-norm solution derived in Chapter (**??**), because the constraint imposed by the data, and represented by Eqn. (3.20), is imposed as a "hard"constraint (*i.e.,* is satisfied exactly). What is preferable, is to find a solution which *balances* the satisfaction of the constraint imposed by the data off against the requirement that the solution be close to the "hunch" $\mathbf{m}_h$. To achieve such a solution, the following performance index must be minimized:

$$J = [\mathbf{m} - \mathbf{m}_h]' \mathbf{S}^{-1} [\mathbf{m} - \mathbf{m}_h] + [\mathbf{Am} - \mathbf{d}]' \mathbf{Q}^{-1} [\mathbf{Am} - \mathbf{d}] \tag{3.31}$$

The solution $\hat{\mathbf{m}}$ which minimizes J is readily found to be

$$\hat{\mathbf{m}} = \left[\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A} + \mathbf{S}^{-1}\right]^{-1} \left(\mathbf{S}^{-1}\mathbf{m}_h + \mathbf{A}'\mathbf{Q}^{-1}\mathbf{d}\right) \tag{3.32}$$

The covarience of error $\mathbf{E}$ associated with the above estimate $\hat{\mathbf{m}}$ is readily shown to be

$$\begin{aligned} \mathbf{E} = & \quad [\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A} + \mathbf{S}^{-1}]^{-1} \mathbf{S}^{-1} \left[[\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A} + \mathbf{S}^{-1}]^{-1}\right]' \\ + & \quad [\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A} + \mathbf{S}^{-1}]^{-1} \mathbf{A}'\mathbf{Q}^{-1}\mathbf{A} \left[[\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A} + \mathbf{S}^{-1}]^{-1}\right]' \end{aligned} \tag{3.33}$$

3.6 Model Resolution

A crucial question associated with underdetermined inverse problems is the fact that the inherent non-uniqueness of the solution implies that it is impossible to completely *resolve* the parameters associated with the solution.

Researchers faced with the prospects of solving an underdetermined inverse problem such as those described in the previous section or in Chapter (**??**) might wish to determine *in advance* of their data-collection effort what aspects of the desired solution $\mathbf{m}$ will be well resolved, or well-determined, by the data. To make this determination, it is necessary to construct the *model resolution* matrix $\mathbf{R}$.

Following along the same lines of reasoning as used to develop the data independence matrix $\mathbf{D}$ for overdetermined problems, we consider the degradation of a known "exact" model $\mathbf{m}^o$ when it is used to create data that is then inverted to determine a "retrodicted" model $\mathbf{m}^r$. Using the minimum-norm inverse developed in Chapter (**??**), we find

$$\mathbf{m}^r = \mathbf{A}' \left[\mathbf{A}\mathbf{A}'\right]^{-1} \mathbf{A}\mathbf{m}^o \tag{3.34}$$

Defining the model resolution matrix $\mathbf{R} = \mathbf{A}' \left[\mathbf{A}\mathbf{A}'\right]^{-1} \mathbf{A}$, our interest is in the question of how closely $\mathbf{R}$ resembles the identity matrix $\mathbf{I}$. For overdetermined problems, $\mathbf{R} = \mathbf{I}$. For underdetermined $\mathbf{R} \neq \mathbf{I}$.

As with the data-independence matrix, the model-resolution matrix depends only on $\mathbf{A}$ (and on $\mathbf{S}$ and $\mathbf{Q}$ if known) which, in turn, depends only on the physics of the problem. It is thus possible, and profitable, to determine $\mathbf{R}$ in advance of any data collection to evaluate how useful the data collection exercise will ultimately be in resolving the unknown model parameters.

Example: The Lunar-Basalt Problem with One Data Point

We revisit the truncated lunar-basalt isochron problem discussed in Chapter (**??**) to derive the model-resolution matrix associated with the determination of the slope and intercept of the isochron. Using MATLAB® we find:

```
R=A'*inv(A*A')*A

>>R =

0.0009    0.0296
0.0296    0.9991
```

Clearly, the intercept β is most accurately resolved, whereas the slope α is least well resolved.

3.7 Bibliography

Nyquist, L. E., C.-Y. Shih, J. L. Wooden, B. M. Bansal, and H. Wisemann, 1979. The Sr and Nd isotopic record of Apollo 12 basalts: Implications for lunar geochemical evolution. *Proc. 10th Lunar Planet. Sci. Conf.*, 77-114.

Chapter 4

Singular Value Decomposition: Geometric Interpretation

4.1 Introduction

The goal of this chapter is to lay a conceptual groundwork for the powerful tool in linear algebra known as the *singular value decomposition* (SVD) of a matrix. Our interest in the SVD stems from the fact that it has has become the main tool currently used to solve overdetermined, underdetermined and mixed linear inverse problems. We will begin with a review of the linear algebra of square, symmetric matrices, and seek a practical, geometric understanding ofeigenvalues and eigenvectors. The concepts will then be generalized to non-symmetric, rectangular matrices such as those commonly encountered in overdetermined and underdetermined inverse problems in the geophysical sciences.

4.2 Geometrical Interpretations of Linear Operators

Linear algebra is the study of linear operators which map vectors between two vector spaces. Matrices are convienent representations linear operators, although much of linear algebra (including the notion of SVD) was developed before matrices were used to represent linear operators. (A nice historical perspective on the development of SVD during the 19'th century is available in Stewart [1993]. Stewart reviews the mathematical works of Gauss (-) Beltrami (1835 - 1899), Jordan (1838-1921) and Sylvester (1814 - 1897) that were instrumental in leading to our modern-day view of linear algebra.) While much of linear algebra can be developed and understood in purely abstract terms, we shall develop here a concrete, geometrical interpretation of linear algebra in an effort to gain greater insight into the mathematical nature of the inverse problems in which we are interested, and into the workings of the SVD.

We consider a linear operator represented by the matrix $\mathbf{A}$ which maps vectors $\mathbf{v} \in \mathcal{R}^N$ into vectors $\mathbf{u} \in \mathcal{R}^M$. We denote $\mathcal{R}^N$ the *domain* of $\mathbf{A}$ and $\mathcal{R}^M$ the *range* of $\mathbf{A}$. When $N = M$, the matrix $\mathbf{A}$ is square (*i.e.,* it has the same number of rows and columns, and is referred to as an $N \times N$ matrix).

4.2.1 Mappings of the Unit Sphere

A useful way to think about what linear operators do is developed by considering how the image of a sphere of unit radius in the domain of $\mathbf{A}$ is deformed as a result of its mapping into the range of $\mathbf{A}$. For the time being, we restrict our attention to square matrices $\mathbf{A} : \mathcal{R}^N \rightarrow \mathcal{R}^N$. First, let's define what is meant by a sphere of unit radius in $\mathcal{R}^N$:

Definition: The unit sphere in $\mathcal{R}^N$ is the set of all vectors $\mathbf{v}$ which satisfy $\|\mathbf{v}\| = 1$ where $\|\mathbf{v}\| = \sqrt{\mathbf{v}'\mathbf{v}} = \sqrt{\sum_{i=1}^{N} v_i^2}$

An example of how the image of a unit sphere in $\mathcal{R}^2$ is distorted by a linear

operator is seen in Fig. (4.1) which displays the result for the square matrix

$$\mathbf{A} = \begin{pmatrix} 0.0492 & 1 \\ 0.1127 & 1 \end{pmatrix} \tag{4.1}$$

(This is the square matrix derived from the radiometric datating problem defined in Chapter (1) where only the isotopic ratios of two mineral separates, Px and Ilm, are considered.) Notice that the figure of the unit sphere (really a unit circle, since this example involves a mapping $\mathbf{A} : \mathcal{R}^2 \rightarrow \mathcal{R}^2$) is distorted into an elliptical shape. (It can be proven that the image of the circle after mapping is an elipse because of the fact that $\mathbf{v}'\mathbf{A}'\mathbf{A}\mathbf{v} = c$, where c is a constant chosen to ensure that $\mathbf{v}'\mathbf{v} = 1$, is a quadratic function of the components of $\mathbf{v}$. This quadratic function formally corresponds to the equation of an ellipse or elliptical surface.)

The geometrical viewpoint suggested by Fig. (4.1) suggests that two special categories of linear operators can be defined:

Catagory A (pure strain). The mapping $\mathbf{A} : \mathcal{R}^N \rightarrow \mathcal{R}^N$ maps the unit sphere into an ellipse. There exist at least $2N$ vectors of unit length (in the set comprising the unit sphere), denoted by $\{\mathbf{e}_i\}_{i=1}^N$ and $\{-\mathbf{e}_i\}_{i=1}^N$, which suffer a change in length but are not rotated, i.e., $\mathbf{A}\mathbf{e}_i = \lambda_i \mathbf{e}_i$ where λ_i is a positive real number. (In continuum mechanics, one would regard a mapping in this category as a representation of a pure strain.) We remark that the vectors $\{\mathbf{e}_i\}_{i=1}^N$ are mutually orthogonal, *i.e.,* $\mathbf{e}_i'\mathbf{e}_j = 0$ if $i \neq j$. Figure (4.2) displays such a case.

Category B (pure rotation). The mapping $\mathbf{A} : \mathcal{R}^N \rightarrow \mathcal{R}^N$ maps the unit sphere to itself (no stretching or shrinking of the length of any vector) with all vectors being rotated about an axis of rotation. Figure (4.3) displays and example of this case.

Having asserted that the above two categories describe the "end-member" distortions brought about by a general linear mapping of the unit sphere, we develop means by which the two categories can be recognized by simple inspection of the matrix $\mathbf{A}$. We prove below that $\mathbf{A}$ belongs to category A if $\mathbf{A}$ is a *symmetric* matrix (*i.e.,* $\mathbf{A}' = \mathbf{A}$). Likewise, $\mathbf{A}$ belongs to category B if $\mathbf{A}$ is

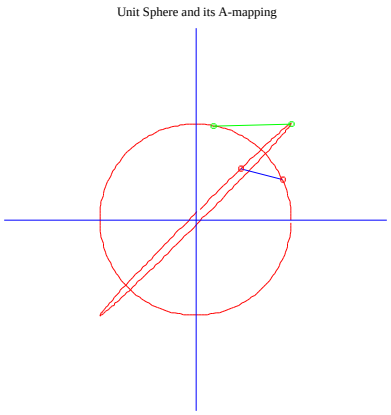

Figure 4.1: A mapping of the unit sphere in $\mathcal{R}^2$ into its image in $\mathcal{R}^2$ by the linear operator represented by the matrix $\mathbf{A} : \mathcal{R}^2 \rightarrow \mathcal{R}^2$. The original position of a point on the unit sphere, and its image after the mapping is shown in two examples by open circles connected by line segments.

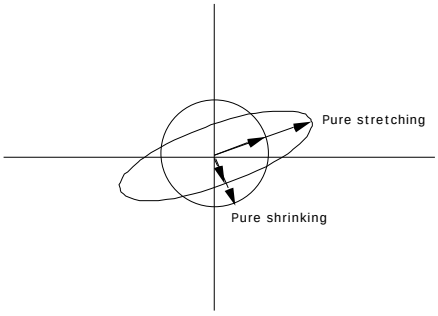

Figure 4.2: A mapping from $\mathcal{R}^2 \rightarrow \mathcal{R}^2$ which does not rotate. (Note, the two pairs of vectors aligned with the principal axes of the ellipse do not rotate. All other vectors will rotate.

a *unitary* matrix, (*i.e.,* its columns are a set of orthonormal column vectors of unit length. (We remind the reader that the length of a vector $\mathbf{v}$ in $\mathcal{R}^N$ to be $\|\mathbf{v}\| = \sqrt{\sum_{i=1}^{N} v_i^2}$, and that a set of vectors $\{\mathbf{v}_i\}$ is orthonormal if $\mathbf{v}_i'\mathbf{v}_j = \delta_{ij}$.)

Symmetric Matrices Belong to Category A

We expect the ellipse which is the image of the unit sphere after mapping to have principal axes which are orthogonal and complete. (The term *complete* referrs to the fact that the principal axes could serve as a coordinate system for the vector space in which the ellipse resides.) By definition, category A mappings must imply that there are N vectors $\{\mathbf{v}_i, i = 1, N\}$ in $\mathcal{R}^N$ which originally lie on the unit sphere, *i.e.,* $\mathbf{v}_j{}'\mathbf{v}_i = \delta_{ji}$ and which become the principal axes of the ellipse,

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i \tag{4.2}$$

Here the scalar coefficients $\{\lambda_i, i = 1, N\}$ are the factors which determines how much the $\{\mathbf{v}_i, i = 1, N\}$ are stretched or shrunk to reach the principal axes of the elliptical image of the unit sphere. (When we discuss the unit sphere and its elliptical image, we must recongnize that these geometrical figures are abstract if the vector spaces in which they reside have dimensionality greater than 3.)

Let's multiply Eqn. (4.2) by $\mathbf{v}'_j$:

$$\mathbf{v}'_j\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}'_j\mathbf{v}_i \tag{4.3}$$

Next, let's do the same thing, but switch the indices i and j:

$$\mathbf{v}'_i\mathbf{A}\mathbf{v}_j = \lambda_j \mathbf{v}'_j\mathbf{v}_i \tag{4.4}$$

In the above expression, we have made use of the commutaion of the vector product $\mathbf{v}'_i\mathbf{v}_j = \mathbf{v}'_j\mathbf{v}_i$. We can also write the left-hand side of Eqn. (4.4) in a manner which swaps the order of the appearance of the vectors $\mathbf{v}_i$ and $\mathbf{v}_j$:

$$\begin{aligned} \mathbf{v}'_i\mathbf{A}\mathbf{v}_j &= (\mathbf{A}\mathbf{v}_j)'\mathbf{v}_i \\ &= \mathbf{v}'_j\mathbf{A}'\mathbf{v}_i \end{aligned} \tag{4.5}$$

(Recall the commutation rules associated with taking the transpose of vectors and matrices: if $\mathbf{a}$ and $\mathbf{b}$ are two vectors and $\mathbf{L}$ is a matrix, then $\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a}$ and $(\mathbf{La})' = \mathbf{a}'\mathbf{L}'$.) If we subtract Eqn. (2.0.4) from Eqn. (2.0.3), and make use of Eqn. (2.0.5) we get

$$\mathbf{v}'_j(\mathbf{A} - \mathbf{A}')\mathbf{v}_i = (\lambda_i - \lambda_j)\mathbf{v}'_j\mathbf{v}_i \tag{4.6}$$

We have already asserted (using our notion of the geometry of an ellipse) that the vectors which align with the principal axes of the ellipse are orthonormal. In addition, when $i = j$ the factor $(\lambda_i - \lambda_j) = 0$. Thus, Eqn. (4.6) implies

$$\mathbf{v}'_i(\mathbf{A} - \mathbf{A}')\mathbf{v}_j = 0 \tag{4.7}$$

for all i and j. Equation (4.7) states that the vector $(\mathbf{A} - \mathbf{A}')\mathbf{v}_j$ is orthogonal to *all* of vectors $\{\mathbf{v}_i, i = 1, N\}$. Recall that the set $\{\mathbf{v}_i, i = 1, N\}$ is complete (one would say that the vectors $\{\mathbf{v}_i, i = 1, N\}$ completely *span* the vector space in which they reside). The only vector which can be orthogonal to all the $\{\mathbf{v}_i, i = 1, N\}$ is the zero vector (a vector having zeros for each and every component). Thus if $(\mathbf{A} - \mathbf{A}')\mathbf{v}_j = \mathbf{0}$, we must conclude that $\mathbf{A}' = \mathbf{A}$ (or $A_{ij} = A_{ji}$). The matrix $\mathbf{A}$ must therefore be symmetric. ■

Unitary Matrices Belong to Category B

Pure rotations *preserve* the angle between two vectors and do not change their length. Thus we can consider a set of orthonormal vectors $\{\mathbf{v}_i, i = 1, N\}$ which span $\mathcal{R}^N$ and assert that their image once transformed by the linear operator $\mathbf{A}$ will also be orthonormal and will span $\mathcal{R}^N$. We denote the image of $\{\mathbf{v}_i, i = 1, N\}$ under the transformation $\mathbf{A}$ by $\{\mathbf{u}_i, i = 1, N\}$. By definition

$$\mathbf{u}_j = \mathbf{A}\mathbf{v}_j \tag{4.8}$$

and

$$\begin{aligned} \mathbf{u}'_k\mathbf{u}_j = \delta_{kj} \quad &= (\mathbf{A}\mathbf{v}_k)'(\mathbf{A}\mathbf{v}_j) \\ &= \mathbf{v}'_k(\mathbf{A}'\mathbf{A})\mathbf{v}_j \end{aligned} \tag{4.9}$$

but $\mathbf{v}'_k\mathbf{v}_j = \mathbf{u}'_k\mathbf{u}_j = \delta_{kj}$, so

$$\mathbf{v}'_k\mathbf{v}_j = \mathbf{v}'_k(\mathbf{A}'\mathbf{A})\mathbf{v}_j \tag{4.10}$$

or

$$\mathbf{v}_j = (\mathbf{A}'\mathbf{A})\mathbf{v}_j \tag{4.11}$$

Thus,

$$\mathbf{A}'\mathbf{A} = \mathbf{I} \tag{4.12}$$

Note that Eqn. (4.12) implies that the columns of $\mathbf{A}$ are composed of orthonormal basis vectors (each column is a complete, separate basis vector). ■

4.3 Eigenvalues and Eigenvectors of a Symmetric Matrix

If $\mathbf{A}$ is a symmetric linear operator (matrix) it belongs to category A. The geometry of the ellipse, and Eqn. (4.2), suggest that there is a set of *eigenvectors*, $\{\mathbf{v}_i, i = 1, \ldots, N\}$, which belong to $\mathcal{R}^N$ and *eigenvalues*, $\{\lambda_i, i = 1, \ldots, N\}$, which can be associated with the matrix $\mathbf{A}$. How do we determine these eigenvectors and eigenvalues? There are many ways to perform such a determination. In this chapter, we adopt a relatively inefficient, but intuitive, technique that is based on our geometric picture of the unit sphere and its distorted image. We will also show that the set of eigenvectors $\{\mathbf{v}_i, i = 1, N\}$ are mutually orthogonal (this confirms our geometric notion that the principal axes of an ellipse are indeed perpendicular) and that the eigenvalues $\{\lambda_i, i = 1, \ldots, N\}$ are all positive.

An appealing way to find the principal axes of an ellipse is to recognize that the principal axes of the ellipse represent vectors in which the distance between the ellipse and the origin is extremized (maximized or minimized in some sense). We can determine the principal axes of an N-dimensional *hyperellipse*, for example, by finding the extrema of the following measure of the length, J, of a vector $\mathbf{u} = \mathbf{A}\mathbf{v}$ which lies on the elliptical image of the unit sphere. In other words, we seek the extrema of J, where

$$J = (\mathbf{A}\mathbf{v})'(\mathbf{A}\mathbf{v}) \tag{4.13}$$

subject to the constraint that

$$\mathbf{v}'\mathbf{v} = 1 \tag{4.14}$$

(*i.e.,* $\mathbf{v}$ lies on the unit sphere).

We can enforce the constraint represented by Eqn. (4.14) by using a *Lagrange multiplier* μ:

$$\tilde{J} = (\mathbf{A}\mathbf{v})'(\mathbf{A}\mathbf{v}) + \mu(1 - \mathbf{v}'\mathbf{v}) \tag{4.15}$$

The *Euler-Lagrange conditions* for the extremization of $\tilde{J}$ are generated by differentiating $\tilde{J}$ with respect to each *component* of $\mathbf{v}$ and with respect to μ, the unknown Lagrange multiplier.

$$\frac{\partial \tilde{J}}{\partial \mu} = 1 - \mathbf{v}'\mathbf{v} = 0 \tag{4.16}$$

$$\frac{\partial \tilde{J}}{\partial v_k} = 2(A_{ik}A_{ij}v_j - \mu v_k) = 0 \tag{4.17}$$

here, v_k denotes the kth component of $\mathbf{v}$, and A_{ij} denotes the ijth component of $\mathbf{A}$.

To satisfy Eqn. (4.17) we must solve

$$\mathbf{A}'\mathbf{A}\mathbf{v} - \mu\mathbf{v} = \mathbf{0} \tag{4.18}$$

or

$$(\mathbf{A}'\mathbf{A} - \mu\mathbf{I})\mathbf{v} = \mathbf{0} \tag{4.19}$$

In general, there is no way to solve Eqn. (4.19) with a *non-trivial* (non-zero) vector $\mathbf{v} \neq \mathbf{0}$ unless the matrix $\mathbf{A}'\mathbf{A} - \mu\mathbf{I}$ is *defective* in some sense. In particular, Eqn. (4.19) can be satisfied for a non-zero $\mathbf{v}$ when

$$\det(\mathbf{A}'\mathbf{A} - \mu\mathbf{I}) = 0 \tag{4.20}$$

The left-hand side of Eqn. (4.20) expresses a polynomial of degree N in the variable μ, thus Eqn. (4.20) expresses the condition that μ is a root of the *characteristic* polynomial of the matrix $\mathbf{A}'\mathbf{A}$. Note that μ is guaranteed to be

a real positive number because $\mathbf{A}'\mathbf{A}$ is a symmetric, square matrix. You can prove this for yourself. These roots are called the eigenvalues of the operator $\mathbf{A}'\mathbf{A}$.

Assuming that the N roots of the characteristic polynomial (4.20), $\{\mu_l, l = 1, N\}$, are determined, the $\mathbf{v}_l$'s can then be determined by solving Eqn. (4.19). This entails performing the LU-decomposition of $(\mathbf{A}'\mathbf{A} - \mu\mathbf{I})$ which, as we have ensured by choosing μ in such a manner that $\det(\mathbf{A}'\mathbf{A} - \mu\mathbf{I}) = 0$, is not possible.

We get around this problem by substituting $\mathbf{v}'$ for the lth row of the matrix $(\mathbf{A}'\mathbf{A} - \mu\mathbf{I})$ and replace the lth zero component on the right-hand side of Eqn. (4.19) with 1. This substitution enforces the constraint that $\mathbf{v}'\mathbf{v} = 1$. Since there are N roots to apply this substitution, there will be N vectors produced by this proceedure. These N vectors, $\{\mathbf{v}_l, l = 1, N\}$, are referred to as the *eigenvectors* of the operator $\mathbf{A}'\mathbf{A}$.

Notice that nothing so far has depended on the assumption that $\mathbf{A}$ is symmetric. We shall next make use of this assumption to show that $\mathbf{A}$ can have eigenvalues and eigenvectors. By assumption, $\mathbf{A}$ is symmetric ($\mathbf{A}' = \mathbf{A}$), thus

$$\begin{aligned} \det(\mathbf{A}'\mathbf{A} - \mu\mathbf{I}) &= \det(\mathbf{A}^2 - \mu\mathbf{I}) \\ &= \det\left(\left[\mathbf{A} + \sqrt{\mu}\mathbf{I}\right]\ \left[\mathbf{A} - \sqrt{\mu}\mathbf{I}\right]\right) \\ &= \det(\mathbf{A} + \sqrt{\mu}\mathbf{I})\ \det(\mathbf{A} - \sqrt{\mu}\mathbf{I}) \\ &= 0 \end{aligned} \tag{4.21}$$

where we have made use of the fact that the determinant is a linear operator on matrices.

To satisfy Eqn. (4.21), either $\det(\mathbf{A} + \sqrt{\mu}\mathbf{I}) = 0$ or $\det(\mathbf{A} - \sqrt{\mu}\mathbf{I}) = 0$. If we choose one convention, say $\det(\mathbf{A} - \sqrt{\mu}\mathbf{I}) = 0$, then we recognize that $\lambda_i = \sqrt{\mu_i}$, for $i = 1, \ldots, N$ are roots of a characteristic polynomial for $\mathbf{A}$. Note that we must demand that the $\mu_i > 0$, $i = 1, \ldots, N$ in order for the square-root to yield a *real*-valued λ_i. We are assured that the $\mu_i > 0$ due to the fact that the characteristic polynomial associated with the symmetric matrix $\mathbf{A}'\mathbf{A}$ does indeed have only positive real roots. (NEED TO PROVE THIS. Refer to a text on linear algebra for a proof of this assertion.)

Following the method described above, we can choose a set of eigenvectors $\{\mathbf{u}_i, i = 1, N\}$ such that $\mathbf{u'}_i\mathbf{u}_j = \delta_{ij}$ and

$$\mathbf{A}\mathbf{u}_i = \lambda_i\mathbf{u}_i \tag{4.22}$$

If we multiply Eqn. (4.22) by $\mathbf{A}$ we see (using the symmetry of $\mathbf{A}$) that

$$\mathbf{A}\mathbf{A}\mathbf{u}_i - \lambda_i\mathbf{A}\mathbf{u}_i = \mathbf{A'}\mathbf{A}\mathbf{u}_i - \lambda_i^2\mathbf{u}_i = \mathbf{A'}\mathbf{A}\mathbf{u}_i - \mu_i\mathbf{u}_i \tag{4.23}$$

which tells us that $\mathbf{u}_i = \mathbf{v}_i$.

Orthagonality of Eigenvectors

We can also affirm the assumption made previously that the vectors $\{\mathbf{v}_i$, $i = 1, \ldots, N\}$ are mutually orthogonal. (This previous assumption followed from our geometric intuition that the principal axes of an ellipse are mutually orthogonal.) To demonstrate this, take the dot-product between two members of the set $\mathbf{v}_i$:

$$\begin{aligned}
\mathbf{v'}_i\mathbf{v}_k &= \tfrac{(\mathbf{A}\mathbf{v}_i)'\mathbf{A}\mathbf{v}_k}{\lambda_i\lambda_k} \\
&= \tfrac{\mathbf{v}_i'\mathbf{A'}\mathbf{A}\mathbf{v}_k}{\lambda_i\lambda_k} \\
&= \frac{\mu_k}{\lambda_i\lambda_k}\mathbf{v}_i'\mathbf{v}_k \\
&= \frac{\lambda_k}{\lambda_i}\mathbf{v}_i'\mathbf{v}_k
\end{aligned} \tag{4.24}$$

Note that we have made use of familiar commutation rule: $(\mathbf{A}\mathbf{v})' = \mathbf{v}'\mathbf{A}'$. For Eqn. (4.24) to hold true when $i \neq k$, either $\mathbf{v}_i'\mathbf{v}_k = 0$ or $\lambda_k/\lambda_i = 1$. If $\lambda_k/\lambda_i \neq 1$, Eqn. (4.24) implies that $\mathbf{v}_i$ and $\mathbf{v}_k$ are orthogonal. If $\lambda_k/\lambda_i = 1$, we say that the matrix $\mathbf{A}$ has degenerate eigenvalues and eigenvectors. In this circumstance, we may select two vectors, $\mathbf{v}_i \neq 0$ and $\mathbf{v}_k \neq 0$, which are orthogonal and satisfy Eqn. (4.22).

4.4 SVD of General, Nonsymmetric, Square Matrices

What happens when the matrix $\mathbf{A} : \mathcal{R}^N \rightarrow \mathcal{R}^N$ is not symmetric, *i.e.,* does not conform precisely to category A defined above? Consideration of this question leads to the concept of the singular-value decomposition (SVD).

We begin by noting the fact that the unit sphere is mapped to an ellipsoid regardless of the fact that $\mathbf{A}$, by assumption, is not symmetric. We may thus consider the conditions which define the extrema of J defined in Eqn (4.15). Following the same steps as followed previously for the case of symmetric matrices we come to Eqn. (4.19). The existance of solutions of Eqn. (4.19) implies that there are two sets of orthonormal vectors, $\{\mathbf{v}_i\}_{i=1}^N$ and $\{\mathbf{u}_i\}_{i=1}^N$, such that

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{v}_i \tag{4.25}$$

$$\mathbf{A}'\mathbf{v}_i = \lambda_i \mathbf{u}_i \tag{4.26}$$

and

$$\lambda_i^2 = \mu_i \tag{4.27}$$

for $i = 1, \ldots, N$. We refer to the vectors $\{\mathbf{v}_i\}_{i=1}^N$ and $\{\mathbf{u}_i\}_{i=1}^N$, and the set of scalar quantities $\{\lambda_i\}_{i=1}^N$ as the SVD of the matrix $\mathbf{A}$. Equations (4.25) - (4.27) suggest a convienent notation. If we define the matrix $\mathbf{V}$ such that it's i'th column is the vector $\mathbf{v}_i$, and likewise define the matrix $\mathbf{U}$ such that it's i'th column is the vector $\mathbf{u}_i$, then

$$\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Lambda} \tag{4.28}$$

where $\mathbf{\Lambda}$ is a matrix which has the λ_i's on its diagonal and zeros everywhere else. Noting the fact that $\mathbf{V}\mathbf{V}' = \mathbf{I}$, we can express $\mathbf{A}$ in a canonical form which will be referred to as the SVD:

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}' \tag{4.29}$$

Example

Let's compute the $\{\mu_i, i = 1, 2\}$, $\{\lambda_i, i = 1, 2\}$, $\{\mathbf{v}_i, i = 1, 2\}$, and the $\{\mathbf{u}_i, i = 1, 2\}$ for the matrix given in Eqn. (4.1). This is a non-symmetric matrix, so we cannot expect $\mathbf{v}_i = \mathbf{u}_i$.

First observe that, according to Eqns. (4.25)-(4.27),

$$\mathbf{A}'\mathbf{A}\mathbf{v}_i - \lambda_i^2 \mathbf{v}_i = \mathbf{0} \tag{4.30}$$

and

$$\mathbf{A}\mathbf{A}'\mathbf{u}_i - \lambda_i^2 \mathbf{u}_i = \mathbf{0} \tag{4.31}$$

This suggests that we can use the MATLAB® -native eigenvalue and eigenvector routines to determine the $\{\mathbf{v}_i, i = 1, 2\}$ and $\{\mathbf{u}_i, i = 1, 2\}$:

```
>> [V,D]=eig(A'*A)
```

```
   V =
```

$$\begin{matrix} 0.9967 & 0.0808 \\ -0.0808 & 0.9967 \end{matrix} \tag{4.32}$$

```
   D =
```

$$\begin{matrix} 0.0020 & 0 \\ 0 & 2.0131 \end{matrix}$$

```
>> [U,D]=eig(A*A')
```

```
   U =
```

$$\begin{matrix} 0.7089 & 0.7053 \\ -0.7053 & 0.7089 \end{matrix}$$

```
   D =
```

$$\begin{matrix} 0.0020 & 0 \\ 0 & 2.0131 \end{matrix}$$

The matrix $\mathbf{D}$ above contains the eigenvalues (μ_i) on its diagonal. Thus, the values of λ_i can be found by taking the square root of the components of $\mathbf{D}$. The values of $\{\mathbf{v}_i, i = 1, 2\}$ and $\{\mathbf{u}_i, i = 1, 2\}$ are to be found in the

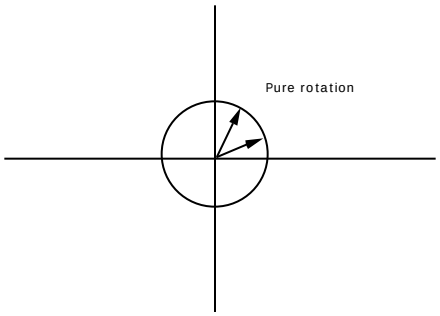

Figure 4.3: All vectors in this mapping from $\mathcal{R}^2 \rightarrow \mathcal{R}^2$ are rotated in the same direction by the same angle of rotation. No stretching or shrinking occurs.

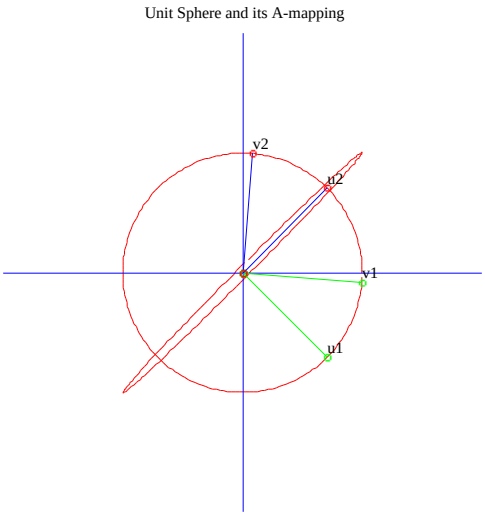

Figure 4.4: The $\{\mathbf{v}_i, i = 1, 2\}$ and $\{\mathbf{u}_i, i = 1, 2\}$ for the matrix $\mathbf{A}$ expressed in Eqn. (4.1).

columns of $\mathbf{V}$ and $\mathbf{U}$. Figure (4.4) displays the location of the $\{\mathbf{v}_i, i = 1, 2\}$ and $\{\mathbf{u}_i, i = 1, 2\}$.

Notice that both the $\{\mathbf{v}_i, i = 1, 2\}$ and $\{\mathbf{u}_i, i = 1, 2\}$ given above are orthonormal basis vectors. They are not perpendicular to each other, however. Can you take this exercise further by verifying that Eqns. (4.25) and (4.26) are true?

Let's determine the SVD of the matrix $\mathbf{A}$ given by Eqn. (4.1). The MATLAB® -native SVD routine does this job easily:

```
>> [U,S,V]=svd(A)  U =
```

$$\begin{matrix} 0.7053 & 0.7089 \\ 0.7089 & -0.7053 \end{matrix}$$

```
S =
```

$$\begin{matrix} 1.4188 & 0 \\ 0 & 0.0448 \end{matrix}$$

```
V =
```

$$\begin{matrix} 0.0808 & -0.9967 \\ 0.9967 & 0.808 \end{matrix}$$

One can compare the columns of $\mathbf{U}$ and $\mathbf{V}$ with derived in the SVD routine with those derived in the previous section. Also note that $\mathbf{S}= \mathbf{\Lambda}$ is the square-root of $\mathbf{D}$ found in the previous section.

4.5 SVD of Rectangular Matrices

All of the results discussed so far in this chapter have applied to square, $N \times N$, matrices. The SVD also exists when $\mathbf{A}$ is a rectangular, $M \times N$, matrix that maps $\mathcal{R}^N \rightarrow \mathcal{R}^M$.

Following the techniques of the previous section, we can consider the eigenvalues and eigenvectors associated with the two symmetric, square ma-

trices $\mathbf{A}'\mathbf{A} : \mathcal{R}^N \rightarrow \mathcal{R}^N$ and $\mathbf{A}\mathbf{A}' : \mathcal{R}^M \rightarrow \mathcal{R}^M$. In particular there exist $\{\mathbf{u}_i, i = 1, \ldots, N\}$, $\{\mathbf{v}_i, i = 1, \ldots, M\}$, $\{\mu_i, i = 1, \ldots, N\}$, and $\{\gamma_i, i = 1, \ldots, M\}$ such that,

$$(\mathbf{A}'\mathbf{A} - \mu_i\mathbf{I})\mathbf{v}_i = \mathbf{0} \qquad i = 1, \ldots, N \tag{4.33}$$

$$(\mathbf{A}\mathbf{A}' - \gamma_j\mathbf{I})\mathbf{u}_j = \mathbf{0} \qquad j = 1, \ldots, M \tag{4.34}$$

For the sake of argument, and without loss of generality, let's assume that $N < M$ and that $\mathbf{A}$ represents the matrix associated with an overdetermined least-squares problem. In this circumstance, there will be $M - N$ more eigenvalues γ_j associated with Eqn. (4.34) than eigenvalues μ_j associated with Eqn. (4.33). Thus, for a subset $\{\gamma_k, k = 1, \ldots, N\}$ of the eigenvlaues $\{\gamma_k, k = 1, \ldots, M\}$, we can prove that $\gamma_k = \mu_k$. Operating on Eqn. (4.33) with $\mathbf{A}$, we obtain

$$\begin{aligned} \mathbf{0} &= \mathbf{A}\mathbf{A}'\mathbf{A}\mathbf{v}_i - \mu_i\mathbf{A}\mathbf{v}_i \\ &= \mathbf{A}\mathbf{A}'(\mathbf{A}\mathbf{v}_i) - \mu_i(\mathbf{A}\mathbf{v}_i) \\ &= \mathbf{A}\mathbf{A}'\mathbf{u}_k - \mu_i\mathbf{u}_k \end{aligned} \tag{4.35}$$

but $\mathbf{A}\mathbf{A}'\mathbf{u}_k = \gamma_k\mathbf{u}_k$ from Eqn. (4.34), so

$$\mathbf{A}\mathbf{A}'\mathbf{u}_k - \mu_i\mathbf{u}_k = (\gamma_k - \mu_i)\mathbf{u}_k = \mathbf{0} \tag{4.36}$$

The only way for Eqn. (4.36) to be satisfied for non-zero $\mathbf{v}_k$ is for $\gamma_k = \mu_i$. We might as well adopt an indexing convention such that $k = i$, so $\gamma_i = \mu_i$.

Having identified the $\{\mathbf{v}_i, i = 1, \ldots, N\}$, $\{\mathbf{u}_i, i = 1, \ldots, M\}$, and $\{\mu_i = \gamma_i, i = 1, \ldots, N\}$, we can relate the two sets of vectors together in the following manner

$$\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{u}_i \qquad i = 1, \ldots, N \tag{4.37}$$

$$\mathbf{A}'\mathbf{u}_i = \lambda_i\mathbf{v}_i \qquad i = 1, \ldots, N \tag{4.38}$$

$$\mathbf{A}'\mathbf{u}_i = \mathbf{0} \qquad i = N, \ldots, M \tag{4.39}$$

where $\lambda_i = \sqrt{\gamma_i}$, $i = 1, \ldots, N$. One can verify that the relationship between $\mathbf{u}_i$ and $\mathbf{v}_i$ implied by Eqns. (4.37) and (4.38) satisfies the definitions of $\mathbf{u}_i$ and $\mathbf{v}_i$ given in Eqns. (4.33) and (4.34).

Using the above relations, we define the SVD for the rectangular $M \times N$ matrix $\mathbf{A}$ as follows:

$$\mathbf{AV} = \mathbf{U}\mathbf{\Lambda} \tag{4.40}$$

where, as before, the kth columns of $\mathbf{V}$ and $\mathbf{U}$ are composed of the components of the vectors $\{\mathbf{v}_k, k = 1, N\}$ and $\{\mathbf{u}_k, k = 1, M\}$ respectively, and

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \lambda_N \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \tag{4.41}$$

Notice that $\mathbf{V}$ is a square $N \times N$ matrix and $\mathbf{U}$ is a square $M \times M$ matrix. The matrix $\mathbf{\Lambda}$ is a $M \times N$ rectangular matrix that has λ_i's on its 'diagonal'.

From Eqn. (4.40) we can deduce the SVD of the rectangular matrix $\mathbf{A}$:

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}' \tag{4.42}$$

where we have made use of the fact that $\mathbf{VV}' = \mathbf{I}$. Notice the similarity between the above expression and that derived for square matrices in Eqn. (4.29).The beauty of singular value decomposition is that the formula for the decomposition of $\mathbf{A}$ is true for both the underdetermined and overdetermined inverse problems.

4.6 The Moore-Penrose Inverse

The SVD is a powerful tool in the solution of linear inverse problems of all types. Before showing how the SVD can be used to solve least-squares and minimum-norm problems, we pause to describe the Moore-Penrose inverse (MP inverse) of a general linear operator. We shall prove that the SVD provides an easy means to generate the MP inverse.

Consider the matrix equation

$$\mathbf{A}\,\mathbf{m} = \mathbf{d} \tag{4.43}$$

where $\mathbf{A}$ is an $M \times N$ matrix (M rows and N columns), $\mathbf{m} \in \mathcal{R}^N$ is a model vector, and $\mathbf{d} \in \mathcal{R}^M$ is a data vector. Our interest is in determining a model m from data d. If $M < N$, the equations are underdetermined. If $M > N$, the equations are overdetermined. In some circumstances, the nature of the equations may be *mixed*, with some elements of $\mathbf{m}$ being underdetermined, and other elements of $\mathbf{m}$ being overdetermined. The matrix $\mathbf{A}^+$ is said to be the MP inverse of $\mathbf{A}$ if

$$\begin{aligned} \mathbf{A}\mathbf{A}^+\mathbf{A} &= \mathbf{A} & \left[\mathbf{A}\mathbf{A}^+\right]' &= \mathbf{A}\mathbf{A}^+ \\ \mathbf{A}^+\mathbf{A}\mathbf{A}^+ &= \mathbf{A}^+ & \left[\mathbf{A}^+\mathbf{A}\right]' &= \mathbf{A}^+\mathbf{A} \end{aligned} \tag{4.44}$$

The MP inverse of the matrix $\mathbf{A}$ in Eqn. (4.43) is easily expressed in terms of the SVD. Suppose that $N \leq M$ (overdetermined problem), then the SVD of $\mathbf{A}$ generates the following set of singular values

$$\begin{aligned} \lambda_1 \geq \lambda_2 \geq \ldots \lambda_K &> 0 \\ \lambda_{K+1} = \ldots = \lambda_N &= 0 \end{aligned} \tag{4.45}$$

where $K \leq N$ is an arbitrary number which designates the number of non-zero singular values. The following expression can be easily verified by substitution into Eqns. (4.44) to be the MP inverse of $\mathbf{A}$:

$$\mathbf{A}^+ = \mathbf{V}\mathbf{\Lambda}^+\mathbf{U}' \tag{4.46}$$

where

$$\mathbf{\Lambda}^+ = \left(\frac{\mathbf{L}^+}{\mathbf{Z}}\right) : \mathcal{R}^M \rightarrow \mathcal{R}^N \tag{4.47}$$

is an $N \times M$ matrix, $\mathbf{L}^+$ is an $N \times N$ square diagonal matrix with elements $(\lambda_1^{-1}, \ldots, \lambda_K^{-1}, 0, \ldots, 0)$ on its diagonal and zeros on its off-diagonal elements, and $\mathbf{Z}$ is an $N \times (M - N)$ matrix of zeros.

In the case of an underdetermined problem, where $N \geq M$ and $K \leq M$, the expression given in Eqn. (4.46) is also a MP inverse. In this circumstance the matrix $\mathbf{\Lambda}^+$ is defined by

$$\mathbf{\Lambda}^+ = (\mathbf{L}|\mathbf{Z}) \tag{4.48}$$

where $\mathbf{\Lambda}^+$ is an $N \times M$ matrix, $\mathbf{L}$ is the $M \times M$ square diagonal matrix with elements $(\lambda_1^{-1}, \ldots, \lambda_K^{-1}, 0, \ldots, 0)$, and $\mathbf{Z}$ is a $(N-M) \times M$ matrix of zeros.

4.7 Solving Overdetermined Linear Problems with SVD

The previous section demonstrates how the MP inverse of a linear operator is easily derived from the SVD of a general, rectangular matrix $\mathbf{A}$. We now revisit the least-squares inverse methods discussed in Chapter (1) to demonstrate the fact that the MP inverse defined above in Eqn. (4.46) is precisely the same as the least-squares inverse defined in Eqn. (1.8).

Recalling Eqn. (**??**), the least-squares solution of an overdetermined problem is

$$\mathbf{m} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{d} \tag{4.49}$$

Substituting the SVD for $\mathbf{A}$ in Eqn. (4.49)gives

$$\begin{aligned} \mathbf{m} &= ((\mathbf{U}\mathbf{\Lambda}\mathbf{V}')'\mathbf{U}\mathbf{\Lambda}\mathbf{V}')^{-1}(\mathbf{U}\mathbf{\Lambda}\mathbf{V}')'\mathbf{d} \\ &= (\mathbf{V}\mathbf{\Lambda}'\mathbf{U}'\mathbf{U}\mathbf{\Lambda}\mathbf{V}')^{-1}(\mathbf{U}\mathbf{\Lambda}\mathbf{V}')'\mathbf{d} \\ &= (\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{V}\mathbf{\Lambda}\mathbf{U}'\mathbf{d} \\ &= \mathbf{V}\mathbf{\Lambda}^+\mathbf{U}'\mathbf{d} \end{aligned} \tag{4.50}$$

where we have made use of the relations $\mathbf{V}'\mathbf{V} = \mathbf{I}$, $\mathbf{U}\mathbf{U}' = \mathbf{I}$, $(\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}$ is diagonal and thus commutes with $\mathbf{V}$, and $(\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda} : \mathcal{R}^M \rightarrow \mathcal{R}^N$ is given by

$$(\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda} = \mathbf{\Lambda}^+ = \begin{pmatrix} \frac{1}{\lambda_1} & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & \frac{1}{\lambda_2} & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & \frac{1}{\lambda_3} & \ldots & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & 0 & \frac{1}{\lambda_N} & 0 & 0 & \ldots & 0 \end{pmatrix} \tag{4.51}$$

The expression in Eqn. (4.50) is generally preferred over that in Eqn. (4.49) for expressing the solution of overdetermined least-squares inverse

problem. Careful inspection of Eqn. (4.50) gives a sense of the geometrical interpretation of the least-squares inverse. The first matrix to operate on $\mathbf{d}$ is $\mathbf{U}'$. This operation simply re-expresses the data $\mathbf{d}$ in terms of the M eigenvectors $\{\mathbf{u}_i, i = 1, \ldots, M\}$ which span the data space $\mathcal{R}^M$, and which are stored as rows of $\mathbf{U}'$. Now that $\mathbf{d}$ is expressed in terms of the eigenvectors, the next operation, multiplication by $\mathbf{\Lambda}^{-1}$ takes the data and maps it (in an intelligent way) to the 'model 'space $\mathcal{R}^N$ where the desired solution $\mathbf{m}$ resides. The last operation, multiplication by $\mathbf{V}$, re-expresses the image of the data vector (which is now expressed in terms of the eigenvectors $\{\mathbf{v}_i, i = 1, \ldots, N\}$ that span the space $\mathcal{R}^N$) into its expression in terms of the original (physical) basis vectors which span $\mathcal{R}^N$.

Example

Let's compute the slope and intercept of the lunar basalt isocrhon discussed in Chapter 1 using SVD instead of the least-squares inverse formula expressed in Eqn. (**??**). We shall make use of the MATLAB® SVD routine:

```
>>d

   d =
                    0.7010
                    0.6999
                    0.7020
                    0.7049

>>A

   A =
                    0.0296    1.0000
                    0.0054    1.0000
                    0.0492    1.0000
                    0.1127    1.0000

>>[U,S,V]=svd(A)
```

```
U =
```

$$\begin{matrix} 0.4995 & 0.2474 & -0.4303 & -0.7100 \\ 0.4989 & 0.5518 & -0.1700 & 0.6463 \\ 0.5000 & 0.0012 & 0.8505 & -0.1633 \\ 0.5016 & -0.7965 & -0.2502 & 0.2270 \end{matrix}$$

```
S =
```

$$\begin{matrix} 2.0024 & 0 \\ 0 & 0.0795 \\ 0 & 0 \\ 0 & 0 \end{matrix}$$

```
V =
```

$$\begin{matrix} 0.0492 & -0.9988 \\ 0.9988 & 0.0492 \end{matrix}$$

Next, let's compute **m** using the SVD approach:

```
>> U'

   ans =
```

$$\begin{matrix} 0.4995 & 0.4989 & 0.5000 & 0.5016 \\ 0.2474 & 0.5518 & 0.0012 & -0.7965 \\ -0.4303 & -0.1700 & 0.8505 & -0.2502 \\ -0.7100 & 0.6463 & -0.1633 & 0.2270 \end{matrix}$$

```
% This script computes the inverse of S, the rectangular matrix %
of eigenvalues:SINV=zeros(2,4);SINV(1,1)=1.0/S(1,1);SINV(2,2)=1.0/S(2,2);%


>> SINV

   SINV =
```

$$\begin{matrix} 0.4994 & 0 & 0 & 0 \\ 0 & 12.5767 & 0 & 0 \end{matrix}$$

```
>>m=V*SINV*U'*d
```

```
m =
```

$$\begin{matrix} 0.0469 \\ 0.6996 \end{matrix}$$

The solution $\mathbf{m}$ (slope=0.0469, intercept=0.6996) is precisely that which we computed in Chapter 1 using Eqn. (**??**).

4.8 Data Independence and Model Resolution

The concepts of the MP inverse and SVD allow a compact development of the notion of data independence and model resolution discussed in previous chapters.

Data Independence

In the case of an overdetermined inverse problem, we can construct the data independence operator $\mathbf{D}$ following the methods of Chapter ():

$$\begin{aligned} \mathbf{D} &= \mathbf{A}\mathbf{A}^{+} \\ &= \mathbf{U}\mathbf{\Lambda}\mathbf{V}'\ \mathbf{V}\mathbf{\Lambda}^{+}\mathbf{U}' \\ &= \mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}^{+}\mathbf{U}' \\ &= \mathbf{U}_{\mathbf{K}}\mathbf{U}'_{\mathbf{K}} \end{aligned} \tag{4.52}$$

where the matrix $\mathbf{U_K}$ is a rectangular $M \times K$ matrix consisting of the first K columns of $\mathbf{U}$, and where K is the number of non-zero singular values λ_i.

Model Resolution

Similarly, in the underdetermined case, we can construct the model resolution matrix, $\mathbf{R}$

$$\mathbf{R} = \quad (\mathbf{A}^{+}\mathbf{A})$$

$$
\begin{aligned}
&= \mathbf{V}\mathbf{\Lambda}^{+}\mathbf{U}'\,\mathbf{U}\mathbf{\Lambda}\mathbf{V}' \\
&= \mathbf{V}\mathbf{\Lambda}^{+}\mathbf{\Lambda}\mathbf{V}' \\
&= \mathbf{V}_{\mathbf{K}}\mathbf{V}'_{\mathbf{K}}
\end{aligned} \tag{4.53}
$$

where, again, the matrix $\mathbf{V_K}$ is the $N \times K$ rectangular matrix consisting of the first K columns of $\mathbf{V}$.

4.9 Model Covarience

The covarience matrix of model errors can be readily expressed using the notation of the SVD. Using $\langle\cdot\rangle$ to denote the expectation value, we have

$$\langle\hat{\mathbf{m}}\rangle = \mathbf{A}^{+}\langle\mathbf{d}\rangle \tag{4.54}$$

Errors in the derived model will similarly be related to errors in the data:

$$(\hat{\mathbf{m}} - \langle\hat{\mathbf{m}}\rangle) = \mathbf{A}^{+}(\mathbf{d} - \langle\mathbf{d}\rangle) \tag{4.55}$$

The covariance matrix $\mathbf{C}$ of the model errors is thus

$$\mathbf{C} = \langle(\hat{\mathbf{m}} - \langle\hat{\mathbf{m}}\rangle)(\hat{\mathbf{m}} - \langle\hat{\mathbf{m}}\rangle)'\rangle = \mathbf{A}^{+}\langle(\mathbf{d} - \langle\mathbf{d}\rangle)(\mathbf{d} - \langle\mathbf{d}\rangle)'\rangle(\mathbf{A}^{+})' = \mathbf{A}^{+}\mathbf{Q}(\mathbf{A}^{+})' \tag{4.56}$$

In situations where $\mathbf{Q}$ is diagonal, and all diagonal elements are the same, say σ^2, *i.e.,*

$$\mathbf{Q} = \sigma^2\mathbf{I} \tag{4.57}$$

we see that

$$\mathbf{C} = \sigma^2\mathbf{A}^{+}(\mathbf{A}^{+})' \tag{4.58}$$

or more compactly

$$
\begin{aligned}
\mathbf{C} = &\quad \sigma^2\mathbf{A}^{+}(\mathbf{A}^{+})' \\
= &\quad \sigma^2\mathbf{V}\mathbf{\Lambda}^{+}\mathbf{U}'\,\mathbf{U}(\mathbf{\Lambda}^{+})'\mathbf{V} \\
= &\quad \sigma^2\mathbf{V}\mathbf{\Lambda}^{+}(\mathbf{\Lambda}^{+})'\mathbf{V}' \\
= &\quad \sigma^2\mathbf{V}_{\mathbf{K}}\mathbf{\Lambda}_{\mathbf{K}}^{-\mathbf{2}}\mathbf{V}'_{\mathbf{K}}
\end{aligned} \tag{4.59}
$$

where $\mathbf{\Lambda}^{-2}$ is a square $K \times K$ diagonal matrix having diagonal elements $\lambda_1^{-2}, \lambda_2^{-2}, \ldots, \lambda_K^{-2}$.

4.10 Testing Data Sufficiency

Situations arise from time to time in which the number of independent data measurements appear to overdetermine the desired end-result model but, in practice, do not. The SVD provides a useful test to foresee circumstances in which this situation arises. Suppose, for example, the lunar-basalt data were clustered in such a manner that the four $^{87}Rb/^{86}Sr$ measurements were clustered very close in numerical value rather than spread over an interval. The fact that there were four independent measurements might suggest that least-squares would provide an accurate and reliable evaluation of $\mathbf{m} = [\alpha\beta]'$. The SVD of the matrix $\mathbf{A}$ would suggest otherwise. The fact that the our $^{87}Rb/^{86}Sr$ measurements are clustered implies that the four rows of $\mathbf{A}$ will be nearly the same. In this circumstance, the SVD of $\mathbf{A}$ will yield two singular values λ_1 and λ_2 in which one was indistinguisable from zero (or, in other words, where $\lambda_1 >> \lambda_2$). If the smaller singular value λ_2 is indistinguishable from zero, then neither $\mathbf{D}$ nor $\mathbf{R}$ will be the identity operator, as suggested by Eqns. (4.52) and (4.53). In the case of a lunar basalt analysis with clustered $^{87}Rb/^{86}Sr$ measurements, K=1, $N = 2$ and $M = 4$.

The above example suggests that an additional role of the SVD in inverse methods is the insight it gives into the nature of the problem (*i.e.,* whether the problem is overdetermined or underdetermined). In the circumstance suggested above, the count of non-zero singular values, K, provides the means to detect underdeterminacy even in circumstances where, on the face of it, the number of data points exceeds the number of undetermined parameters.

4.11 Summary

In this chapter we have accomplished an important bit of 'dirty work'. We have managed to develop a conceptual, geometric view of the linear-algebra of square matrices, and have extended it to include rectangular matrices. In doing this, we have developed the concepts of eigenvectors and eigenvalues, and have derived the very useful singular-value decomposition (SVD). We have seen how the SVD provides a means for finding the least-squares inverse

of a rectangular matrix that is equivalent to the more brute-force method derived in the previous chapter.

4.12 Bibliography

Stewart, G. W., 1993. On the early history of the singular value decomposition. *SIAM Review*, **35**, 551-566.

4.13 Laboratory Exercises

Using SVD to Solve Overdetermined, Underdetermined and Mixed Linear Inverse Problems

In this lab you will re-compute the slope and intercept of the lead-isotope isochron which determines the age of the earth [Patterson, 1956] using SVD as the technique for inverting a rectangular (overdetermined) matrix.

Problem 1. Use the MATLAB® routines to compute the SVD of the matrix **A** you generated in problem 4 of Chapter 1.

Problem 2. Using the SVD, compute **m**, the vector containing the slope and intercept of the lead-isotope isochron determined by Patterson's [1956] data.

Problem 3. Show that the MP inverse of an underdetermined problem is the same as the minimum-norm problem derived in Chapter (3).

Chapter 5

Idiosyncratic Line-Fitting Algorithms: Control Method and Simulated Annealing

5.1 Introduction

We revisit the line-fitting problem of Chapter (1) to introduce two additional methods for solving overdetermined, least-squares inverse problems: control methods and simulated annealing. Our motivation for introducing these methods is the fact that the problem of finding the slope and intercept of an isochron in radiometric dating is complicated by the fact that isotopic measurements appear in both the data vector $\mathbf{d}$ and the linear operator $\mathbf{A}$ of the following linear prolbem (see § 1.7):

$$\mathbf{Am} = \mathbf{d} \tag{5.1}$$

In the lunar basalt example of Chapter (1), the data vector $\mathbf{d} \in \mathcal{R}^4$ consisted of four observed $^{87}\mathrm{Sr}/^{86}\mathrm{Sr}$ values associated with four minaral separates:

$$\mathbf{d} = \begin{bmatrix} S_{wr} \\ S_{Plag} \\ S_{Px} \\ S_{Ilm} \end{bmatrix} \tag{5.2}$$

and the 4×2 matrix $\mathbf{A} : \mathcal{R}^2 \rightarrow \mathcal{R}^4$ contained four observed $^{87}\mathrm{Rb}/^{86}\mathrm{Sr}$ ratios.

The least-squares solution to Eqn. (5.1) which takes account of errors $\underline{\epsilon}$ in $\mathbf{d}$, namely $\hat{\mathbf{m}} = (\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A})^{-1}\,\mathbf{A}'\mathbf{Q}^{-1}\mathbf{d}$ does not account for the fact that there is also uncertainty associated with the elements of $\mathbf{A}$. To adequately account for these additional errors, we must abandon Eqn. (5.1) and the linear-algebraic methods associated with solving it in favour of a more primitive, non-linear approach to finding $\hat{\mathbf{m}}$ given the $^{87}\mathrm{Sr}/^{86}\mathrm{Sr}$ and $^{87}\mathrm{Rb}/^{86}\mathrm{Sr}$ data.

The complication outlined above has been dealt with in radiometric geochronology by the least-squares methods developed by York [1966]. York's work emphasizes the complexity of even the most simple least-squares line fitting problem. We shall not review York's methods, but rather introduce more modern techniques for accomplishing the same fundamental goal.

The first method we shall introduce is known as *control*. The control method was invented by engineers and applied mathmaticians (see Bryson and Ho [1975] or Bellman [1967]) to solve constrained optimization problems associated with the manoever of spacecraft, the operation of chemical-processing plants and other such practical concerns. Problems which are amenable to control methods are those in which the minimum of a least-squares performance index is sought subject to constraints which may take the form of differential equations. The essential ingredient of the control method is its systematic means to determine the gradient of a least-squares performance index J in the space of unknown parameters to be determined. The advantage of control methods over the linear-algebraic methods described in Chapter (1) is that they can be extended quite readily to the solution of non-linear least-squares problems.

The second method we shall discuss is called *simulated annealing* (see chapter 10 of Press *et al.* [1989]). Simulated annealing, like the control

method, is appropriate for problems involving the minimization of a non-linear, least-squares performance index. Unlike the control method, however, simulated annealing appeals to random trial and error as a means to find the solution. The appealing name, "simulated annealing", stems from an analogy that can be drawn between the mathematical process of minimization and the physical process associated with the annealing of metal as it cools from a liquid state. Unlike the control method, simulated annealing requires very little mathematical set-up to impliment an algorithm for solving a least-squares inverse problem.

5.2 Radiometric Dating Redux

Consider the lunar basalt isochron fitting problem described in § (1.7). Let the vector $\mathbf{X} \in \mathcal{R}^4$ denote the four observed $^{87}\mathrm{Rb}/^{86}\mathrm{Sr}$ ratios associated with the four mineral separates. Let $\mathbf{Y} \in \mathcal{R}^4$ denote the four observed $^{87}\mathrm{Sr}/^{86}\mathrm{Sr}$ ratios. Our problem is to choose a slope α and intercept β such that the following least-squares performance index J is minimized

$$J = [\mathbf{X} - \mathbf{x}]' \, \mathbf{S} \, [\mathbf{X} - \mathbf{x}] + [\mathbf{Y} - \mathbf{y}]' \, \mathbf{Q} \, [\mathbf{Y} - \mathbf{y}] \tag{5.3}$$

subject to the four constraints

$$\alpha x_i - y_i + \beta = 0 \qquad \text{for } i = 1, \ldots, 4 \tag{5.4}$$

Pairs of compontents (x_i, y_i) of the vectors $\mathbf{x} \in \mathcal{R}^4$ and $\mathbf{y} \in \mathcal{R}^4$ represent the four points which lie on the isochron and which are, in some sense, *closest* to the corresponding data points (X_i, Y_i). The matrices $\mathbf{S}$ and $\mathbf{Q}$ express the covarience of the errors in measurements of $\mathbf{X}$ and $\mathbf{Y}$, respectively; *i.e.,*

$$\mathbf{S} = \langle (\mathbf{X} - \langle \mathbf{X} \rangle) \, (\mathbf{X} - \langle \mathbf{X} \rangle)' \rangle \tag{5.5}$$

$$\mathbf{Q} = \langle (\mathbf{Y} - \langle \mathbf{Y} \rangle) \, (\mathbf{Y} - \langle \mathbf{Y} \rangle)' \rangle \tag{5.6}$$

To minimize J subject to the 4 constraints represented by Eqn. (5.4), we employ a Lagrange-multiplier vector $\underline{\lambda}$ to augment J:

$$H = [\mathbf{X} - \mathbf{x}]' \, \mathbf{S}^{-1} \, [\mathbf{X} - \mathbf{x}] + [\mathbf{Y} - \mathbf{y}]' \, \mathbf{Q}^{-1} \, [\mathbf{Y} - \mathbf{y}] + 2\underline{\lambda}' \, [\alpha \mathbf{x} - \mathbf{y} + \beta \mathbf{1}] \tag{5.7}$$

where

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{5.8}$$

Our goal now is to minimize H by judiciously choosing the 14 variables $x_i, i = 1, \ldots, 4$, $y_i, i = 1, \ldots, 4$, $\lambda_i, i = 1, \ldots, 4$, α and β.

The Euler-Lagrange equations expressing the conditions that must be met for H to be minimum are derived by considering the variation δH. Following the usual practice, we obtain:

$$-\mathbf{S}^{-1}\left[\mathbf{X} - \mathbf{x}\right] + \alpha\underline{\lambda} = \mathbf{0} \tag{5.9}$$

$$-\mathbf{Q}^{-1}\left[\mathbf{Y} - \mathbf{y}\right] - \underline{\lambda} = \mathbf{0} \tag{5.10}$$

$$\alpha\mathbf{x} - \mathbf{y} + \beta\mathbf{1} = \mathbf{0} \tag{5.11}$$

$$\underline{\lambda}'\mathbf{x} = 0 \tag{5.12}$$

and

$$\underline{\lambda}'\mathbf{1} = \sum_{i=1}^{4} \lambda_i = 0 \tag{5.13}$$

Equations (5.9)-(5.13) represent 14 equations for the 14 unknowns. It is not easy to solve Eqns. (5.9)-(5.13) because they are nonlinear (due to vector products between two unknowns such as $\underline{\lambda}$ and $\mathbf{x}$ in Eqn. (5.12), for example).

One strategy for minimizing H is to recognize that the left-hand sides of Eqns. (5.9)-(5.13) represent the gradient of H with respect to each of the 14 unknown variables. An iterative, down-gradient search technique may be utilized to find the solution which minimizes H using the ability to readily evaluate the gradient of H. Such an algorithm might be constructed as follows.

5.2.1 Control Algorithm

Step 1-

Guess $\mathbf{x}$, α and β. Denote them by $\mathbf{x}^{[1]}$, $\alpha^{[1]}$ and $\beta^{[1]}$.

Step 2-

Compute $\mathbf{y}^{[1]}$ and $\underline{\lambda}^{[1]}$ using Eqns. (5.11) and (5.10), respectively.

Step 3-

Compute ∇H using the expression derived from the left-hand sides of Eqns. (5.9), (5.12) and (5.13):

$$\nabla H = 2 \begin{bmatrix} -\mathbf{S}\left[\mathbf{X} - \mathbf{x}^{[1]}\right] + \alpha^{[1]}\underline{\lambda}^{[1]} \\ (\underline{\lambda}^{[1]})'\mathbf{x}^{[1]} \\ (\underline{\lambda}^{[1]})'\mathbf{1} \end{bmatrix} \tag{5.14}$$

Step 4-

Test to determine if $\nabla H = \mathbf{0}$. If so, then the search algorithm stops because $\mathbf{x}^{[1]}$, $\mathbf{y}^{[1]}$, $\alpha^{[1]}$ and $\beta^{[1]}$ have minimized H. If not, proceed to the next step.

Step 5-

Using a down-gradient search algorithm, obtain an improved guess $\mathbf{x}^{[2]}$, $\alpha^{[2]}$ and $\beta^{[2]}$. For efficiency, select a search algorithm that makes use of ∇H computed from the previous step. (Numerous algorithms are available for this step. We recommend the conjugate gradient method, described in the appendix to this chapter, for its versatility.)

step 6-

Proceed to Step 2.

5.3 Example: Lunar Basalt Isochron

We apply the control algorithm described above as a means to radiometrically date the lunar basalt discussed in § (1.7). In this example,

$$\mathbf{X} = \begin{bmatrix} 0.0296 \\ 0.00537 \\ 0.0492 \\ 0.1127 \end{bmatrix} \tag{5.15}$$

$$\mathbf{Y} = \begin{bmatrix} 0.70096 \\ 0.69989 \\ 0.70200 \\ 0.70490 \end{bmatrix} \tag{5.16}$$

$$\mathbf{S} = \left(\frac{1}{2}\right)^2 \begin{bmatrix} 0.0004^2 & 0 & 0 & 0 \\ 0 & 0.00005^2 & 0 & 0 \\ 0 & 0 & 0.0004^2 & 0 \\ 0 & 0 & 0 & 0.0009^2 \end{bmatrix} \tag{5.17}$$

$$\mathbf{Q} = \left(\frac{1}{2}\right)^2 \begin{bmatrix} 0.00007^2 & 0 & 0 & 0 \\ 0 & 0.00009^2 & 0 & 0 \\ 0 & 0 & 0.00005^2 & 0 \\ 0 & 0 & 0 & 0.00006^2 \end{bmatrix} \tag{5.18}$$

The data for $\mathbf{S}$ and $\mathbf{Q}$ are derived from table 3 (part II) of Nyquist *et al.* [1979]. (The 2σ error levels are divided by two and squared to obtain the diagonal components of the covarience matrices.)

For an initial guess, we choose $\mathbf{x}^{[1]} = \mathbf{X}$, $\alpha^{[1]} = 0.01$, and $\beta^{[1]} = 0.699$. The folowing MATLAB® algorithm was used to search for the solution $\mathbf{m} = [\mathbf{x}' \ \alpha \ \beta]'$ which minimizes H:

```
% This program determines the slope and intercept of the
%
X=[0.0296
.00537
.0492
0.1127];
```

```
%
Y=[0.70096
0.69989
0.702
0.7049];
%
S=.25e4*[.0004
.00005
.0004
.0009].^ 2;
%
S=diag(S,0);
Sinv=inv(S)
%
Q=.25e4*[.00007
.00009
.00005
.00006].^ 2;
%
Q=diag(Q,0);
Qinv=inv(Q)
%
beta=0.699;
alpha=0.01;
guess= [X
alpha
beta];
%
options(1)=0; % set to 1 for verbose output
options(9)=1; % set to 1 for check of analytic gradient
options(2)=1.e-5; % stopping criteria for m
options(3)=1.e-5; % stopping criteria for H
options(14)=500; % max number of iterations in search
%
m = fminu('H',guess,options,'gradH',X,Y,S,Q) % minimization
%
%
```

Observe that $\mathbf{S}$ and $\mathbf{Q}$ have been multiplied by a scaling factor of 1.0×10^4 to better condition the algorithm for rapid convergence to the solution. The value of this scaling factor was found experimentally. The MATLAB® toolbox function `fminu` calls two functions `H` and `gradH`. These two functions represent the objective function H and its gradient ∇H, respectively; and are listed as follows:

```
function [h] = H(guess,X,Y,Sinv,Qinv)
xtrial=guess(1:4);
ytrial=guess(5)*xtrial+guess(6)*ones(4,1);
h= (X-xtrial)'*Sinv*(X-xtrial) + (Y-ytrial)'*Qinv*(Y-ytrial);

function [dh] = gradH(guess,X,Y,Sinv,Qinv)
xtrial=guess(1:4);
ytrial=guess(5)*xtrial+guess(6)*ones(4,1);
lambda=-Qinv*(Y-ytrial);
alpha=guess(5);
dh=[-2*Sinv*(X-xtrial)+2*alpha*lambda
2*lambda'*xtrial
2*lambda'*ones(4,1)];
```

Normally, these functions are created by the user and then stored as m-files in MATLAB® .

The above MATLAB® routines generated the following solution $\mathbf{m}$:

```
m =

0.0295
0.0054
0.0493
0.1125
0.0469
0.6996
```

Recall that the first four components of $\mathbf{m}$ are the components of $\mathbf{x}$, the fifth component of $\mathbf{m}$ is the slope of the isochron, and the last component of $\mathbf{m}$ is the intercept of the isochron. Figure (5.1) displays the isochron associated with the above solution.

5.4 Error Analysis Associated with the Control Method

In Chapter (3) we derived the linear relationship between the covarience of $\mathbf{m}$, $\mathbf{E} = \langle (\mathbf{m} - \langle \mathbf{m} \rangle)(\mathbf{m} - \langle \mathbf{m} \rangle)' \rangle$, and the covarience of the data $\mathbf{Q}$. The expanded view of the isochron problem presented in this chapter necessitates that we deal with a non-linear relationship between $\mathbf{E}$ and the two covarience matrices $\mathbf{S}$ and $\mathbf{Q}$.

Non-linearity in the relationship between $\mathbf{m}$ and the data $\mathbf{X}$ and $\mathbf{Y}$ suggests that $\mathbf{E}$ cannot be written as an explicit function of $\mathbf{S}$ and $\mathbf{Q}$. To obtain $\mathbf{E}$ in this circumstance, we resort to a statistical technique. We generate N separate vectors $\{\mathbf{m}_n\}_{n=1}^N$ from N separate renditions of the data $\tilde{\mathbf{X}}_n$ and $\tilde{\mathbf{Y}}_n$, $n = 1, \ldots, N$ which differ from the expected value of the data $\mathbf{X}$ and $\mathbf{Y}$ by random errors $\mathbf{r}_x$ and $\mathbf{r}_y$, respectively. The random errors are assumed to be normally distributed, to have zero mean, and a covarience equal to that of the data, *i.e.,*

$$\langle \mathbf{r}_x \rangle = \mathbf{0} \tag{5.19}$$

$$\langle \mathbf{r}_y \rangle = \mathbf{0} \tag{5.20}$$

$$\langle (\mathbf{r}_x - \langle \mathbf{r}_x \rangle)(\mathbf{r}_x - \langle \mathbf{r}_x \rangle)' \rangle = \mathbf{S} \tag{5.21}$$

$$\langle (\mathbf{r}_y - \langle \mathbf{r}_y \rangle)(\mathbf{r}_y - \langle \mathbf{r}_y \rangle)' \rangle = \mathbf{Q} \tag{5.22}$$

The N separate solutions to the isochron problem generated by the N separate perturbations of the data are then analysed statistically to obtain an estimate of $\mathbf{E}$:

$$\langle \mathbf{m} \rangle = \frac{1}{N} \sum_{n=1}^{N} \mathbf{m}_n \tag{5.23}$$

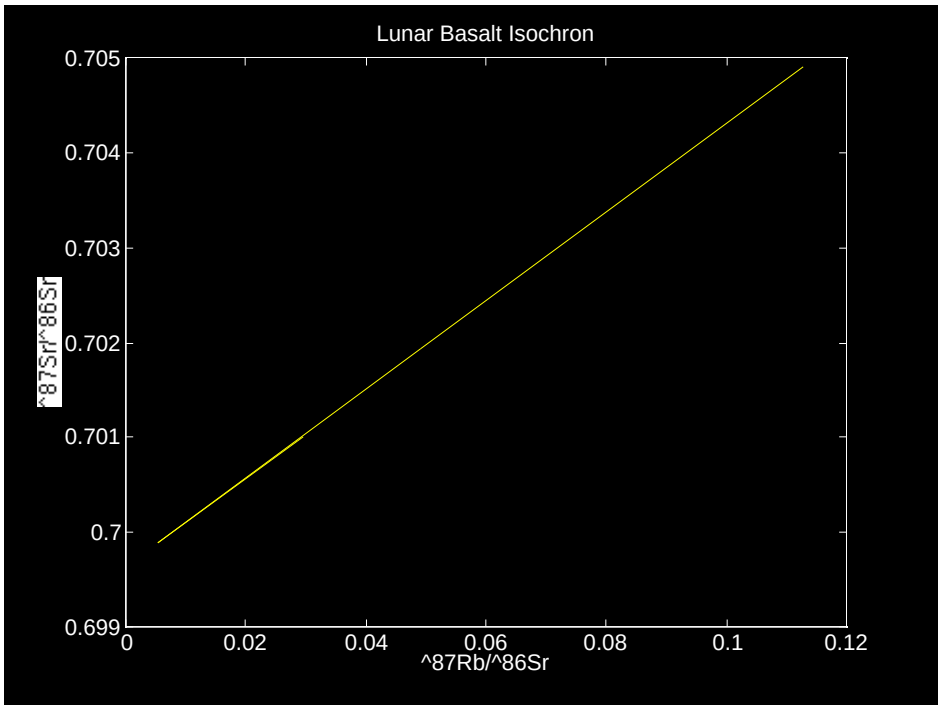

Figure 5.1: The lunar basalt isochron found by control methods.

$$\langle (\mathbf{m} - \langle \mathbf{m} \rangle)(\mathbf{m} - \langle \mathbf{m} \rangle)' \rangle = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{m}_i - \langle \mathbf{m} \rangle)(\mathbf{m}_i - \langle \mathbf{m} \rangle)' \qquad (5.24)$$

Lunar Basalt Example

The above empirical means of determining **E** was employed for the Lunar basalt isochron problem using the following MATLAB® algorithm:

```
% This algorithm determines the covarience matrix E.
%
% First, set up random data
%
X=[0.0296
.00537
.0492
0.1127];
Y=[0.70096
0.69989
0.702
0.7049];
S=1.0e4*.25*[.0004
.00005
.0004
.0009].  ^ 2;
S=diag(S,0);
Sinv=inv(S);
Q=1.0e4*(.25*[.00007
.00009
.00005
.00006].  ^ 2);
Q=diag(Q,0);
Qinv=inv(Q);
%
options(1)=0; % set to 1 for verbose output
options(9)=0; % set to 1 for check of analytic gradient
```

```
options(2)=1.e-5; % stopping criteria for m
options(3)=1.e-5; % stopping criteria for H
options(14)=100; % max number of iterations in search
%
N=10;
XR=zeros(4,N);
YR=zeros(4,N);
MR=zeros(6,N);
for n=1:N
XR(:,n)=X+randn(4,1).*sqrt(1.0e-4*diag(S));
YR(:,n)=Y+randn(4,1).*sqrt(1.0e-4*diag(Q));
end
%
for n=1:N
n
beta=0.6996;
alpha=0.0469;
guess= [XR(:,n)
alpha
beta];
MR(:,n) = fminu('H',guess,options,'gradH',XR(:,n),YR(:,n),Sinv,Qinv)
% minimization algorithm
end
%
% Now do statistics on MR:
%
meanM=[mean(MR')]'
standard=std(MR')
E=cov(MR')
```

What is of central interest is the standard deviation of the individual components of $\{\mathbf{m}_n\}_{n=1}^N$. The above algorithm gives

```
standard =

1.0e-03 *
```

```
0.1987    0.0260    0.1589    0.5928    0.4898    0.0254
```

The standard deviation of the slope α is approximately 4.272×10^{-4}, or about 1.1% of the derived value of α. This level of uncertainty is slightly higher than that derived in Chapter (3).

It should be pointed out that a significant problem can crop up if the above monte-carlo method is used to establish the model covarience $\mathbf{E}$. If the search algorithm is not sufficiently sensitive, and the down-gradient search stops before $\mathbf{m}$ is significantly changed from its value at the initial guess, the variance of $\mathbf{m}$ can be perceived to be artificially small. To safeguard against this possibility, one must use special care to see to it that the model covarience does not change when parameters controling the down-gradient search are made more sensitive.

5.5 Radiometric Dating Redux[2]: Simulated Annealing

Here we solve the isochron problem using a modern computational technique called simulated annealing. This method is similar to the control method described in the previous sections in that it accounts for uncertainty in both $\mathbf{X}$ and $\mathbf{Y}$. Unlike the control method, or the linear least-squares method discussed in Chapter (1), simulated annealing does not make use of Euler-Lagrange conditions as a means to search for the minimum of J. Instead, simulated annealing employes a random selection process to perturb a trial solution vector $\mathbf{m} = [\mathbf{x}' \ \alpha \ \beta]'$ and simple trial and error to select those perturbations which yield a smaller J. Here we define J in such a manner as to incorporate esxplicitly the linear constraint of the previous section:

$$J = [\mathbf{X} - \mathbf{x}]' \, \mathbf{S} \, [\mathbf{X} - \mathbf{x}] + [\mathbf{Y} - \alpha\mathbf{x} - \beta\mathbf{1}]' \, \mathbf{Q} \, [\mathbf{Y} - \alpha\mathbf{x} - \beta\mathbf{1}]' \qquad (5.25)$$

For a given guess $\mathbf{m}$, a perturbed guess $\tilde{\mathbf{m}}$ is constructed using a sequence of N random numbers ξ_n, $n = 1, \ldots, N$, where N is the dimensionality of $\mathbf{m}$.

$$\tilde{\mathbf{m}} = \mathbf{m} + \begin{bmatrix} \xi_1 \Delta x \\ \xi_2 \Delta x \\ \vdots \\ \xi_{N-1} \Delta \alpha \\ \xi_N \Delta \beta \end{bmatrix} \tag{5.26}$$

where Δx, $\Delta \alpha$, and $\Delta \beta$ are appropriately chosen ranges over which the random perturbations are to vary. The random numbers ξ_n, $n = 1, \ldots, N$ are assumed to have a normal (bell-shaped) probability of mean 0 and standard deviation 1.

The simulated annealing algorithm revolves around the question of whether to accept the randomly perturbed guess $\tilde{\mathbf{m}}$ as a *replacement* of the initial guess $\mathbf{m}$. Defining J_o and J' to be the values of J associated with $\mathbf{m}$ and $\tilde{\mathbf{m}}$, respectively, the decision to accept $\tilde{\mathbf{m}}$ is based on two variables: $P(J_o, J')$ and ξ_d, where

$$P(J_o, J') = e^{\frac{(J_o - J')}{\theta}} \tag{5.27}$$

and where $\xi_d \in [0, 1]$ is a random number that has a uniform distribution in the interval [0,1]. The parameter θ is referred to as the annealing temperature. Its significance will become clear below.

According to the simulated annealing algorithm, a perturbed guess $\tilde{\mathbf{m}}$ is accepted unconditionally when $P(J_o, J') > 1$. This makes sense because $J' < J_o$ implies that the perturbed guess is closer to the minimum of J than the original guess. The crucial aspect of the simulated annealing algorithm which makes it attractive in some applications is that the perturbed guess is also accepted when $\xi_d < P(J_o, J')$, *i.e.,* even when $J' > J_o$. The point of accepting some of the perturbed guesses, even when they are worse than the original guess, is that they introduce a randomness to the search process which helps the algorithm from being "caught" near isolated local minima of J. This role of the parameter θ is to determine the frequency at which "bad" guesses are occasionally accepted. When θ is large, $P(J_o, J')$ will be near 1 even when J' is much greater than J_o. Since ξ_d is distributed uniformly on the interval [0,1] (equal probability for all values between 0 and 1), there is

a greater chance of accepting bad guesses when θ is large. Conversely, when θ is small, there $P(J_o, J')$ will be closer to 0 for a given $(J_o - J')$, and the chance of accepting bad guesses is reduced.

Typically, during an effort to minimize J, the "annealing temperature" θ will begin relatively large, and reduce as J is reduced. Initially, a greater acceptance rate of bad guesses is beneficial to the algorithm because it helps avoid local minima in J which would otherwise "trap" the guess $\mathbf{m}$. As J is reduced toward its anticipated minima, it becomes less beneficial to accept bad guesses on the premise that local minima in J are affecting the guess. Thus, the parameter θ is reduced as J is reduced.

The name of the algorithm, simulated annealing, reflects the analogy between the minimization algorithm and the metallugical process of annealing metal. To manufacture metal with optimum hardness characteristics, the crystal structure of the atoms must be in near perfect order, *i.e.,* the free energy of the metal must be minimized. Often, as metals cool, the atoms get "caught" in imperfect orderings which do not minimize the free energy. The metalurgist can raise the temperature of such an imperfectly ordered metal to allow sufficient random thermal motions that anneal out the imperfect orderings. The parallels between the minimization of free energy by increasing the annealing temperature and the minimization of J by increasing the parameter θ are what motivate the term "simulated annealing" in the current context.

Example: Lunar Basalt Isochron by Simulated Annealing

The following MATLAB® algorithm is used to generate the guess $\mathbf{m}$ which minimizes J defined in Eqn. (**??**) using the lunar basalt data of Nyquist *et al.* [1979] summarized in Eqns. (5.15)-(5.18):

```
% This program determines the slope and intercept
% of the lunar basalt isochron using
% simulated annealing.
%
X=[0.0296
```

```
.00537
.0492
0.1127];

Y=[0.70096
0.69989
0.702
0.7049];

S=(.5*[.0004
.00005
.0004
.0009]).  ^ 2;
S=diag(S,0);
Sinv=inv(S);

Q=(.5*[.00007
.00009
.00005
.00006]).^ 2;
Q=diag(Q,0);
Qinv=inv(Q);

beta=0.6990;
alpha=0.04;
m= [X
alpha
beta];
dm= [.0001*X
.0001
.00001];
theta=1.e-2;
%
%
N=1000;
History=zeros(N,2);
counter=0;
```

```
while counter <= N
if counter == 500
theta=theta/10;
end
mtilde=m+randn(6,1).*dm;
counter=counter+1;
History(counter,1)=H(m,X,Y,Sinv,Qinv);
History(counter,2)=m(5);
P=exp( (H(m,X,Y,Sinv,Qinv) - H(mtilde,X,Y,Sinv,Qinv)) / theta);
if P >= 1
m=mtilde;
elseif rand(1,1) <= P
m=mtilde;
end
end
%
%
plot(History(:,1))
```

The result of the above algorithm is shown in Figs. (5.2) and (5.3). The algorithm is able to converge to the answer derived previously using control methods; but the convergence is slow and is not monotonic.

In some practical circumstances, slow convergence, such as that demonstrated by Figs. (5.2) and (5.3) may not be a limiting performance issue when selecting a method for solving a particular inverse problem. The computational cost of computing J and generating random numbers ξ_i, $i = 1, \ldots, N$, and ξ_d may be much less than the cost of solving Euler-Lagrange equations or impliming a control algorithm. In such circumstances, simulated annealing may offer a viable means of achieving a solution efficiently and with minimum programming cost.

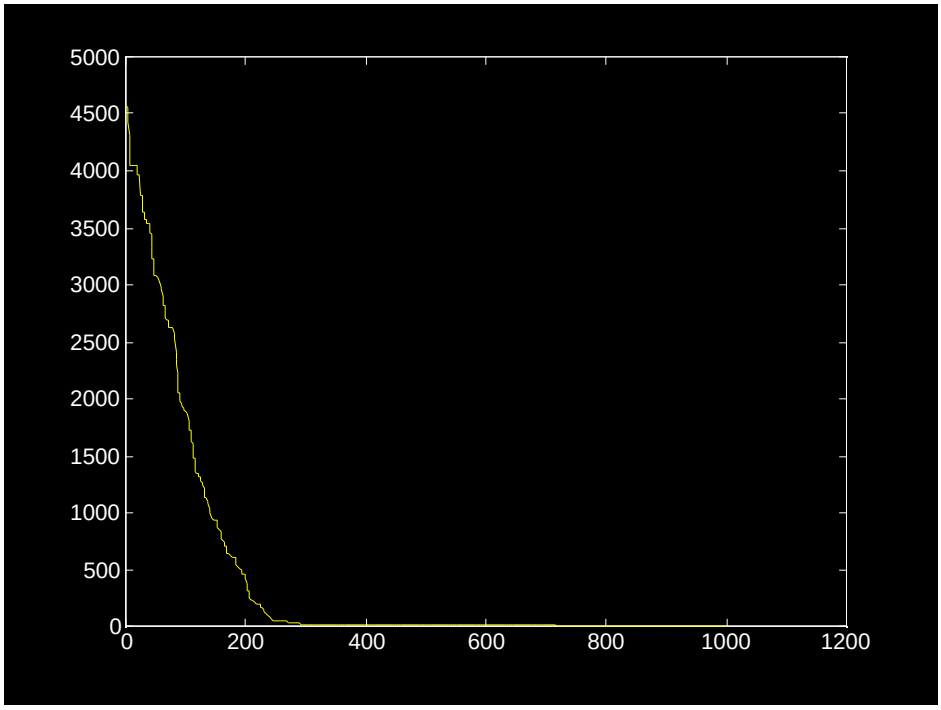

Figure 5.2: The performance index J as a function of iteration count in the effort to solve the lunar basalt isochron problem using a simulated annealing algorithm.

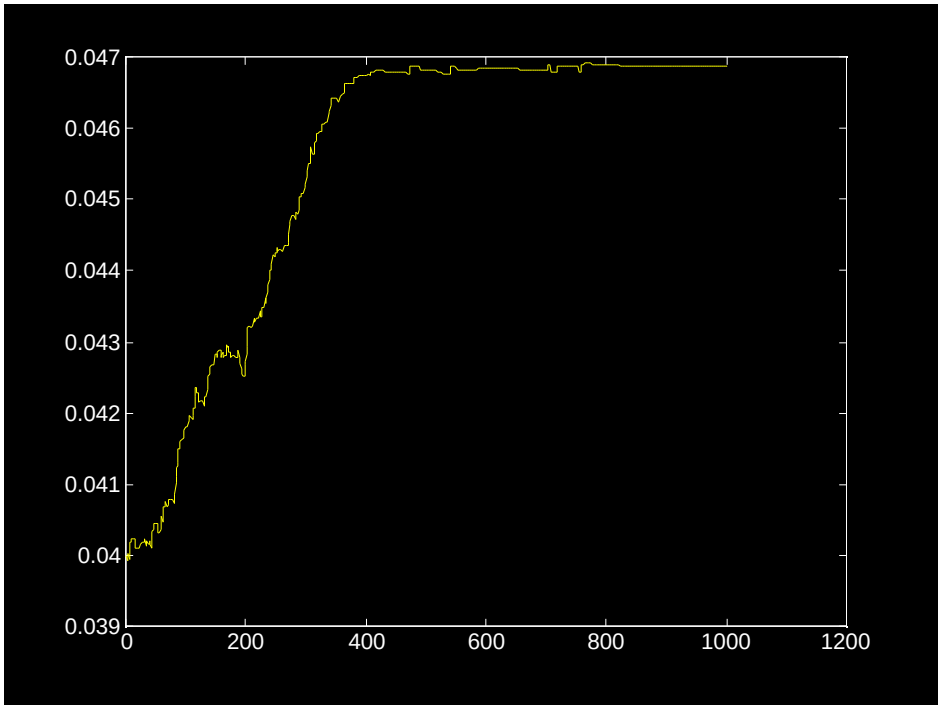

Figure 5.3: The slope of the lunar-basalt isochron as a function of iteration count in the effort to solve the lunar basalt isochron problem using a simulated annealing algorithm.

5.6 Bibliography

Bellman, R., 1967. *Introduction to the Mathematical Theory of Control Processes (Volumes 1 and 2).* (Academic Press, New York, 245 pp., Vol. 1, 301 pp., Vol 2.)

Bryson, A. E., Jr., and Y.-C. Ho, 1975. *Applied Optimal Control.* (J. Wiley & Sons, New York, 481 pp.)

Nyquist, L. E., C.-Y. Shih, J. L. Wooden, B. M. Bansal, and H. Wisemann, 1979. The Sr and Nd isotopic record of Apollo 12 basalts: Implications for lunar geochemical evolution. *Proc. 10th Lunar Planet. Sci. Conf.*, 77-114.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1989. *Numerical Recipes.* (Cambridge University Press, Cambridge, U.K., 702 pp.)

York, D., 1966. Least-squares fitting of a straight line. *Canadian Journal of Physics*, **44**, 1079-1086.

Part II

Applications

Chapter 6

Sea-Floor Spreading Models and Ocean Bathymetry

6.1 Overview

One of the crucial developments in geophysical sciences that shape our modern view of the earth was the discovery of sea-floor spreading (plate tectonics) in the early 1960's. A key step in this discovery was the explanation of the curious bathymetry of the ocean floor. According to the sea-floor spreading hypothesis, oceanic crust is created from a liquid, isothermal magma at the mid-ociean ridges, and is destroyed by subduction in subduction zones usually located far from the mid-ocean ridges. Near the mid-ocean ridges, the ocean floor is shallow because the rock that comprises the oceanic crust is hot. Far from the mid-ocean ridges, the ocean floor is deep because the rock is cold. We shall study the process which determines the depth of the ocean basins as a function of distance from the mid-ocean ridge. Our goal will be to learn how least-squares inverse methods can be used to fit models of sea-floor spreading to bathymetric data derived from the Pacific. Before doing so, however, we will derive the solution for conductive cooling of a semi-infinite solid, and use it to repeat the calculation of the age of the earth by Kelvin in 1864 discussed in Chapter 1.

6.2 Sea-Floor Spreading and Continental Drift

One of the crucial scientific discoveries in the twentieth century was the fact that the ocean floor moves horizontally across the surface of the globe like a great conveyor belt, carrying the continents with it. The importance of this movement is realized when one considers that it is responsible for most of the geochemical processes which, over the long term, are necessary for recycling the earth's crust and making the chemical composition of the ocean and atmosphere habitable. The effect of sea-floor spreading we shall consider here concerns only the ocean bathymetry. In particular, we want to know why the ocean basins have a depth that appears to be minimum near their center and to increase with the square-root of the distance on either side. For a more complete discussion of sea-floor spreading and its effect on the earth, the reader should consult a text on geology.

A key step in the discovery of sea-floor spreading involved the analysis of the geologic age and depth of the ocean floor. As shown in Figs. (6.1) and (6.2), a typical transect across an ocean such as the North Atlantic shows a curious deepening and aging of the ocean floor away from the center of the ocean basin. The age vs. distance relationship suggests that ocean crust is created at the mid-ocean ridge, and moves away at a constant rate. The deepening can be explained by the thermal contraction of the oceanic crust as it cools with increasing ages. Recall that ocean crust is being constantly created by the solidification of hot, molten basalts in the seam of the mid-ocean ridges. Thus, as the oceanic crust moves away on either side of the ridge, it will progressively cool and sink deeper into the earth's mantle.

Several groups of marine geophysicists [Davis and Lister, 1974; Parsons and Sclater, 1977], realized that the depth of the sea-floor varied linearly with the square-root of its geologic age (at least for the first 70 million years or so). This relationship suggested that the cause of depth variation was thermal contraction associated with conductive cooling which, as will soon be shown, is a function of $\sqrt{t}$. This relationship is shown in Fig. (6.3), which displays typical depth values plotted as a function of the square-root of the age.

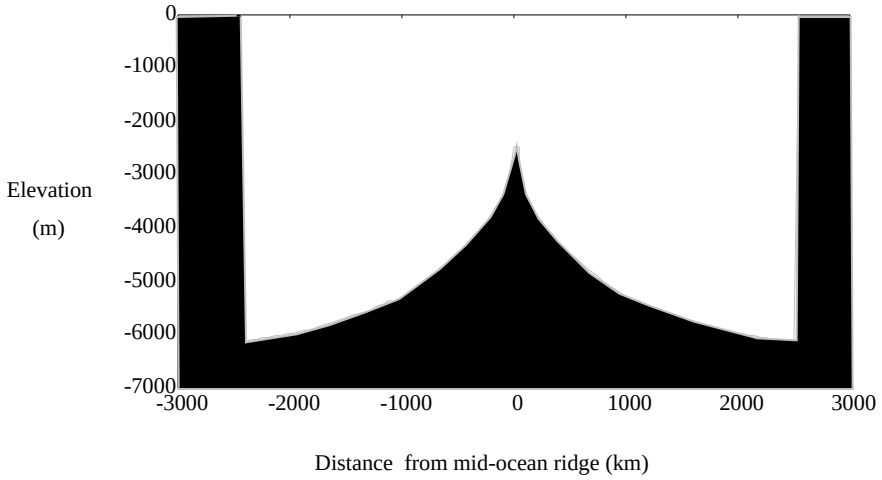

Figure 6.1: Ocean Bathymetry in typical cross section.

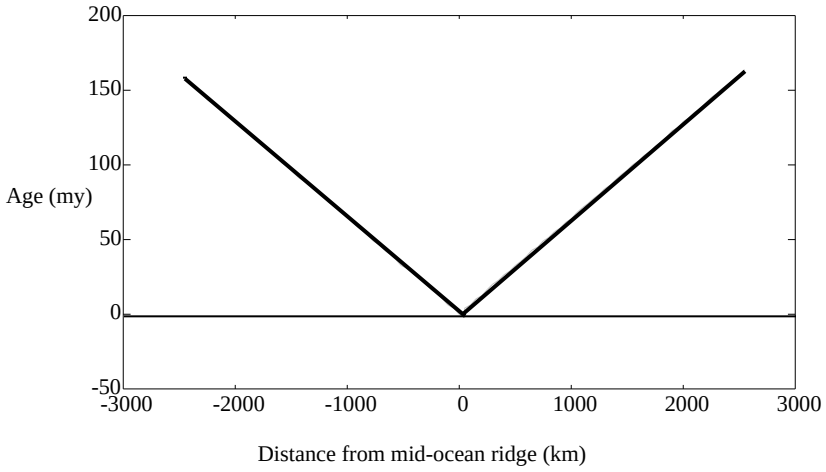

Figure 6.2: Schematic transect of the geological age of the ocean floor across a typical ocean basin. This age is determined typically by bio-stratigraphic analysis of the ocean sediments that are piled up atop the basaltic bedrock.

Several conductive-cooling models for the oceanic crust have been proposed to explain the relationship shown in Fig. (6.3). The simplest, proposed by Davis and Lister [1974], makes use of a solution Lord Kelvin [1864], the greatest physicist of the 19^{th} century, derived for the conductive cooling of the semi-infinite solid. This solution is derived below in the context of its original use by Kelvin, the estimation of the age of the earth as a solid planet from measurements of its geothermal properties.

6.2.1 Kelvin's Solution for the Cooling of the Earth

In 1864, Lord Kelvin (Sir William Thompson) presented an estimate of the age of the Earth based on the geothermal temperature gradient measured in Scotland [Kelvin, 1864; see also Carslaw and Jeager, 1988, p. 85]. He assumed that the Earth was assembled in a molten state, and began to cool by conductive heat transfer. Using an estimate for the temperature of molten rock, the present-day geothermal gradient, and the solution to the conductive heat transfer equation, discussed below, he determined the time elapsed since the Earth was in a molten state. His estimate, 94 million years, contradicted the prevailing scientific view of his day that the Earth was many Billions of years old. We now know that the earth is over 4.5 billion years old. The error in Kelvin's estimate was due to the fact that he had not considered the effects of heat generated within the Earth by radioactive decay or of convection in the mantle. The story of Kelvin's work is interesting, nevertheless, and provides a valuable insight into the working of modern science [Richter, 1986]. I review Kelvin's work here because it provides a necessary result for the derivation of the bathymetric profiles of the ocean.

Kelvin [1864] treated the Earth as a semi-infinite solid occupying $z < 0$. Conductive heat transfer in this geometry is described by the following equations [Carslaw and Jeager, 1988]:

$$\theta_t = \kappa\theta_{zz} \tag{6.1}$$

$$\theta(0, z) = \theta_o \tag{6.2}$$

$$\theta(t, 0) = \theta_s = 0 \tag{6.3}$$

$$\theta_z(t, z \to -\infty) \quad \to 0 \tag{6.4}$$

where θ is temperature, t is time, z is elevation with respect to the planar surface of the semi-infinite solid, and subscripts t and zz denote single and double partial differentiation with respect to the subscripted variable, respectively. The surface temperature θ_s is taken to be 0 C (roughly the atmospheric temperature in Scotland on a cold day).

There are several ways to solve (6.1) - (6.4) discussed in Carslaw and Jeager [1988]. We shall use the Laplace transform method [Arfken, 1970, p. 688]. First, a few words about the Laplace transform.

6.2.2 Laplace Transform

Let $\mathcal{L}(f(t)) = \tilde{f}(s)$ be the Laplace transform of $f(t)$. By definition,

$$\mathcal{L}(f(t)) = \int_0^\infty \; e^{-st} f(t) dt \tag{6.5}$$

One might wonder why the Laplace transform would be useful in solving a problem such as that defined by (6.1) - (6.4). The utility of the Laplace transform is appreciated when one considers how it transforms the time-derivative term in (6.1):

$$\begin{aligned}
\mathcal{L}(\theta_t) &= \int_0^\infty e^{-st} \frac{\partial \theta}{\partial t} dt \\
&= \int_0^\infty \frac{\partial}{\partial t} \left(e^{-st} \theta \right) dt + \int_0^\infty s e^{-st} \theta dt \\
&= -\theta_o + s\mathcal{L}(\theta) \\
&= -\theta_o + s\tilde{\theta}
\end{aligned} \tag{6.6}$$

We see that the advantage gained by applying the Laplace transform to (6.1) is that it eliminates the time-derivative term (thus, converting a partial differential equation into an ordinary differential equation where only derivatives

with respect to z appear) and folds-in the initial condition at the same time. Taking the Laplace transform of (6.1) gives:

$$s\tilde{\theta} - \theta_o = \kappa\tilde{\theta}_{zz} \tag{6.7}$$

This equation is called the subsidiary equation. Notice that it is simply a second-order, non-homogeneous ordinary differential equation for the function $\tilde{\theta}(z)$.

The Laplace-transformed boundary conditions which go along with (6.7 are written

$$\tilde{\theta}(s,0) = 0 \tag{6.8}$$

$$\tilde{\theta}_z(s, z \to -\infty) \to 0 \tag{6.9}$$

6.2.3 Solution of the Subsidiary Equation

The general solution to the subsidiary equation may be written as the sum of two independent solutions of the homogeneous form of the subsidiary equation and a particular solution which satisfies the non-homogeneous form of the subsidiary equation:

$$\tilde{\theta}(s,z) = Ae^{\sqrt{\frac{s}{\kappa}}z} + Be^{-}\sqrt{\frac{s}{\kappa}}z + \frac{\theta_o}{s} \tag{6.10}$$

The boundary conditions imply $B = 0$ and $A = -\frac{\theta_o}{s}$, thus

$$\tilde{\theta}(s,z) = \frac{\theta_o}{s} - \frac{\theta_o e^{\sqrt{\frac{s}{\kappa}}z}}{s} \tag{6.11}$$

Now that we have $\tilde{\theta}(s,z)$, our problem becomes one of inverting the Laplace transform for $\theta(t,z)$.

6.2.4 Inverse Laplace Transform

The inverse Laplace transform is defined using the so-called Bromwich integral:

$$\mathcal{L}^{-1}(\tilde{f}(s)) = f(t) = \frac{1}{2\pi i} \int_{-i\infty+\gamma}^{i\infty+\gamma} \tilde{f}(s) e^{st} ds \tag{6.12}$$

where γ is a small positive real number and $i = \sqrt{-1}$. Clearly, the Bromwich integral represents a contour integration on the complex plain. Figure (6.4) displays the path of integration associated with the Bromwich integral.

If you need to invert a Laplace transform, the first thing you try is to go to a table of inverse transforms and look up your function to see if you can avoid the tedious work of evaluating the Bromwich integral. If you are unlucky, you are forced to take on the integration of the Bromwich integral without the help of a table. In the present circumstances, we are unlucky.

6.2.5 Integrating the Bromwich Integral

Our goal is to evaluate

$$\theta(t, z) = \frac{\theta_o}{2\pi i} \int_{-i\infty+\gamma}^{i\infty+\gamma} \left(\frac{1}{s} - \frac{e^{\sqrt{\frac{s}{\kappa}} z}}{s} \right) e^{st} ds \tag{6.13}$$

One of the tricks of complex analysis at our disposal is Cauchy's integral theorem. Cauchy's theorem states that the integral of a function over any closed contour in the complex plain is identically zero when the function has no poles (singularities like $\frac{1}{s}$ as $s \to 0$) enclosed by the contour or branch cuts which cross the contour. The integrand of the above equation contains a pole at $s = 0$ and a branch cut along the negative part of the real axis. (The branch cut comes from the fact that we desire to make the function $\sqrt{s}$ single valued on the complex plain.) We may thus imagine a closed contour which contains, as one of its parts, the contour of the Bromwich integral and which avoids enclosing the poles or crossing the branch cut of the integrand in the above equation. A diagram showing this closed contour is shown in

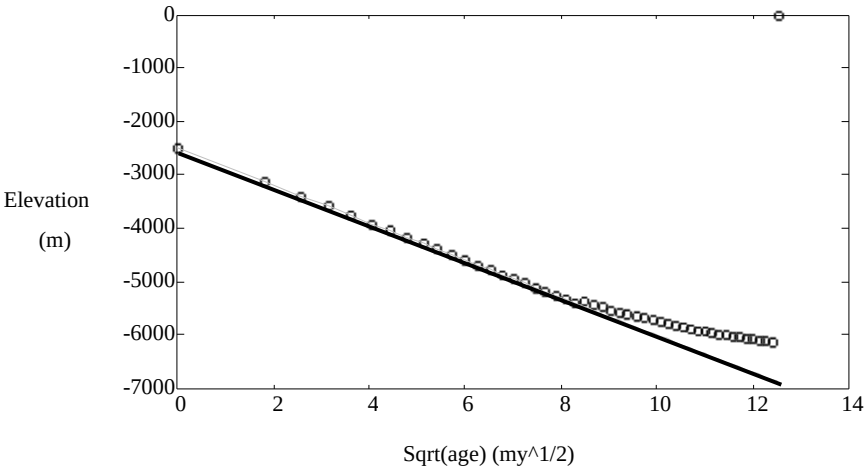

Figure 6.3: Schematic plot of depth (open circles) vs. the square-root of the geologic age. A linear relationship is demonstrated for the ocean floor that is younger than about 70-million years.

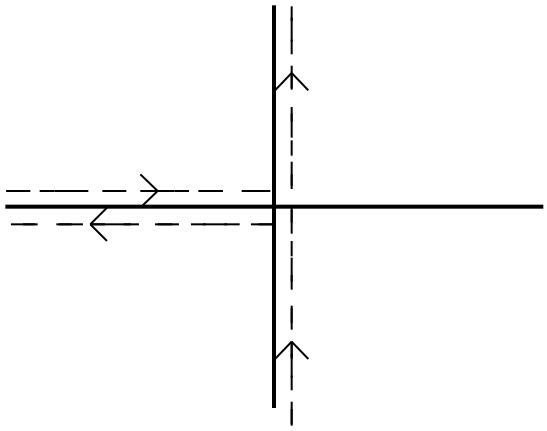

Figure 6.4: The contour required to invert the Laplace transform may, in the present problem, be deformed to the "keyhole" contour shown above.

Fig. (6.4). Observe that Cauchy's integral theorem allows us to equate the above integral which represents half of the contour integration with -1 times the integral over the "keyhole" contour which excludes the branch cut along the negative real axis and the pole at $s = 0$. We find it easier to perform the integration along this keyhole contour. The inverse Laplace transform thus reduces to

$$\begin{aligned}\theta(t,z) &= \frac{-\theta_o}{2\pi i}\Big[\int_{\infty}^{0}\left(\frac{e^{-i\pi}}{r}-\frac{e^{-i\pi}}{r}e^{e^{i\frac{\pi}{2}}\frac{z\sqrt{r}}{\sqrt{\kappa}}}\right)e^{re^{i\pi}t}dr \\ &\quad+\int_{0}^{\infty}\left(\frac{e^{i\pi}}{r}-\frac{e^{i\pi}}{r}e^{e^{-i\frac{\pi}{2}}\frac{z\sqrt{r}}{\sqrt{\kappa}}}\right)e^{re^{-i\pi}t}dr \\ &\quad+\lim_{r\to 0}\int_{-\pi}^{\pi}\left(\frac{e^{-i\phi}}{r}-\frac{e^{-i\phi}}{r}e^{e^{i\frac{\phi}{2}}\frac{z\sqrt{r}}{\sqrt{\kappa}}}\right)e^{re^{i\phi}t}d\phi\Big] \qquad (6.14)\end{aligned}$$

where we have made use of polar coordinates (r, ϕ) to represent s and $\sqrt{s}$:

$$s = re^{i\phi} \qquad (6.15)$$

$$\sqrt{s} = \sqrt{r}e^{i\frac{\phi}{2}} \qquad (6.16)$$

making note of the identities $e^{-i\pi} = -1$, $e^{-i\phi} = -1$, $e^{i\frac{\phi}{2}} = i$, and $e^{-i\frac{\phi}{2}} = -i$, the above integral over the keyhole contour is rewritten as

$$\begin{aligned}\theta(t,z) &= \frac{-\theta_o}{2\pi i}\Big[\int_{\infty}^{0}\left(\frac{-1}{r}-\frac{-1}{r}e^{i\frac{z\sqrt{r}}{\sqrt{\kappa}}}\right)e^{-rt}dr \\ &\quad+\int_{0}^{\infty}\left(\frac{-1}{r}-\frac{-1}{r}e^{-i\frac{z\sqrt{r}}{\sqrt{\kappa}}}\right)e^{-rt}dr \\ &\quad+\lim_{r\to 0}\int_{-\pi}^{\pi}\left(\frac{e^{-i\phi}}{r}-\frac{e^{-i\phi}}{r}e^{e^{i\frac{\phi}{2}}\frac{z\sqrt{r}}{\sqrt{\kappa}}}\right)e^{re^{i\phi}t}d\phi\Big] \qquad (6.17)\end{aligned}$$

We observe that the third integral term on the right-hand side of the above equation is zero when the limit of $r \to 0$ is taken. We also observe that the limits of integration on the first integral term may be reversed to give

$$\begin{aligned}\theta(t,z) &= \frac{-\theta_o}{2\pi i}\Big[\int_0^\infty \left(\frac{1}{r} - \frac{1}{r}e^{i\frac{z\sqrt{r}}{\sqrt{\kappa}}}\right)e^{-rt}dr \\ &\quad + \int_0^\infty \left(\frac{-1}{r} - \frac{-1}{r}e^{-i\frac{z\sqrt{r}}{\sqrt{\kappa}}}\right)e^{-rt}dr\Big] \\ &= \frac{\theta_o}{2\pi i}\Big[\int_0^\infty \left(\frac{e^{i\frac{z\sqrt{r}}{\sqrt{\kappa}}}}{r} - \frac{e^{-i\frac{z\sqrt{r}}{\sqrt{\kappa}}}}{r}\right)e^{-rt}dr\Big] \qquad (6.18)\end{aligned}$$

We again change coordinates using $\rho = \sqrt{r}$, $d\rho = \frac{dr}{2\sqrt{r}}$, and $dr = 2\rho d\rho$ to give

$$\begin{aligned}\theta(t,z) &= \frac{\theta_o}{2\pi i}\int_0^\infty \left(\frac{e^{i\frac{z\rho}{\sqrt{\kappa}}}}{\rho^2} - \frac{e^{-i\frac{z\rho}{\sqrt{\kappa}}}}{\rho^2}\right)e^{-\rho^2 t}2\rho d\rho \\ &= \frac{\theta_o}{\pi i}\int_0^\infty \frac{1}{\rho}\left(e^{i\frac{z\rho}{\sqrt{\kappa}}-\rho^2 t} - e^{-i\frac{z\rho}{\sqrt{\kappa}}-\rho^2 t}\right)d\rho \qquad (6.19)\end{aligned}$$

The above integral is manupulated using the following identity

$$\int_0^\infty \frac{1}{\rho}e^{\frac{-i\rho z}{\sqrt{\kappa}}-\rho^2 t}d\rho = \int_{-\infty}^0 \frac{-1}{\rho}e^{\frac{i\rho z}{\sqrt{\kappa}}-\rho^2 t}d\rho \qquad (6.20)$$

to give

$$\theta(t,z) = \frac{\theta_o}{\pi i}\int_{-\infty}^\infty \frac{1}{\rho}e^{i\frac{z\rho}{\sqrt{\kappa}}-\rho^2 t}d\rho \qquad (6.21)$$

We now define $\zeta = \frac{z}{\sqrt{4\kappa t}}$ and $x = \sqrt{t}\rho$. With these new variables, the argument of the exponential function in the integrand of the above equation becomes

$$\begin{aligned}\tfrac{iz\rho}{\sqrt{\kappa}} - t\rho^2 &= 2i\zeta\sqrt{t}\rho - t\rho^2 \\ &= 2i\zeta x - x^2 \\ &= 2i\zeta x - x^2 - \zeta^2 + \zeta^2 \\ &= (\zeta + ix)^2 - \zeta^2 \qquad (6.22)\end{aligned}$$

We also note that

$$\frac{d\rho}{\rho} = \frac{dx}{x} \tag{6.23}$$

Thus, the integral we are evaluating becomes

$$\theta(t,z) = \frac{\theta_o}{\pi i} \int_{-\infty}^{\infty} e^{(\zeta+ix)^2-\zeta^2} \frac{dx}{x} \tag{6.24}$$

This integral is too difficult to evaluate as it stands, but we can make progress towards its evaluation by considering the ζ-derivative of $\theta(t,z)$:

$$\begin{aligned} \frac{\partial\theta(t,z)}{\partial\zeta} &= \frac{\theta_o}{\pi i} \int_{-\infty}^{\infty} \left(2(\zeta+ix)-2\zeta\right) e^{(\zeta+ix)^2-\zeta^2} \frac{dx}{x} \\ &= \frac{2\theta_o}{\pi i} \int_{-\infty}^{\infty} ix e^{(\zeta+ix)^2-\zeta^2} \frac{dx}{x} \\ &= \frac{2\theta_o}{\pi} e^{-\zeta^2} \int_{-\infty}^{\infty} e^{(\zeta+ix)^2} dx \end{aligned} \tag{6.25}$$

We find that this integral for the ζ-derivative of $\theta(t,z)$ is easy to evaluate if we define two variables, u and v, such that

$$u^2 = -(\zeta+ix)^2 \tag{6.26}$$

and

$$v^2 = u^2 \tag{6.27}$$

with $du = dv = -dx$. With this change of variables, and with the identity $Y = \sqrt{YY}$, the integral for the ζ-derivative of $\theta(t,z)$ becomes

$$\begin{aligned} \frac{\partial\theta(t,z)}{\partial\zeta} &= \frac{-2\theta_o}{\pi} e^{-\zeta^2} \int_{-\infty}^{\infty} e^{-u^2} dx \\ &= \frac{-2\theta_o}{\pi} e^{-\zeta^2} \sqrt{\int_{-\infty}^{\infty} e^{-u^2} du \int_{-\infty}^{\infty} e^{-v^2} dv} \\ &= \frac{-2\theta_o}{\pi} e^{-\zeta^2} \sqrt{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(u^2+v^2)} du\ dv} \end{aligned}$$

$$= \frac{-2\theta_o}{\pi} e^{-\zeta^2} \sqrt{\int_0^{\infty} \int_0^{2\pi} e^{-r^2} r d\phi \; dr}$$
$$= \frac{-2\theta_o}{\pi} e^{-\zeta^2} \sqrt{2\pi \int_0^{\infty} \frac{-1}{2} \frac{\partial}{\partial r} \left(e^{-r^2}\right) dr}$$
$$= \frac{-2\theta_o}{\pi} e^{-\zeta^2} \sqrt{\pi} \tag{6.28}$$

We now know the ζ-derivative of $\theta(t, z)$, so it is an easy matter to determine $\theta(t, z)$ (here, we make use of the definition of the error function):

$$\begin{aligned} \theta(t, z) &= \frac{-2\theta_o}{\sqrt{\pi}} \int_0^{-\zeta} e^{-\xi^2} d\xi \\ &= \theta_o \text{erf}(-\zeta) \\ &= \theta_o \text{erf}\left(\frac{-z}{\sqrt{4\kappa t}}\right) \end{aligned} \tag{6.29}$$

Observe that the error function is antisymmetric about the $z = 0$ level.

The error function, $\text{erf}(x)$, is a well-known special function that is tabulated in various mathematical handbooks [Abramowitz and Stegun, 1964; Press *et al.*, 1989]. In particular, the student will find that it is implimented as a function in MATLAB® .

6.2.6 Geothermal Gradient

The geothermal gradient at $z = 0$ is given by the derivative of (6.29) with respect to z

$$\begin{aligned} \theta_z(t, 0) &= \frac{\partial}{\partial z} \left(\frac{2\theta_o}{\sqrt{\pi}} \int_0^{\frac{z}{\sqrt{4\kappa t}}} e^{-\xi^2} d\xi \right) \Big|_{z=0} \\ &= -\frac{\theta_o}{\sqrt{\pi \kappa t}} \end{aligned} \tag{6.30}$$

The cooling history of the upper 100 km of a semi-infinite solid with a diffusivity of $\kappa = 1.18 \times 10^{-6}$ m^2 s^{-1} occupying the region $z < 0$ is displayed in Fig. (6.5). Each curve represents $\theta(t, z)$ at 10-million year intervals starting with an initial temperature of $\theta_o(z) = 3871$ C. (The values for κ, θ_s and θ_o are taken from Kelvin's [1864] analysis.) The surface temperature is assumed constant at 0 C for the entire cooling history. Note that significant deviation from the initial temperature profile occurs in a relatively thin upper crust of the Earth, according to this model. This suggests that the cooling half-space model may be adequate for describing the early stages of a more complicated Earth model such as the model of the oceanic crust we will discuss below.

6.2.7 Kelvin's Edinburgh Calculation

Kelvin [1864] used (6.30) to estimate the age of the Earth, T_e, from measurements of the geothermal gradient made near Edinburgh, Scotland:

$$T_e = \frac{-1}{\pi\kappa}\left(\frac{\theta_o}{\theta_z(T_e, z=0)}\right)^2 \tag{6.31}$$

Using measurements to evaluate the right-hand side of (6.31), in particular $\theta_z(T_e, z = 0) = -1/27$ C m^{-1}, Kelving determined that $T_e \approx 94 \times 10^6$ years. A plot of $\theta_z(t, z)$ for the first 100-million years of the cooling history of the semi-infinite solid shown in Fig. (6.5) is displayed in Fig. (6.6).

Kelvin's [1864] analysis was flawed for two reasons. He did not account for the generation of heat within the earth due to the decay of radioactive elements, and he was unaware of convective cooling processes associated with mantle convection. Radioactivity and mantle convection were not discovered until the next century, so Kelvin had no way of knowing about these flaws. As suggested by Richter [1986], Kelvin's method, despite its flaws, was important because it represented the first time the laws of physics were applied to something so large and seemingly inscrutible as the Earth.

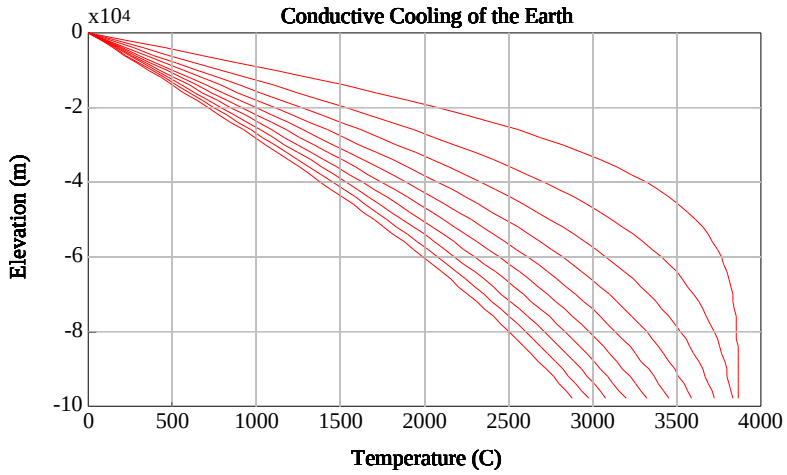

Figure 6.5: Conductive cooling of a semi-infinite solid occupying $z < 0$.

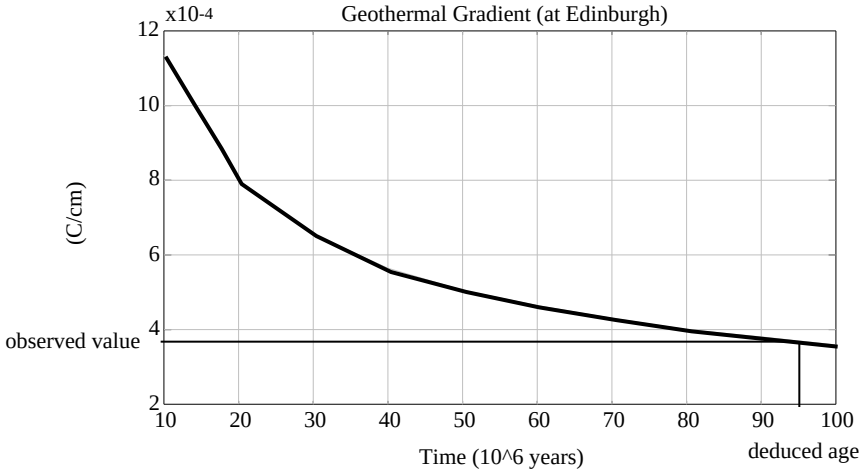

Figure 6.6: History of geothermal gradient at $z = 0$ for a semi-infinite solid occupying $z < 0$. Values of $\theta_z(T_e, z = 0)$ and T_e described by Kelvin [1864] in his determination of the age of the Earth are indicated by the lines.

6.3 Theory of Sea-Floor Subsidence

We are now ready to develop a theory which explains the bathymetry of the ocean. Following Davis and Lister [1974], we adopt (6.29) as a satisfactory approximation to the temperature profile of a column of oceanic crust as it cools from its initial molten state. We assume that subsidence $\Delta d(t)$ of the sea floor from its initial elevation at the mid-ocean ridge $d(t = 0)$ is determined by two processes: thermal contraction, and isostatic depression due to increasing water load above the subsiding sea floor.

Thermal contraction $\Delta h(t)$ is related to $\theta(t, z)$ by the thermal expansion coefficient α

$$\Delta h(t) = -\alpha\theta_o \int_{-\infty}^{0} \left(1 - \mathrm{erf}(\frac{-z}{2\sqrt{\kappa t}})\right) \mathrm{d}z \tag{6.32}$$

Here the role of the integral is to sum the temperature change over what we take to be the infinite depth of the oceanic crust. We know, of course, that the oceanic crust is of limited thickness. The minus sign appears in (6.32) due to the fact that the temperature of the oceanic crust is cooling with time, and is thus contracting vertically. We use the expression for an infinitely thick crust here because we know that during the brief time interval oceanic crust actually resides on the surface of the earth (up to about 200 million years), there is little difference between the heat lost from a plate and that lost from a semi-infinite solid. We thus avoid the complexity of dealing with finite thickness by taking $-\infty$ as the upper limit on the integral of (6.32). Before evaluating this integral, we consider the effect of isostatic depression.

Isostatic depression due to sea-water loading, $\Delta g(t)$, is determined from $\Delta d(t)$ (to be determined later) by assuming that deep below the Earth's surface there exists a horizontal *compensation level* that is parallel to the sea-surface (*i.e.,* the geoid). Gravitational equilibrium requires that the the total mass of water *and* oceanic crust above is a constant that is independent of location. In other words,

$$\rho_m \Delta g(t) = \rho_w \Delta d(t) \tag{6.33}$$

where ρ_w and ρ_m are the densities of seawater and mantle material, respectively, and $\rho_w \Delta d(t)$ is the extra load caused by sea-water filling the void

caused by the thermal contraction of the oceanic crust. The net change in ocean depth $\Delta d(t)$ is the sum of $\Delta h(t)$ and $\Delta g(t)$. This gives

$$\Delta d(t) = \frac{1}{1 - \rho_w/\rho_m} \Delta h(t) \tag{6.34}$$

We are now ready to determine $\Delta d(t)$ by evaluating the integral in (6.32). This is somewhat tricky and is done as follows. First, change the variable of integration from z to $x = -z/(2\sqrt{\kappa t})$

$$\begin{aligned} \int_{-\infty}^{0} \left(1 - \operatorname{erf}\left(\tfrac{-z}{\sqrt{4\kappa t}}\right)\right) dz &= -2\sqrt{\kappa t} \int_{\infty}^{0} (1 - \operatorname{erf}(x))\, dx \\ &= 2\sqrt{\kappa t} \int_{0}^{\infty} (1 - \operatorname{erf}(x))\, dx \end{aligned} \tag{6.35}$$

We next make use of the definition of the complementary error function ($\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-\xi^2} d\xi$)

$$\begin{aligned} 2\sqrt{\kappa t} \int_{0}^{\infty} (1 - \operatorname{erf}(x))\, dx &= 2\sqrt{\kappa t} \int_{0}^{\infty} \operatorname{erfc}(x) dx \\ &= \frac{4\sqrt{\kappa t}}{\sqrt{\pi}} \int_{0}^{\infty} \int_{x}^{\infty} e^{-\xi^2} d\xi\, dx \end{aligned} \tag{6.36}$$

We recognize that the domain of integration in the above double integral is the wedge contained within the region of the first quadrant of the (ξ, x)-plain enclosed by the positive ξ axis and the line $x = \xi$. We can reverse the order of integration, without changing this domain of integration, to obtain a simplification:

$$\begin{aligned} \frac{4\sqrt{\kappa t}}{\sqrt{\pi}} \int_{0}^{\infty} \int_{x}^{\infty} e^{-\xi^2} d\xi\, dx &= \frac{4\sqrt{\kappa t}}{\sqrt{\pi}} \int_{0}^{t} \int_{0}^{\infty} e^{-\xi^2} d\xi\, dx \\ &= \frac{4\sqrt{\kappa t}}{\sqrt{\pi}} \int_{0}^{\infty} \xi e^{-\xi^2} d\xi \end{aligned}$$

$$
\begin{aligned}
&= \frac{4\sqrt{\kappa t}}{\sqrt{\pi}} \int_0^\infty \frac{-1}{2} \frac{\partial}{\partial \xi} \left(e^{-\xi^2} \right) d\xi \\
&= \frac{-2\sqrt{\kappa t}}{\sqrt{\pi}} e^{-\xi^2} \Big|_0^\infty \\
&= \frac{2\sqrt{\kappa t}}{\sqrt{\pi}}
\end{aligned} \tag{6.37}
$$

We thus achieve the following expression for $\Delta d(t)$:

$$
\Delta d(t) = \frac{-2\rho_m \alpha \theta_o \sqrt{\kappa t}}{\sqrt{\pi}(\rho_m - \rho_w)} \tag{6.38}
$$

where $\Delta d(t) < 0$ denotes *increasing* depth. Assuming a mid-ocean ridge depth of d_m, the depth function $d(t)$ can be written

$$
d(t) = d_m + \frac{-2\rho_m \alpha \theta_o \sqrt{\kappa t}}{\sqrt{\pi}(\rho_m - \rho_w)} \tag{6.39}
$$

(Again, remember that depths are intended to be negative numbers, thus $d_m < 0$ and $d(t)$ will become increasingly negative as $t \to \infty$. You will be asked to find the parameters α and θ_o which give the best fit to ocean bathymetric data in the laboratory exercises associated with this chapter.

6.4 A Plate Model of Oceanic Crust

The conductive cooling model presented in the previous section describes the ocean bathymetry for young (less than 70 million years) oceanic crust with reasonable accuracy (see Fig. 6.3). For older crust, the actual ocean depth is more shallow than that predicted by (6.39). This inaccuracy is a consequence of a thermal-cooling model that is too simple. Parsons and Sclater [1977] proposed that the oceanic crust should be modeled as a plate of fixed thickness, and that it should sit above an astheonsphere that has fixed temperature due to vigorous mantle convection. The advantage of Parsons and Sclater's model is that it captures the asymptotic behavior of ocean

depth as the age becomes very large without sacrificing the ability to explain the depth of young oceanic crust.

Parsons and Sclater [1977] were concerned primarily with the asymptotic behavior of the ocean crust after it has cooled for a very long time. They thus considered the oceanic crust to be a plate of fixed final thickness a_o to be reached as $t \rightarrow \infty$. The geometry of this plate is summarized in Fig. (6.7).

The equations which govern the conductive cooling of this plate are

$$\theta_t = \kappa \theta_{zz} \qquad -a(t) < z < 0 \tag{6.40}$$

$$\theta(0, z) = \theta_o \qquad -a(t) < z < 0 \tag{6.41}$$

$$\theta(t, 0) = \theta_s = 0 \tag{6.42}$$

$$\theta(t, -a(t)) = \theta_o \tag{6.43}$$

where $a(t)$ is the plate thickness. To account for the changing thickness of the plate, it is convenient to adopt a stretched vertical coordinate $\zeta = z/a(t)$ so that the domain of (6.40 - 6.43 can be treated as the fixed interval $0 > \zeta > -1$. To perform this coordinate transformation on the governing equations, we note that

$$\frac{\partial}{\partial t} \rightarrow \frac{\partial}{\partial t} - \frac{\dot{a}\zeta}{a}\frac{\partial}{\partial \zeta} \tag{6.44}$$

and

$$\frac{\partial^2}{\partial z^2} \rightarrow \frac{1}{a^2}\frac{\partial^2}{\partial \zeta^2} \tag{6.45}$$

where $\dot{a}$ is the time derivative of a. To simplify the above equations, we adopt a non-dimensional time variable

$$t \rightarrow \frac{a^2}{\kappa} t \tag{6.46}$$

$$\theta \rightarrow \theta_o \theta \tag{6.47}$$

and note that $\frac{\dot{a}}{a} << \frac{\kappa}{a^2}$ for our problem. These simplifications allow us to rewrite (6.40) - (6.43) as

$$\theta_t = \theta_{\zeta\zeta} \qquad -1 < \zeta < 0 \tag{6.48}$$

$$\theta(0,\zeta) = 1 \qquad -1 < \zeta < 0 \tag{6.49}$$

$$\theta(t,0) = 0 \tag{6.50}$$

$$\theta(t,-1) = 1 \tag{6.51}$$

Equations (6.48) - (6.51) are readily solved by the separation of variables method. First we note that the asymptotic solution when $t \to \infty$ is $\theta \to -\zeta$. The full solution can be written as the sum of this asymptotic, steady-state solution and a transient solution, $\tilde{\theta}$ which satisfies homogenous (=0) boundary conditions and a slightly different initial condition ($=1+\zeta$)

$$\tilde{\theta} = T(t)Z(\zeta) \tag{6.52}$$

Equation (6.48) becomes

$$\frac{T'}{T} - \frac{Z''}{Z} = 0 \tag{6.53}$$

where primes denote differentiation. Noting that (6.53) requires that a function of t only (T'/T) cancel a function of ζ only (Z''/Z), we must conclude that both terms in (6.53) must be scalar quantities. In other words,

$$\frac{T'}{T} = \lambda = \frac{Z''}{Z} \tag{6.54}$$

Solutions Z_n which satisfy the homogeneous boundary conditions are of the form

$$Z_n(\zeta) = b_n \sin(n\pi\zeta) \qquad n = 1, \ldots, \infty \tag{6.55}$$

Corresponding solutions T_n are of the form

$$T_n(t) = e^{-(n\pi)^2 t} \qquad n = 1, \ldots, \infty \tag{6.56}$$

The full solution may be written as linear combinations of the $T_n(t) \cdot Z_n(\zeta)$'s:

$$\theta(t,\zeta) = -\zeta + \sum_{n=1}^{\infty} b_n e^{-(n\pi)^2 t} \sin(n\pi\zeta) \tag{6.57}$$

The constants $\{b_n, n = 1, \ldots, \infty\}$ may be evaluated by enforcing the initial condition $\theta(0, \zeta) = 1$, which implies that

$$\begin{aligned} \tilde{\theta}(0,\zeta) = &\quad 1 + \zeta \\ = &\quad \sum_{n=1}^{\infty} b_n \sin(n\pi\zeta) \qquad -1 < \zeta < 0 \end{aligned} \tag{6.58}$$

The b_n's are evaluated by standard Fourier series techniques. First, we note that

$$b_n = 2 \int_{-1}^{0} (1 + \zeta) \sin(n\pi\zeta) \mathrm{d}\zeta \tag{6.59}$$

The integrand may be broken into two terms which are readily integrated:

$$\begin{aligned} 2 \int_{-1}^{0} \sin(n\pi\zeta) \mathrm{d}\zeta = &\quad \frac{-2}{n\pi} \cos(n\pi\zeta)|_{-1}^{0} \\ = &\quad \frac{-2}{n\pi}(1 - (-1)^n) \\ = &\quad \begin{cases} \frac{-4}{n\pi} & \text{if n is odd} \\ 0 & \text{if n is even} \end{cases} \end{aligned} \tag{6.60}$$

for $n = 1, \ldots, \infty$. Also,

$$\begin{aligned} & 2 \int_{-1}^{0} \zeta \sin(n\pi\zeta) \mathrm{d}\zeta \\ = &\quad 2 \int_{-1}^{0} \mathrm{d} \left(\frac{-\zeta \cos(n\pi\zeta)}{n\pi} \right) - 2 \int_{-1}^{0} \frac{-\cos(n\pi\zeta)}{n\pi} \mathrm{d}\zeta \\ = &\quad \frac{-2}{n\pi}(\zeta \cos(n\pi\zeta))|_{-1}^{0} - \frac{-2}{(n\pi)^2} \sin(n\pi\zeta)|_{-1}^{0} \\ = &\quad \begin{cases} \frac{2}{n\pi} & \text{if n is odd} \\ \frac{-2}{n\pi} & \text{if n is even} \end{cases} \end{aligned} \tag{6.61}$$

Combining the intermediate results presented in (6.60) and (6.61) we find that

$$b_n = \frac{-2}{n\pi} \qquad n = 1, \ldots, \infty \tag{6.62}$$

Thus the complete solution to the plate model is

$$\theta(t,\zeta) = -\zeta + \sum_{n=1}^{\infty} \frac{-2}{n\pi} e^{-(n\pi)^2 t} \sin(n\pi\zeta) \tag{6.63}$$

In dimensional form (recall (6.46) and (6.47)), this expression is

$$\theta(t,\zeta) = \theta_o \left(-\frac{z}{a} + \sum_{n=1}^{\infty} \frac{-2}{n\pi} e^{\frac{-(n\pi)^2 \kappa t}{a^2}} \sin(\frac{n\pi z}{a}) \right) \tag{6.64}$$

The geothermal heat flux at $z = 0$, $q(t,0)$, is readily determined by taking the z-derivative of $\theta(t,z)$

$$q(t,0) = -k\theta_o \left(-\frac{1}{a} + \sum_{n=1}^{\infty} \frac{-2}{a} e^{\frac{-(n\pi)^2 \kappa t}{a^2}} \right) \tag{6.65}$$

where k is the thermal conductivity of the oceanic crust.

The thermal subsidence is again determined by summing a thermal contraction contribution and an isostatic depression contribution.

$$\Delta h(t) = \alpha\theta_o \int_{-a}^{0} \left(1 + \frac{z}{a} - \sum_{n=1}^{\infty} \frac{-2}{n\pi} e^{\frac{-(n\pi)^2 \kappa t}{a^2}} \sin(\frac{n\pi z}{a}) \right) \mathrm{d}z \tag{6.66}$$

This expression is easily evaluated by noting that

$$\int_{-a}^{0} \left(1 + \frac{z}{a} \right) \mathrm{d}z = \frac{a}{2} \tag{6.67}$$

and

$$\int_{-a}^{0} \frac{-2}{n\pi} \sin(\frac{n\pi z}{a}) \mathrm{d}z = \begin{cases} 0 & \text{if n is even} \\ \frac{4a}{(n\pi)^2} & \text{if n is odd} \end{cases} \tag{6.68}$$

The result is

$$\Delta h(t) = \frac{\alpha\theta_o a}{2} \left(1 - \sum_{n \text{ odd}} \frac{8}{(n\pi)^2} e^{\frac{-(n\pi)^2 \kappa t}{a^2}} \right) \tag{6.69}$$

Making use of (**??**), we derive the depth anomaly

$$\Delta d(t) = \frac{\alpha\rho_m\theta_o a}{2(\rho_m - \rho_w)} \left(1 - \sum_{n \text{ odd}} \frac{8}{(n\pi)^2} e^{\frac{-(n\pi)^2 \kappa t}{a^2}} \right) \tag{6.70}$$

We note that the asymptotic subsidence at $t \to \infty$ is given by (this is the result when all the exponential terms in the sum have decayed to zero)

$$\Delta d_s = \frac{\alpha \rho_m \theta_o a}{2(\rho_m - \rho_w)} \tag{6.71}$$

Thus,

$$d(t) = d_m + \Delta d_s \left(1 - \sum_{n \text{ odd}} \frac{8}{(n\pi)^2} e^{\frac{-(n\pi)^2 \kappa t}{a^2}} \right) \tag{6.72}$$

We note as a reminder that (6.72) is approximate in the sense that we did not account for the fact that a changes with time. This change, as argued previously, is so small compared to the size of a (typically 100 km or so), that the approximation is satisfactory for practical application

Parsons and Sclater [1977] demonstrated that the plate model for ocean crust subsidence was superior to the semi-infinite solid model derived by Davis and Lister [1974] because it captured the otherwise anomalous behavior of the ocean floor at large geologic ages shown in Fig. (6.3) without losing the satisfactory attributes of the $\sqrt{t}$-dependence for young ages. The advantage of the plate model is reflected in the fact that as the ocean crust ages, it becomes less like a semi-infinite body and more like a plate with a finite amount of heat to be dissipated. Eventually, the plate is able to attain a steady-state temperature depth profile (the linear term in (6.64)). Thus at great age, the plate reaches a constant asymptotic elevation, and this is in agreement with the very old ocean crust in Fig. (6.3).

Parsons and Sclater [1977] suggested that two simple empirical formulae could be derived from the solution (6.72)

$$d(t) = 2500 + 350\sqrt{t} \ \text{ m} \qquad \text{for } 0 < t < 70\text{m.y.} \tag{6.73}$$

and

$$d(t) = 6400 - 3200 \ e^{\frac{-t}{62.8}} \ \text{ m} \qquad \text{for } t > 70\text{m.y.} \tag{6.74}$$

These results perform reletively well in explaining the depth/age relationships for the ocean floor around the Earth. Recent revisions of the Parsons and Sclater model [Stein and Stein, 1992] suggest that improvements to the Parsos and Sclater model can be made by using inverse methods to fit the expressions

for thermal subsidence and geothermal heat flow in (6.72) and (6.65) to the observations from the world ocean. Parameters to be fit include θ_o, α, and a. This will be the objective in Lab 5.

6.5 Bibliography

Abramowitz, M. and I. A. Stegun, 1964. *Handbook of Mathematical Functions*, (National Bureau of Standards, Applied Mathematics Series, 55, U. S. Government Printing Office, Washington, DC, 1046 pp.)

Carslaw, H. S. and J. C. Jaeger, 1988. *Conduction of Heat in Solids.* (Clarendon Press, Oxford, 510 pp.)

Davis, E. E. and C. R. B. Lister, 1974. Fundamentals of ridge crest topography, *Earth and Planetary Science Letters*, **21**, 405-413.

Kelvin, 1864. The secular cooling of the Earth, *Transactions of the Royal Society of Edinburgh*, **23**, 157.

Menard, H. W., 1986. *The Ocean of Truth.* (Princeton University Press, Princeton, 351 pp.)

Parsons, B. and J. G. Sclater, 1977. An analysis of the variation of ocean floor bathymetry and heat flow with age, *Journal of Geophysical Research*, **82**, 803-827.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1989. *Numerical Recipes.* (Cambridge University Press, Cambridge, U.K., 702 pp.)

Richter, F. M., 1986. Kelvin and the Age of the Earth, *Journal of Geology*, **94**, 395-401.

Stein, C. A. and S. Stein, 1992. A model for the global variation in oceanic depth and heat flow with lithospheric age. *in press.*

6.6 Lab: Fitting an Oceanic Crust Model to Oceanic Depth and Heat Flow Data

Stein and Stein [1992] revisited the problem of fitting an oceanic plate model to ocean bathymetry and heat flow data in order to obtain a better understanding of the *anomalies* which occur in regions (such as the Darwin Rise in the Pacific Ocean) where the heat-flow process may be more complicated than that which is treated by the Parsons and Sclater [1977] model. We will repeat their analysis here:

6.6.1 Cooling Half Space Model

Define the performance index [Stein and Stein, 1992], J, using the N observed values of depth, $\{d_i, i = 1, \ldots, N\}$, and M observed values of heat flow, $\{q_i, i = 1, \ldots, M\}$, of a combined North Atlantic/North Pacific data set (these data are provided as vectors **d** and **q** in the MATLAB® data file associated with this lab)

$$J = \frac{1}{N} \sum_{i=1}^{N} \frac{(d_i - \hat{d}_i)^2}{\sigma_{d_i}^2} + \frac{1}{M} \sum_{j=1}^{M} \frac{(q_j - \hat{q}_j)^2}{\sigma_{q_j}^2} \tag{6.75}$$

where variables with ˆ-s denote model-predicted quantities, $\{\sigma_{d_i}, i = 1, \ldots, N\}$ are the standard deviations for the depth data, and $\{\sigma_{q_i}, i = 1, \ldots, M\}$ are the standard deviations for the heat flux data (provided as MATLAB® vectors **sigmad** and **sigmaq**).

Problem 1. Plot $\{d_i, i = 1, \ldots, N\}$ and $\{q_i, i = 1, \ldots, M\}$ as a function of the age of the oceanic crust at the location where the data were measured. (Ages for **d** and **q** are included as MATLAB® variables **aged** and **ageq**.) Include error-bars on the plots (use the MATLAB® error-bar plotting function).

Problem 2. Fit the semi-infinite solid model of the oceanic crust presented in the previous section to the sub-set of N' and M' depth and heat flow

data elements which are associated with an age less than 70-million years. To make this task easier, convert the non-linear relations between the model unknowns, α and θ_o, and the depth and heat-flow variables, $d(t)$ and q(t), given by (**??**) and (**??**), to linear relations using $\alpha = \alpha^{(0)} + \Delta\alpha$ and $\theta_o = \theta_o^{(0)} + \Delta\theta_o$. Assume,

$$d(t) \approx d_m - \frac{2\rho_m \theta_o^{(0)} \alpha^{(0)} \sqrt{\kappa t}}{\sqrt{\pi}(\rho_m - \rho_w)} (1 + \frac{\Delta\alpha}{\alpha^{(0)}} + \frac{\Delta\theta_o}{\theta_o^{(0)}}) \tag{6.76}$$

$$q(t) \approx \frac{k\theta_o^{(0)}}{\sqrt{\pi\kappa t}} (1 + \frac{\Delta\theta_o}{\theta_o^{(0)}}) \tag{6.77}$$

where $\alpha^{(0)} = 3.28 \times 10^{-5}$ (C^{-1}) and $\theta_o^{(0)} = 1333$ (C) are the estimates of α and θ_o derived by Parsons and Sclater [1977] in their earlier study of ocean bathymetry. The quantities $\Delta\theta_o$ and $\Delta\alpha$ are the unknowns to be determined by fitting the model (6.76) and (6.77) to the data according to the performance index expressed in (6.75) (suitably redefined to account only for data associated with ages less than 70-million years). Use $\kappa = 8.047 \times 10^{-7}$ m^2 s^{-1}, $\rho_m = 3330$ kg m^{-3}, $k = 3.138$ W m^{-1} C^{-1}, $\rho_w = 1000$ kg m^{-3}, and $d_m = 2600$ m [Stein and Stein, 1992]. Graph your results by plotting the data as discrete points and the best-fitting model as a solid line.

6.6.2 Cooling Plate Model

Noting the poor fit between model and data in Problem 2 for data associated with ocean crustal ages over 70-million years, we abandon the cooling half-space model in favour of the cooling plate model derived in the previous section.

Problem 3. The heat flux and ocean depth predicted by the plate model involve infinite sums of terms which have factors which decrease with increasing summation index. Determine the upper cut-off value for the summation index. Use as as criterion for this determination the idea that the next term in the series beyond the cut-off term would change the truncated series, if it were included in the sum, by less than a factor of 10^{-8}. Assume that all data will involve ages greater than 5×10^5 years.

Problem 4. Repeat the model-fitting analysis of Problem 2, except use all the data and the plate model represented by (6.65) and (6.72). The additional unknown to be included in your analysis is Δa which is the first-order correction to an assumed oceanic plate thickness of $a_o = 125 \times 10^3$ m. Plot your results. Discuss any systematic deviations between your best-fit ocean bathymetry and the data.

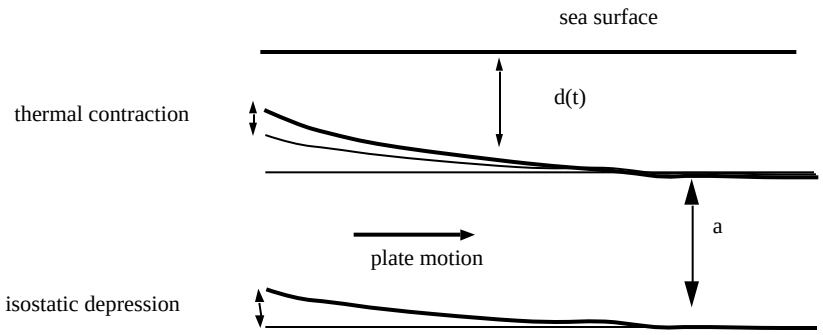

Figure 6.7: Schematic plot of Parsons and Sclater's [1977] oceanic plate geometry.

Chapter 7

Borehole Paleothermometry

7.1 Overview

Success in the search for an accurate paleothermometer has come from application of inverse methods to unusual observations. The CLIMAP project (CLIMAP, 1976), for example, mapped the sea-surface temperature (SST) of the world ocean during the last glacial maximum period (approximately 18,000 years ago) using observations of plankton in the sediments dredged from the ocean floor. The surface temperature on the Greenland and Antarctic ice sheets is reconstructed for past times using the oxygen-isotope composition of ice measured in ice cores. Here we examine the problem of detecting past climate change through the measurement and analysis of temperature profiles in rock boreholes.

This problem is of considerable interest because it provides a means to suppliment historical temperature data (*i.e.,* surface temperature recorded daily at meteorological stations) over land surfaces. By gaining a more complete record of this history, it may be possible to determine whether the climate over land surfaces has warmed during the last century. As discussed by Pollack and Chapman (1993), the rise of atmospheric CO_2 during the last 200 years is expected to have produced some warming over the last century.

A problem arises when one tries to check this prediction because most of the historical temperature data come from a relatively few locations that are biased towards Northern Hemisphere centers of population. This problem appears to have been overcome by the analysis of temperature profiles in boreholes of approximately 100 m depth which are routinely measured in remote parts of the globe (Pollack and Chapman, 1993; Lewis, 1992).

We will examine the borehole paleothermometry problem here because it gives insight into the benefits and drawbacks of least-squares inverse methods applied to physical systems which involve diffusion. As we shall see, the diffusive nature of heat transfer in the upper crust of the earth limits the ability to detect past climate change. Another reason for studying the paleothermometry problem is that the solution may be developed with finite-difference representation and with continuous representation in parallel. This parallel development helps to illustrate the underlying similarity between the mathematics of integral equations and that of linear algebra.

7.2 An Ideal Borehole Paleothermometry Problem

Let $\theta(z,t)$ denote the deviation of the temperature profile from steady state in an infinite-half space earth (Figure 7.1) subject to a surface temperature history $T_s(t)$ which is uniform at 0° C for time $t < 0$ and which is non-zero and irregular for $0 < t < t_f$. Suppose that this temperature profile is observed at $t = t_f$, and the function representing the temperature-depth observation is $\theta_b(z)$. (We assume, for this example, that it is possible to measure the temperature profile in the infinite-half space earth.) The paleothermometry problem is succinctly stated as follows: determine $T_s(t)$ for $0 < t < t_f$ from $\theta_b(z)$.

Many physical factors important in geothermal heat flow are disregarded in this simple, idealized problem. For example, groundwater may introduce a convective element of heat transport in porous rocks and soils. Likewise, groundwater in polar regions may freeze (forming permafrost) thereby lib-

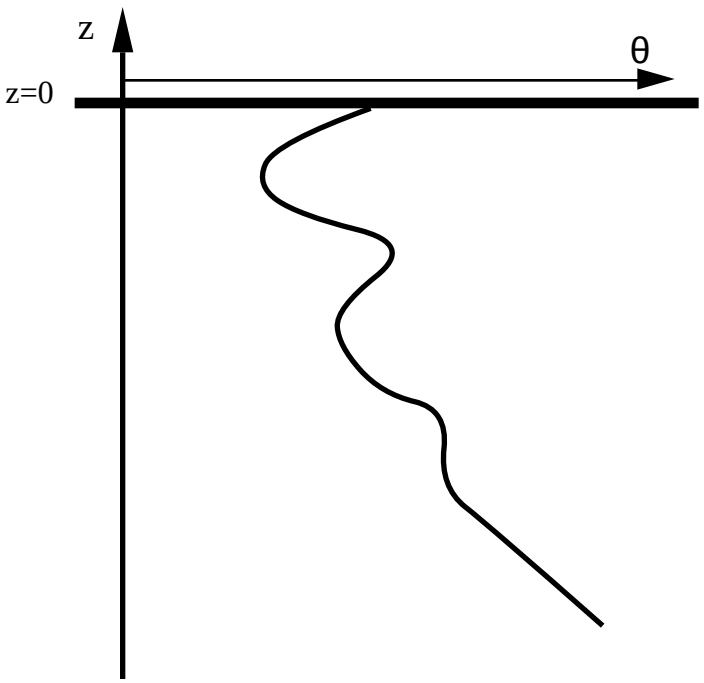

Figure 7.1: A temperature-depth profile $\theta_b(z)$ measured in an ideal half-space earth. Departure of $\theta_b(z)$ from the steady-state geothermal gradient is assumed to result entirely from past surface temperature changes at $z = 0$.

erating latent heat. An additional consideration is the presense of surface snow or standing water. A few centimeters of snow cover is often capable of insulating the ground surface from extremely cold atmospheric temperatures. We disregard these, and other, considerations to avoid complexity which might obscure the inverse methods introduced here.

7.3 Solution of the Forward Problem

The equations which govern the evolution of $\theta(z,t)$ are (Carslaw and Jeager, 1988)

$$\begin{aligned} \theta_t &= \kappa\theta_{zz} \quad z<0, 0<t<t_f && (7.1)\\ \theta(z,t=0) &= 0 && (7.2)\\ \theta_z(z\to-\infty,t) &= 0 && (7.3)\\ \theta(z=0,t) &= T_s(t) \quad 0<t<t_f && (7.4) \end{aligned}$$

For simplification, we nondimensionalize the above equations by adopting new time and vertical distance coordinates that have been suitably scaled, it i.e.,

$$\begin{aligned} t &= Tt' && (7.5)\\ z &= Zz' && (7.6) \end{aligned}$$

If we choose $T = Z^2/\kappa$, then (7.1)-(7.4) are simplified by virtue of the fact that the thermal diffusivity κ no longer appears in Eqn. (7.1):

$$\begin{aligned} \theta_t &= \theta_{zz} \quad z<0, 0<t<t_f && (7.7)\\ \theta(z,t=0) &= 0 && (7.8)\\ \theta_z(z\to-\infty,t) &= 0 && (7.9)\\ \theta(z=0,t) &= T_s(t) \quad 0<t<t_f && (7.10) \end{aligned}$$

where the primes on t and z have been dropped for notational simplicity. Effectively, by re-scaling t and z using Eqns. (7.5) and (7.6), we are able to define a more convienent way of measuring time and space (a stopwatch and ruler with more convienent units of measurement).

7.3.1 Green's Function Approach

The solution $\theta(z, t_f)$ to Eqns. (7.7)-(7.10) is formally determined by use of a Green's function $G(z, t; \xi)$ which satisfies the following "backward" heat-diffusion problem:

$$\begin{aligned} G_t &= -G_{zz} && z < 0, 0 < t < t_f && (7.11)\\ G(z, t = t_f) &= -\delta(z - \xi) && && (7.12)\\ G_z(z \to -\infty, t) &= 0 && && (7.13)\\ G(z = 0, t) &= 0 && 0 < t < t_f && (7.14) \end{aligned}$$

where $\delta(z - \xi)$ is the delta function. Notice that Eqn. (7.11) represents the adjoint form of Eqn. (7.7). Equations (7.11)-(7.14) are referred to as the backward heat-diffusion problem because the transformation $\tau = t_f - t$ yields the following problem which is equivalent to Eqns. (7.7)-(7.10):

$$\begin{aligned} G_\tau &= G_{zz} && z < 0, 0 < \tau < t_f && (7.15)\\ G(z, \tau = 0) &= -\delta(z - \xi) && && (7.16)\\ G_z(z \to -\infty, \tau) &= 0 && && (7.17)\\ G(z = 0, \tau) &= 0 && 0 < \tau < t_f && (7.18) \end{aligned}$$

Notice that the function $G(z, t; \xi)$ has three arguments. The argument which follows the semicolon denotes the location where $G(z, t_f; \xi)$ has a non-zero (infinite) value at $t = t_f$.

The expression which gives $\theta(z, t)$ in terms of $G(z, t; \xi)$ is found by manipulating the following integral

$$0 = \int_0^{t_f} \int_{-\infty}^{0} \left[\theta(G_t + G_{zz}) + G(\theta_t - \theta_{zz})\right] \, dz dt \tag{7.19}$$

Performing integration by parts, we get

$$0 = \int_{-\infty}^{0} \theta G \, |_0^{t_f} \, dz + \int_0^{t_f} \kappa(\theta G_z - G\theta_z)|_{-\infty}^{0} dt \tag{7.20}$$

Consequently, using boundary conditions,

$$0 = -\int_{-\infty}^{0} \theta(z, t_f) \cdot \delta(z - \xi)\ \mathrm{d}z + \int_{0}^{t_f} T_s(t) G_z(0, t; \xi)\ dt \tag{7.21}$$

Thus,

$$\theta(\xi, t_f) = \int_{0}^{t_f} T_s(t) G_z(0, t; \xi) \mathrm{d}t \tag{7.22}$$

The solution of the forward problem at any time t_f is the integral transform of the surface-temperature history expressed by Eqn. (7.22). The Green's function $G_z(0, t; \xi)$ is used to convey the essence of the heat-transfer *physics* that governs this particular problem. What is important to realize at this point, is that the solution to *any* linear heat-diffusion problem can be described in terms of an integral equation like Eqn. (7.22). This property will provide an important advantage below when we describe the inverse heat-diffusion problem.

7.3.2 Determination of $G_z(0, t; \xi)$ using Duhammel's Theorem

To determine $G_z(0, t; \xi)$, we make use of the analysis of the previous chapter concerning the sea-floor cooling problem:

$$Q_t = Q_{zz} \quad z < 0, t > 0 \tag{7.23}$$

$$Q(z, t = 0) = 0 \tag{7.24}$$

$$Q_z(z \to -\infty, t) = 0 \tag{7.25}$$

$$Q(z = 0, t) = 1 \quad t > 0 \tag{7.26}$$

The solution of this problem was found to be

$$Q(z, t) = \mathrm{erf}\left(\frac{-z}{2\sqrt{t}}\right) \tag{7.27}$$

We also make use of Duhammel's theorem (Carslaw and Jeager, 1988, p. 31; p. 62) to define the Green's function in terms of the analytic expression for $Q(z, t)$.

Theorem 7.1 (Duhammel)

$$\theta(z,t) = \int_0^t T_s(t')Q_t(z,t-t')dt'$$

where $Q(z,t-t') = \text{erf}\left(\frac{-z}{2\sqrt{t-t'}}\right)$

Proof. To show the above identity, consider the following three heat-diffusion problems for the temperature fields $R(z,t)$, $S(z,t)$ and $T(z,t)$ in the domain $z<0$, $t>0$:

$$\begin{array}{ccc} R_t = R_{zz} & S_t = S_{zz} & T_t = T_{zz} \\ R(0,t) = \begin{cases} 0 & t<t' \\ 1 & t \geq t' \end{cases} & S(0,t) = \begin{cases} 0 & t<t'+dt' \\ 1 & t \geq t'+dt' \end{cases} & T(0,t) = \begin{cases} 0 & t<t' \\ 1 & t' \leq t \leq t'+dt' \\ 0 & t>t'+dt' \end{cases} \\ R(z,0) = 0 & S(z,0) = 0 & T(z,0) = 0 \\ R_z(-\infty,t) = 0 & S_z(-\infty,t) = 0 & T_z(-\infty,t) = 0 \end{array}$$

Galilean invarience allows us to recognize the fact that the solutions $S(z,t)$ and $R(z,t)$ are related to the solution $Q(z,t)$. In fact, $S(z,t)$ and $R(z,t)$ are the same as $Q(z,t)$ if the time variable in S and R is replaced by $t-t'$ and $t-t'-dt'$, respectively:

$$S(z,t) = Q(z,t-t') \quad R(z,t) = Q(z,t-t'-dt')$$

Linearity of the heat-diffusion operator and boundary conditions allows us to compose a solution for $T(z,t)$ by taking the difference between $S(z,t)$ and $R(z,t)$:

$$T(z,t) = S(z,t) - R(z,t) = Q(z,t-t') - Q(z,t-t'-dt') \approx Q_t(z,t-t')dt' \tag{7.28}$$

Linearity also permits us to multiply $T(z,t)$ by $T_s(t')$ to yield a solution to the following heat-diffusion problem,

$$T_t = T_{zz}$$

$$T(0,t) = \begin{cases} 0 & t < t' \\ T_s(t') & t' \leq t \leq t' + dt' \\ 0 & t > t' + dt' \end{cases}$$
$$T(z,0) = 0$$
$$T_z(-\infty, t) = 0$$

The solution of this problem is $T(z,t) = T_s(t')Q_t(z, t-t')$. Again, linearity allows us to superimpose the solutions to the above problem for a continuous range of t' to compose $\theta(z,t)$:

$$\theta(z,t) = \int_0^t T_s(t')Q_t(z, t-t')dt' \tag{7.29}$$

which is the desired result. ■

Using Eqn. (7.27) and the definition of the error function, erf(), given in Chapter (5), we find

$$Q_t(z, t-t') = \frac{ze^{\frac{-z^2}{4(t-t')}}}{2\sqrt{\pi}(t-t')^{\frac{3}{2}}} \tag{7.30}$$

Equating $G_z(0,t';z)$ with $Q_t(z, t_f - t')$ gives

$$G(0,t';z) = \frac{e^{\frac{-z^2}{4(t_f-t')}}}{\sqrt{\pi}(t_f-t')^{\frac{1}{2}}} \tag{7.31}$$

Thus, the solution to the forward problem is given by

$$\theta(z,t_f) = \int_0^{t_f} T_s(t')\frac{ze^{\frac{-z^2}{4(t_f-t')}}}{2\sqrt{\pi}(t_f-t')^{\frac{3}{2}}}\mathrm{d}t' \tag{7.32}$$

We shall make use of this formal expression to derive the solution of the inverse problem.

7.4 A Family of Inverse Probems

Before outlining our approach to the inverse heat diffusion problem, it is important to distinguish between three related inverse problems. The first problem is termed *ill posed* and is intractable. The second and third probems are least squares problems. We shall solve the third problem in two separate manners using either the mathematical apparatus of integral equations and finite differences.

7.4.1 Inverse Problem 1.

Given borehole data $\theta_b(z)$ and a definition of the heat transfer physics (*i.e.,* a specification of $G_z(0,t;z)$), find $T_s(t)$ using the relation

$$\theta_b(z) = \int_0^{t_f} T_s(t) G_z(0,t;z) dt \tag{7.33}$$

□ A schematic view of the temporal and spatial domain associated with this inverse problem is given in Fig. (7.2).

To see why Inverse Problem 1 is ill posed, recall the problem of fitting an isochron to the lead-isotope data in Chapter (1). That simple line-fitting problem was ill posed because there were more than two lead-isotope data points, and they did not happen to be colinear. We found that there could be no solution to the line-fitting problem that was exact; at best, we could only determine an isochron which satisfies the data in the *least squares* sense.

We run into a similar difficulty with Problem 1. Unless the borehole data, $\theta_b(z)$, satisfies the strict mathematical properties required of solutions of the heat-transfer equations (*i.e.,* that the function $\theta_b(z)$ be smooth in some restricted sense), Inverse Problem 1 has no solution. For example, there is no surface temperature history that is capable of introducing a step discontinuity in the temperature depth profile of a homogeneous, semi-infinite material. Thus, if $\theta_b(z)$ posesses a step discontinuity, say as a result of measurement error, Inverse Problem 1 cannot be solved.

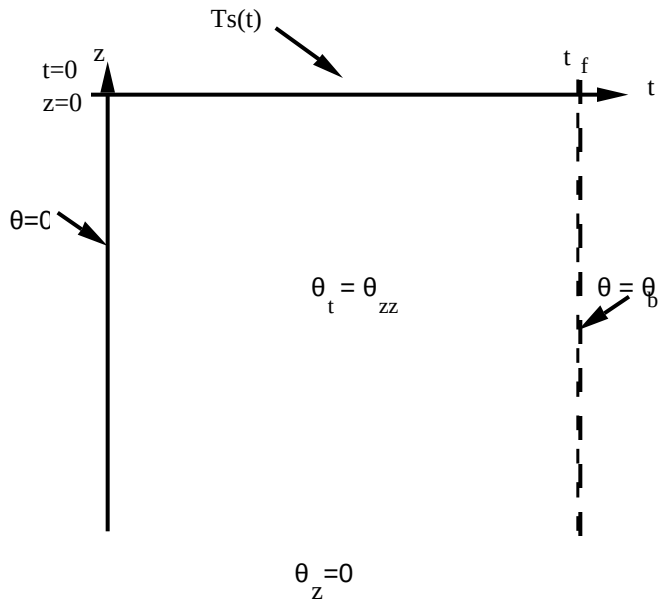

Figure 7.2: The domain in which the inverse borehole temperature problem is solved. Boundary conditions are known on three sides of the "box" (at $t = 0$, at $t = t_f$, and at $z = -\infty$), and are unknown on one side (at $z = 0$). Within the box, the constraint $\theta_t = \theta_{zz}$ applies.

In the world of borehole paleothermometry, borehole data $\theta_b(z)$ is rarely free of serious measurement error, and is determined by the interpolation of discrete point measurements of temperature and depth. Most interpolation functions that are used to express the data as a continuous profile lack the strict mathematical properties required of solutions of the heat-transfer equations. Interpolation polynomials, for example, are not analytic (differentiable up to an infinite number of times). This is why we are forced to adopt a least squares approach to the paleothermometry problem.

Another difficulty with Inverse Problem 1 is that Eqn. (7.33) is an integral equation of the *first kind*, that is, $T_s(t)$ appears only under the integral sign (Courant and Hilbert, 1953, p. 159). This kind of integral equation can have a *null space*, *i.e.,* there might be a function (or functions) $\tilde{T}_s(t) \neq 0$ such that

$$\int_0^{t_f} \kappa \mathcal{G}_z(0,t;\xi)\ \tilde{T}_s(t)\mathrm{d}t = 0 \tag{7.34}$$

By setting the right-hand side of Eqn. (7.34) to zero, we mean 0 Kelvin; in other words, the history $\tilde{T}_s$ has *no* effect on the borehole temperature profile. If solutions satisfying Eqn. (7.34) exist, then arbitrary improvement in accuracy of the borehole measurements can never overcome the fact that an arbitrary factor times $\tilde{T}_s(t)$ can be added to $T_s(t)$ without changing the predicted borehole profile.

On intuitive grounds, this situation is physically unlikely; *i.e.,* it is unrealistic to think that there are climate histories which have literally no effect on borehole temperatures at time t_f. What is intuitively more reasonable, is that *finite-precision arithmetic* and *limited borehole-measurement sensitivity* makes the possibility expressed in Eqn. (7.34) a certainty in practical situations. Finite-precision arithmetic (such as that which is performed on a computer) suggests that some temperature histories $\tilde{T}_s(t)$ cannot be operated on by the integral transform expressed by Eqn. (7.33) without serious arithmetic error. In fact, this arithmetic error may make the associated borehole temperature profile that results from $\tilde{T}_s$ to be zero. Limited-measurement sensitivity implies that there are an infinite number of borehole temperature profiles $\tilde{\theta}(z)$ that cannot be distinguished from zero. The $\tilde{T}_s(t)$ associated with these profiles, or which cannot be transformed accurately by Eqn.

(7.33), constitute the *null space* of Inverse Problem 1. In this circumstance, *no* inverse method can guide us in the choice between the infinite number of equally satisfactory solutions to Eqn. (7.33) which differ only by their projection into the null space of Eqn. (7.33).

7.4.2 Inverse Problem 2.

Approximate the solution of Inverse Problem 1 in a least-squares sense. □

Here we define a least-squares performance index, and make use of the expression (7.33)

$$J = \int_{-\infty}^{0} \left[\int_{0}^{t_f} G_z(0,t;z)\ T_s(t)\mathrm{d}t - \theta_b(z) \right]^2 dz \tag{7.35}$$

Henceforth, the distinction between profiles $\theta_b(z,t)$ which do satisfy the mathematical properties of solutions to the heat-transfer problem and profiles which don't is no longer important. This is indeed the main benefit of adopting a least-squares approach to the inverse heat-diffusion problem.

Inverse Problem 2 is solved by choosing a $T_s(t)$ which minimizes J. This is normally done by using the calculus of variations to determine an Euler-Lagrange condition which, when satisfied, ensures $\delta J = 0$ under arbitrary variations of $T_s(t)$, where δJ is the variation of J. A second version of Inverse Problem 2 is defined by introducing an auxilliary constraint, such as a climatology. We will focus on this second version because it provides a means to deal with difficulties which arise from finite-precision arithmetic.

7.4.3 Inverse Problem 3.

Approximate the solution of Inverse Problem 1 in a least-squares sense, but add subsidiary performance conditions, such as a cost function that depends on deviations between $T_s(t)$ and "climatology" $\phi(t)$. □

In this circumstance, the performance index J of Inverse Problem 2 is modified as follows

$$J = (1-\alpha) \int_{-\infty}^{0} \left[\int_{0}^{t_f} G_z(0,t;z)\ T_s(t)dt - \theta_b(z) \right]^2 dz + \alpha \int_{0}^{t_f} \left[T_s(t) - \phi(t) \right]^2 dt \tag{7.36}$$

The "mixture" parameter $0 < \alpha < 1$ describes the trade-off between the fit to borehole data and the fit to climatology. Observe that when $\alpha = 0$, borehole misfit is the only consideration used to determine the surface-temperature history (Inverse Problem 3 becomes Inverse Problem 2); and, when $\alpha = 1$, borehole misfit becomes irrelevant, only deviation from climatology is important. (We note that α could be defined as a function of time. For simplicity, we have refrained from doing so here.)

To minimize the J defined in Eqn. (7.36), we determine the Euler-Lagrange equation, namely the condition under which $\delta J = 0$ for arbitrary δT_s. We begin by rewriting Eqn. (7.36) in a manner which isolates explicit dependence of J on $T_s(t)$:

$$\begin{aligned} J = \ & (1-\alpha) \left\{ \int_{-\infty}^{0} \left[\int_{0}^{t_f} G_z(0,t;z)\ T_s(t)dt - \theta_b(z) \right] \right. \\ & \left. + \left[\int_{0}^{t_f} G_z(0,t';z)\ T_s(t')dt' - \theta_b(z) \right] dz \right\} \\ & + \alpha \int_{0}^{t_f} \left[T_s(t) - \phi(t) \right]^2 dt \end{aligned} \tag{7.37}$$

Expanding the products expressed in the integrands, we get

$$\begin{aligned} J \ = (1-\alpha) \{ & \int_{-\infty}^{0} \int_{0}^{t_f} \int_{0}^{t_f} T_s(t) T_s(t') G_z(0,t;z) G_z(0,t';z) dt dt' dz \\ & -2 \int_{-\infty}^{0} \int_{0}^{t_f} T_s(t) G_z(0,t;z) \theta_b(z) dt dz \end{aligned}$$

$$+\int_{-\infty}^{0}\theta_b^2(z)dz\}$$

$$+\alpha\int_0^{t_f}\left[T_s^2(t)-2T_s(t)\phi(t)+\phi^2(t)\right]dt \tag{7.38}$$

In the second term of the first integral of Eqn. (7.38) we have made a substitution of the dummy variable of integration from t' to t. This does not change J. We may also swap the order of integration in Eqn. (7.38) as follows

$$\begin{aligned}J \;&=(1-\alpha)\{\int_0^{t_f}[\int_0^{t_f}\int_{-\infty}^{0}T_s(t)T_s(t')G_z(0,t;z)G_z(0,t';z)dzdt'\\&\quad-2\int_{-\infty}^{0}T_s(t)G_z(0,t;z)\theta_b(z)dz]dt\\&\quad+\int_{-\infty}^{0}\theta_b^2(z)dz\}\\&\quad+\alpha\int_0^{t_f}\left[T_s^2(t)-2T_s(t)\phi(t)+\phi^2(t)\right]dt\end{aligned} \tag{7.39}$$

Observe that $G_z(0,t;z)$ and $\theta_b(z)$ are known functions of z, thus the integrations over z in Eqn. (7.39) may be done at this stage. If we define

$$K(t,t')=\int_{-\infty}^{0}G_z(0,t;z)G_z(0,t';z)dz \tag{7.40}$$

and

$$\Theta_b(t)=\int_{-\infty}^{0}G_z(0,t;z)\theta_b(z)dz \tag{7.41}$$

Eqn. (7.39) becomes

$$J \;=(1-\alpha)\{\int_0^{t_f}[\int_0^{t_f}T_s(t)T_s(t')K(t,t')dt'$$

$$
\begin{aligned}
&-2T_s(t)\Theta_b(z)]dt \\
&+\int_{-\infty}^{0} \theta_b^2(z)dz\} \\
+\alpha &\int_0^{t_f} \left[T_s^2(t) - 2T_s(t)\phi(t) + \phi^2(t)\right] dt
\end{aligned}
\tag{7.42}
$$

We are now ready to take the variation:

$$
\begin{aligned}
\delta J \quad &= 2(1-\alpha)\{ \int_0^{t_f} [\int_0^{t_f} \delta T_s(t) T_s(t') K(t,t')dt' \\
&\qquad -2\delta T_s(t)\Theta_b(z)]dt\} \\
&+2\alpha \int_0^{t_f} \left[\delta T_s(t) T_s(t) - \delta T_s(t)\phi(t)\right] dt
\end{aligned}
\tag{7.43}
$$

To ensure $\delta J = 0$ for arbitrary $\delta T_s(t)$, we must insist that the integrand of the integral over t be zero. This gives the Euler-Lagrange condition,

$$
(1-\alpha)\int_0^{t_f} T_s(t')K(t,t')dt' + \alpha T_s(t) = (1-\alpha)\Theta_b(t) + \alpha\phi(t) \tag{7.44}
$$

Equation (7.44) is an integral equation of the *second kind* (Courant and Hilbert, 1953, p. 112), and the kernel $K(t,t')$ is symmetric. The solution of Eqn. (7.44) constitutes what we define here to be the *least-squares* solution to the paleothermometry problem.

In focussiong our attention on Inverse Problem 3, we have subtly changed the mathematical nature of the paleothermometry problem. Inverse Problem 1 leads to an integral equation of the first kind, which may have no solution due to the incompatibility between the data, $\theta_b(z)$, and solutions of heat-transfer equations (*i.e.,* analyticity). Inverse Problem 3, however, leads to an integral equation of the second kind which always has a solution (or many solutions). Inverse Problems 1 and 3 also differ by the fact that $K(t,t')$ is a symmetric kernel whereas $G_z(0,t;z)$ is not. In addition, the data in Inverse

Problem 1, $\theta_b(z)$, can be "rough"; whereas, the data in Inverse Problem 3, $\Theta_b(t)$, is "smoothed" by the integral transform implied by Eqn. (7.41). From what we know about singular value decomposition (SVD), we expect the kernel $K(t,t')$ to have eigenvalues that are near zero. This motivates the introduction of climatology $\eta(t)$ and a non-zero α. Even with all the advantages gained by abandoning Inverse Problem 1 in favour of Inverse Problem 3, we still must keep in mind the fact that the kernel $K(t,t')$ may have singularities (*i.e.,* may not be square-integrable), and so may be very difficult to work with.

7.5 Least Squares Solution: Continuous Case

If the symmetric kernel $K(t,t')$ in Eqn. (7.44) is square-integrable (continuous, for example), then we could exploit the fact that it has *eigenvalues* $\{\mu_i\}_{i=1}^{\infty}$ and *eigenfunctions* $\{\psi_i(t)\}_{i=1}^{\infty}$ which span the Hilbert space of continuous functions on the interval $0 < t < t_f$ (Courant and Hilbert, 1953, p. 122) where

$$\int_0^{t_f} K(t,t')\psi_i(t)\mathrm{d}t' \;\; = \mu_i\psi_i(t) \tag{7.45}$$

$$\int_0^{t_f} \psi_i(t')\psi_j(t')dt' = \delta_{ij} \tag{7.46}$$

where $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ij} = 1$ if $i = j$. By square integrability we mean

$$\int_0^{t_f}\int_0^{t_f} [K(t,t')]^2 \mathrm{d}t \; \mathrm{d}t' \leq M \tag{7.47}$$

where $M < \infty$ is a fixed bound. Inspection of the functional form of $G_z(0,t;z)$ assures us that the $K(t,t')$ is *not* square integrable. Strictly speaking, we cannot use the mathematical machinery associated with the eigenvalues and eigenfunctions.

The fact that $K(t,t')$ is not square integrable is an indication of the great difficulty of the inverse heat-diffusion problem. Even with the benefits gained

by adopting a least-squares approach, the problem is still intractable owing to the singularities we anticipate in $K(t,t')$. Once a numerical method is adopted, however, the question of square-integrability is no longer crucial to the development of the solution. We shall thus proceed and make use of the eigenvalue and eigenfunction approach in spite of the lack of square-integrability. We must keep in mind, however, that an essential element of the continuous version of the problem is being lost when we do this.

We proceed formally from now on and denote with $\{\mu_i\}_{i=1}^{\infty}$ and $\{\psi_i(t)\}_{i=1}^{\infty}$ the eigenvalues and eigenfunctions associated with $K(t,t')$ (Courant and Hilbert, 1953, p. 122). Then, if we expand T_s, Θ_b and ϕ as follows

$$T_s(t) = \sum_{i=1}^{\infty} a_i \psi_i(t) \tag{7.48}$$

$$\Theta(t) = \sum_{i=1}^{\infty} b_i \psi_i(t) \tag{7.49}$$

$$\phi(t) = \sum_{i=1}^{\infty} c_i \psi_i(t) \tag{7.50}$$

substitute these expressions into Eqn. (**??**), and make use of the orthogonality of the eigenfunctions, we get an expression which relates the unknown coefficents a_i to the known coefficients b_i and c_i:

$$[(1-\alpha)\mu_i + \alpha]a_i = (1-\alpha)b_i + \alpha c_i \qquad \text{for } i = 1, \ldots, \infty \tag{7.51}$$

The solution $T_s(t)$ thus becomes

$$T_s(t) = \sum_{i=1}^{\infty} \frac{(1-\alpha)b_i + \alpha c_i}{(1-\alpha)\mu_i + \alpha} \psi_i(t) \tag{7.52}$$

We have now produced a description of the formal solution to the continuous version of the least-squares paleothermometry problem with climatology (Inverse Problem 3).

We can now address the question of trade-off between the fit to borehole data and the fit to climatology expressed in the definition of J for Inverse Problem 3. The mixture parameter, α, determines the importance of the coefficients of the *climatology*, c_i, relative to the coefficients of the *borehole*

data b_i in determining the coefficients of the surface-temperature history a_i. An objective choice of α will depend on an evaluation of the relative confidence ascribed to the borehole temperature measurements (which determine the b_i's) and the climatology (which determine the c_i's). For example, one might choose $\frac{\alpha}{(1-\alpha)} = \frac{\sigma_b^2}{\sigma_c^2}$, where σ_b^2 is a measure of temperature-measurement uncertainty and σ_c^2 is a measure of the uncertainty in the climatology.

Another important role is played by the mixture parameter α, and this must also be considered when selecting the numerical value of α. It serves to *regularize* the problem of determining $T_s(t)$. Even without specifying the symmetric kernel $K(t, t')$, we can anticipate that the eigenvalues μ_i will tend to converge to 0 as $i \to \infty$ (Courant and Hilbert, 1953, p. 130). Thus, when $\alpha = 0$ (*i.e.,* when we *force* our solution to depend *only* on achieving the best fit between predicted borehole temperatures and data), the series expansion for $T_s(t)$ could become divergent because of zeros, or extremely small numbers, in the denominator of the expansion coefficients defined by Eqn. (7.52). By specifying $\alpha > 0$, the denominator remains finite when $\mu_i \to 0$. The series expressions thus is well-behaved from an arithmetic standpoint. The benefit of regularizing the problem is *offset* by the fact that the retrodicted surface-temperature history depends *more* on the climatology $\phi(t)$ and *less* on the borehole temperature data.

7.6 Discrete Version of the Forward Problem

The above analysis is impractical from the standpoint that an integral equation must be solved for an unknown continuous function. Typically in the analysis of geophysical data, numerical methods are preferred. We thus redevelop the least squares solution to the paleothermometry problem using the finite difference technique. We shall see that there are many parallels between the finite-difference and continuous forms of the least-squares solution.

The finite difference version of the forward problem Eqns. (7.7)-(7.10) can be developed by defining finite-difference versions of the borehole data

$\underline{\theta}_b$ (down to a finite depth in the earth), the temperature depth profile $\underline{\theta}^{[n]}$ (at time $t = n\Delta t$ and down to a finite depth), and the surface-temperature history $\mathbf{T}_s$:

$$\underline{\theta}_b = \begin{pmatrix} \theta_b(0) \\ \theta_b(-\Delta z) \\ \vdots \\ \theta_b(-(M-1)\Delta z) \end{pmatrix} \tag{7.53}$$

$$\underline{\theta}^{[n]} = \begin{pmatrix} \theta(0,(n-1)\Delta t) \\ \theta(-\Delta z,(n-1)\Delta t) \\ \vdots \\ \theta(-(M-1)\Delta z,(n-1)\Delta t) \end{pmatrix} \tag{7.54}$$

$$\mathbf{T}_s = \begin{pmatrix} T_s(0) \\ T_s(\Delta t) \\ \vdots \\ T_s((N-1)\Delta t) \end{pmatrix} \tag{7.55}$$

where $\Delta z = D_o/(M-1)$, D_o is the depth of the bottom of the borehole, $\Delta t = t_f/(N-1)$, M is the number of grid points in the vertical where the temperature is measured, and N is the number of time steps. We make one alteration to the forward problem to accomodate a finite depth, D_o, of borehole measurement. Instead of requiring $\theta_z \to 0$ as $z \to -\infty$, we require $\theta_z = 0$ at $z = -D_o$.

The continuous partial differential equation (7.7) is converted to an algebraic equation using the following finite-difference expressions for the derivatives of θ:

$$\theta_t(-(j-1)\Delta z,(n+1/2)\Delta t) \to \frac{\underline{\theta}_j^{[n+1]} - \underline{\theta}_j^{[n]}}{\Delta t} \tag{7.56}$$

$$\theta_{zz}(-(j-1)\Delta z, n\Delta t) \to \frac{\underline{\theta}_{j+1}^{[n+1]} + \underline{\theta}_{j-1}^{[n+1]} - 2\underline{\theta}_j^{[n+1]}}{\Delta z^2} \tag{7.57}$$

$$\theta_z(-D_o, n\Delta t) = 0 \to \underline{\theta}_M^{[n+1]} - \underline{\theta}_{M-1}^{[n+1]} = 0 \tag{7.58}$$

With these changes, the algebraic equations representing *implicit* time stepping of (7.7) may be written

$$\mathbf{A}\underline{\theta}^{[n]} - \mathbf{B}\underline{\theta}^{[n-1]} = \mathbf{C}T_s^{[n-1]} \tag{7.59}$$

where $T_s^{[n-1]}$ is the $(n-1)$th component of $\mathbf{T}_s$, and

$$\mathbf{A} = \frac{1}{\Delta z^2} \cdot \begin{pmatrix} \Delta z^2 & 0 & 0 & 0 & \dots \\ -1 & \frac{\Delta z^2}{\Delta t} + 2 & -1 & 0 & \dots \\ 0 & -1 & \frac{\Delta z^2}{\Delta t} + 2 & -1 & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \dots & 0 & -1 & \frac{\Delta z^2}{\Delta t} + 2 & -1 \\ \dots & 0 & 0 & -1 & 1 \end{pmatrix} \tag{7.60}$$

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ 0 & \frac{1}{\Delta t} & 0 & 0 & \dots \\ 0 & 0 & \frac{1}{\Delta t} & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \dots & 0 & 0 & \frac{1}{\Delta t} & 0 \\ \dots & 0 & 0 & 0 & 0 \end{pmatrix} \tag{7.61}$$

and,

$$\mathbf{C} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{7.62}$$

Equation (7.59) is said to be an *implicit* representation of the diffusion operator because the diffusion term is evaluated at $t = n\Delta t$ (corresponding to time step $n+1$) whereas the corresponding evaluation of the time-derivative term is at $t = (n-1/2)\Delta t$. An implicit representation is preferable to an explicit representation (where the diffusion term is evaluated at time $t = (n-1)\Delta t$ or at time step n) because it is unconditionally stable to numerical perturbations. An explicit formulation will *blow up* (computer-hacker jargon for unbounded arithmetic growth) if the time-step size Δt is not small enough.

The solution of the forward problem in the finite difference formulation is found by recursively applying Eqn. (7.59). This is a tedious operation, but the result is readily seen by inspecting the first few steps of the recursive application:

$$\underline{\theta}^{[2]} = \qquad \mathbf{A}^{-1}\mathbf{B}\underline{\theta}^{[1]} + \mathbf{A}^{-1}\mathbf{C}T_s(0)$$

$$\begin{aligned}\underline{\theta}^{[3]} &= \mathbf{A}^{-1}\mathbf{B}\underline{\theta}^{[2]} + \mathbf{A}^{-1}\mathbf{C}T_s(\Delta t) \\ &= [\mathbf{A}^{-1}\mathbf{B}]^2\underline{\theta}^{[1]} + \mathbf{A}^{-1}\mathbf{B}[\mathbf{A}^{-1}\mathbf{C}T_s(0)] + \mathbf{A}^{-1}\mathbf{C}T_s(\Delta t) \qquad (7.63) \\ &\vdots\end{aligned}$$

The recursion formula is

$$\underline{\theta}^{[k]} = [\mathbf{A}^{-1}\mathbf{B}]^{k-1}\underline{\theta}^{[1]} + \sum_{n=1}^{k-1}[\mathbf{A}^{-1}\mathbf{B}]^{k-n-1}\mathbf{A}^{-1}\mathbf{C}T_s((n-1)\Delta t) \qquad (7.64)$$

The initial condition is homogeneous, *i.e.*, $\underline{\theta}^{[1]} = \underline{0}$, thus we may write the solution to the forward problem as follows

$$\underline{\theta}^{[N]} = \mathbf{G}\mathbf{T}_s \qquad (7.65)$$

where the finite-difference analogue $\mathbf{G}$ to the continuous Green's function $G(0,t;z)$ is given by

$$\mathbf{G}_{mn} = \begin{cases} \left[(\mathbf{A}^{-1}\mathbf{B})^{N-n}\mathbf{A}^{-1}\mathbf{C}\right]_m & n < N-1 \\ 0 & n = N \end{cases} \qquad (7.66)$$

Zeros appear in the last column of $\mathbf{G}$ because the temperature profile at time $(N-1)\Delta t$ does not depend on $T_s((N-1)\Delta t)$. Observe that $\mathbf{G}$ is a mapping from $\mathcal{R}^N$ to $\mathcal{R}^M$; thus, $\mathbf{G}$ is a rectangular matrix if $N \neq M$.

7.7 Discrete Version of the Least-Squares Solution

The path set by our consideration of the continuous versions of Inverse Problems 1 - 3 guides us in our deliberations on the discrete analogues of Inverse problems 1 - 3. For example, the discrete version of Inverse Problem 1

$$\underline{\theta}_b = \mathbf{G}\mathbf{T}_s \qquad (7.67)$$

where $\underline{\theta}_b$ is the discrete version of the measured borehole temperatures, is immediately recognized as ill-posed because $\mathbf{G}$ is a rectangular matrix which

cannot be inverted. The discrete version of Inverse Problem 3 is solved by choosing a $\mathbf{T}_s$ that minimizes the following least-squares performance index

$$J = (1-\alpha)(\mathbf{G}\mathbf{T}_s - \underline{\theta}_b)'(\mathbf{G}\mathbf{T}_s - \underline{\theta}_b) + \alpha(\mathbf{T}_s - \underline{\phi})'(\mathbf{T}_s - \underline{\phi}) \tag{7.68}$$

where $\underline{\phi}$ is the discretized version of the climatology $\phi(t)$,

$$\underline{\phi} = \begin{pmatrix} \phi(0) \\ \phi(\Delta t) \\ \vdots \\ \phi((N-1)\Delta t) \end{pmatrix} \tag{7.69}$$

It is readily shown that the minimum of J is achieved when

$$(1-\alpha)\mathbf{G}'\mathbf{G}\mathbf{T}_s + \alpha\mathbf{T}_s = (1-\alpha)\mathbf{G}'\underline{\theta}_b + \alpha\underline{\phi} \tag{7.70}$$

If we define

$$\mathbf{K} = \mathbf{G}'\mathbf{G} \tag{7.71}$$

$$\mathbf{\Theta}_b = \mathbf{G}'\underline{\theta}_b \tag{7.72}$$

Equation (7.70) becomes

$$(1-\alpha)\mathbf{K}\mathbf{T}_s + \alpha\mathbf{T}_s = (1-\alpha)\mathbf{\Theta}_b + \alpha\underline{\phi} \tag{7.73}$$

We remark, in analogy to what we have said before concerning Inverse Problems 1 and 3, that Eqn. (7.73) offers several advantages over Eqn. (7.67). Foremost of these advantages is the fact that the matrix $\mathbf{K}$ is square and symmetric; whereas, $\mathbf{G}$ is not.

To solve Eqn. (7.73) for the unknown surface-temperature history $\mathbf{T}_s$, we must perform three steps. First, we must select a climatology $\underline{\phi}$. Second, we must choose an α. Third, we must find a way to solve the linear-algebra problem posed by Eqn. (7.73). Often, the last step will be difficult because of the fact that the square, symmetric matrix $\mathbf{K}$ may be ill-conditioned. (By ill-conditioned, we mean that it's determinant is close to zero. We shall delve into this point below.)

7.8 Least-Squares Solution: Discrete Case

The solution to Eqn. (7.73) can be expressed formally in terms of the eigenvalues and eigenvectors associated with the symmetric matrix $\mathbf{K} : \mathcal{R}^N \rightarrow \mathcal{R}^N$. Let $\{\underline{\Psi}_i\}_{i=1}^N$ and $\{\mu\}_{i=1}^N$, respectively, denote the eigenvectors and eigenvalues of $\mathbf{K}$. In other words, let

$$\mathbf{K}\underline{\Psi}_i = \mu_i \underline{\Psi}_i \tag{7.74}$$

The symmetry of $\mathbf{K}$ assures us that the eigenvectors $\{\underline{\Psi}_i\}_{i=1}^N$ are complete and can be made orthonormal, and will span the vector space $\mathcal{R}^N$. (By complete, we mean that *any* arbitrary vector in $\mathcal{R}^N$ can be expressed as a linear combination of the vectors in the set $\{\underline{\Psi}_i\}_{i=1}^N$.) Thus we may expand $\mathbf{T}_s$, $\underline{\phi}$ and $\mathbf{\Theta}_b$ as follows

$$\mathbf{T}_s = \sum_{i=1}^{N} a_i \underline{\Psi}_i \tag{7.75}$$

$$\mathbf{\Theta}_b = \sum_{i=1}^{N} b_i \underline{\Psi}_i \tag{7.76}$$

$$\underline{\phi} = \sum_{i=1}^{N} c_i \underline{\Psi}_i \tag{7.77}$$

where the coefficients $\{a_i\}_{i=1}^N$, $\{b_i\}_{i=1}^N$, and $\{c_i\}_{i=1}^N$ (note that we have chosen 'b' for borehole data and 'c' for climatology) are given by

$$a_i = \mathbf{T}'_s \underline{\Psi}_i \tag{7.78}$$

$$b_i = \mathbf{\Theta}'_b \underline{\Psi}_i \tag{7.79}$$

$$c_i = \underline{\phi}' \underline{\Psi}_i \tag{7.80}$$

Substituting these series expansions into Eqn. (7.73), and making use of Eqn. (7.74), gives a relation between the unknown coefficients $\{a_i\}_{i=1}^N$ and the known coefficients $\{b_i\}_{i=1}^N$ and $\{c_i\}_{i=1}^N$

$$[(1-\alpha)\mu_i + \alpha]a_i = (1-\alpha)b_i + \alpha c_i \tag{7.81}$$

The solution of the discrete form of the paleothermometry problem is thus

$$\mathbf{T}_s = \sum_{i=1}^{N} \frac{(1-\alpha)b_i + \alpha c_i}{(1-\alpha)\mu_i + \alpha} \underline{\Psi}_i \tag{7.82}$$

The advantage gained by expressing $\mathbf{T}_s$ in terms of the eigenvectors of $\mathbf{K}$ as opposed to the equally valid, but less formal, expression

$$\mathbf{T}_s = [(1-\alpha)\mathbf{K} + \alpha\mathbf{I}]^{-1}\{(1-\alpha)\mathbf{\Theta}_b + \alpha\underline{\phi}\} \tag{7.83}$$

is that the mathematical properties of $\mathbf{T}_s$ can be immediately appreciated. The fact that $\{\underline{\Psi}_i\}_{i=1}^N$ are independent of the borehole data and climatology tells us how the solution depends on the *physics* of the heat-transfer process. The eigenvectors depend solely on the matrix $\mathbf{K}$ which depends, in turn, on a statement of the discrete form of the forward problem. Typically, these eigenvectors are oscillatory; thus we can expect the solution to the paleothermometry problem to also oscillate even in situations where we know that the "true" surface-temperature history does not.

7.9 Limits to Resolution of Past Thermal History

An important issue to be addressed in borehole paleothermometry concerns the fact that diffusion makes the surface-temperature history of the distant past very difficult to recover, *i.e.,* there is limited "thermal memory". The quantification of this source of uncertainty is a key step in the analysis of a paleothermometry problem [Dahl-Jensen *et al.*, 1993; MacAyeal *et al.*, 1993]. One approach to estimating the uncertainty involves the model-resolution matrix $\mathbf{R}$ which has been discussed in the context of the underdetermined least-squares problems of Chapter ().

Suppose that we have the ability to measure a discrete version of the borehole temperature profile with perfect precision and resolution (in other words, suppose that we have data which is free of measurement error). Would

our inverse method produce the "exact" surface-temperature history? Let's denote the exact history by $\mathbf{T}_s^e$ and the history inferred from the "perfect" data by $\mathbf{T}_s^{inf}$. The thermal memory issue is quantified by finding the relation between $\mathbf{T}_s^{inf}$ and $\mathbf{T}_s^e$ is provided by the series expansion for $\mathbf{T}_s^{inf}$ given in Eqn. (7.82). For this purpose, we revise consider a solution of the inverse problem in which climatology is not accounted for (*i.e.,* $\alpha = 0$). Using Eqn. (7.82) we have

$$\mathbf{T}_s^{inf} = \sum_{i=1}^{N} \frac{b_i}{\mu_i} \underline{\Psi}_i \tag{7.84}$$

In practical circumstances, the matrix $\mathbf{K}$ may be ill-conditioned to the point where we can only recover $M < N$ of it's eigenvalues. In this situation, we would have

$$\mathbf{T}_s^{inf} = \sum_{i=1}^{M} \frac{b_i}{\mu_i} \underline{\Psi}_i \tag{7.85}$$

We recall that, by definition, the error-free borehole temperature data is related to the exact surface temperature history by $\mathbf{\Theta}_b = \mathbf{G}'\mathbf{G}\mathbf{T}_s^e$, and write b_i, using Eqn. (7.79), as follows:

$$\begin{aligned} b_i = & \quad \mathbf{\Theta}_b' \underline{\Psi}_i \\ = & \quad [\mathbf{G}'\mathbf{G}\mathbf{T}_s^e]' \underline{\Psi}_i \\ = & \quad [\mathbf{K}\mathbf{T}_s^e]' \underline{\Psi}_i \end{aligned} \tag{7.86}$$

With the above identity, Eqn. (7.85) becomes

$$\begin{aligned} \mathbf{T}_s^{inf} = & \quad \sum_{i=1}^{M} \frac{[\mathbf{K}\mathbf{T}_s^e]' \underline{\Psi}_i}{\mu_i} \underline{\Psi}_i \\ = & \quad \mathbf{R}\mathbf{T}_s^e \end{aligned} \tag{7.87}$$

where $\mathbf{R}$, the model-resolution matrix, is defined by

$$\mathbf{R} = \sum_{i=1}^{M} \underline{\Psi}_i \underline{\Psi}_i' \tag{7.88}$$

and where we have made use of the symmetry of $\mathbf{K}$ and the relation $\mathbf{K}\underline{\Psi}_i = \mu_i \underline{\Psi}_i$. We note, in passing, that the expression in Eqn. (7.88) is similar to

the expression we could use to write the identity matrix $\mathbf{I}$, *i.e.,*

$$\mathbf{I} = \sum_{i=1}^{N} \underline{\Psi}_i \underline{\Psi}_i' \tag{7.89}$$

The fact that the upper limit of the summation index in Eqn. (7.88) is M and not N indicates that $\mathbf{R}$ will not, in general, be the same as the identity matrix. Indeed, the difference between $\mathbf{R}$ and the identity matrix represents a measure of the lack of resolution brought about by the limitation to thermal memory.

It is interesting to note that the off-diagonal spread in the model-resolution matrix, depends only on the properties of $\mathbf{K}$ (*i.e.,* it's eigenvalues, eigenvectors, and M the number of significant eigenvalues). Thus, the degree to which $\mathbf{T}_s^e$ is degraded by diffusive heat transfer can be evaluated in advance of any field project without having to have any borehole data actually in hand.

7.10 Uncertainty

Having described a crucial source of uncertainty, limited thermal memory, we touch on the question of estimating the covarience of the derived surface-temperature history $\mathbf{T}_s$. Following the proceedure in § (3.3.2), we assume the following relationships between $\hat{\mathbf{T}}_s$, $\hat{\underline{\theta}}_b$, $\hat{\underline{\phi}}$ and their respective errors $\underline{\zeta}$, $\underline{\epsilon}$, and $\underline{\gamma}$:

$$\underline{\zeta} = \quad \hat{\mathbf{T}}_s - \mathbf{T}_s \tag{7.90}$$

$$\underline{\epsilon} = \quad \hat{\underline{\theta}}_b - \underline{\theta}_b \tag{7.91}$$

$$\underline{\gamma} = \quad \hat{\underline{\phi}} - \underline{\phi} \tag{7.92}$$

where the $\hat{\cdot}$ denotes quantities that are measured or derived from measurements. We assume at the outset that the covarience matrices $\mathbf{Q}$ and $\mathbf{S}$

associated with $\underline{\epsilon}$ and $\underline{\gamma}$, respectively, are known:

$$\langle \underline{\epsilon}\underline{\epsilon}' \rangle = \mathbf{Q} \tag{7.93}$$

$$\langle \underline{\gamma}\underline{\gamma}' \rangle = \mathbf{S} \tag{7.94}$$

and that

$$\langle \underline{\epsilon}\underline{\gamma}' \rangle = \mathbf{0} \tag{7.95}$$

Our goal is to estimate $\mathbf{E} = \langle \mathbf{T_s} \mathbf{T'_s} \rangle$, the covarience of the derived surface-temperature history.

From Eqn. (7.83), we have

$$\underline{\zeta} = [(1-\alpha)\mathbf{K} + \alpha\mathbf{I}]^{-1}\{(1-\alpha)\mathbf{G}'\underline{\epsilon} + \alpha\underline{\gamma}\} \tag{7.96}$$

Substitution of Eqn. (7.96) into the expression for $\mathbf{E}$ yields the result

$$\begin{aligned} \mathbf{E} = \; & (1-\alpha)^2 [(1-\alpha)\mathbf{K} + \alpha\mathbf{I}]^{-1} \mathbf{G}'\mathbf{Q}\mathbf{G} \left[[(1-\alpha)\mathbf{K} + \alpha\mathbf{I}]^{-1} \right]' \\ & + \alpha^2 [(1-\alpha)\mathbf{K} + \alpha\mathbf{I}]^{-1} \mathbf{S} \left[[(1-\alpha)\mathbf{K} + \alpha\mathbf{I}]^{-1} \right]' \end{aligned} \tag{7.97}$$

It will be seen, from the exercises associated with this chapter, that $\mathbf{E}$ is not diagonal and suggests that errors in the inferred surface-temperature history increase with increasing age.

7.11 Paleothermometry by Control Methods

In the preceeding development, the solution to the paleothermometry problem (Problem 2 or Problem 3 of §(7.4) has been presented using the mathematical formalism associated with integral equations (continuous case) and linear algebra (discrete case). MacAyeal it et al. [1992] suggest that control methods might also work well in solving the paleothermometry problem. Here we develop a control-method approach, and show that it is formally equivalent to the solutions derived previously in this chapter. We also suggest ways in which the control method might offer technical advantages over the approaches derived above.

7.11.1 Continuous Version of a Control Method

We formally derive the continuous version of a control-method approach by minimizing the following performance index J

$$J = \frac{1}{2} \int_{-\infty}^{0} dz \, (\theta(z, t_f) - \theta_b)^2 \tag{7.98}$$

subject to Eqns. (7.7)-(7.10) as constraints. These constraints may be conveniently accounted for in the minimization of J through the use of a Lagrange-multiplier function $\lambda(z, t)$:

$$J = \frac{1}{2} \int_{-\infty}^{0} dz \, (\theta(z, t_f) - \theta_b)^2 + \int_{0}^{t_f} dt \int_{-\infty}^{0} dz \lambda \, [\theta_t - \theta_{zz}] \tag{7.99}$$

At this stage, we do not represent the boundary conditions or initial condition associated with the forward problem using the Lagrange-multiplier approach (although, this would be possible to do). We must therefore keep in mind when we use the calculus of variations to derive the Euler-Lagrange conditions which ensure a minimization of J that variations of θ at the boundaries or the initial time $t = 0$ may not be allowed.

Application of the calculus of variations gives an expression for the variation of J, δJ:

$$\begin{aligned} \delta J \;\; = & \int_{-\infty}^{0} dt 2 \delta\theta(z, t_f) \, (\theta(z, t_f) - \theta_b) \\ & + \int_{0}^{t_f} dt \int_{-\infty}^{0} dz \delta\lambda \, [\theta_t - \theta_{zz}] \\ & + \int_{0}^{t_f} dt \int_{-\infty}^{0} dz \lambda \, [\delta\theta_t - \delta\theta_{zz}] \end{aligned} \tag{7.100}$$

The last term on the right-hand side of Eqn. (7.100) may be simplified by

integration by parts:

$$\int_0^{t_f} dt \int_{-\infty}^{0} dz\lambda \left[\delta\theta_t + \delta\theta_{zz}\right] = \int_{-\infty}^{0} dz \left[\lambda\delta\theta\right]_0^{t_f} - \int_0^{t_f} dt \left[\lambda\delta\theta_z\right]_{-\infty}^{0} + \int_0^{t_f} dt \left[\lambda_z\delta\theta\right]_{-\infty}^{0} - \int_0^{t_f} dt \int_{-\infty}^{0} dz\delta\theta \left[\lambda_t + \lambda_{zz}\right] \tag{7.101}$$

Recalling the boundary and initial conditions of the forward problem, we recognize that variations in the boundary values and initial conditions are not allowed, *i.e.,* $\delta\theta(z, t = 0) = \delta\theta_z(-\infty, t) = 0$. Combining these simplifications, substituting Eqn. (7.101) into Eqn. (7.100), and insisting that $\delta J = 0$ for arbitrary $\delta\lambda$ and $\delta\theta$ we arrive at the following set of Euler-Lagrange equations:

$$\theta_t = \theta_{zz} \tag{7.102}$$
$$\theta(z, t = 0) = 0 \tag{7.103}$$
$$\theta(z = 0, t) = T_s(t) \tag{7.104}$$
$$\theta_z(z \to -\infty, t) = 0 \tag{7.105}$$

$$\lambda_t = -\lambda_{zz} \tag{7.106}$$
$$\lambda(z, t = t_f) = -\left(\theta(z, t = t_f) - \theta_b\right) \tag{7.107}$$
$$\lambda(z = 0, t) = 0 \tag{7.108}$$
$$\lambda_z(z = -\infty, t) = 0 \tag{7.109}$$

$$\lambda_z(z = 0, t) = 0 \tag{7.110}$$

Equations (7.102)-(7.105) may be recognized as the forward problem. Equations (7.106)-(7.109) represent the adjoint form of the forward problem, and this leads to the convention of referring to λ as the "adjoint trajectory". Equation (7.110) represents the extra condition needed ultimately to determine T_s. A cursory look at the adjoint equations (Eqns. 7.106 - 7.109) might cause alarm. The normal diffusion term λ_{zz} appears on the right-hand side of Eqn. (7.106) with a minus sign. The governing equation for λ is thus the backward diffusion equation which is notoriously ill conditioned. A more careful examination, however, reassures us that the adjoint problem will not

be intractable. In particular, the initial condition in the adjoint equation is *not* applied at $t = 0$, but is applied at the terminal time $t = t_f$. This location of the initial condition on λ circumvents the otherwise ill-conditioned nature of the backward diffusion equation.

To see the correspondence between the solution to the paleothermometry problem derived previously and represented by Eqn. (7.44) (with $\alpha = 0$) and that obtained when the above Euler-Lagrange equations are solved, we proceed to solve Eqns. (7.102)-(7.110). The first step is to notice the correspondence between the adjoint problem (Eqns. 7.106 - 7.109) and Eqns. (7.11)-(7.14) which define the Green's function $G(z, t; \zeta)$. Making use of this Green's function, we can immediately write the solution to the adjoint problem:

$$\lambda(z,t) = \int_{-\infty}^{0} d\zeta \, (\theta(\zeta, t_f) - \theta_b) \, G(z,t;\zeta) = 0 \tag{7.111}$$

With this expression, Eqn. (7.110) may be written

$$\lambda_z(z=0,t) = \int_{-\infty}^{0} d\zeta \, (\theta(\zeta, t_f) - \theta_b) \, G_z(0,t;\zeta) \tag{7.112}$$

Substituting the Green's function representation of $\theta(\zeta, t_f)$ given by Eqn. (7.22), *i.e.,*

$$\theta(\zeta, t_f) = \int_{0}^{t_f} dt' T_s(t') G_z(0, t'; \zeta) \tag{7.113}$$

into Eqn. (7.112), we obtain

$$\int_{-\infty}^{0} d\zeta \int_{0}^{t_f} dt' T_s(t') G_z(0,t';\zeta) G_z(0,t;\zeta) - \int_{-\infty}^{0} d\zeta \theta_b(\zeta) G_z(0,t;\zeta) = 0 \tag{7.114}$$

Recalling the definitions of K and $\Theta_b(t)$ given in Eqns. (7.40) and (7.41), we obtain a result which is identical to that derived previously:

$$\int_{0}^{t_f} dt' K(t,t') T_s(t') = \Theta_b(t) \tag{7.115}$$

Having shown the correspondence between the control-method approach and the least-squares approach derived previously, we next consider the practical advantages associated with solving a discrete version of the paleothermometry problem using a control method.

7.11.2 Discrete Version of a Control Method

An excellent description of how to apply the control method to ice-sheet borehole temperature data is provided by Firestone [1992]. Here we provide only a cursory sketch on how a control method might be implimented to solve the discrete version of the problem developed in §(7.7). Our goal is to chose a vector $\mathbf{T}_s$ (having components that represent the surface temperature at discrete time steps) which minimizes

$$J = \left[\underline{\theta}^{[N]} - \underline{\theta}_b\right]' \left[\underline{\theta}^{[N]} - \underline{\theta}_b\right] \tag{7.116}$$

subject to the $N-1$ constraints

$$\mathbf{A}\underline{\theta}^{[n]} - \mathbf{B}\underline{\theta}^{[n-1]} = \mathbf{C}T_s^{[n-1]} \tag{7.117}$$

for $n = 2, \ldots, N$. A convenient way to enforce the constraints represented by Eqn. (7.117) is to use a sequence of Lagrange-multiplier vectors $\underline{\lambda}^{[n]}$, $n = 2, \ldots, N$. Doing so requires us to minimize the following augmented performence index:

$$J = \left[\underline{\theta}^{[N]} - \underline{\theta}_b\right]' \left[\underline{\theta}^{[N]} - \underline{\theta}_b\right] + \sum_{n=2}^{N} \left[\underline{\lambda}^{[n]}\right]' \left[\mathbf{A}\underline{\theta}^{[n]} - \mathbf{B}\underline{\theta}^{[n-1]} - \mathbf{C}T_s^{[n-1]}\right] \tag{7.118}$$

There are several approaches that we may choose from for the minimization of J as defined in Eqn. (7.118). The approach we have adopted so far in this Chapter is to use calculus to define the derivatives of J with respect to the unknown (free) parameters, to set these derivatives to zero, and solve using linear algebra the resulting Euler-Lagrange equations. Another approach is to view J as a multi-variate function of the unknowns $T_s^{[n]}, n = 1, \ldots, N-1$, and use a standard search algorithm to find it's minimum. Search algorithms we might appeal to in finding the minimum might include steepest decent,

conjugate gradient, or many of the other methods that have been customized for various purposes. (A good review of search algorithms useful in finding the minima of functions is provided by Press *et al.* [1989]. The optimization toolbox implimented by MATLAB® contains numerous algorithms that can be convienently programed in solving an inverse problem such as that represented by the minimization of J discussed here.)

The method we will adopt to minimize J as defined in Eqn. (7.118) is iterative. We develop a "down-gradient search" strategy which is summarized as follows: Step 1. We guess $T_s^{[n]},\ n = 1, \ldots, N-1$. Step 2. We compute J. Step 3. We compute $\partial J/\partial T_s^{[n]}$ for all the $n = 1, \ldots, N-1$. Step 4. We test to see if $\partial J/\partial T_s^{[n]} = 0$ (or close to zero). Step 5. If the test in Step 4 is affirmative, then the process stops, if not, the process continues by estimating an improvement to the initial guess $T_s^{[n]},\ n = 1, \ldots, N-1$ using an algorithm (not discussed here) that makes use of $\partial J/\partial T_s^{[n]}$. Step 6. We substitute the improved guess (from Step 5) for the initial guess formed in Step 1 and proceed to Step 2.

The hardest step is Step 5. It is hard because we must improve the original guess using only our knowledge of which "direction" in the vector space of all possible $\mathbf{T}_s$ is "downhill". In other words, we have only the particular values of J and $\partial J/\partial T_s^{[n]}$ for $n = 1, \ldots, N$ associated with our initial guess to work with in improving the guess. Despite the difficulty of this problem, many "canned" software routines are available for this task, and can be implimented with great ease using MATLAB® . We shall not consider the algorithms associated with such software procedures here. Instead we shall simply give reference to the section on conjugate-gradient methods in Press *et al.* [1989].

To impliment this search algorithm, we make use of the Euler-Lagrange equations associated with J. First and second, we make use of the Euler-Lagrange equations generated when we require that $\partial J/\partial \underline{\theta}^{[n]} = 0,\ n = 2, \ldots, N$ and $\partial J/\partial \underline{\lambda}^{[n]} = 0\ n = 1, \ldots, N-1$:

$$\mathbf{A}\underline{\theta}^{[n]} - \mathbf{B}\underline{\theta}^{[n-1]} = \mathbf{C}T_s^{[n-1]} \quad \text{for } n = 2, \ldots, N \tag{7.119}$$

$$\underline{\theta}^{[1]} = \mathbf{0} \tag{7.120}$$

$$\mathbf{A}'\underline{\lambda}^{[n-1]} - \mathbf{B}'\underline{\lambda}^{[n]} = \mathbf{0} \text{ for } n = N, \ldots, 2 \tag{7.121}$$

$$\mathbf{A}'\underline{\lambda}^{[N]} = -\left[\underline{\theta}^{[N]} - \underline{\theta}_b\right] \tag{7.122}$$

Equations (7.119) and (7.120) represent the forward problem, and Eqns. (7.121) and (7.122) represent the adjoint problem.

In Step 2 (described above), we must calculate J from a guessed $\mathbf{T}_s$. Equations (7.119) and (7.120) are what we use to impliment Step 2. We plug the values of the guessed $T_s^{[n]},\ n = 1, \ldots, N$ into Eqns. (7.119) and (7.120), solve for $\underline{\theta}^{[n]},\ n = 2, \ldots, N$ (our interest being predominantly to determine $\underline{\theta}^{[N]}$). Once $\underline{\theta}^{[N]}$ is calculated, we plug it into Eqn. (7.116) to give the explicit value of J associated with the initial guess.

Equations (7.121) and (7.122) represent the adjoint problem, and are needed to compute $\underline{\lambda}^{[n]},\ n = N-1, \ldots, 1$. (We will postpone for the moment a discussion of why we might be interested in knowing $\underline{\lambda}^{[n]},\ n = N-1, \ldots, 1$.) To compute $\underline{\lambda}^{[n]},\ n = N-1, \ldots, 1$, we take the mismatch between $\underline{\theta}^{[N]}$ (derived from solving the forward problem) and $\underline{\theta}_b$ to form the initial condition for $\underline{\lambda}^{[N]}$ represented by Eqn. (7.121). We then recursively apply Eqn. (7.120) to generate $\underline{\lambda}^{[n]},\ n = N-1, \ldots, 1$.

Armed with the $\underline{\lambda}^{[n]},\ n = N-1, \ldots, 1$, we proceed to Step 3, the calculation of $\partial J/\partial T_s^{[n]}$ for $n = 1, \ldots, N-1$. Taking the derivative of the expression in Eqn. (7.118) we obtain

$$\frac{\partial J}{\partial T_s^{[n]}} = \mathbf{C}'\underline{\lambda}^{[n+1]} \text{ for } n = 1, \ldots, N-1 \tag{7.123}$$

Armed with the derivatives expressed in the above formula, we then proceed with Steps 4 and 5.

One might wonder what advantage has been gained from calculating $\underline{\lambda}^{[n]},\ n = N-1, \ldots, 1$ using the adjoint problem. The answer concerns the work we would have had to perform to compute $\partial J/\partial T_s^{[n]}$ for $n = 1, \ldots, N-1$ using a finite-difference algorithm (*e.g.,* by computing J N times with slightly perturbed values of the initial guess $\mathbf{T}_s$ and forming difference approximations to the derivatives of J with respect to the components of $\mathbf{T}_s$).

To impliment a finite-difference algorithm, we would have to compute J N times. This would necessitate running the forward problem N times to determine the values of $\underline{\theta}^{[N]}$ needed to substitute into Eqn. (7.116). In contrast, by the control method described above, the $N-1$ derivatives of J with respect to the $N-1$ $T_s^{[n]}$ variables necessitates solving the forward problem once and solving the adjoint problem once. Considering the great similarity between the adjoint and forward problems in terms of computational difficulty, the control method effectively achieves what the finite-difference method achieves with a factor of $2/N$ less computational work. This is the main advantage for adopting the control method over other, less sophisticated methods.

7.12 Bibliography

Abramowitz, M. and I. A. Stegun, 1964. *Handbook of Mathematical Functions*, (National Bureau of Standards, Applied Mathematics Series, 55, U. S. Government Printing Office, Washington, DC, 1046 pp.)

Birch, F., 1948. The effect of Pleistocene climatic variations upon geothermal gradients. *American Journal of Science*, **246**, 729-760.

Broecker, W. S., G. Bond, M. Klas, G. Bonani, and W. Wolfli, 1990. A salt oscillator in the glacial Atlantic? 1. The concept. *Paleoceanography*, **5**, 469-477.

Carslaw, H. S. and J. C. Jaeger, 1988. *Conduction of Heat in Solids.* (Clarendon Press, Oxford, 510 pp.)

CLIMAP, 1976. The surface of the ice-age earth. *Science*, **191**, 1131-1137.

Dahl-Jensen, D., S. J. Johnsen, W. S. B. Paterson and C. Ritz, 1993. Comments on "Paleothermometry by control methods" by MacAyeal and others, *Journal of Glaciology*, **39**, 421-423.

Firestone, J., 1992. *Resolving the Younger Dryas Event through Borehole Thermometry.* (Phd Dissertation, University of Washington, 159 pp.)

Imbrie, J. and K. P. Imbrie, 1979. *Ice Ages: Solving the Mystery*, (Harvard University Press, Cambridge, Mass., 224 pp.)

Johnsen, S. J. and nine others, 1992. Irregular glacial interstadials recorded in a new Greenland ice core. *Nature*, **359**, 311-313.

Lachenbruch, A. H. and B. V. Marshall, 1986. Changing climate: geothermal evidence from permafrost in the Alaskan Arctic. *Science*, **234**, 689-696.

Lewis, T. (editor), 1992. Climatic change inferred from underground temperatures (special issue). *Global and Planetary Change*, **6**.

MacAyeal, D. R., J. Firestone, and E. Waddington, 1991. Paleothermometry by control methods. *Journal of Glaciology*, **37**, 326-338.

MacAyeal, D. R., J. Firestone, and E. Waddington, 1993. Paleothermometry redux. *Journal of Glaciology*, **39**, 423-431.

Pollack, H. N. and C. S. Chapman, 1993. Underground records of changing climate. *Scientific American*, **268**, 44-50.

7.13 Laboratory Exercise: Warming Anomaly in Permafrost of Northern Alaska

Recent analysis of geothermal temperature profiles in Northern Alaska by Lachenbruch and Marshall [1986] suggests that the climate has been warming during the last several decades to a century. Is this the first signal of the coming CO_2-greenhouse effect?

Problem 1 Assume that the Earth's climate has warmed over the last 100 years according to the time series displayed in Pollack and Chapman [1993]. Determine the present-day temperature anomaly in a borehole of depth 300 m that is associated with this warming time series. (Data will be found in the MATLAB® load file associated with this lab.)

Problem 2 Using the temperature anomaly generated in Problem 1, $\theta_b(z)$, determine a surface-temperature history $T_s^{inf}(t)$ using the least-squares inverse method described in this chapter. For the pre-concieved surface-temperature climatology, $\phi(t)$, use a running 10-year mean of the time-series used to generate the temperature anomaly. Use $\alpha = 0.5$. Compare your result with the actual surface temperature history from Pollack and Chapman [1993].

Problem 3 Perform the analysis in Problem 2, except use $\alpha = 0$, 0.25, and 0.75. Compare your results with the actual surface temperature history and with the result of Problem 2.

Problem 4 Compute the model resolution matrix $\mathbf{R}$ for the best-fitting and worst-fitting results of Problems 2 and 3. Display the matrix by graphing the value of its components along select rows as a function of the column number. By inspection of these graphs, how does the retrodiction of "recent" temperature history compare with the retrodiction of "ancient" temperature history?

Problem 5 Using the same data as that used by Lachenbruch and Marshall [1986], compute the surface temperature history of Northern Alaska. Do you think that Lachenbruch and Marshall's claim of climate warming is justified?

Chapter 8

Kalman Smoother: Estimation of Atmospheric Trace-Gas Emissions

8.1 Overview

In many scientific research areas, particularly in those which involve systems like the atmosphere and ocean which may be monitored for operational forecasting reasons, we are faced with the problem of blending noisy observations with a description of the system's dynamics to achieve an improved estimate of the system's state. In huricane forecasting, for example, we might wish to use observations about the prior drift of the storm center to update a forecast model. In chemical oceanography, we might use a time series of noisy tracer-concentration measurements and a numerical ocean circulation model to better describe the time-average ocean circulation regime. In atmospheric chemistry, we might wish to estimate the source of a particular chemical constituent, such as the now-outlawed chloroflorocarbons, using only measurements of the consituent's concentration in the stratosphere.

All of these problems may be treated using a class of inverse methods

known as "filtering, smoothing and prediction". One specific member of this class of methods known as the Kalman filter has risen to prominence in the meteorological and oceanographic literature. We shall focus on this method here and use use as a vehicle to explore this method a simple atmospheric chemistry problem.

8.2 An Idealized Atmospheric-Chemistry Model

For the purpose of teaching the conceptual basis of the Kalman filter, we will consider a idealized diffusive model of chemical transport in the atmosphere. In this model, we disregard vertical variation of chemical constituents, and proceed as if the atmosphere were simply a spherical surface as depicted in Fig. (8.1).

Let $c(t, \theta)$ be the vertically and zonally averaged concentration of a particular chemical constituent as a function of time t and latitude θ. Let $s(t, \theta)$ be the emission (or absorbtion, if less than zero) of this chemical due to natural (*e.g.,* photochemical) or anthropogenic processes. The diffusion equation which governs c is assumed, for the sake of illustration, to be

$$c_t = \frac{D}{R_e \cos\theta}(\cos\theta c_\theta)_\theta + s \tag{8.1}$$

where D is the diffusivity, R_e is the mean radius of the earth, and subscripts denote partial differentiation with respect to the subscripted variable. For a unique solution to the above equation, an initial condition must be specified. (Boundary conditions are not necessary because the domain, the earth's atmosphere, is considered to be a two-dimensional spherical surface.) Here, we require

$$c = 0 \quad \text{at} \quad t = 0 \tag{8.2}$$

It is best to simplify the above description of chemical tracer diffusion as much as possible prior to investigating the properties of it's solution. The first simplification is to adopt non-dimensional variables, specifically

$$t \quad \rightarrow \frac{R_e^2}{D} \tag{8.3}$$

$$c \rightarrow C_o c \tag{8.4}$$
$$s \rightarrow \frac{C_o D}{R_e^2} s \tag{8.5}$$

The second step is to change coordinates from θ to $x = \sin\theta$. The resulting statement of the dimensionless atmospheric diffusion problem is

$$c_t = ((1-x^2)c_x)_x + s \tag{8.6}$$
$$c = 0 \text{ at } t = 0 \tag{8.7}$$

8.3 A Simple Inverse Problem

The inverse problem we will focus on as a means of developing the Kalman filter may be defined as follows. Suppose that both the source and concentration of a particular chemical consituent are ovserved over a time interval $[0, T]$. Let these observations be denoted by $s^{obs}(x,t)$ and $c^{obs}(x,t)$, respectively. Furthermore, assume that the errors associated with these two observations (*e.g.,* errors due to measurement inaccuracy and the effects of over-simplified physics) have known statistics. In other words, let

$$s^o = s + \xi(x,t) \tag{8.8}$$
$$c^o = c + \zeta(x,t) \tag{8.9}$$

where ξ and ζ represent random errors which possess the following statistical properties

$$\langle \xi(x,t) \rangle = 0 \tag{8.10}$$
$$\langle \zeta(x,t) \rangle = 0 \tag{8.11}$$
$$\langle \xi(x,t)\zeta(x',t') \rangle = 0 \tag{8.12}$$
$$\langle \xi(x,t)\xi(x',t') \rangle = Q(x,x')\delta(t-t') \tag{8.13}$$
$$\langle \zeta(x,t)\zeta(x',t') \rangle = R(x,x')\delta(t-t') \tag{8.14}$$

where the expectation operator $\langle a(x,t) \rangle$ of an arbitrary variable $a(x,t)$ is defined by

$$\langle a(x,t) \rangle = \int_{-\infty}^{\infty} P_a(\lambda, x, t) a(x,t) d\lambda \tag{8.15}$$

and where $P_a(\lambda, x, t)$ is the probability that the random variable $a(x,t)$ will have value λ at position x and at time t. Eqns. (8.15) and (8.16) indicate that the errors in the observations of the source and concentration have zero mean. Eqn. (8.12) indicates that errors in the observations of the source are uncorrelated with errors in the observations of the concentration. Eqns. (8.13) and (8.14) indicate that errors in the observations of source and concentration are uncorrelated in time, but correlated in space according to the correlation functions $Q(x,x')$ and $R(x,x')$. We assume Q and R to be known to us at the outset of the analysis.

The inverse problem we wish to solve is this: Find an estimate of $s(x,t)$ and $c(x,t)$ that is based on the observations and that satisfies Eqns. (8.6) and (8.7). To make this inverse problem challenging, we assume at the outset that s^o and c^o do not satisfy Eqns. (8.6) and (8.7).

8.4 Green's Function Approach to the Forward Problem

To gain greater appreciation for the difficulty of the inverse problem defined above, we develop a Green's function approach to the solution of Eqns. (8.6) and (8.7). We begin this development by separating the variables x and t to split the homogeneous form ($s = 0$) of Eqn. (8.6) into a pair of ordinary differential equations involving one variable only. Accordingly, we write

$$c(x,t) = X(x)T(t) \tag{8.16}$$

where X and T are functions of x only and t only, respectively. Substitution of (8.16) into Eqn. (8.6), and division by XT, gives

$$\frac{T'}{T} = \gamma = \left(1 - x^2\right)\frac{X''}{X} - 2x\frac{X'}{X} \tag{8.17}$$

where γ is an undetermined constant, and primes denote differentiation. (We know that γ is a constant, and not a function of x or t, because the far left and far right sides of Eqn. (8.17) are functions of t only and of x only, respectively. Functions of t and of x can only be equal when they are constants.) Eqn.

(8.17) is satisfied when the following two ordinary differential equations are satisfied:

$$\left(1-x^2\right)X'' - 2xX' - \gamma X = 0 \tag{8.18}$$

$$T' - \gamma T = 0 \tag{8.19}$$

8.4.1 Legendre Equation

Eqn. (8.18) is known as Legendre's equation. It is solved by assuming that X has the form of a power series

$$X(x) = \sum_{n=0}^{\infty} a_n x^{n+\alpha} \tag{8.20}$$

where the $a_n,\ n = 0, \ldots, \infty$ are unknown coefficients and α is an unknown exponent. Substitution of (8.20) into Eqn. (8.18), and the requirement that coefficient of each power of x be zero separately, gives the following three conditions:

$$\alpha(\alpha-1)a_0 = 0 \tag{8.21}$$

$$\alpha(1+\alpha)a_1 = 0 \tag{8.22}$$

$$a_{n+2} = \frac{(n+\alpha)(n+\alpha+1)+\gamma}{(n+\alpha+2)(n+\alpha+1)}a_n \quad \text{for} \quad n = 2, \ldots, \infty \tag{8.23}$$

Eqn. (8.23) suggests that both a_0 and a_1 cannot be zero simuletaneously without resulting in the trivial solution where all the a_n are zero. We thus consider separately two cases where either $a_0 \neq 0$ or $a_1 \neq 0$. These two cases will yield series representations of X which either involve only even powers of x or odd powers of x (depending on the choice of α). We make note of the fact that, regardless of which coefficient a_0 or a_1 is taken to be non-zero, the choice of $\alpha = 0$ will ensure that Eqns. (8.21) and (8.22) are satisfied.

To ensure that the power series representation of X is convergent when $x = \pm 1$, we must insist that the series be truncated at some finite level. In other words, we must choose the undefined constant γ in a manner such that for some integer l

$$a_{l+2} = 0 \tag{8.24}$$

For $\alpha = 0$, Eqn. (8.23) demonstrates the above condition is met when $\gamma = -l(l+1)$.

8.4.2 Legendre Polynomials

We recognize the series representation for X for each l to be one of the Legendre polynomials [Arfken, 1970]. Denoting these polynomials as $P_l(x)$, Eqn. (8.18) can be written

$$\left(1 - x^2\right) P_l'' - 2xP_l' + l(l+1)P_n = 0 \tag{8.25}$$

The Legendre polynomials arise commonly in diffusion problems involving spherical geometry. Two of the properties which make these polynomials particularly useful is that they are orthonormal and complete. In other words,

$$\int_{-1}^{1} P_n(x)P_m(x)dx \quad = \tfrac{2}{2n+1}\delta_{nm} \tag{8.26}$$

$$\sum_{n=0}^{\infty} \frac{2n+1}{2} P_n(x)P_n(x') \quad = \delta(x - x') \tag{8.27}$$

These two properties of the Legendre polynomials tell us that any arbitrary function $f(x)$ defined on the interval $x \in [-1, 1]$ can be expressed as a series of P_n's

$$f(x) \qquad = \int_{-1}^{1} f(x')\delta(x - x')dx' \tag{8.28}$$

$$= \int_{-1}^{1} f(x') \sum_{n=0}^{\infty} \tfrac{2n+1}{2} P_n(x)P_n(x')dx' \tag{8.29}$$

$$= \sum_{n=0}^{\infty} f_n P_n(x) \tag{8.30}$$

where the expansion coefficients f_n are given by

$$f_n = \frac{2n+1}{2} \int_{-1}^{1} f(x') P_n(x') dx' \tag{8.31}$$

We will make use of this ability to express arbitrary functions as series of P_n's to simplify the inverse problem.

A text on mathematical methods may be consulted to see the details associated with the definitions of the Legendre polynomials [*e.g.,* Arfken, 1970]. The first 5 Legendre polynomials are listed here for reference,

$$\begin{aligned}
P_0(x) &= 1 \\
P_1(x) &= x \\
P_2(x) &= \tfrac{1}{2}\left(3x^2 - 1\right) \\
P_3(x) &= \tfrac{1}{2}\left(5x^3 - 3x\right) \\
P_4(x) &= \tfrac{1}{8}\left(35x^4 - 30x^2 + 3\right) \\
P_5(x) &= \tfrac{1}{8}\left(63x^5 - 70x^3 + 15x\right)
\end{aligned}$$

8.4.3 Time Dependence

Having determined the general form of $X(x)$ in terms of Legendre polynomials, we next solve Eqn. (8.19) for $T(t)$. The constant γ was required to be $-l(l+1)$ inorder to make the power series solution of Eqn. (8.18) converge, thus Eqn. (8.19) may be written

$$T' = -l(l+1)T \tag{8.32}$$

The solution is

$$T(t) = C_l e^{-l(l+1)t} \tag{8.33}$$

where C_l is an coefficient determined by the initial conditions.

8.4.4 General Solution to Homogeneous Forward Problem

Using the results of the previous two subsections, the general solution to the homogeneous form of the forward problem (where $s = 0$) may be expressed as

$$c(x,t) = \sum_{l=0}^{\infty} C_l e^{-l(l+1)t} P_l(x) \tag{8.34}$$

We will make use of this expression to determine the Green's function necessary to solve the non-homogeneous form of the forward problem (where $s \neq 0$).

8.4.5 The Inhomogeneous Forward Problem

To solve Eqns. (8.6) and (8.7), we define the Green's function G(x,t;x') which satisfies

$$G_t = -\left(\left(1-x^2\right) G_x\right)_x \tag{8.35}$$

$$G(x,t';x') = \delta(x-x') \tag{8.36}$$

where x' is a parameter (the location where $G(x,t')$ is a delta-function), and t' is the time when a terminal condition is applied in lieu of an initial condition. We construct the solution to the inhomogeneous forward problem by examining the following integral which we know to be zero due to Eqns. (8.6) and (8.35):

$$\int_0^{t'} \int_{-1}^{1} \left\{ G\left(c_t - \left(\left(1-x^2\right) c_x\right)_x - s\right) + c\left(G_t + \left(\left(1-x^2\right) G_x\right)_x\right)\right\} dxdt \tag{8.37}$$

Integration by parts, and use of the initial and terminal conditions on c and G (Eqns. (8.7) and (8.36)) yields the following simplifications:

$$\int_0^{t'} \int_{-1}^{1} \left\{ Gc_t + cG_t \right\} dxdt$$

$$= \int_0^{t'} \int_{-1}^{1} (Gc)_t \, dxdt$$

$$= \int_{-1}^{1} \left(c(x,t')G(x,t';x') - c(x,0)G(x,0;x') \right) dx$$

$$= c(x',t') \tag{8.38}$$

$$\int_0^{t'} \int_{-1}^{1} \left(-G \left(\left(1-x^2 \right) c_x \right)_x + c \left(\left(1-x^2 \right) G_x \right)_x \right) dxdt$$

$$= \int_0^{t'} \int_{-1}^{1} \left(\left(-G \left(1-x^2 \right) c_x \right)_x + \left(c \left(1-x^2 \right) G_x \right)_x \right) dxdt$$

$$+ \int_0^{t'} \int_{-1}^{1} \left(\left(G_x \left(1-x^2 \right) c_x \right) - \left(c_x \left(1-x^2 \right) G_x \right) \right) dxdt$$

$$= 0 \tag{8.39}$$

Use of the above simplifications in Eqn. (8.37), gives the Green's function form of the solution to the inhomogeneous forward problem:

$$c(x',t') = \int_0^{t'} \int_{-1}^{1} G(x,t;x') s(x,t) dxdt \tag{8.40}$$

Eqn. (8.40) represents the linear integral operator which determines the tracer concentration at (x',t') from the source emissions $s(x,t)$ at times $0 < t < t'$.

The solution to the homogeneous forward problem given by Eqn. (8.34) can be used to express $G(x,t;x')$ in terms of the Legendre polynomials. First, however, we must recognize that the transformation $\tau = t' - t$ is necessary to render Eqns. (8.35) and (8.36) into a form suitable for application of Eqn. (8.34). After making this transformation, we have

$$G(x,\tau;x') = \sum_{l=0}^{\infty} G_l(x') e^{-l(l+1)(t'-t)} P_l(x) \tag{8.41}$$

where the G_l are determined from the initial condition (at $\tau = 0$) as in Eqn. (8.31):

$$G_l(x') = \frac{2l+1}{2} \int_{-1}^{1} G(x, \tau = 0; x') P_l(x) dx \tag{8.42}$$

Using Eqn. (8.27) to express $G(x, \tau = 0; x')$, and recognizing the orthagonality condition Eqn. (8.26), we obtain

$$G_l(x') = \frac{2l+1}{2} P_l(x') \tag{8.43}$$

which ultimately yields

$$G(x, \tau; x') = \sum_{l=0}^{\infty} \frac{2l+1}{2} e^{-l(l+1)(t'-t)} P_l(x) P_l(x') \tag{8.44}$$

Eqns. (8.44) and (8.40) provide a complete formal description of the solution to the inhomogeneous atmospheric tracer diffusion problem.

8.4.6 Spectral Form of the Solution

The form of the Green's function indicated by Eqn. (8.44) and the completeness of the Legendre polynomials suggests that the complexities of spatial (x) dependence in the problem can be eliminated by considering only the time dependent coefficients for the Legendre-polynomial series representations of c and s. Using series representations for c and s in Eqn. (8.40) and the identity for G given in Eqn. (8.44), we obtain

$$c_l(t) = \int_{0}^{t} s_l(t') g_l(t, t') dt' \tag{8.45}$$

where

$$g_l(t, t') = e^{-l(l+1)(t'-t)} \tag{8.46}$$

We shall work with this spectral representation of the solution to the inhomogeneous forward problem in our development of the continuous form of the solution to the inverse problem.

8.5 Inverse Problem Restated

Suppose that we have observations $s_l^o(t)$ and $c_l^o(t)$ over the time period $t \in [0, T]$ and for all l. Suppose further that we know the covarience of the observation errors in advance, *i.e.,*

$$s_l^o(t) = s_l(t) + \xi_l(t) \tag{8.47}$$
$$c_l^o(t) = c_l(t) + \zeta_l(t) \tag{8.48}$$

where,

$$\langle \xi_l(t)\xi_l(t')\rangle = Q_l\delta(t-t') \tag{8.49}$$
$$\langle \zeta_l(t)\zeta_l(t')\rangle = R_l\delta(t-t') \tag{8.50}$$
$$\langle \xi_l(t)\zeta_l(t')\rangle = 0 \tag{8.51}$$

The inverse problem is to determine estimates $\hat{s}_l(t)$ and $\hat{c}_l(t)$ which satisfy Eqn(8.45) and which are, in some sense, more accurate than the observations s_l^o and c_l^o.

One way to achieve these estimate is to apply the Kalman filter [Bryson and Ho, 1975] to the observations. The Kalman filter yields a linear function of s_l^o and c_l^o which minimizes the following measure of accuracy denoted by J:

$$J = \int_0^T \langle (\hat{s}_l(t) - s_l(t))^2 \rangle dt \tag{8.52}$$

8.6 Continuous Version of the Kalman Filter

We derive the Kalman filter by assuming that $\hat{s}_l(t)$ is equal to $s_l^o(t)$ plus a correction that depends on the misfit between $c_l^o(t)$ and $\int_0^t s_l^o(t')g_l(t,t')dt'$,

$$\hat{s}(t) = s^o(t) + \int_o^T \mathfrak{K}(t,t') \left\{ c^o(t') - \int_0^{t'} s^o(t'')g(t',t'')dt'' \right\} dt' \tag{8.53}$$

Observe that subscripts l are dropped to simplify the notation. It should be understood that the above definition holds for each $s_l,\ l = 1, \ldots, \infty$. The kernel $\mathfrak{K}(t, t')$ is an undetermined function of t and t' referred to as the *gain function*. It will be determined below by applying the calculus of variations to minimize J defined in Eqn. (8.52). Before proceeding to determine $\mathfrak{K}(t, t')$, we note the following simplification:

$$\hat{s} - s = \hat{s} - s^o + s^o - s \tag{8.54}$$

Using Eqns. (8.53) and (8.47), we obtain

$$\hat{s}(t) - s(t) = \int_0^T \mathfrak{K}(t, t') \left\{ c^o(t') - \int_0^{t'} g(t', t'') s^o(t'') \right\} dt' + \xi(t) \tag{8.55}$$

We further make use of Eqn. (8.48), and assume that the true, exact functions s and c satisfy Eqn. (8.45):

$$c^o(t) = c(t) + \zeta(t) = \int_0^t g(t, t') s(t') dt' + \zeta(t) \tag{8.56}$$

Substitution of Eqn. (8.56) into Eqn. (8.55), and making further use of Eqn. (8.47), we obtain

$$\hat{s}(t) - s(t) = \int_0^T \mathfrak{K}(t, t') \left\{ \zeta(t') - \int_0^{t'} g(t', t'') \xi(t'') dt'' \right\} dt' + \xi(t) \tag{8.57}$$

Substitution of the above result into the definition of J given by Eqn. (8.52), and taking care to keep dummy variables of integration t', τ and τ' distinct, gives J in the following form

$$\begin{aligned} J = \int_0^T \langle & \left[\int_0^T \mathfrak{K}(t, t') \left\{ \zeta(t') - \int_0^{t'} g(t', t'') \xi(t'') dt'' \right\} dt' + \xi(t) \right] \\ & \cdot \left[\int_0^T \mathfrak{K}(t, \tau) \left\{ \zeta(\tau) - \int_0^{\tau} g(\tau, \tau') \xi(\tau') d\tau' \right\} d\tau + \xi(t) \right] \rangle dt \end{aligned} \tag{8.58}$$

The expectation operator $\langle\cdot\rangle$ commutes with the linear integral operators in Eqn. (8.58), thus the above expression (in Eqn. (8.58)) reduces to the sum of four terms

$$J = \int_0^T dt \int_0^T dt' \int_0^T d\tau \mathfrak{K}(t,t')\mathfrak{K}(t,\tau)\langle\zeta(t')\zeta(\tau)\rangle$$
$$+ \int_0^T dt \int_0^T dt' \int_0^T d\tau \int_0^{t'} dt'' \int_0^{\tau} d\tau' \mathfrak{K}(t,t')\mathfrak{K}(t,\tau)g(t',t'')g(\tau,\tau')E(\xi(t'')\xi(\tau'))$$
$$-2\int_0^T dt \int_0^T dt' \int_0^{t'} dt'' \mathfrak{K}(t,t')g(t',t'')\langle\xi(t'')\xi(t)\rangle$$
$$+ \int_0^T dt \langle\xi(t)\xi(t)\rangle \tag{8.59}$$

Use of the definitions (8.49)-(8.51) and careful integration over the range of variables which appear as arguments to the δ-functions simplifies the above expression for J

$$J = \int_0^T dt \int_0^T dt' \mathfrak{K}(t,t')\mathfrak{K}(t,t')R$$
$$+ \int_0^T dt \int_0^T dt' \int_0^T d\tau \mathfrak{K}(t,t')\mathfrak{K}(t,\tau)K(t',\tau)Q$$
$$-2\int_0^T dt' \int_0^{t'} dt'' \mathfrak{K}(t'',t')g(t',t'')Q$$
$$+ \int_0^T dt Q \tag{8.60}$$

where

$$K(t',\tau) = \int_0^{t'} dt'' g(t',t'')g(\tau,t'') \tag{8.61}$$

The expression in Eqn. (8.60) may be simplified further by introducing an integration over τ in the first term on the right hand side (necessitating the multiplication of the integrand by $\delta(t'-\tau)$) and by boosting the upper limit of integration over dt'' in the third term to T (necessitating the multiplication of the integrand by the Heaviside function $H(t''-t')$ which is zero when t'' is greater than t' and 1 otherwise). The result is

$$\begin{aligned} J &= \int_0^T dt \int_0^T dt' \int_0^T d\tau \mathfrak{K}(t,t')\mathfrak{K}(t,\tau)\left\{R\delta(t'-\tau) + QK(t',\tau)\right\} \\ &\quad -2\int_0^T dt' \int_0^T dt'' \mathfrak{K}(t'',t')g(t',t'')H(t''-t')Q + \int_0^T dtQ \end{aligned} \tag{8.62}$$

We are now ready to employ the calculus of variations to define the Euler-Legrange condition needed to define $\mathfrak{K}(t,t')$. Taking the variation of J with respect to $\mathfrak{K}(t,t')$ (and recognizing that the variable of integration t'' in the second term in the above equation is a dummy variable and may be replaced with t), we obtain

$$\begin{aligned} \delta J = 2\int_0^T dt \int_0^T dt' \int_0^T d\tau \delta\mathfrak{K}(t,t')\mathfrak{K}(t,\tau)\left\{R\delta(t'-\tau) + QK(t',\tau)\right\} \\ -2\int_0^T dt' \int_0^T dt \delta\mathfrak{K}(t,t')g(t',t)H(t-t')Q \end{aligned} \tag{8.63}$$

For $\delta J = 0$ for arbitrary $\delta\mathfrak{K}(t,t')$, the integrand of the integrals over t and t' in the above expression must be zero. This is the Euler-Lagrange condition we need to determine $\mathfrak{K}$:

$$\int_0^T d\tau \mathfrak{K}(t,\tau)\left\{R\delta(t'-\tau) + QK(t',\tau)\right\} - g(t',t)H(t-t')Q = 0 \tag{8.64}$$

The above expression is an integral equation of the second kind (owing to the δ-function in the integrand on the left-hand side).

The inverse problem is now solved, at least formally. Equation (8.64) is solved to produce the gain kernel $\mathfrak{K}(t,t')$ which, in turn, is used in Eqn. (8.53) with the data ($s^o(t)$ and $c^o(t)$) to define the estimate $\hat{s}(t)$. This estimate is then stuffed into the Green's function representation of the solution to the forward problem (Eqn. (8.45)) to yield the estimate $\hat{c}(t)$.

8.7 Discrete Version of the Kalman Filter

In the previous section we defined the Kalman filter as the linear integral operator given in Eqn. (8.53) where the kernel $\mathfrak{K}$ is chosen to minimize the expectation value of the error denoted by J in Eqn. (8.52). In this section, we derive a finite-difference version of the inverse problem which may be of use in circumstances where continuous function analysis is impractical.

8.7.1 Finite-Difference version of the Forward Problem

As stated previously, the tracer diffusion problem representing the forward problem is

$$c_t = \left(\left(1-x^2\right)c_x\right)_x + s \tag{8.65}$$

$$c(x,t=0) = 0 \tag{8.66}$$

As suggested by the analysis of the continuous version of the problem, it is appropriate to express both c and s in terms of the Legendre polynomials

$$c = \sum_{l=0}^{\infty} c_l(t) P_l(x) \tag{8.67}$$

$$s = \sum_{l=0}^{\infty} s_l(t) P_l(x) \tag{8.68}$$

Noting the fact that

$$\left(\left(1-x^2\right)\frac{dP_l}{dx}\right)_x = -l(l+1)P_l(x) \tag{8.69}$$

the forward problem represented by Eqns. (8.65) and (8.66) reduces to a system of ordinary differential equations for the $c_l(t)$'s:

$$\dot{c}_l = -l(l+1)c_l + s_l \tag{8.70}$$

$$c_l(t=0) = 0 \tag{8.71}$$

which hold for $l = 1, \ldots, \infty$.

A finite difference version of Eqns. (8.70) and (8.71) is

$$\frac{c_l^{n+1} - c_l^n}{\Delta t} + l(l+1)c_l^{n+1} \quad = s_l^{n+1} \tag{8.72}$$

$$c_l^{n=1} \qquad = 0 \tag{8.73}$$

In the above equations, superscripts denotes the discrete time level. Using vector notation,

$$\mathbf{c}_l = \begin{bmatrix} c_l^1 \\ c_l^2 \\ \vdots \\ c_l^n \\ \vdots \\ c_l^N \end{bmatrix} \tag{8.74}$$

$$\mathbf{s}_l = \begin{bmatrix} s_l^1 \\ s_l^2 \\ \vdots \\ s_l^n \\ \vdots \\ s_l^N \end{bmatrix} \tag{8.75}$$

where $(N-1)\Delta t = T$ and Δt is the time-step size.

The solution to the forward problem expressed in Eqns. (8.70) and (8.71) may be expressed as

$$\mathbf{c}_l = \mathbf{G}_l \mathbf{s}_l \tag{8.76}$$

where

$$\mathbf{G}_l = \mathbf{A}_l^{-1}\mathbf{B} \tag{8.77}$$

and

$$\mathbf{A}_l = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ & \ddots & & & & & & & \\ & & & & 0 & \frac{-1}{\Delta t} & \left(\frac{1}{\Delta t} + l(l+1)\right) & & \\ & & & & & & & \ddots & \end{bmatrix} \tag{8.78}$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ & & & & \ddots \end{bmatrix} \tag{8.79}$$

8.7.2 Inverse Problem in Discrete Form

The discrete, finite difference version of the inverse problem we wish to consider is to determine estimates of $\mathbf{s}_l$ and $\mathbf{c}_l$, denoted by $\hat{\mathbf{s}}_l$ and $\hat{\mathbf{c}}_l$, which are linear functions of observations $\mathbf{s}_l^o$ and $\mathbf{c}_l^o$, which satisfy Eqns. (8.76) and (8.77), and which minimize the expectation value of the trace J of the covarience matrix $\mathbf{E}_l$ defined by

$$\mathbf{E}_l = \langle(\hat{\mathbf{s}}_l - \mathbf{s}_l)(\hat{\mathbf{s}}_l - \mathbf{s}_l)'\rangle \tag{8.80}$$

In other words, we wish to select a linear combination of $\hat{\mathbf{s}}_l$ and $\hat{\mathbf{c}}_l$ such that

$$J = \sum_{i=1}^{N} (E_l)_{ii} \tag{8.81}$$

is minimized.

As in the continuous version of the inverse problem, we know in advance the covarience matricies representing observation error:

$$\mathbf{s}_l^o = \mathbf{s}_l + \underline{\xi}_l \tag{8.82}$$

$$\mathbf{c}_l^o = \mathbf{c}_l + \underline{\zeta}_l \tag{8.83}$$

with covarience

$$\langle \underline{\xi}_l \underline{\xi}_l' \rangle = \mathbf{Q}_l \tag{8.84}$$
$$\langle \underline{\zeta}_l \underline{\zeta}_l' \rangle = \mathbf{R}_l \tag{8.85}$$
$$\langle \underline{\xi}_l \underline{\zeta}_l' \rangle = 0 \tag{8.86}$$

where the matricies $\mathbf{Q}_l$ and $\mathbf{R}_l$ are diagonal (zeros on off-diagonal elements) but may have different values for each of the diagonal elements depending on the nature of the observational noise.

8.7.3 Matrix Form of the Kalman Filter

We assume at the outset that $\hat{\mathbf{s}}$ is a linear function of $\mathbf{s}^o$ and the misfit between $\mathbf{c}^o$ and $\mathbf{G}\mathbf{s}^o$ (henceforth, we shall drop subscripts l for notational simplicity),

$$\hat{\mathbf{s}} = \mathbf{s}^o + \mathbf{K}(\mathbf{c}^o - \mathbf{G}\mathbf{s}^o) \tag{8.87}$$

where the matrix $\mathbf{K}$ is the gain matrix to be determined by minimization of J.

We proceed as in the continuous case by first noting the identity

$$\begin{aligned}
\hat{\mathbf{s}} - \mathbf{s} &= \hat{\mathbf{s}} - \mathbf{s}^o + \mathbf{s}^o - \mathbf{s} \\
&= \mathbf{K}\left(\mathbf{c}^o - \mathbf{G}\mathbf{s}^o\right) + \underline{\xi} \\
&= \mathbf{K}\left(\mathbf{c} + \underline{\zeta} - \mathbf{G}\mathbf{s}^o\right) + \underline{\xi} \\
&= \mathbf{K}\left(\underline{\zeta} - \mathbf{G}\left(\mathbf{s}^o - \mathbf{s}\right)\right) + \underline{\xi} \\
&= \mathbf{K}\left(\underline{\zeta} - \mathbf{G}\underline{\xi}\right) + \underline{\xi} \\
&= \left(\mathbf{I} - \mathbf{K}\mathbf{G}\right)\underline{\xi} + \mathbf{K}\underline{\zeta}
\end{aligned} \tag{8.88}$$

The index we wish to minimize thus becomes (recalling that the expectation operator commutes with the matrix operations)

$$\begin{aligned}
J &= \mathrm{tr}\left(\left(\mathbf{I} - \mathbf{K}\mathbf{G}\right)\langle \underline{\xi}\underline{\xi}' \rangle \left(\mathbf{I} - \mathbf{G}'\mathbf{K}'\right) + \mathbf{K}\langle \underline{\zeta}\underline{\zeta}' \rangle \mathbf{K}'\right) \\
&= \mathrm{tr}\left(\left(\mathbf{I} - \mathbf{K}\mathbf{G}\right)\mathbf{Q}\left(\mathbf{I} - \mathbf{G}'\mathbf{K}'\right) + \mathbf{K}\mathbf{R}\mathbf{K}'\right)
\end{aligned} \tag{8.89}$$

In component notation,

$$J = (\delta_{ij} - K_{ik}G_{ka})Q_{ab}(\delta_{bi} - G_{lb}K_{il}) + K_{ik}R_{kl}K_{il} \tag{8.90}$$

Taking the derivative of J with respect to the $\alpha\beta$'th component of $\mathbf{K}$, and setting to zero gives

$$\frac{1}{2}\frac{\partial J}{\partial K_{\alpha\beta}} = -G_{\beta a}Q_{ab}(\delta_{b\alpha} - G_{lb}K_{\alpha l}) + R_{\beta l}K_{\alpha l} = 0 \tag{8.91}$$

Recognizing the fact that the zero on the right-hand side of the above equation defines the $\alpha\beta$'th element of a matrix, we derive the following matrix representation of Eqn. (8.91)

$$-\mathbf{GQ}(\mathbf{I} - \mathbf{G}'\mathbf{K}') + \mathbf{R}\mathbf{K}' = \mathbf{0} \tag{8.92}$$

This matrix equation can be readily solved to find $\mathbf{K}$:

$$\mathbf{K}' = \left[\mathbf{GQG}' + \mathbf{R}\right]^{-1}\mathbf{GQ} \tag{8.93}$$

To solve the inverse problem, *i.e.,* to determine $\hat{\mathbf{s}}$ and $\hat{\mathbf{c}}$, the gain matrix specified above is substituted into Eqn. (8.87) to determine $\hat{\mathbf{s}}$ which, in turn, is substituted into Eqn. (8.76) to determine $\hat{\mathbf{c}}$. The linear operator defined by Eqn. (8.87) is referred to as the Kalman filter, and is named after the famous applied mathmatician R. E. Kalman.

8.7.4 Estimate Covarience

An explicit expression for the the covarience matrix $\mathbf{E}$ defined by Eqn. (8.80) is readily obtained from the Kalman filter derived above. First we note that

$$\begin{aligned}\mathbf{E} &= (\mathbf{I} - \mathbf{KG})\mathbf{Q}(\mathbf{I} - \mathbf{G}'\mathbf{K}') + \mathbf{KRK}' \\ &= \mathbf{K}(\mathbf{GQG}' + \mathbf{R})\mathbf{K}' - \mathbf{KGQ} + \mathbf{Q}(\mathbf{I} - \mathbf{G}'\mathbf{K}')\end{aligned} \tag{8.94}$$

Using Eqn. (8.93), we observe that

$$(\mathbf{GQG}' + \mathbf{R})\mathbf{K}' = \mathbf{GQ} \tag{8.95}$$

Substitution of the above identity into Eqn. (8.94) gives

$$\mathbf{E} = \mathbf{Q}(\mathbf{I} - \mathbf{G}'\mathbf{K}') \tag{8.96}$$

Future analysis (that I intend to perform) will show that the covarience $\mathbf{E}$ is *reduced* in magnitude by the use of the Kalman filter. In other words, treatment of the data by the Kalman filter yields a better estimate of the tracer source function than the original observation.

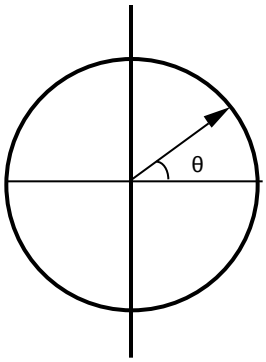

Figure 8.1: Geometry of atmospheric tracer problem.

Chapter 9

Backus-Gilbert Method: Free Oscillations of Lake Michigan

9.1 Introduction

In this chapter, we will use the frequency spectrum of the free oscillations of a shallow, narrow lake (Lake Michigan) to determine its longitudinal depth profile. The formulation of this problem and it's solution is intended to introduce the techniques used to infer the internal density structure of the earth from seismic data, and the famous inverse method developed by G. Backus and F. Gilbert [1967, 1968] for solving this internal density structure problem. The Backus-Gilbert method can be differentiated from the minimum-norm technique discussed previously in that its objective is to optimize the resolution of undetermined model parameters. As with the Kalman smoother described in Chapter (8), fitting data takes a back seat to optimization of a fundamental skill property of the inverse method (*e.g.,* the model-covarience matrix or the model-resolution matrix).

At the conclusion of this chapter, we shall investigate the formal differences between the Backus-Gilbert method and the minimum-norm method described in Chapter (2). We shall prove that the minimum-norm inverse is

identical to the Backus-Gilbert inverse in circumstances where the number of non-zero singular values resulting from the SVD K is equal to the dimension of the undetermined model space N. In circumstances where $K < N$, the Backus-Gilbert inverse is not a minimum-norm inverse. This does not mean that the Backus-Gilbert inverse is less appropriate than the minimum-norm inverse. It does, however, suggest that a somewhat arbitrary penalty associated with model simplicity is being paid when the Backus-Gilbert inverse is chosen over the minimum-norm inverse.

9.2 Free-Oscillations of a Long, Narrow Lake

In the mid 1960's, two exceptionally gifted seismologists named George Backus and Freeman Gilbert derived a technique for determining the density structure and elastic properties of the earth from observations of the earth's vibrational frequencies. To be specific, they used the first ten or so vibration frequencies (corresponding to the most grave spatial structure of the vibrating planet) to determine the planet's density, bulk modulus, shear modulus, and the local quality factor (a variable which determines the attenuation due to non-elastic properties), all as a function of radial position. The astounding aspect of their work, aside from the creative use of data, was the inventive nature of the inverse method used to solve the inverse problem. This inverse method has become widely used in the geophysical community, and is known as the "Backus-Gilbert" method.

Our goal will be to understand the Backus-Gilbert method in the same context as it was originally invented. To reach this goal, however, we will abandon the original problem posed by Backus and Gilbert (1967 and 1968), and formulate a similar, but much simpler problem involving a long and narrow lake (such as Lake Michigan). The physics of planetary free oscillations is too demanding for our purposes; thus we replace the elastic planet with a lake and consider a much simpler system which possesses the same intrinsic property: the tendency to oscillate at discrete frequencies.

To this end, we now consider the infinatessimally small free-oscillations of the surface of a long, narrow, and shallow lake such as Lake Michigan. We

emphasize that the lake must be long, narrow, and shallow so that a one-dimensional analysis which neglects the rotation of the earth can be used to describe the mass continuity and momentum balance of the water in the lake. (For an in-depth analysis of lacustrian wave dynamics, refer to Hutter [1993].) In actuality, Lake Michigan probably does not satisfy these conditions, but we shall proceed under the assumption that earth rotation may be neglected. The momentum and mass continuity equations for this body of water can be simplified to the following form (again, a strict development of the conditions which must be met for such simplification to be valid is provided elsewhere [Hutter, 1993]):

$$\frac{\partial v}{\partial t} = -g\frac{\partial \eta}{\partial x} \tag{9.1}$$

$$\frac{\partial \eta}{\partial t} = -\frac{\partial}{\partial x}(hv) \tag{9.2}$$

where $v(x,t)$ is the z-independent horizontal velocity (directed longitudinally along the lake), $\eta(x,t)$ is the perturbation of the free surface of the lake about its position of rest, $h(x)$ is the depth, $g = 9.81$ m/sec^2 is the gravitational acceleration, x is the longitudinal spatial coordinate, and t is time. Equations (9.1) and (9.2) apply whenever the depth of the lake is small compared to the horizontal scale of the motions being considered, when the lake water is homogeneous, and when the perturbations to the free surface (or velocities) are sufficiently small that non-linear effects need not be considered. The boundary condition to be applied at $x = 0$ and $x = L$, where 0 and L are the x-coordinates of the lake's ends, is the no-flux condition

$$v = 0 \tag{9.3}$$

Using Eqn. (9.1), the no-flux boundary condition implies that

$$\frac{\partial \eta}{\partial x} = 0 \quad \text{at } x = 0, L \tag{9.4}$$

As shown in Fig. (1), the lake is taken to be a long, canal-like body of depth $h(x)$. These two equations can be combined into a single, second order partial differential equation, known as a wave equation, by taking the time derivative of Eqn. (9.2) and using (9.1):

$$\frac{\partial^2 \eta}{\partial t^2} = g\frac{\partial}{\partial x}\left(h\frac{\partial \eta}{\partial x}\right) \tag{9.5}$$

Equation (9.5) is termed "separable" because $h(x)$ is a function of x-only. In this circumstance, η can be written as a products of two functions, $X(x)$ and $T(t)$, which are functions of a single variables only, x or t, respectively:

$$\eta(x,t) = X(x)T(t) \tag{9.6}$$

By writing η this way, we can reduce the partial differential equation (9.5) to two separate ordinary differential equations involving functions of a single variable only:

$$\frac{\mathrm{d}^2T}{\mathrm{d}t^2} = -\omega^2 T \tag{9.7}$$

$$g\frac{\mathrm{d}}{\mathrm{d}x}\left(h\frac{\mathrm{d}X}{\mathrm{d}x}\right) + \omega^2 X = 0 \tag{9.8}$$

Equation (9.7) can be easily solved in terms of $T(t)$:

$$T(t) = A\cos(\omega t) + B\sin(\omega t) \tag{9.9}$$

The constants A and B are normally determined by an initial condition. In the study of free oscillations, we need not concern ourselves with initial conditions; thus we leave A and B undetermined.

The frequency ω is yet to be determined. All of the information available in the time-dependent equation (9.7) is now exhausted. We thus look to Eqn. (9.8) to provide an evaluation of ω. The theory of differential equations tells us that there exists a class of functions $\{X_n(x)\}_{n=1}^{\infty}$ which are *eigenfunctions* of the operator $\left[g\frac{\partial}{\partial x}\left(h(x)\frac{\partial}{\partial x}\right) + \omega^2\right]$ and which satisfy the boundary conditions:

$$\frac{\mathrm{d}X}{\mathrm{d}x} = 0 \qquad \text{at } x = 0, L \tag{9.10}$$

Associated with each eigenfunction X_n is an eigenfrequencies ω_n which can be grouped together in ascending order, $\{\omega_n\}, n = 1, ..., \infty$ with $\omega_n \leq \omega_n + 1$. In this circumstance, the function $T(t)$ associated with a particular $X_n(x)$ is also associated with a corresponding ω_n. Thus the functions $T_n(t)$ are also given the subscript n.

Orthagonality of Eigenfunctions

An important property of the $X_n(x)$ is their *orthogonality*:

$$\int_0^L X_n X_m \, dx = 0 \quad \text{if } n \neq m \tag{9.11}$$

$$\int_0^L X_n X_m \, \mathrm{d}x = 1\mathrm{m}^3 \quad \text{if } n = m \tag{9.12}$$

Here we remark that the eigenfunctions have dimensions of length. This will make the analysis to come somewhat awkward. Use of non-dimensional variables would eliminate this awkward character.

Proof. We consider the integral of the product of X_m with Eqn. (9.8) evaluated with X_n:

$$\int_0^L X_m \cdot \left\{ \omega_n^2 X_n + g \frac{\mathrm{d}}{\mathrm{d}x} \left(h \frac{\mathrm{d}X_n}{\mathrm{dx}} \right) \right\} \mathrm{d}x = 0 \tag{9.13}$$

After itegration by parts, we obtain:

$$\omega_n^2 \int_0^L X_m \cdot X_n \, \mathrm{d}x + g \left\{ \int_0^L \frac{\mathrm{d}}{\mathrm{d}x} \left(X_m \cdot h \cdot \frac{\mathrm{d}X_n}{\mathrm{d}x} \right) \mathrm{d}x - \int_0^L \frac{\mathrm{d}X_m}{\mathrm{d}x} \cdot h \cdot \frac{\mathrm{d}X_n}{\mathrm{d}x} \, \mathrm{d}x \right\} = 0 \tag{9.14}$$

The first part of the second term (contained in brackets) is zero, because of the boundary conditions stated by Eqn. (9.10). The remaining part of the second term (contained in brackets) can be integrated by parts again, to give:

$$\omega_n^2 \int_0^L X_m \cdot X_n \, \mathrm{d}x - g \left\{ \int_0^L \frac{\mathrm{d}}{\mathrm{d}x} \left(X_n \cdot h \cdot \frac{\mathrm{d}X_m}{\mathrm{d}x} \right) \mathrm{d}x - \int_0^L X_n \cdot \frac{\mathrm{d}}{\mathrm{d}x} \left(h \cdot \frac{\mathrm{d}X_m}{\mathrm{d}x} \right) \mathrm{d}x \right\} = 0 \tag{9.15}$$

Boundary conditions can again be used to assure that the first part of the second term (contained in brackets) is zero. Finally, Eqn. (9.8) can again be

used to rewrite the derivatives of X_m in terms of ω_m^2:

$$(\omega_n^2 - \omega_m^2) \int_0^L X_m \cdot X_n \, \mathrm{d}x = 0 \tag{9.16}$$

By assumption, $\omega_n^2 \neq \omega_m^2$ when $n \neq m$. Thus, for the expression on the left-hand side of Eqn. (9.16) to be zero when $n \neq m$, the integral must be zero. This proves that X_n and X_m are orthogonal. ■

9.3 Eigenfunctions and Eigenfrequencies of a Flat-Bottomed Lake

If the lake depth is constant, $h(x) = h_o$, the $\{\omega_n\}_{n=1}^{\infty}$ and $\{X_n\}_{n=1}^{\infty}$ can be determined without difficulty. In this case, Eqn. (9.8) reduces to

$$gh_o \frac{\mathrm{d}^2 X_n}{\mathrm{d}x^2} + \bar{\omega}^2 X = 0 \tag{9.17}$$

where the overbar denotes the eigenfunctions and eigenfrequencies associated with a flat-bottomed lake. The solution of this equation, subject to the boundary conditions (9.3), is

$$\bar{X}_n(x) = \sqrt{\frac{2}{L}} \cos\left(\frac{n\pi x}{L}\right) \tag{9.18}$$

The frequencies associated with these eigenmodes, $\bar{\omega}_n$, are

$$\bar{\omega}_n = n\pi \frac{\sqrt{gh_0}}{L} \tag{9.19}$$

The fraction appearing on the right-hand side of Eqn. (9.19) deserves comment. The numerator $\sqrt{gh_o}$ represents the magnitude of the phase (and group) velocity of shallow-water gravity waves on a shallow lake. The frequencies thus can be interpreted in terms of the time taken for a shallow-water gravity wave to cross the length of the lake. This interpretation is relevant because the free-oscillations of the Lake can be thought of as the constructive interference between travelling waves which reflect off the two closed ends of the lake.

Completeness

It is important to record for future reference that the eigenfunctions associated with a flat-bottomed lake are not *complete* in the sense that an arbitrary function of x, say $h(x)$, cannot necessarily be expressed as a linear combination of the $\bar{X}_n$'s. The $\bar{X}_n$'s are all cosine functions. Sine functions must be added to the set $\{\bar{X}_n\}_{n=1}^{\infty}$ to yield a complete set of functions that, for example, could represent $h(x)$ as a linear combination.

9.3.1 A Discretized Analysis of the Flat-Bottomed Lake

It is worth mentioning briefly a numerical technique for determining the frequencies and eigenfunctions. Such a technique would be needed if $h(x)$ were not a simple function of x. For the case of a flat bottomed lake, the finite-difference version of Eqn. (9.17) and the boundary condition (9.3) can be written:

$$\frac{gh_o}{\Delta x^2}\left(\bar{X}_{i+1} - 2\bar{X}_i + \bar{X}_{i-1}\right) + \bar{\omega}^2\bar{X}_i = 0 \quad \text{for } i = 2, ..., l-1 \tag{9.20}$$

$$\bar{X}_2 - \bar{X}_1 = 0 \tag{9.21}$$

$$\bar{X}_l - \bar{X}_{l-1} = 0 \tag{9.22}$$

where the subscripts denote the gridpoint at which X is evaluated, and l is the number of gridpoints. The index i runs from 1 to l; thus the value of Δx is $L/(l-1)$. Equations (9.20)-(9.22) can be represented using matrix notation as follows:

$$\left(\mathbf{A} + \bar{\omega}^2\mathbf{I}\right)\bar{\mathbf{X}} = \mathbf{0} \tag{9.23}$$

where the matrix $\mathbf{A}$ is a $l \times l$ square matrix with elements given by:

$$A_{ii} = \frac{-2gh_o}{\Delta x^2} \quad \text{if } i = 2, ..., l-1 \tag{9.24}$$

$$A_{ii-1} = \frac{gh_o}{\Delta x^2} \quad \text{for } i = 2, ..., l-1 \tag{9.25}$$

$$A_{ii+1} = \frac{gh_o}{\Delta x^2} \quad \text{for } i = 2, ..., l-1 \tag{9.26}$$

$$A_{11} = 1 \tag{9.27}$$

$$A_{12} = -1 \tag{9.28}$$

$$A_{ll} = 1 \tag{9.29}$$

$$A_{l-1\ l} = -1 \tag{9.30}$$

$$A_{ij} = 0 \quad \text{otherwise} \tag{9.31}$$

and $\mathbf{I}$ is the $l \times l$ identity matrix, $\bar{\mathbf{X}}$ is the column vector containing the values of $\bar{X}$ at the l gridpoints, and $\bar{\omega}^2$ is a scalar.

Equation (9.23) can be recognized as the problem which must be solved for $-\bar{\omega}^2$ to find the eigenvalues and eigenvectors of the matrix $\mathbf{A}$. (Notice that the eigenvalues λ_n of $\mathbf{A}$ are identified with $-\bar{\omega}_n^2$.) As demonstrated below, it is easy to determine the eigenfrequencies and eigenvectors of $\mathbf{A}$ using the MATLAB® routines set up for this purpose. One problem that routinely crops up with a finite-difference approach is that the accuracy of the eigenfrequencies ω_n degrades as $n \to l$, where l is the mumber of eigenvalues of $\mathbf{A}$. This is a well-known flaw of centered finite-difference methods applied to wave-propagation and free-oscillation problems.

Example: Free-Oscillations of a Flat-Bottomed Lake Michigan

We use the above finite-difference formulation to estimate the $\bar{\omega}_n$ for Lake Michigan. We take $h_o = 150$m, $L = 650$ km, $g = 9.81$ m s^{-1}, and $l = 50$. The following MATLAB® routine is used to generate the eigenvalues and eigenfunctions of the matrix $\mathbf{A}$ above:

```
g=9.81;
h_naut=150;
L=650e3;
omega_bar = [pi*[1:50]*sqrt(g*h_naut)/L]';
T_bar= (2*pi)*ones(size(omega_bar))./omega_bar;
%The above determines the unperturbed frequencies
Ngrid=50;
```

```
dx=L/(Ngrid-1);
A=zeros(Ngrid,Ngrid);
A=A+2*g*h_naut/dx2̂ *eye(size(A));
A=A-diag(ones(Ngrid-1,1)*(g*h_naut/dx^ 2),-1);
A=A-diag(ones(Ngrid-1,1)*(g*h_naut/dx^ 2),1);
A(1,2)=-1/dx;
A(Ngrid,Ngrid-1)=-1/dx;
A(1,1)=1/dx;
A(Ngrid,Ngrid)=1/dx;
omega=sqrt(eig(A));
%Discard zero-frequency mode (steady solution):
omega=sort(omega);
omega=omega(2:Ngrid);
plot(omega); hold on; plot(omega_bar)
title('Analytic vs.  Finite-Difference Frequencies')
xlabel('Mode Number')
ylabel('1/sec')
pause
[V,D]=eig(A);
hold off
clg
[s,I]=sort(diag(D));
plot(V(:,I(2)))
hold on
plot(V(:,I(3)))
title('First Two Eigenmodes of Lake Michigan')
xlabel('Normalized distance along lake (L=50 units)')
ylabel('Free-surface elevation (m)')
```

Figures (9.1) and (9.2) display the results of the above MATLAB® algorithm. Figure (9.1) suggests that the finite-difference method for determining the frequencies of free oscillation will be accurate only for the lowest frequency modes. Figure (9.2) demonstrates the finite-difference representation of the cosine solutions given by Eqn. (9.18) for two of the most grave modes ($n = 1, 2$).

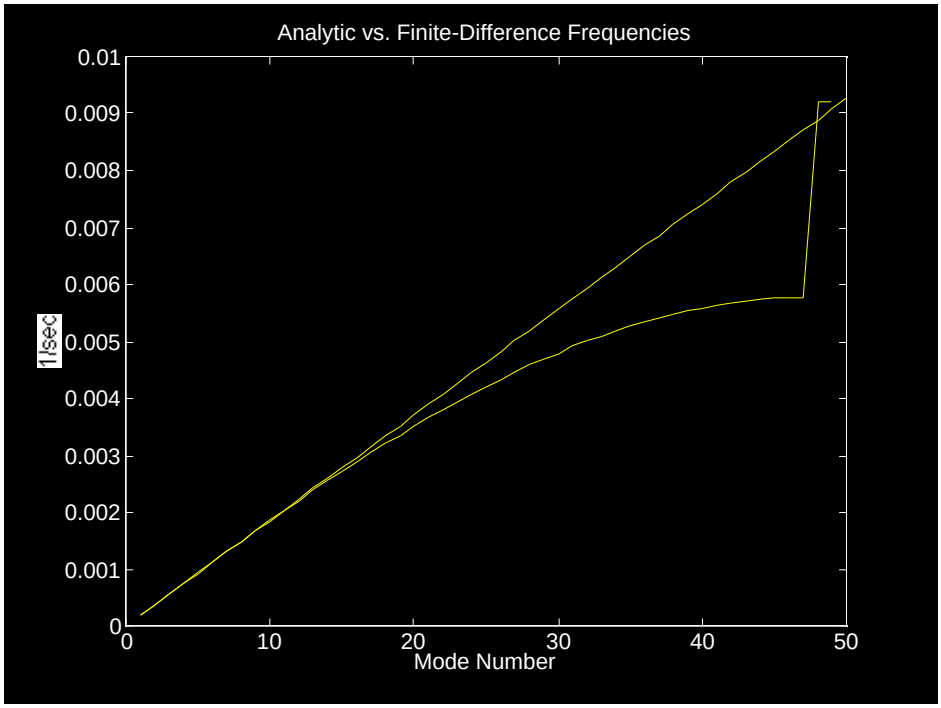

Figure 9.1: The frequencies of free-oscillation $\bar{\omega}_n$ plotted against n for a flat-bottomed Lake Michigan of depth 150 m and length 650 km. The straight line represents the exact analytic result given by Eqn. (9.19). The curved line represents the result that comes from the finite-difference algorithm. (Note that the finite-difference approach creates 50 eigenvalues. The first is zero and is associated with the X =constant solution. The above graph reflects the fact that the zero-frequency eigenvalue and eigenfunction were discarded.) The finite-difference calculated $\bar{\omega}_n$ fall below the analytic values as n becomes large. Note that the slope of the finite-difference line goes to zero for large n. This defect is associated with the fact that grid-point-to-grid-point oscillations will contaminate the finite difference calculation when forced with freely propagating waves.

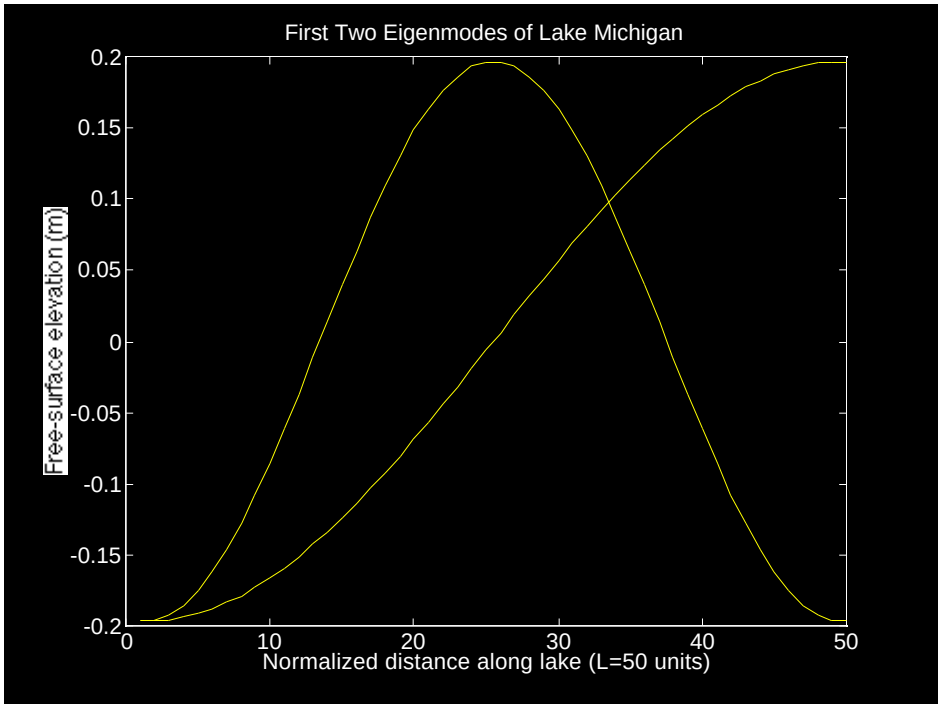

Figure 9.2: The two most grave eigenfunctions ($n = 1, 2$) of free-oscillation $\bar{X}_n$ plotted against distance (where $L = 50$ plotting units) for a flat-bottomed Lake Michigan of depth 150 m and length 650 km. Both functions exhibit zero slope at $x = 0, L$ as required by the no-flux boundary conditions.

9.4 An Inverse Problem

Given the M lowest frequencies of free oscillation, ω_n, $n = 1, \ldots, M$, observed using the time-series analysis of a tide guage record, determine the $h(x) = h_o + \zeta(x)$ depth profile of the lake. In other words, determine $\zeta(x)$ from the partial frequency spectrum $\{\omega_n\}_{n=1}^{M}$. ■

As mentioned before, this inverse problem is similar to that embraced by Backus and Gilbert [1968]. Their interest was in the reconstruction of the internal structure of the earth. The data they used to make this reconstruction was a subset of the frequency spectrum of the earth's free oscillations. We have chosen to focus on the free oscillations of a long narrow lake (Lake Michigan) to avoid the complexity of the physics associated with elastic deformation of the earth. The added complexity of the original problem solved by Backus and Gilbert adds nothing to the mathematical nature of the Backus-Gilbert inverse which we will develop below.

9.5 Linearization of the Inverse Problem

What makes the above inverse problem so difficult is that the observed eigenfrequencies, ω_n, are related by Eqn. (9.8) to two unknown functions of x: the depth perturbation $\zeta(x)$ which is what we want to determine, and the set of M eigenfunctions $\{X_n\}_{n=1}^{M}$. The relationship between the observed ω_n and $\zeta(x)$ can be simplified when $\zeta(x)$ is small, *i.e.,* when $\zeta(x)/h_o << 1$. In this circumstance, the unknown X_n's in Eqn. (9.8) may be linearized about the known $\bar{X}_n$'s associated with a flat-bottomed lake of depth h_o. As we shall see below, performing this linearization yields a linear relationship between ζ and the observed ω_n which only involves the functions $\bar{X}_n$.

We begin with Eqn. (9.8), which we write again

$$(hX_n')' + \frac{\omega_n^2}{g} X_n = 0 \tag{9.32}$$

The variation of the above equation may be expressed in terms of the varia-

tions $\delta h = \zeta(x)$, δX_n, $\delta\omega_n$ and their derivatives:

$$(\delta h X_n')' + (h\delta X_n') + \frac{\omega_n^2}{g}\delta X_n + 2\delta\omega_n\frac{\omega_n}{g}X_n = 0 \tag{9.33}$$

Multiplying by X_m and integrating over $[0, L]$ we obtain

$$\int_0^L X_m \left(\delta h X_n'\right)' dx + \int_0^L X_m \left(h\delta X_n'\right)' dx + \int_0^L \frac{\omega_n^2}{g} X_m \delta X_n dx + 2\delta\omega_n \frac{\omega_n}{g}\int_0^L X_m X_n dx = 0 \tag{9.34}$$

The first term on the left-hand side of Eqn. (9.34) may be integrated by parts,

$$\begin{aligned}\int_0^L X_m \left(\delta h X_n'\right)' dx &= \int_0^L \left(X_m \delta h X_n'\right)' dx - \int_0^L \delta h X_m' X_n' dx \\ &= 0 - \int_0^L \delta h X_m' X_n' dx\end{aligned} \tag{9.35}$$

Where we have made use of the assumption that variations in X_n' are zero at the boundaries $x = 0, L$ due to the the boundary conditions $\bar{X}_n' = \bar{X}_m' = 0$ at $x = 0, L$. Integration by parts twice on the second term of the left-hand side of Eqn. (9.34), and use of the boundary conditions again gives

$$\int_0^L X_m \left(h\delta X_n'\right)' dx = \int_0^L \left(h X_m'\right)' \delta X_n dx \tag{9.36}$$

Substitution of Eqns. (9.35) and (9.36) into Eqn. (9.34), and rearrangement of terms, gives

$$-\int_0^L \delta h X_m' X_n' dx + \int_0^L \left[\left(h X_m'\right) + \frac{\omega_n^2}{g} X_m\right]\delta X_n dx + 2\delta\omega_n \frac{\omega_n}{g}\int_0^L X_m X_n dx = 0 \tag{9.37}$$

Taking $m = n$, and making use of the relation $(hX_n')' + \frac{\omega_n^2}{g} X_n = 0$, gives a relation between δh and $\delta\omega_n$:

$$\frac{\frac{g}{2\omega_n}\int_0^L \delta h X_n' X_n' dx}{\int_0^L X_n X_n dx} = \delta\omega_n \tag{9.38}$$

We can use the above equation to determine the relationship between small deviations from a flat-bottomed bathymetry and the deviations in frequency from those expressed in Eqn. (9.19) by setting $\delta h = \zeta(x)$, $\delta\omega_n = \Delta\omega_n = \omega_n - \bar{\omega}_n$, where ω_n is the observed frequency and $\bar{\omega}_n$ is the computed frequency for a flat-bottomed lake of depth h_o, and by replacing X_n with $\bar{X}_n$. This gives the relation,

$$\frac{\frac{g}{2\bar{\omega}_n}\int_0^L \zeta(x) \bar{X}_n' \bar{X}_n' dx}{\int_0^L \bar{X}_n \bar{X}_n dx} = \Delta\omega_n \tag{9.39}$$

which holds for each of the M observations of $\Delta\omega_n$. Note that the orthonormality of the X_n's, $\int_0^L \bar{X}_n \bar{X}_n dx = 1$ m^2 (dimensional form) can allow the above expression to be simplified. We retain the integral in the denominator on the left-hand side of the above expression as a reminder of the dimensions that are associated with the $\bar{X}_n$'s.

We have achieved a powerful result in deriving Eqn. (9.39). The unknown perturbation $\zeta(x)$ is linearly related to the observations $\Delta\omega_n$, $n = 1, \ldots, M$. We will exploit this linear relationship to derive the Backus-Gilbert inverse. Before doing so, however, it is important to restate Eqn. (9.39) in such a way as to demonstrate explicitly the underdetermined nature of the inverse problem.

9.6 A Fourier-Series Approach

Before developing a means to invert Eqn. (9.39) for the unknown ζ, we can gain considerable insight by recognizing that the integral operator in Eqn. (9.39) is reminicent of the projection operator which determines the coefficients of the Fourier-series representation of $\zeta(x)$. We begin with an analysis of the kernel $\bar{X}_n'\bar{X}_n'$.

The equation for the $\bar{X}_n$'s (Eqn. 9.17) may be written

$$X'' + \rho^2 X = 0 \tag{9.40}$$

where $\rho^2 = \frac{\omega^2}{gh_o}$. The solutions, as stated before, are

$$\bar{X}_n(x) = \sqrt{\frac{2}{L}} \sin(\frac{n\pi x}{L}) \tag{9.41}$$

Our interest is in the product $\bar{X}_n'\bar{X}_n'$ which appears in Eqn. (9.39). Defining,

$$U = X'X' \tag{9.42}$$

we can show that

$$\begin{aligned} U' &= -2\rho^2 X'X \\ U'' &= -2\rho^2 U + 2\rho^4 X^2 \end{aligned}$$

Taking a third derivative of U, and making use of the expression for U' given above, we arrive at a differential equation for U:

$$\left(U'' + 4\rho^2 U\right)' = 0 \tag{9.43}$$

The general solution of the above equation is of the form

$$U = A\left[1 + B\cos 2\rho x + C\sin 2\rho x\right] \tag{9.44}$$

Boundary conditions, $X' = 0$ at $x = 0, L$, imply

$$\begin{aligned} U(0) &= 0 \\ U(L) &= 0 \\ U'(0) &= 0 \end{aligned}$$

These boundary conditions allow us to determine $B = -1$, $C = 0$, and $\rho = \frac{n\pi}{L}$. We thus find a set of eigenfunctions U_n:

$$U_n(x) = A\left[1 - \cos\frac{2n\pi x}{L}\right] \tag{9.45}$$

The constant $A = \frac{n^2\pi^2}{L^3}$ is chosen to equate U with $\bar{X}_n'\bar{X}_n'$. (Note that the dimensional units of $\bar{X}_n$ are assumed to be $\mathrm{m}^{-1/2}$ for the constant A to make dimensional sense. This awkward aspect of the analysis could have been avoided by converting to dimensionless variables at the outset.)

The linear relation between ζ and $\Delta\omega_n$ in Eqn. (9.39) may now be restated in terms of the $\{U_n\}_{n=1}^M$:

$$\frac{\dfrac{n^2\pi^2 g}{2L^3\bar{\omega}_n}\displaystyle\int_0^L \zeta U_n dx}{\displaystyle\int_0^L \bar{X}_n\bar{X}_n dx} = \Delta\omega_n \tag{9.46}$$

We can identify the above integral operator with the operator necessary to determine the n'th coefficient of the expansion of ζ as a series of terms involving the U_n. What is important to realize at this stage is the fact that the set $\{U_n\}_{n=1}^M$ do not form a complete set of functions which span the interval $[0, L]$. This is due to the restriction imposed on the U_n by the boundary conditions. To be more explicit, substitution of Eqn. (9.45) into Eqn. (9.46) gives

$$\frac{\dfrac{n^2\pi^2 g}{2L^3\bar{\omega}_n}\displaystyle\int_0^L \zeta\left(1 - \cos\frac{2n\pi x}{L}\right)dx}{\displaystyle\int_0^L \bar{X}_n\bar{X}_n dx} = \Delta\omega_n \tag{9.47}$$

Making use of the Fourier-series expansion,

$$\zeta(x) = \zeta_o + \sum_{i=0}^{\infty}\alpha_i\cos\frac{n\pi x}{L} + \sum_{i=0}^{\infty}\beta_i\sin\frac{n\pi x}{L} \tag{9.48}$$

and making use of the orthonormality (in dimensional form) of the $\bar{X}_n$'s, the above equation becomes

$$\frac{gn^2\pi^2}{2\bar{\omega}_n L^2}\left(\zeta_o - \frac{\alpha_{2n}}{2}\right) = \Delta\omega_n \tag{9.49}$$

for $n = 1, \ldots, M$. (The awkward nature of not having used dimensionless variables is aleviated in the above expression. A quick check shows that the dimensions of the above expression are balanced across the equals sign.)

Equation (9.49) provides important guidance to the solution of the inverse problem. First, it shows us that the relation between the unknown ζ and the data $\Delta\omega_n$ is *linear*. Second, it shows us that the observations $\Delta\omega_n$, $n = 1, \ldots, M$, constrain only $M + 1$ unknown scalar parameters in the Fourier-series expansion of $\zeta(x)$. This tells us what we can expect from the data: only the first M even-numbered coefficients α_{2n} of the cosine expansion of ζ and the constant ζ_o are related to the data. The structure of ζ not related to the data is undetermined, and cannot be determined no matter how accurately we are able to measure the frequencies of free oscillation. We can anticipate the result of a minimum-norm solution: the odd-numbered coefficients α_{2n-1} of the cosine expansion up to $n = M$, all the coefficients α_n for $n > 2M$, and all the coefficients β_n of the sine expansion will be zero.

9.7 A Minimum-Norm Solution

The inverse problem posed above reduces to the determination of $M + 1$ coefficients of the Fourier-series expansion of ζ using the M equations represented by Eqn. (9.49). This problem may be written in matrix notation as follows:

$$\mathbf{Am} = \mathbf{d} \tag{9.50}$$

where $\mathbf{m} \in \mathcal{R}^{M+1}$ is the unknown vector of expansion coefficients

$$\mathbf{m} = \begin{bmatrix} \zeta_o \\ \alpha_2 \\ \vdots \\ \alpha_{2M} \end{bmatrix} \tag{9.51}$$

the vector $\mathbf{d} \in \mathcal{R}^M$ contains the data

$$\mathbf{d} = \begin{bmatrix} \Delta\omega_1 \\ \Delta\omega_2 \\ \vdots \\ \Delta\omega_M \end{bmatrix} \tag{9.52}$$

and the $M \times (M+1)$ matrix $\mathbf{A} : \mathcal{R}^{M+1} \rightarrow \mathcal{R}^M$ is defined by

$$A_{ij} = \begin{cases} \dfrac{n^2\pi^2 g}{2L^2\bar{\omega}_i} & \text{if } j = 1 \\ \dfrac{-n^2\pi^2 g}{4L^2\bar{\omega}_i} & \text{if } j = i+1 \\ 0 & \text{otherwise} \end{cases} \tag{9.53}$$

The minimum-norm solution of Eqn. (9.50) is obtained using the methods developed in Chapter (2):

$$\mathbf{m} = \mathbf{A}' \left[\mathbf{A}\mathbf{A}'\right]^{-1} \mathbf{d} \tag{9.54}$$

The model-resolution matrix associated with the above minimum-norm solution is

$$\mathbf{R}_{mn} = \mathbf{A}' \left[\mathbf{A}\mathbf{A}'\right]^{-1} \mathbf{A} \tag{9.55}$$

9.8 Example: Minimum-Norm Solution with Lake Michigan Data

To demonstrate a minimum-norm solution to the above inverse problem, we estimate 11 coefficients of the Fourier-series representation of ζ using observations of the 10 $\Delta\omega_n$'s associated with the 10 lowest frequencies of Lake Michigan's free oscillations. The demonstration is entirely theoretical, *i.e.,* the observations will be generated using a known bathymetry function

$\zeta(x)$, then an estimated bathymetry will be derived from the observations. The known bathymetry function for this example is:

$$h(x) = h_o + \zeta(x) \tag{9.56}$$

where

$$\begin{aligned} \zeta(x) &= \zeta_o + \sum_{n=1}^{20} \alpha_n \cos\frac{n\pi x}{L} + \sum_{n=1}^{20} \beta_n \sin\frac{n\pi x}{L} \\ &= 12 + 5\cos\frac{\pi x}{L} - 3\cos\frac{2\pi x}{L} + 7\sin\frac{6\pi x}{L} \end{aligned} \tag{9.57}$$

Our demonstration will proceed as follows. First, we will linearize the inverse problem around a flat-bottom bathymetry with $h_o = 150$ m. This will allow us to generate the $\bar{\omega}_n$'s and $\bar{X}_n$'s. Second, we will use the above expression for ζ to generate "perturbed" frequencies of free oscillation, ω_n. These steps will give us the data, $\Delta\omega_n$, $n = 1, \ldots, 10$. Third, we will "invert" the data using the minimum-norm inverse to obtain an estimate of ζ denoted by $\hat{\zeta}$:

$$\hat{\zeta}(x) = \hat{\zeta}_o + \sum_{n=1}^{10} \alpha_{2n} \cos\frac{2n\pi x}{L} \tag{9.58}$$

Finally, we will compare ζ with our minimum-norm estimate $\hat{\zeta}$ and compute the model-resolution matrix $\mathbf{R}_{mn}$.

Finite-Difference Generation of $\Delta\omega_n$

To generate the data, we adopt the finite-difference approach described in § (9.3.1). The finite-difference version of Eqn. (9.8) with variable depth $h(x)$ is written:

$$\frac{g}{2\Delta x^2}\left[(h_{i+1} + h_i)\, X_{i+1} - (h_{i+1} + 2h_i + h_{i-1})\, X_i + (h_i + h_{i-1})\, X_{i-1}\right]$$

$$+\omega^2 X_i = 0 \tag{9.59}$$

for $i = 2, \ldots, l-1$, with boundary conditions,

$$\bar{X}_2 - \bar{X}_1 = 0 \tag{9.60}$$

$$\bar{X}_l - \bar{X}_{l-1} = 0 \tag{9.61}$$

As before, the above finite-difference equations may be expressed by

$$\left(\mathbf{A} + \omega^2 \mathbf{I}\right) \mathbf{X} = \mathbf{0} \tag{9.62}$$

where the $l \times l$ matrix $\mathbf{A}$ is given by:

$$A_{ij} = \begin{cases} \frac{-g}{2\Delta x^2}\left(h_{i+1} + 2h_i + h_{i-1}\right) & \text{if } i = j \\ \frac{g}{2\Delta x^2}\left(h_{i+1} + h_i\right) & \text{if } j = i+1 \\ \frac{g}{2\Delta x^2}\left(h_i + h_{i-1}\right) & \text{if } j = i-1 \\ 1 & \text{if } (i,j) = (1,1),\ (l,l) \\ -1 & \text{if } (i,j) = (1,2),\ (l,l-1) \end{cases} \tag{9.63}$$

The following MATLAB® script was used to generate the eigenvalues of $\mathbf{A}$ defined above for the bathymetry function given in Eqn. (9.57). For consistency, the finite-difference algorithm instead of the exact analytic expression was used to calculate the $\bar{\omega}_n$. With this practice, the inherent inaccuracy of the finite-difference determination of the ω_n does not adversely affect the determination of $\hat{\zeta}$.

```
g=9.81;
h_naut=150;
L=650e3;
Ngrid=50;
%
h=h_naut*ones(Ngrid,1);
dx=L/(Ngrid-1);
A=zeros(Ngrid,Ngrid);
for i=2:Ngrid-1
A(i,i)=g*(h(i+1)+2*h(i)+h(i-1))/(2*dx^ 2);
A(i,i-1)=-g*(h(i-1)+h(i))/(2*dx^ 2);
A(i,i+1)=-g*(h(i+1)+h(i))/(2*dx^ 2);
end
A(1,2)=-1/dx;
```

```
A(Ngrid,Ngrid-1)=-1/dx;
A(1,1)=1/dx;
A(Ngrid,Ngrid)=1/dx;
omega_bar=sqrt(eig(A));
%Discard zero-frequency mode (steady solution):
omega_bar=sort(omega_bar);
omega_bar=omega_bar(2:Ngrid);
%
% The above determines the unperturbed frequencies in a computationally
% consistent fashion.
%
Ngrid=50;
x=linspace(0,L,Ngrid)';
h=(h_naut+12)*ones(Ngrid,1)+5*cos(pi/L*x)-3*cos(2*pi/L*x)+7*sin(6*pi/L*x);
dx=L/(Ngrid-1);
A=zeros(Ngrid,Ngrid);
for i=2:Ngrid-1
A(i,i)=g*(h(i+1)+2*h(i)+h(i-1))/(2*dx^ 2);
A(i,i-1)=-g*(h(i-1)+h(i))/(2*dx^ 2);
A(i,i+1)=-g*(h(i+1)+h(i))/(2*dx^ 2);
end
A(1,2)=-1/dx;
A(Ngrid,Ngrid-1)=-1/dx;
A(1,1)=1/dx;
A(Ngrid,Ngrid)=1/dx;
omega_pert=sqrt(eig(A));
%Discard zero-frequency mode (steady solution):
omega_pert=sort(omega_pert);
omega_pert=omega_pert(2:Ngrid);
plot(omega_pert); hold on; plot(omega_bar)
title('Perterbed vs.  Unperturbed Bathymetry')
xlabel('Mode Number')
ylabel('1/sec')
Delta_omega(1:10)=omega_pert(1:10)-omega_bar(1:10);
pause
hold off
clg
```

```
plot(Delta_omega)
title('Observed Frequency Difference')
xlabel('Mode Number')
ylabel('1/sec')
```

The data $\{\Delta\omega_n\}_{n=1}^{10}$ are displayed in Fig. (9.57)

The Minimum-Norm Inverse

Using the $\{\Delta\omega_n\}_{n=1}^{10}$ generated above, we now turn out attention to solving Eqn. (9.50) with $M = 10$ and $N = M + 1 = 11$. The following MATLAB® routine was used to construct $\mathbf{A} : \mathcal{R}^{M+1} \to \mathcal{R}^M$ and find the minimum-norm solution $\mathbf{m} \in \mathcal{R}^{M+1}$ with data $\mathbf{d} \in \mathcal{R}^M$.

```
% This routine finds the minimum-norm inverse of Am=d
d=Delta_omega';
% A is a 10 row by 11 column matrix:
g=9.81;
L=650e3;
A=zeros(10,11);
for i=1:10
A(i,1)=g*i^ 2*pi^ 2/(2*L^ 2*omega_bar(i));
A(i,i+1)=-g*i^ 2*pi^ 2/(4*L^ 2*omega_bar(i));
end
%
m=A'*inv(A*A')*d
```

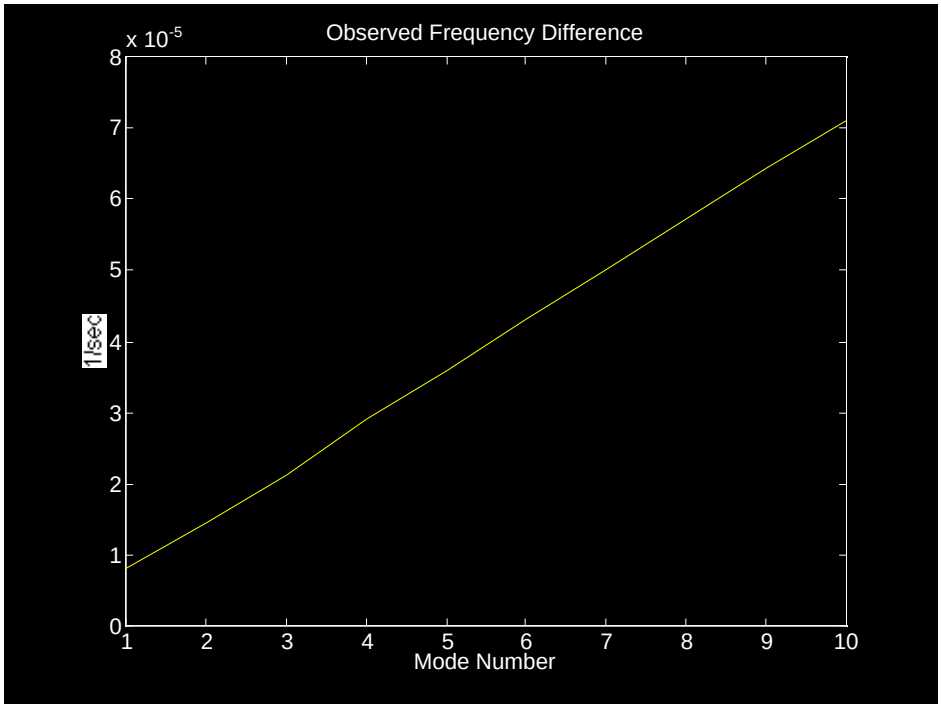

Figure 9.3: The $\Delta\omega_n$, $n = 1, \ldots, 10$, generated by a finite difference method using the bathymetry function given in Eqn. (9.57).

The solution obtained by the above, minimum-norm algorithm is

$$\hat{\mathbf{m}} = \begin{bmatrix} 11.5828 \\ -3.5864 \\ -0.7416 \\ -0.0386 \\ -0.6121 \\ -0.4402 \\ -0.3304 \\ -0.2155 \\ -0.0885 \\ 0.0528 \\ 0.2090 \end{bmatrix} \tag{9.64}$$

This minimum-norm solution can be compared with what $\mathbf{m}$ should have been for the ζ expressed in Eqn. (9.57):

$$\mathbf{m} = \begin{bmatrix} 12 \\ -3 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{9.65}$$

A comparison between $\zeta(x)$ given by Eqn. (9.57) and the $\hat{\zeta}(x)$ constructed using the Fourier-coefficients derived from $\hat{\mathbf{m}}$ above is shown in Fig. (9.4). It is clear from the figure that the retrodicted $\hat{\zeta}(x)$ differs greatly from the known $\zeta(x)$ used to generate the data. This difference emphasizes the consequence of the fact that the data constrain only the 10 *even* terms in the cosine expansion of $\zeta(x)$.

The model-resolution matrix $\mathbf{R}_{mn}$ associated with the minimum-norm solution is displayed in Fig. (9.5). The example worked here suggests that the minimum-norm inverse does a good job in resolving the $\mathbf{m}$.

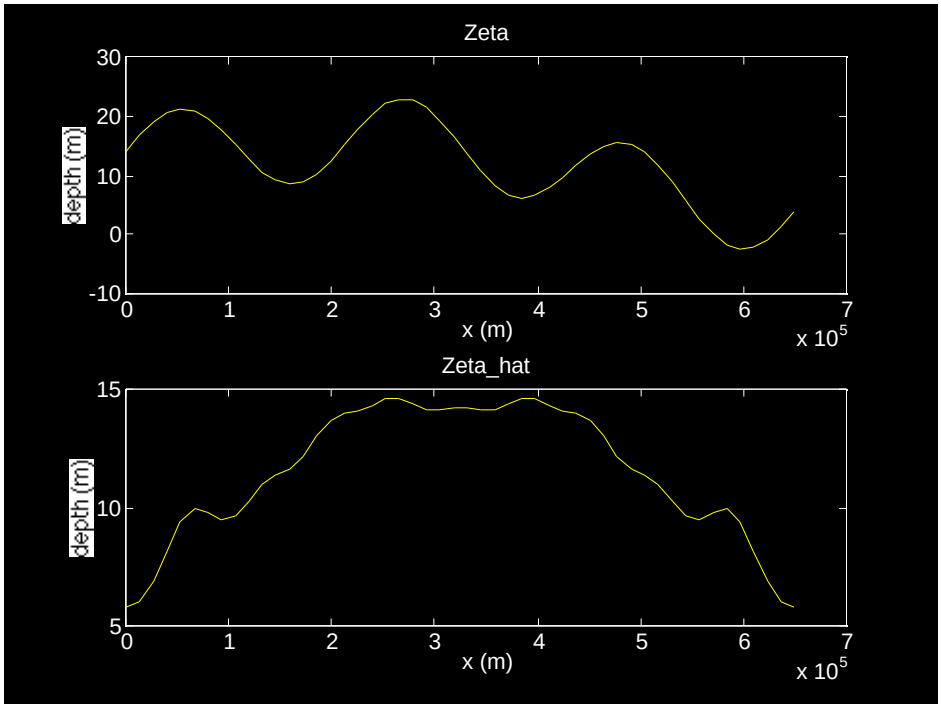

Figure 9.4: A comparison between the $\zeta(x)$ used to generate the data $\Delta\omega_n$, $n = 1, \ldots, 10$, and the $\hat{\zeta}(x)$ resulting from the minimum-norm inverse.

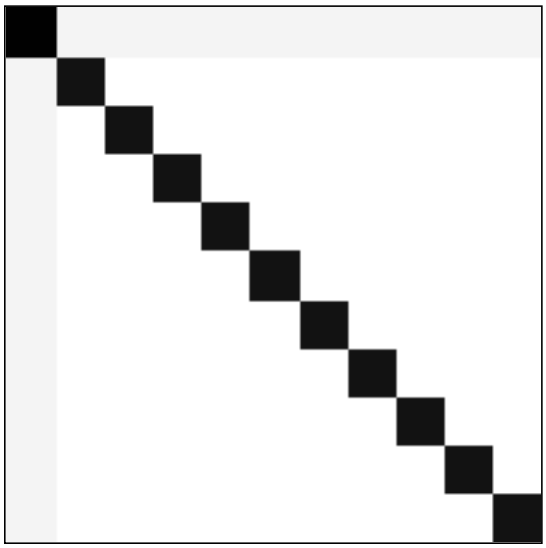

Figure 9.5: The model-resolution matrix $\mathbf{R}_{mn}$ associated with the minimum-norm inverse. Each pixel of the above image represents an element R_{ij} of $\mathbf{R}_{mn}$. The darker the color, the higher the value of the corresponding element. The dark swath down the diagonal of the matrix indicates that the minimum-norm inverse does a fairly good job of resolving ζ_o and the even-numbered coefficients of the cosine expansion of $\zeta(x)$.

9.9 Dirichlet Spread of the Model-Resolution Matrix

The Backus-Gilbert inverse differs from the minimum-norm inverse by virtue of the fact that it is designed to optimize a particular quality of the model-resolution matrix known as the *Backus-Gilbert spread*. Before explaining what the Backus-Gilbert spread is, we define the *Dirichlet spread* and show that the minimum-norm solution optimizes (minimizes) this quantitative measure of the model-resolution matrix.

Definition. The Dirichlet spread, S_d of a $N \times N$ model-resolution matrix $\mathbf{R}$ is defined by the following measure of how $\mathbf{R}$ differs from the identity matrix $\mathbf{I}$:

$$S_d = \sum_{i,j=1}^{N} (R_{ij} - \delta_{ij})^2 \tag{9.66}$$

■

Minimization of S_d

It is easy to show that the minimum-norm inverse $\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1} = \mathbf{A}^{-mn}$ minimizes S_d.

Proof: Let $\mathbf{A}^\star$ denote the inverse of $\mathbf{A}$ which minimizes the Dirichlet spread, S_d of the model-resolution matrix. Our goal is to prove that $\mathbf{A}^\star = \mathbf{A}^{-mn}$. The model-resolution matrix associated with $\mathbf{A}^\star$ is $\mathbf{A}^\star\mathbf{A}$. Thus,

$$\begin{aligned} S_d &= \|\mathbf{A}^\star\mathbf{A} - \mathbf{I}\|^2 \\ &= \sum_{i,j=1}^{N} (R^\star_{ij} - \delta_{ij})^2 \\ &= \sum_{i,j=1}^{N} \Big\{ (\sum_{\alpha=1}^{M} a_{j\alpha}A_{\alpha i}) - \delta_{ij} \Big\}^2 \\ &= \sum_{i=1}^{N} \Big[\sum_{j=1}^{N} \Big\{ (\sum_{\alpha=1}^{M} a_{j\alpha}A_{\alpha i}) - \delta_{ij} \Big\}^2 \Big] \end{aligned} \tag{9.67}$$

where a_{lk} is the l,k'th element of $\mathbf{A}^\star$.

Differentiating with respect to the unknown element $\alpha_{j\beta}$, and noting the fact that $a_{j\beta}$ enters just once in the above expression, we obtain

$$\frac{\partial S_d}{\partial a_{j\beta}} = 2\sum_{i=1}^{N}\left[\left\{(\sum_{\alpha=1}^{M} a_{j\alpha}A_{\alpha i}) - \delta_{ij}\right\}A_{\beta i}\right] \tag{9.68}$$

If the above expression is set to zero, we obtain a condition that must be satisfied by $\mathbf{A}^\star$ to minimize S_d:

$$\sum_{\alpha=1}^{M}[\sum_{i=1}^{N} A_{\beta i}A_{\alpha i}]\, a_{j\alpha} = A_{\beta j} \tag{9.69}$$

which, in matrix notation is,

$$(\mathbf{A}\mathbf{A}')(\mathbf{A}^\star)' = \mathbf{A} \tag{9.70}$$

i.e.,

$$(\mathbf{A}^\star)' = (\mathbf{A}\mathbf{A}')^{-\mathbf{1}}\mathbf{A} \tag{9.71}$$

Taking the transpose gives the result we desire:

$$\begin{aligned}\mathbf{A}^\star &= \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-\mathbf{1}} \\ &= \mathbf{A}^{-mn}\end{aligned} \tag{9.72}$$

In other words, the minimum-norm inverse minimizes the Dirichlet spread of the model-resolution matrix. ■

9.10 The Backus-Gilbert Spread of the Model-Resolution Matrix

Backus and Gilbert proposed a slightly different measure of spread and designed an inverse method for the purpose of optimizing model resolution. The Backus-Gilbert spread S_{bg} of the model-resolution matrix $\mathbf{R}$ is defined as follows:

Definition. The Backus-Gilbert spread, S_{bg} of a $N \times N$ model-resolution matrix $\mathbf{R}$ is defined by the following measure of how $\mathbf{R}$ differs from the identity matrix $\mathbf{I}$:

$$\begin{aligned} S_d &= \sum_{i,j=1}^{N} (i-j)^2 \left(R_{ij} - \delta_{ij}\right)^2 \\ &= \sum_{i,j=1}^{N} (i-j)^2 \left(R_{ij}^2 - 2R_{ij}\delta_{ij} + \delta ij^2\right) \\ &= \sum_{i,j=1}^{N} (i-j)^2 R_{ij}^2 \end{aligned} \tag{9.73}$$

■

The Backus-Gilbert spread differs from the Dirichlet spread S_d because of the appearance of the term $(i-j)^2$. The introduction of this term suggests that S_{bg} measures not only the deviation of the model-resolution matrix from the identity matrix, but also the tendency of the model-resolution matrix to be diagonally dominant (have it's largest values arrayed along the diagonal).

9.11 Derivation of the Backus-Gilbert Inverse

The Backus-Gilbert inverse for the general problem

$$\mathbf{Am} = \mathbf{d} \tag{9.74}$$

is defined as the $N \times M$ matrix $\mathbf{G} : \mathcal{R}^M \rightarrow \mathcal{R}^N$, which gives

$$\tilde{\mathbf{m}} = \mathbf{Gd} \tag{9.75}$$

The model-resolution matrix $\mathbf{R}_{bg}$ associated with the Backus-Gilbert inverse is defined as

$$\mathbf{R}_{bg} = \mathbf{GA} \tag{9.76}$$

The Backus-Gilbert inverse is chosen to minimize the S_{bg} of $\mathbf{R}_{bg}$ subject

to the auxilliary constraint that rows of $\mathbf{R}_{bg}$ add up to one, *viz.*

$$\sum_{j=1}^{N} R_{ij} - 1 = 0 \tag{9.77}$$

for $i = 1, \ldots, N$. This extra constraint is added for two reasons. First, the S_{bg} does not reference the diagonal elements of $\mathbf{R}_{bg}$ at all, so they would be undetermined without the constraint. Second, it is desireable for the rows of $\mathbf{R}_{bg}$ to represent weighted-averaging operators which act on $\mathbf{m}$ to obtain each component of $\tilde{\mathbf{m}}$. Using a Lagrange multiplier vector $\underline{\lambda}$ to enforce the constraint, the condition defining $\mathbf{G}$ is the minimization of the following scalar quantity:

$$\begin{aligned} H &= \sum_{i,j=1}^{N} (i-j)^2 R_{ij}^2 + \sum_{i=1}^{N} 2\lambda_i \left(\sum_{j=1}^{N} R_{ij} - 1 \right) \\ &= (i-j)^2 \left(G_{ik} A_{kj} \right)^2 + 2\lambda_i \left((G_{il} A_{lj}) - 1 \right) \end{aligned} \tag{9.78}$$

where the summation convention of repeated indices has been adopted to avoid excessive notational complexity.

The derivatives of H with respect to the unknown $G_{\alpha\beta}$ and λ_γ are

$$\frac{\partial H}{\partial G_{\alpha\beta}} = 2(\alpha - i)^2 G_{\alpha k} A_{kj} A_{\beta j} + 2\lambda_\alpha A_{\beta j} \tag{9.79}$$

$$\frac{\partial H}{\partial \lambda_\gamma} = 2 \left(G_{\gamma l} A_{lj} - 1 \right) \tag{9.80}$$

Setting the derivatives to zero yields the Euler-Lagrange equations which may be solved to obtain $\mathbf{G}$.

The solution of the Euler-Lagrange equations is tricky in the present situation, because it is difficult to write them in a format which is amenable to linear algebra. We can overcome this difficulty by solving for the rows of $\mathbf{G}$ one at a time. Accordingly, we define a set of N vectors $\mathbf{g}_\alpha \in \mathcal{R}^M$ which

represent the N rows of the matrix $\mathbf{G}$, *i.e.,*

$$\mathbf{G} = \begin{bmatrix} \mathbf{g'_1} \\ \mathbf{g'_2} \\ \vdots \\ \mathbf{g'_N} \end{bmatrix} \tag{9.81}$$

or,

$$\mathbf{g}_\alpha = \begin{bmatrix} G_{\alpha 1} \\ G_{\alpha 2} \\ \vdots \\ G_{\alpha N} \end{bmatrix} \tag{9.82}$$

With this convenient definition, the Euler-Lagrange equations may be broken down into a set of N equations of the form

$$\mathbf{K}_\alpha \mathbf{g}_\alpha + \lambda_\alpha \mathbf{u} = \mathbf{0} \tag{9.83}$$

$$\mathbf{g'}_\alpha \mathbf{u} - 1 = 0 \tag{9.84}$$

where, the matrix $\mathbf{K}_\alpha : \mathcal{R}^M \rightarrow \mathcal{R}^M$ is defined by the relation $[\mathbf{K}_\alpha]_{k\beta} = (\alpha - j)^2 A_{kj} A'_{j\beta}$, and the vector $\mathbf{u} \in \mathcal{R}^M$ is defined by by the relation $u_\beta = A_{\beta j}$.

Dropping momentarily the subscript α for notational convenience, we solve Eqns. (9.83) and (9.84) for λ and $\mathbf{g}$. First, we multiply Eqn. (9.83) by $\mathbf{K}^{-1}$ (we assume $\mathbf{K}^{-1}$ exists because $\mathbf{K}$ is an $M \times M$ square matrix) and solve for an expression which gives $\mathbf{g}$ in terms of λ:

$$\mathbf{g} = -\lambda \mathbf{K}^{-1} \mathbf{u} \tag{9.85}$$

Next, we substitute the above expression into Eqn. (9.84) and solve for λ:

$$\lambda = \left(\frac{-1}{\mathbf{u'K^{-1}u}} \right) \tag{9.86}$$

Observe that $\left(\frac{-1}{\mathbf{u'K^{-1}u}} \right)$ is a scalar. Finally, we substitute the above expression into Eqn. (9.83) to yield an expression for $\mathbf{g}$:

$$\mathbf{g} = \left(\frac{-1}{\mathbf{u'K^{-1}u}} \right) \mathbf{K}^{-1} \mathbf{u} \tag{9.87}$$

The above solution for each of the $\{\mathbf{g}_\alpha\}_{\alpha=1}^{N}$ gives the Backus-Gilbert inverse $\mathbf{G}$.

The model-resolution matrix associated with the Backus-Gilbert inverse is readily shown to be

$$\mathbf{R}_{bg} = \begin{bmatrix} \mathbf{g'_1} \\ \mathbf{g'_2} \\ \vdots \\ \mathbf{g'_N} \end{bmatrix} \mathbf{A} \tag{9.88}$$

Making use of Eqn. (9.87), we obtain

$$\mathbf{R}_{bg} = \begin{bmatrix} \left(\dfrac{-1}{\mathbf{u'K_1}^{-1}\mathbf{u}}\right) \mathbf{K_1}^{-1}\mathbf{u'} \\ \left(\dfrac{-1}{\mathbf{u'K_2}^{-1}\mathbf{u'}}\right) \mathbf{K_2}^{-1}\mathbf{u'} \\ \vdots \\ \left(\dfrac{-1}{\mathbf{u'K_N}^{-1}\mathbf{u}}\right) \mathbf{K_N}^{-1}\mathbf{u'} \end{bmatrix} \mathbf{A} \tag{9.89}$$

9.12 Example: Backus-Gilbert Solution with Lake Michigan Data

We demonstrate the Backus-Gilbert inverse by applying it to the same problem discussed in S (9.8). Using the same $\{\Delta\omega_n\}_{n=1}^{10}$ data generated previously, we use the following MATLAB® routine to generate $\tilde{\mathbf{m}} = \mathbf{Gd}$:

```
% This routine computes the Backus-Gilbert inverse.
% m_tilde = G * d
d=Delta_omega';
% A is a 10 row by 11 column matrix:
g=9.81;
L=650e3;
A=zeros(10,11);
for i=1:10
```

```
A(i,1)=g*i^ 2*pi^ 2/(2*L^ 2*omega_bar(i));
A(i,i+1)=-g*i^ 2*pi^ 2/(4*L^ 2*omega_bar(i));
end
%
% Construct G, the Backus-Gilbert inverse:
%
G=zeros(11,10);
g_alpha_transpos=zeros(10,11);
u=zeros(10,1);
for beta=1:10
for j=1:11
u(beta)=u(beta)+A(beta,j);
end
end
%
for alpha=1:11
K=zeros(10,10);
%
for m=1:10
for beta=1:10
for j=1:11
K(m,beta)=K(m,beta)+ (alpha-j)^ 2*A(m,j)*A(beta,j);
end
end
end
%
g_alpha_transpos(:,alpha)=(1/(u'*inv(K)*u))*inv(K)*u;
%
end
%
G=g_alpha_transpos';
m_tilda=G*d
R_bg=G*A
```

The solution $\tilde{\mathbf{m}}$ obtained from the above Backus-Gilbert algorithm is:

$$\tilde{\mathbf{m}} = \begin{bmatrix} 25.6460 \\ 26.7519 \\ 23.9071 \\ 23.2042 \\ 23.7776 \\ 23.6057 \\ 23.4959 \\ 23.3811 \\ 23.2540 \\ 23.1127 \\ 22.9565 \end{bmatrix} \tag{9.90}$$

The Backus-Gilbert solution is unsatisfactory because it bears virtually no resemblance to what is expected. A comparison of the retrodicted $\tilde{\zeta}(x)$ using the Backus-Gilbert inverse, the minimum-norm result $\hat{\zeta}(x)$, and the true $\zeta(x)$ is provided in Fig. (9.6). For comparison with the minimum-norm model-resolution matrix $\mathbf{R}_{mn}$ shown in Fig. (9.5), the $\mathbf{R}_{bg}$ is shown in Fig. (9.7).

Another defect of the Backus-Gilbert solution is that it fails to satisfy the data. In particular, the expression $\mathbf{A}\tilde{\mathbf{m}} = \tilde{\mathbf{d}}$ yields the following version of the $\Delta\omega_n$'s:

$$\tilde{\mathbf{d}} = 1.0e-04 \begin{bmatrix} 0.0746 \\ 0.1667 \\ 0.2566 \\ 0.3356 \\ 0.4228 \\ 0.5104 \\ 0.5993 \\ 0.6898 \\ 0.7823 \\ 0.8769 \end{bmatrix} \tag{9.91}$$

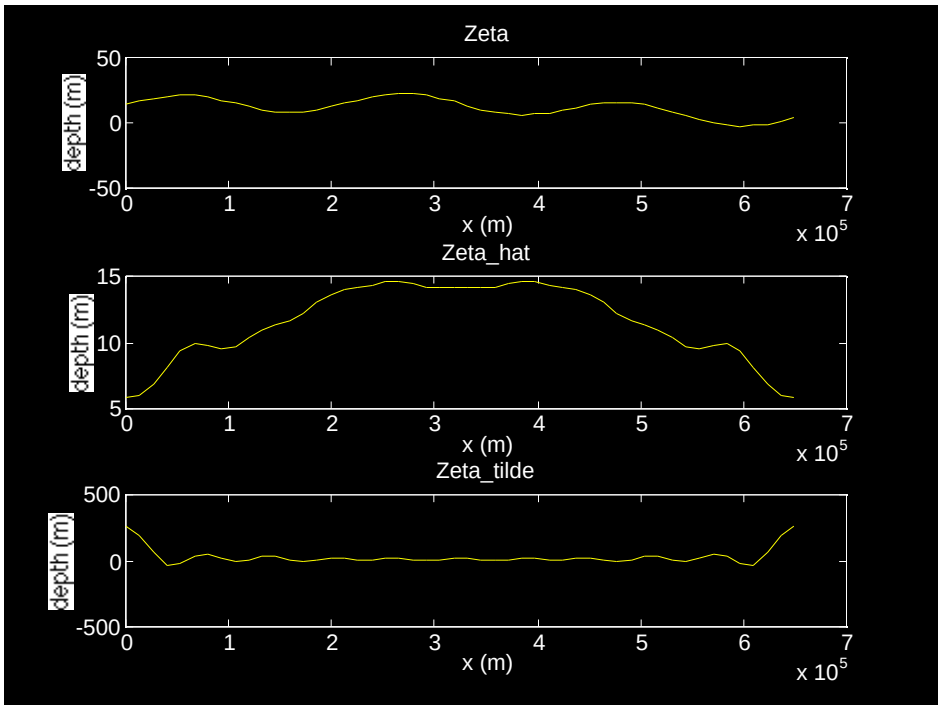

Figure 9.6: The true (top), minimum-norm (middle) and Backus-Gilbert (bottom) derived bathymetry functions $\zeta(x)$. While neither the minimum-norm nor the Backus-Gilbert methods yield an accurate result, the Backus-Gilbert bathymetry is grossly inaccurate, and fails spectacularly in comparison with the minimum-norm result.

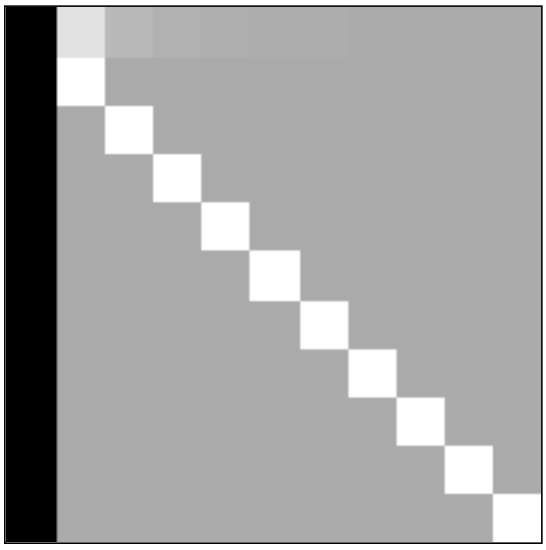

Figure 9.7: The model-resolution matrix $\mathbf{R}_{bg}$ associated with the Backus-Gilbert inverse. Each pixel of the above image represents an element R_{ij} of $\mathbf{R}_{bg}$. The darker the color, the higher the value of the corresponding element. The structure of this matrix suggests that the Backus-Gilbert method does not do a particularly good job of resolving ζ_o and the even-numbered coefficients of the cosine expansion of $\zeta(x)$.

The data vector $\mathbf{d}$, for reference, is:

$$\mathbf{d} = 1.0e-04 \begin{bmatrix} 0.0814 \\ 0.1455 \\ 0.2120 \\ 0.2900 \\ 0.3605 \\ 0.4314 \\ 0.5020 \\ 0.5721 \\ 0.6416 \\ 0.7104 \end{bmatrix} \tag{9.92}$$

It is important to remember that the minimum-norm solution $\hat{\mathbf{m}}$ satisfies the data exactly ($\mathbf{A}\hat{\mathbf{m}} = \mathbf{d}$). The above comparison suggests that a great price has been paid to minimize the Backus-Gilbert spread of the model-resolution matrix.

9.13 An Alternative Definition of the Backus-Gilbert Spread

The poor performance of the Backus-Gilbert method in the above example demonstrates the problems that can arise when the issue of fitting data becomes secondary to the issue of model resolution. Backus and Gilbert [1968] suggested that the criteria used to determine $\mathbf{G}$ could be modified to suit each particular application. In this section we shall show that the subsidiary constraint given by Eqn. (9.77), namely

$$\sum_{j=1}^{N} R_{ij} - 1 = 0 \tag{9.93}$$

was at the root of the poor performance of the Backus-Gilbert method. We repeat the derivation of the Backus-Gilbert inverse and compute $\bar{\mathbf{m}}$ for the Lake-Michigan problem posed above using a different subsidiary constraint:

$$R_{ii} - 1 = 0 \tag{9.94}$$

for $i = 1, \ldots, N$. This alternative subsidiary condition requires the diagonal elements of $\mathbf{R}$ to be 1. No constraints are imposed, however, on the sums of the rows of $\mathbf{R}$.

Following the approach used previously, the Backus-Gilbert inverse $\mathbf{G}$ is found by minimizing the following performance index:

$$\begin{aligned} H &= \sum_{i,j=1}^{N} (i-j)^2 R_{ij}^2 + \sum_{i=1}^{N} 2\lambda_i \left(R_{ii} - 1\right) \\ &= (i-j)^2 \left(G_{ik} A_{kj}\right)^2 + 2\lambda_i \left(\left(G_{il} A_{li}\right) - 1\right) \end{aligned} \tag{9.95}$$

where the summation convention of repeated indices has been adopted to avoid excessive notational complexity. The Euler-Lagrange equations associated with the above definition of H yield the following equations for the $\alpha = 1, \ldots, M$ rows of $\mathbf{G}$

$$\mathbf{K}_\alpha \mathbf{g}_\alpha + \lambda_\alpha \mathbf{u}_\alpha = \mathbf{0} \tag{9.96}$$

$$\mathbf{g'}_\alpha \mathbf{u}_\alpha - 1 = 0 \tag{9.97}$$

where, the matrix $\mathbf{K}_\alpha : \mathcal{R}^M \rightarrow \mathcal{R}^M$ is defined as before by the relation $[\mathbf{K}_\alpha]_{k\beta} = (\alpha - j)^2 A_{kj} A'_{j\beta}$, and the vector $\mathbf{u}_\alpha \in \mathcal{R}^M$ is defined by by the relation $u_\beta = A_{\beta\alpha}$.

The following MATLAB® routine gives the solution to the Lake Michigan bathymetry problem using the solution of Eqns. (9.96) and (9.97) to define $\mathbf{G}$:

```
% This routine computes the Backus-Gilbert inverse.
% m_tilde = G * d
d=Delta_omega';
% A is a 10 row by 11 column matrix:
g=9.81;
L=650e3;
A=zeros(10,11);
for i=1:10
A(i,1)=g*i^ 2*pi^ 2/(2*L^ 2*omega_bar(i));
A(i,i+1)=-g*i^ 2*pi^ 2/(4*L^ 2*omega_bar(i));
end
```

```
G=zeros(11,10);
g_alpha_transpos=zeros(10,11);
for beta=1:10
end
for alpha=1:11
K=zeros(10,10);
u=zeros(10,1);
for beta=1:10
u(beta)=A(beta,alpha);
end
for m=1:10
for beta=1:10
for j=1:11
K(m,beta)=K(m,beta)+ (alpha-j)^ 2*A(m,j)*A(beta,j);
end
end
end
g_alpha_transpos(:,alpha)=(1/(u'*inv(K)*u))*inv(K)*u;
end
G=g_alpha_transpos';
m_bar=G*d
R_bg2=G*A
```

The solution obtained using the above algorithm, $\bar{\mathbf{m}}$, is:

$$\bar{\mathbf{m}} = \begin{bmatrix} 12.8230 \\ -6.3393 \\ 0.1876 \\ 0.5940 \\ -0.3314 \\ -0.0678 \\ -0.0383 \\ -0.0147 \\ 0.0146 \\ 0.0619 \\ 0.2396 \end{bmatrix} \tag{9.98}$$

and comes closer to satisfying the data (but not exactly):

$$\mathbf{A}\bar{\mathbf{m}} = 1.0e-04 \begin{bmatrix} 0.0973 \\ 0.1549 \\ 0.2289 \\ 0.3169 \\ 0.3927 \\ 0.4716 \\ 0.5509 \\ 0.6306 \\ 0.7102 \\ 0.7863 \end{bmatrix} \neq 1.0e-04 \begin{bmatrix} 0.0814 \\ 0.1455 \\ 0.2120 \\ 0.2900 \\ 0.3605 \\ 0.4314 \\ 0.5020 \\ 0.5721 \\ 0.6416 \\ 0.7104 \end{bmatrix} \tag{9.99}$$

The model-resolution matrix associated with this improved version of the Backus-Gilbert inverse is displayed in Fig. (9.8).

9.14 Conclusion

We have focussed on an interesting problem in this chapter, namely, how to extract bathymetric structure from observations of the frequency of free oscillation. What we have learned is that minimum-norm and Backus-Gilbert methods provide a relatively unsatisfactory result. At best, all we can hope to recover from our measurements of frequencies are the even numbered coefficients of the cosine-series expansion of the unknown bathymetry. To get the full bathymetric structure, we anticipate having to *augment* the observed frequencies with observations relating to the eigenmodes X_n's.

An important conclusion can be drawn as a result of the comparison between the Backus-Gilbert inverse and the minimum-norm inverse. The improvement of the Backus-Gilbert method's model-resolution matrix comes at a terrible price: the solution can be extremely inaccurate, and the data are no longer satisfied. It is for these reasons that the Backus-Gilbert method is not recommended for most underdetermined inverse problems.

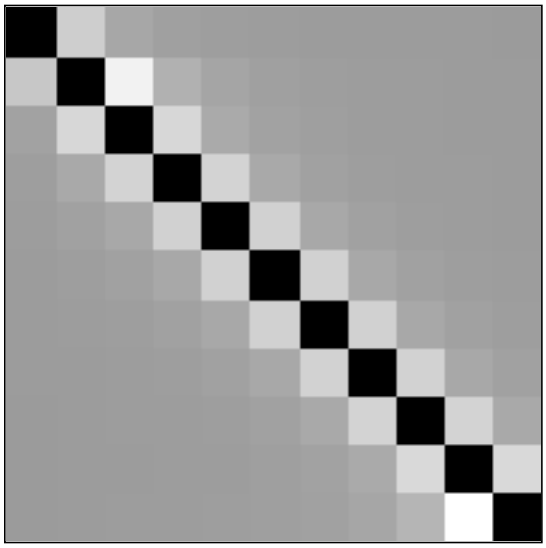

Figure 9.8: An improved model-resolution matrix $\mathbf{R}_{bg2}$ associated with the Backus-Gilbert inverse derived under a subsidiary constraint that $R_i i = 1$, for $i = 1, \ldots, N$. Each pixel of the above image represents an element R_{ij} of $\mathbf{R}_{bg2}$. Clearly, this version of the Backus-Gilbert inverse yields a more satisfactory model resolution matrix than that depicted in Fig. (9.7).

9.15 Bibliography

Backus, G. and F. Gilbert, 1968. The resolving power of gross earth data. *Geophysical Journal of the Royal Astronomical Society*, **16**, 169-205.

Backus, G. and F. Gilbert, 1968. Constructing P-velocity models to fit restricted sets of travel-time data. *Bulletin of the Seismological Society of America*, **59**, 1407-1414.

Hutter, K, 1993. Waves and oscillations in the ocean and in lakes. *Continuum Mechanics in Environmental Sciences and Geophysics*. (CISM Courses and Lectures No. 337, International Centre for Mechanical Sciences, K. Hutter, editor) 80-240.