UNDERSTANDING INVERSE THEORY ×10066

Robert L. Parker

Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093

INTRODUCTION

Much of our knowledge of the Earth's interior is perforce based on the interpretation of measurements made at the surface, rather than direct sampling of the material in the interior. In the past few years there have been great advances in the mathematical aspects of this problem, and the topic has come to be called geophysical inverse theory. To apply these ideas, there must be a valid mathematical model of the physics of the system under study, so that one would be able to calculate the values of observations made on an exactly known structure: the calculation of the behavior of a specified system is the solution of the "forward" or "direct" problem. Frequently it is the forward problem that presents a difficult challenge to the theoretical geophysicist. Illustrations include the mechanism of earthquake rupture or the generation of the Earth's magnetic field; in problems like these, inverse theory is normally quite inappropriate. When the forward problem has been completely solved, there are of course unknown parameters in the mathematical model representing physical properties of the Earth such as Lamé parameters, density or electrical conductivity. The goal of inverse theory is to determine the parameters from the observations or, in the face of the inevitable limitations of actual measurement, to find out as much as possible about them. The quality that distinguishes inverse theory from the parameter estimation problem of statistics (Bard 1974, Rao 1973) is that the unknowns are *functions,* not merely a handful of real numbers. This means that the solution contains in principle an infinite number of variables, and therefore with real data the problem is as underdetermined as it can be. Naturally, there are geophysical problems containing a relatively small number of free parameters: for example, in describing the relative instantaneous motion of N lithospheric plates, we find that the assumption of internal rigidity reduces the number of unknowns to $3N-3$ for the $N-1$ relative angular velocity vectors (McKenzie & Parker 1974). Sometimes, however, unknown structures are conceived in terms of small numbers of homogeneous layers for reasons of computational simplicity rather than on any convincing geophysical or

geological grounds. Such simplification may lead to false confidence in the solution because the true amount of freedom has not been allowed in the parameters.

The theory falls into two distinct parts. One deals with the ideal case in which the data are supposed to be known exactly and as densely as desired, and is of course mainly the province of the applied mathematician. The other treats the practical problems that are created by incomplete and imprecise data. It might be thought that an exact solution to an inverse problem with perfect data would prove extremely useful for the practical case. Actually, this often turns out to be untrue because geophysical inverse problems are almost always unstable in a sense to be defined more precisely later; when this is so, the solution obtained by the analytic technique is very sensitive to the way in which the data set is completed and to the errors in it. In my view analytical studies are more valuable for their results concerning uniqueness and conditions for existence and stability.

The advances mentioned earlier arise primarily from a recognition that practical inverse problems never possess unique solutions and that an honest attempt to interpret the data must appraise the variety of compatible solutions. In the class of linear problems there is now a very satisfactory body of theory for achieving this end, but work on the larger and more prevalent class of nonlinear problems is still in a relatively primitive state and perhaps always will be.

The plan of this article is as follows. In the next section I outline and discuss some mathematical questions in the analysis of idealized data. Although many of the points raised by the geophysical problems are fascinating to the mathematician, I have tried to keep in mind that the aim of our science is to learn about the Earth and therefore I attempt to restrict discussion to matters of practical consequence. After this, I deal directly with the problem of using actual measurements. First, there is a fairly extended account of the linear theory in which I permit myself occasionally to give some of the mathematical details. Here the principles of the celebrated Backus-Gilbert method are described and a particular formulation is given of the spectral expansion method which has made possible the solution of so many practical problems. Finally, there is a section on the analysis of nonlinear systems, consisting of a discussion of the linearization approximation and a brief sketch of some of the commendable efforts to do without approximation.

ANALYSIS OF PERFECT DATA

The task of retrieving model parameters from a complete and precise set of data is clearly a mathematical undertaking. There is no single simple method of attacking the various aspects of inverse theory, but instead methods are drawn from almost every branch of applied mathematics. Therefore, I make no attempt to give any derivations in this section, but rather to highlight the fundamental concepts and the way that they impinge upon the more empirical side of geophysics. I consider the questions of existence, uniqueness, construction, and stability, which are some of the concerns that must be dealt with in a complete solution of an inverse problem. Sabatier (1971) has given a more extended account of these topics in a wide-ranging comparative review.

Mathematicians take the view, logically enough, that before attempting to calculate parameters, one ought to define the class of possible data that are associated with the model.

This is the question of *existence* and, although it receives little attention in the geophysical literature, it is of great importance in testing the assumptions behind any mathematical model. Every model contains simplifications and approximations, some of which may be hard to justify initially; for example, in many inverse problems the idealization of horizontal (or radial) stratification is introduced at the outset. To give a concrete instance, consider the problem of electrical conductivity sounding with the magneto-telluric method (Cagniard 1953). Here recordings are made in one place of orthogonal components of the horizontal electric and magnetic field (E_x and B_y say, with z vertical). The ratio of the Fourier transforms of the two signals is related to the electrical conductivity profile of a horizontally layered medium beneath the observing station. If the medium is truly layered, the orientation of the x and y axes will not affect the results, i.e. the magneto-telluric ratio is isotropic. Mathematically this is a statement of a *necessary condition* for the existence of a solution to the inverse problem of finding $\sigma(z)$ from the ratio. When the data are not isotropic we know that there is no solution to the problem as posed and that one of our model assumptions is false. Under these circumstances it would be foolish to continue with an interpretation based on the original model. Suppose now that an isotropic ratio was indeed observed; does this in itself guarantee a solution of the kind proposed? This is equivalent to asking whether isotropy is a *sufficient condition* for existence. In fact, isotropy is by no means enough to insure a corresponding layered solution: a particularly thorough exploration of the various restraints imposed upon the data by the layering assumption has been given by Weidelt (1972). All of these are further necessary conditions; a sufficient condition for their problem has never been derived to my knowledge. Another geophysical inverse problem which has been considered from this viewpoint is that of body-wave travel times; Gerver & Markusevitch (1966) give a detailed account.

A different mathematical matter of great geophysical importance is the question of *uniqueness*: if it is granted that there is a solution for a given set of data, is there only one such solution? Profound consequences follow if the answer is no, for it is then established that even perfect data (complete and exact) do not contain enough information to recover the Earth's structure. Several courses are then open: one can explore what further assumptions can be plausibly made about the Earth to narrow the class of solutions (perhaps down to uniqueness); alternatively, additional kinds of measurements might be introduced; or finally, one may be willing to tolerate the ambiguity if the class of admissible solutions still contains decisive information about the Earth. The last course is indeed the one that must be followed when actual observations are to be analyzed.

Uniqueness is often rather difficult to prove and, in the absence of a proof, practising geophysicists naturally proceed without one. This has sometimes led to surprises. Consider the problem of modelling the geomagnetic field. For technical reasons the intensity of a magnetic field is far easier to measure accurately than its

three components; these reasons become even more compelling when the instruments are in orbit about the Earth. Therefore our knowledge of the global field depends heavily upon intensity values. The question arises whether one can uniquely[1] recover a harmonic vector field when only its intensity is known. To simplify the problem assume $|\mathbf{B}|$ is known everywhere on the surface of a sphere and that the field has an internal source. Here the forward problem is trivial: calculation of $|\mathbf{B}|$ from $\mathbf{B}$. Backus (1970) was able to devise a doubly infinite family of pairs of vector fields, in which the members of each pair generate identical intensities on a sphere: this counter-example demonstrates the impossibility of a general uniqueness proof. Backus (1968) also proved a series of uniqueness results, however, one being that the vector acceleration of gravity, $\mathbf{g}$, can be determined uniquely from $|\mathbf{g}|$ alone. Models of the geomagnetic field largely based on the intensity observations had of course been constructed, but a problematic inaccuracy in predicting the components was soon discovered (Cain 1971), and the poor performance of models based upon intensity became a major worry (Stern & Bredekamp 1975). It is now widely agreed that lack of uniqueness in the analytic problem is a major factor (called the Backus Ambiguity). In this case the only way to improve the solution appears to be measurement of the component values themselves.

An example of nonuniqueness, known for three centuries, is one that arises in the interpretation of gravitational fields: whereas a knowledge of the density structure of a system completely specifies its external gravitational field, a complete knowledge of the external field does not specify the density structure uniquely. In this case further information can be supplied by our knowledge of geology: for example, in exploration work and crustal studies it may often be assumed that the buried systems consist of relatively homogeneous units, perhaps with known densities; then the problem is to determine the shapes of these units. With several more assumptions about the buried body (e.g. that it is finite in extent and that a vertical line never intersects the body more than once) Smith (1961) was able to prove a uniqueness theorem stating that there is only one uniform body responsible for a given gravity anomaly, when the density is specified.

Uniqueness has been established for a number of geophysical inverse problems, e.g. the body-wave travel-time problem (Gerver & Markusevitch 1966) and the magneto-telluric problem mentioned earlier (Bailey 1970). Perhaps the most important exception is in the normal-mode inverse problem: the frequencies of free oscillation of the Earth have in the past five years yielded a precise, detailed picture of the mechanical structure of the deep interior (e.g. Gilbert & Dziewonski 1975), but there is no proof that the totality of such frequencies actually defines only one Earth model. Progress has been made on some analogous systems, however (Barcilon 1976).

We come next to the matter of *construction*, which, as Sabatier (1971) has remarked, receives an undue degree of attention. The uniqueness and existence of a solution are granted; what is now required is a procedure that will, in a finite number of steps, produce the solution with any specified finite precision. If iterative

[1] There is of course a trivial ambiguity of sign since both $+\mathbf{B}$ and $-\mathbf{B}$ give rise to the same $|\mathbf{B}|$. A single additional measurement of $\mathbf{B}$ serves to resolve this ambiguity.

methods are involved, a global convergence proof is required, and on this point most purely numerical schemes founder.

The classical solution of a geophysical inverse problem is the renowned Herglotz-Wiechert formula for obtaining the velocity-depth function from travel-time distance measurements in seismology (Bullen 1965). The idealized problem is as follows. A surface source emits signals travelling as seismic body waves through a spherically symmetric Earth, in which the continuous velocity v increases monotonically with depth. The time, T, taken for the earliest impulse to travel from the source to an observer at a range Δ (measured as an angle at the center of the Earth) is assumed known everywhere in some finite range, including $\Delta = 0$. If the rate of increase of v with depth is great enough but not too great, T is a monotonically increasing, differentiable function of Δ, and from the ray geometry we have

$$\frac{r_\Delta}{v_\Delta} = \frac{dT}{d\Delta},$$

where r_Δ is the radius of greatest penetration of a ray arriving at range Δ and v_Δ is the velocity at that radius. Then r_Δ can be found from the integral

$$\ln(a/r_\Delta) = \frac{1}{\pi} \int_0^\Delta \cosh^{-1}\left[\frac{dT}{d\Delta'} \Big/ \frac{dT}{d\Delta}\right] d\Delta',$$

where a is the Earth's radius. This formula combined with the first gives a means for constructing v at a radius of r. A similar formula can be developed for a flat Earth model. If the velocity fails to increase quickly enough with depth or there is a decreasing velocity, the derivative $dT/d\Delta$ ceases to exist and the method fails. Indeed there is no longer only one solution in the latter case, and then arrival-time functions of buried sources are needed to find the velocity (Gerver & Markusevitch 1966). This classical method is one that has been used with actual measurements, and it is the basis of a novel technique devised to account for data inadequacy, as we shall see later.

The geophysical inverse problem that seems to have collected the largest number of distinct methods of construction is the magneto-telluric problem, or its close relative, the problem of electromagnetic induction where only magnetic fields are recorded. The method of Siebert (1964) relies on the behavior of the data as the frequency tends to infinity and constructs a power series for the solution. Bailey's approach (Bailey 1970) uses the response data over the whole frequency range to develop a nonlinear integro-differential equation based upon the principle of causality (i.e. that the currents in the Earth must flow after the forcing field has been applied, never before). Weidelt (1972) employs a modification of the Gel'fand-Levitan method (1955). This method has the advantage of a degree of generality, for inverse problems where unknown parameters appear as coefficients in a Sturm-Liouville differential equation can usually be cast into the required form. A Fredholm integral equation is derived from the spectrum of the differential operator. Weston (1972) uses the Gel'fand-Levitan approach, but on data recorded in the time domain, rather than using the frequency-domain Fourier transform.

Having applied some of these methods to field observations, Bailey (1973) concluded that they were not particularly successful. The reason for this, in simple terms, is that the relevant equations describe magnetic and electric fields diffusing into the Earth, and that the information about deep structures is returned to the surface by strongly attenuated fields. Thus, unless the measurements are of astronomical precision, that information is lost. The analytic solutions seem to rely heavily upon the infinite density and precision of the idealized data and are therefore unsuitable for application to actual measurements.

There is a way to predict whether, in a particular problem, the solution depends upon the data in this rather unsatisfactory way: it is by deciding the question of *stability* of the problem. Mathematically, a problem is said to be stable if the solution depends continuously on the data and unstable if it does not. In simple terms this means that for all data sets lying close[2] to a particular set the solutions fall close to each other. This concept was introduced by Hadamard (1902) in connection with the study of boundary-value problems and he designated unstable problems "ill-posed." Very many inverse problems in geophysics are unstable in this sense. The best known example is that of downward continuation of harmonic fields (Bullard & Cooper 1948). Here the data are measurements of a harmonic function (typically a gravity or magnetic anomaly) taken on a particular level: the field values are then required at a deeper level that is nearer the sources. It is easy to show that two fields may differ by an arbitrarily small amount at the upper level, yet be quite different at the lower one, if the difference between the two original signals is confined to short enough wavelengths. This indicates the instability of downward continuation, which has long been recognized. One consequence of this instability is the corresponding lack of stability in any procedure to construct the source structure from potential field observations because downward continuation is implicit in all such processes.

Almost no work has been done on stability in the Western literature on analytical geophysical inverse problems, but the Russian school of applied mathematicians has made some important contributions here (e.g. see Lavrentiev 1967). Numerical methods for mitigating the undesirable effects of instability are discussed when we consider linear inverse problems in the next section.

ANALYSIS OF EXPERIMENTAL DATA: LINEAR PROBLEMS

Introduction

The class of linear inverse problems is particularly simple and it is the class about which the most is known; indeed, nonlinear problems are often treated by making approximations that reduce them locally to linear ones. A linear inverse problem is defined as one in which the data are linear functionals[3] of the model. Fortunately

---

[2] The concept of closeness requires the introduction of a metric onto the space of functions defining the model and the observations. Usually a measure based on a norm is meant, like the two-norm: then the distance between f_1 and f_2 is $\| f_1 - f_2 \|_2$.

[3] Mathematically, a functional is a mapping that maps a set of functions into the real numbers.

we can be more specific, since in geophysics it is almost always true that the model is related to the data via an integral transform:

$$e(x) = \int_I G(x, y)m(y)\,dy, \tag{1}$$

where m is the unknown function we are seeking, e a function representing the observations, and G a kernel derived from theory (for definiteness the independent variables may be thought of as one-dimensional and I as a real interval). Viewed as an equation for m, (1) is a Fredholm integral equation of the first kind. The two analytical questions of greatest importance for practical situations are those of uniqueness and stability.

The problem of uniqueness of solutions to equation (1) can be boiled down to the question: are there any nontrivial solutions $a(y)$ to the equation

$$0 = \int_I G(x, y)a(y)\,dy?$$

If the answer is no, then $m(y)$ is unique. If it is yes, then the class,[4] A, of all such solutions (so that $a \in A$) is called the annihilator of $G(x, y)$. Our knowledge of $e(x)$ can tell us nothing whatsoever about those parts of m that belong to A, and therefore these parts must be deduced from information other than that contained in e. For the complex kernels G of physical processes it is relatively rare that uniqueness can be established; in those cases that can be handled, a common procedure is to transform (1) into one of the well-studied integral transforms, for example:

Laplace: $G(x, y) = e^{-xy}$, $0 \leqq y < \infty$,

Fourier: $G(x, y) = e^{2\pi ixy}$, $-\infty < y < \infty$,

Hankel: $G(x, y) = J_v(xy)(xy)^{1/2}$, $0 \leqq y < \infty$.

All of these have unique solutions for a sufficiently well-behaved class of model functions. Furthermore, many Volterra equations [which contain the Heaviside function $H(x-y)$] can be shown to possess unique solutions, e.g. the Abel equation with

$$G(x, y) = H(x-y)(x-y)^{-v}, \quad 0 < v < 1.$$

Suppose an annihilator exists for a particular $G(x, y)$; its presence will not necessarily be revealed by numerical solutions of (1) because the (necessarily finite-dimensional) representation of $G(x, y)$ may not be singular, even though the true kernel is.

The matter of stability is also of great importance from a practical viewpoint. As we have seen in the previous section, the construction of solutions to an unstable problem is difficult: the smallest error in e may result in a wild excursion of m that bears no relation to the true solution. The numerical inversion of the Fourier and Laplace transforms is a good illustration of the influence of stability. Both

---

[4] A is actually a linear vector space; it is also sometimes called the null space of $G(x, y)$.

transforms possess unique inverses and analytic methods for their construction. However, the inversion of the Fourier transform is stable (in the two-norm) while the inversion of the Laplace transform is not (in any conventional norm); this fact accounts for the relative ease with which numerical inversion can be performed on the Fourier transform, while the Laplace transform is notoriously difficult to invert.

For linear problems the study of stability is not difficult. Roughly speaking, if the kernel G tends to "smooth" m, the inversion is unstable. A fairly widely applicable result is that, if I is finite and G is continuous, inversion for m is not stable.

In addition to downward continuation, some other unstable linear problems of geophysics are inversion of surface strain for fault displacement (Weertman 1965) and calculation of density contrasts arising from isostatic compensation (Dorman & Lewis 1972). Actually, differentiation of a data series is an unstable process also, and since differentiation enters some nonlinear construction methods (e.g. the Gel'fand-Levitan method) they may be unstable also.

Limitations of Experimental Data

The first limitation of actual observations that comes to mind is their imprecision: everyone knows that experimentally determined numbers are inexact. From a mathematical viewpoint, an equally important property is availability of only a finite number of measurements; indeed we shall maintain that it is *always* realistic to replace (1) with

$$e(x_i) = \int_I G(x_i, y)m(y)\,\mathrm{d}y, \qquad i = 1, 2 \ldots N,$$

or

$$e_i = \int_I G_i(y)m(y)\,\mathrm{d}y, \qquad i = 1, 2 \ldots N. \tag{2}$$

This manifestation of observational inadequacy is immediately obvious in some measurements, such as digital samples in the time domain or the determination of free periods of vibration of the Earth. When continuous records are considered, such as seismograms, it is still possible to reduce the continuous curves to a finite list of numbers. One way of achieving this is to construct a Fourier series[5] for the record; we may discard coefficients corresponding to frequencies higher than some finite limit because the sensing and recording instruments cannot respond to arbitrarily high frequency inputs.

The inadequacy of real measurements as expressed by the finiteness of N in (2) is in principle independent of their inaccuracy, and it is useful initially to study the idealized situation where the numbers e_i are perfectly accurate. This enables us to deal at once with the question of uniqueness and existence in regard to $m(y)$ determined from (2). It is easy to show that there are infinitely many different

---

[5] All actual records are finite in duration and of bounded variation—these conditions guarantee the existence of the Fourier series.

solutions $\mathbf{a}(y)$ to

$$\int_I G_i(y)\mathbf{a}(y)\,dy = 0, \qquad i = 1, 2 \ldots N;$$

in fact the annihilator here is infinite-dimensional for any finite collection of $G_i(y)$ whatever. Therefore models constructed from actual measurements can never be unique. Furthermore, if the kernels G_i are linearly independent (as they usually are in practical problems), solutions to (2) always exist.

Experimental error contributes additional indeterminacy to that caused by incompleteness of the data. In the following discussions of various methods, we assume that the statistical errors in e_i can be adequately described by a Gaussian distribution with known parameters. Although this may frequently be a poor approximation, there are two reasons for retaining it: first, the analysis can be carried out exactly under this assumption; second, knowledge of the true statistical distribution in a set of measurements is usually very poor so that an elaborate treatment based on a perfectly general error law seems hardly justified. Here we shall consider only statistically independent data, but there is no great difficulty in generalizing to the case of a joint-normal distribution over the N-data set (Gilbert 1971).

Since the actual solution $m(y)$ cannot be recovered from our measurements, there are in fact two courses open to us: (a) we can derive properties of $m(y)$ that all solutions share, which then must be properties of the true solution; (b) we can introduce assumptions about m to restrict the class of admissible solutions. In our development we begin by following the first course and this leads us naturally to suggestions of what useful yet plausible assumptions might be made.

Backus-Gilbert Formulation

Backus & Gilbert (1968) suggested that a very simple type of model-property be calculated, namely linear functions of m. By taking linear combinations of the data we can compute all functionals of the form

$$l = \int_I F(y)m(y)\,dy, \tag{3}$$

where l is calculated from the data (assumed error-free for the moment). Thus we have

$$l = \sum_{i=1}^N \alpha_i e_i, \tag{4}$$

and $F(y)$ is given by

$$F(y) = \sum_{i=1}^N \alpha_i G_i(y). \tag{5}$$

The weights α_i are of course entirely arbitrary. If these weights could be chosen so that $F(y)$ approached the Dirac function $\delta(y-y_0)$, the value of l would become $m(y_0)$. This is usually not feasible but Backus & Gilbert showed how to construct

functions $F(y)$ that are concentrated as much as possible on a chosen y_0 and are small elsewhere. One way of doing this is to define a delta-function quality measure such as

$$S[\tilde{\delta}] = 12 \int_I [\tilde{\delta}(y, y_0)(y - y_0)]^2 \, dy, \qquad (6)$$

with $\int_I \tilde{\delta}(y, y_0) \, dy = 1$.

Here the approximation $\tilde{\delta}(y, y_0)$ is intended to focus upon the position y_0. The measure S has dimensions of length and is indeed a rough estimator of the width of peak in $\tilde{\delta}$ around y_0; the smaller S can be made, the more closely $\tilde{\delta}(y, y_0)$ resembles the improper function $\delta(y - y_0)$. We now identify $\tilde{\delta}$ with F in (5) and choose α_i so as to minimize S for a particular point y_0. Then the model property l in (4) gives us a smoothed estimate of the true $m(y_0)$, the degree of smoothing depending on the shape of $\tilde{\delta}(y, y_0)$. Further, we can loosely identify the length S with the averaging width associated with our estimate l of $m(y_0)$; when S is large we obtain only a blurred picture of m at y_0, whereas small S indicates good resolution. Recall that these averages are universal properties of any solution to (2) and that such averages of any member of the annihilator would always vanish.

It should also be understood that (6) is only one of a variety of ways to define deviation of $\tilde{\delta}$ from a delta function. One important reason for choosing this particular definition is that it leads to *linear* equations for the coefficients α_i when the smallest S is sought. It shares this almost essential property with several other definitions of "deltaness," for example,

$$W = 12 \int_I \left[\int^y \tilde{\delta}(y', y_0) \, dy' - H(y - y_0) \right]^2 \, dy,$$

where $H(y)$ is the Heaviside function. A general, unifying theory of the criteria for "deltaness" appears in a rather difficult paper by Backus (1970b). He considered different operations for mapping the linear space of functionals containing distributions into a "smoother" Hilbert space and thereby generated a very wide (but not exhaustive) class of quality measures. Sometimes an appropriate choice of criterion can result in substantial reduction in numerical work; for example, by using the W-measure above, Oldenburg (1976) was able to arrive at equations for α_i that were particularly simple to solve.

Before discussion of the role of experimental error in the Backus-Gilbert formulation, it is worthwhile asking how l is related to the true value of $m(y_0)$. In fact, $m(y_0)$ can be arbitrarily far from l. The only circumstance in which the two numbers will be close is that when the model $m(y)$ is itself smooth in some sense. If we are willing to assume a certain degree of smoothness for all admissible models, we can calculate the greatest deviation between our estimator and the true value. These ideas are made precise by Backus (1970a,b) who gives a very general treatment of the problem. Actually this approach has enjoyed almost no practical application, perhaps because of the difficulty in estimating the numbers that quantify smoothness.

When statistical uncertainty is included, we find that the number l in equation (4) has an error that is easily computed when Gaussian statistics describe the random-

ness in e_i. This error is given by

$$\sigma^2 = \mathrm{var}\,[l] = \sum \alpha_i^2 \, \mathrm{var}\,[e_i], \tag{7}$$

when the e_i are uncorrelated. We should like to make σ^2 in (7) and S in (6) simultaneously as small as possible because they both quantify deficiencies in l as an estimator of $m(y_0)$. Backus & Gilbert (1970) showed how to obtain the best possible values of α_i, but a unique set is not possible: there is a single degree of freedom in the choice, and this leads us to the concept of a trade-off diagram.

Consider the plane defined by the variables σ^2 and S (Figure 1). It is possible to show that, for an arbitrary set of α_i (subject only to the condition $\int \tilde{\delta}\,dy = 1$), there is a region $\mathscr{R}$ inside of which all associated pairs (S,σ^2) must lie. Clearly, to do the best job we should like to be on the lower-left edge of the region, because then, for a given value of σ^2, there would be no way to obtain a smaller (i.e. better) value of S. The curve on the lower left side of the region $\mathscr{R}$ is called the *trade-off curve*. Backus & Gilbert proved that decreasing S always increased σ^2 on this part of the boundary of $\mathscr{R}$, so that improved resolution can only be obtained at the price of degraded statistical reliability. There is no "best" point on the curve. In some applications a poor resolution may be acceptable if good statistical precision is thereby guaranteed: for example, when the solution is very flat, one may be content
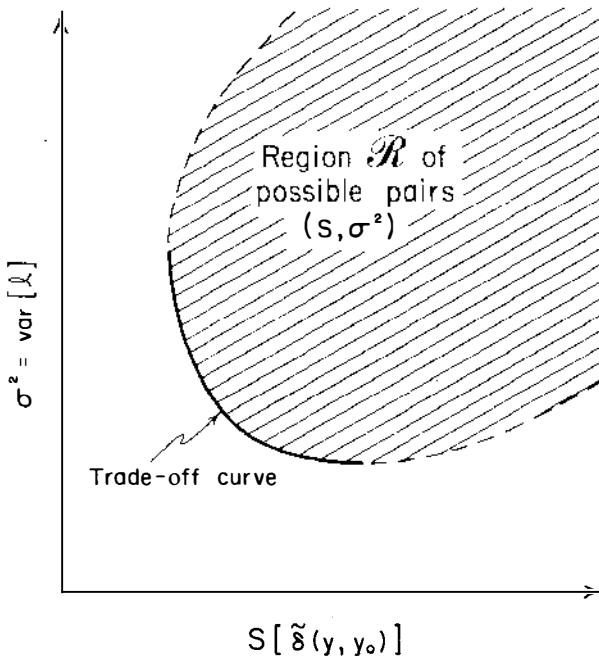

Figure 1   Region of feasible pairs (S, σ^2) shown shaded. The trade-off curve on which the best σ^2 is found for a given S is on the extreme left-lower edge of the region.

to average over a fairly broad interval to obtain precise estimates of the model value, while with rapidly varying solutions one might wish to obtain high resolution, even with relatively poor precision.

The Backus-Gilbert approach is not really a method for finding solutions $m(y)$ to (1); indeed, considered as a function of y_0, l in (4) does not usually satisfy the original data, even when they are error-free. Rather, it is a technique for assessing the significance of a solution. One may judge whether a particular feature (like a jump or spike) is really resolved by the data (Dorman & Lewis 1972; Hobbs 1973); one can compute how much detail is available in principle from a given distribution of observations. To calculate solutions $m(y)$ [especially when (1) is a linearization of a nonlinear problem] a different method has been found that offers distinct advantages, mainly with regard to numerical stability: this is the spectral expansion method.

A Spectral Expansion

The procedure to be described is a particular formulation of the numerical solution to practical linear inverse problems advocated by many authors (Gilbert 1971; Jackson 1972; Wiggins 1972; Jupp & Vozoff 1975), all of whom based the solution ultimately upon the philosophy of Lanczos (1961). Most workers have dealt with finite-dimensional models because this allows a direct description in terms of matrices and makes possible a limited form of the Backus-Gilbert interpretation of the results. The approach developed here, following Gilbert (1971), is in greater harmony with the cardinal factor distinguishing inverse theory from conventional parameter estimation, namely, that the space of unknowns is infinite-dimensional.

Analytic instability appears as extremely poor conditioning in the matrices encountered in a numerical solution, even if nonuniqueness is evaded by reducing the number of unknowns sought. The spectral expansion explicitly isolates those parts of the solution that are well determined by the data and those that are not; furthermore, the statistically reliable component of m (as derived from the known error estimates in the data) are also those with the highest numerical reliability in actual, finite-accuracy computations. These features of the method make it the most suitable when very large data sets are handled, even though it is up to ten times slower computationally than simple matrix inversion.

We treat noisy data from the beginning: consider (2) weighted by the inverse of the standard error of each measurement (recall that statistically independent e_i are assumed with individual Gaussian error distributions of zero mean and standard error σ_i)

$$\frac{e_i}{\sigma_i} = \int_I \frac{G_i(y)}{\sigma_i} m(y) \, dy, \qquad i = 1, 2 \ldots N,$$

which we shall write as

$$e_i' = \int_I G_i'(y) m(y) \, dy, \qquad i = 1, 2 \ldots N. \tag{6}$$

Thus e_i' is dimensionless with unit variance. Define the matrix Γ with elements Γ_{ij}

thus:

$$\Gamma_{ij} = \int_I G_i'(y)G_j'(y)\,dy.$$

Γ is easily shown to be positive-definite and symmetric, so that it may be diagonalized with an orthogonal matrix $\mathbf{O}$ thus:

$$\mathbf{O}^T\mathbf{\Gamma}\mathbf{O} = \mathbf{\Lambda},$$

where

$$\mathbf{\Lambda} = \text{diag}\,\{\lambda_1, \lambda_2 \dots \lambda_N\},$$

and

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \lambda_N > 0.$$

Some of the eigenvalues of Γ may be zero if the $G_i(y)$ are linearly dependent, but we shall ignore this case for the moment. The set of eigenvalues will be called the spectrum of the problem defined by (6). This spectrum is not the same as that of the analytic problem (1), since the values λ_i depend critically upon the estimated errors in the data and the distribution of the observations. Now consider the functions $\psi_i(y)$ defined by

$$\psi_i(y) = \lambda_i^{-1/2} \sum_j O_{ji} G_j'(y).$$

It can easily be verified that $\psi_i(y)$ are an orthonormal set, i.e.

$$\int_I \psi_i\psi_j\,dy = \delta_{ij}.$$

Therefore we may consider an expansion of m in terms of these orthogonal functions

$$m = \sum_{i=1}^N a_i\psi_i + \psi_*, \tag{7}$$

where $\psi_* \in A$ (the annihilator)[6] and $\int \psi_*\psi_i\,dy = 0$. The coefficients of this expansion are obviously

$$a_i = \int_I \psi_i m\,dy$$
$$= \lambda_i^{-1/2} \sum_j O_{ji} e_j' \tag{8}$$

and are common properties of all solutions satisfying the data. It requires a little algebra to show that the standard error of each coefficient a_i is $\lambda_i^{-1/2}$ and that the a_i are *statistically independent*. Thus the expansion of m is in terms of functions whose coefficients increase in uncertainty; after ψ_N we reach ψ_*, whose uncertainty is total.

[6] If any zero eigenvalues occur their (associated) functions should be assimilated into ψ_*, not normalized by the factor $\lambda^{-1/2}$.

When analytically unstable problems are treated in this way, it is almost invariably the case that the functions $\psi_i(y)$ become more oscillatory as i increases; i.e. the largest eigenvalues are associated with the smoothest functions. Then the smoothest parts of m are most accurately determined because their coefficients are most precise. However, such behavior is not universal, because examples can be constructed with the opposite pattern,[7] but these exceptions are so rare in practice that we shall proceed on the assumption that the expansion (7) is in order of decreasing smoothness. What can we say about ψ_*, the part of m quite undetermined by the data? It is possible to verify that this too is an oscillatory function by asking for the smoothest possible member of A: we might define this function to be the one with the least value of $\int (d^2 f/dy^2)^2 \, dy$ and calculate its shape with the usual calculus of variations. No one has ever done this to my knowledge. A geophysical example with relatively smooth members in the annihilator is given by Parker & Huestis (1974), where the annihilator contains all the components that do not average to zero.

Another common property of the spectra obtained in practical problems is the condensation of the eigenvalues towards zero; this means that the spacing between eigenvalues gets smaller as their magnitude decreases. When many of the eigenvalues are small (i.e. a small fraction of λ_1) this indicates a numerically (and physically) very poorly posed problem, since now many of the coefficients exhibit large uncertainties.

Let us now return to (7) and regard it as an expansion in the natural "modes" of the data. Perhaps we should like to "filter out" the noisy components of the series and thus gain some statistical stability in our estimate of m. Clearly we should then discard those functions ψ_i where coefficients exhibit unacceptably large variance; if at the same time this means throwing out high frequency components, our solution will also be a smoothed version of the true m (provided ψ_* is not a smooth function). This is of course exactly analogous to the Backus-Gilbert formulation. The more reliable, truncated version of (7) can be accurately represented as that part of m expansible with the smoother functions ψ_i. We may further say either (a) that the high frequency parts are less interesting, or perhaps (b) that they can be assumed to be of small amplitude anyway.

The precise number of components to be accepted in the expansion is still largely a matter of judgement. One factor that should be taken into account is that the truncated series (7) no longer fits the original data precisely. Because of the random component in e_i, we should not expect or demand exact agreement. If we define χ^2 to be the squared two-norm misfit to the data thus:

$$\chi^2 = \sum_{i=1}^{N} (\tilde{e}'_i - e'_i)^2,$$

where $\tilde{e}'_i$ is the value obtained by substituting the truncated series into the right side of (6), we find after a little algebra that

---

[7] For example, consider the set of kernels $G'_n(y) = n \sin \pi n y$, $n = 1, 2 \ldots N$ with $0 \leq y \leq 1$; here we have $\lambda_n = n^2/2$ and $\psi_n(y) = 2^{-1/2} \sin \pi n y$.

$$\chi^2 = \sum_{j=L+1}^{N} \lambda_j a_j^2;$$

here L terms have been accepted in (7). This equation shows that those coefficients a_i associated with small eigenvalues contribute relatively little to the misfit compared with their contributions to the solution or its uncertainty.

Now χ^2 is the standard statistic, since e_i' has unit variance. It is possible to pick L so as to make χ^2 equal to its expected value; the difficulty here is in selecting the number of degrees of freedom, because the underlying model m has in principle an infinite number of parameters. If one regards the truncated series as the "model," one has $v = N - L$, and L can be picked to make $P(\chi^2) = 0.5$. In fact v so defined is regarded by many (see Richards 1975) to be the number of independent data in the original set: often v is only 10% of N.

Other methods of stabilizing (7) have been suggested; for example one can choose m with the smallest norm, $\| m \|_2$ (here $\sqrt{\sum a_i^2}$), subject to its not exceeding a certain misfit level (Jackson 1973). This has the effect of replacing λ_i by $\lambda_i + C$ in (8), the equation for a_i, where C is a positive constant determined by a nonlinear equation. Evidently this has the same effect of reducing the influence of the functions associated with small eigenvalues. The method of Twomey (1963), which seeks the smoothest model in another sense, has the same sort of properties, as does the method of Franklin (1970).

An Illustration

The following example is based upon a solution to a geophysical inverse problem given by Dorman (1975). The objective is to determine the density structure within the Earth from gravity observations but, as noted earlier, further restrictions of some kind are always required to reduce the large degree of ambiguity in the general problem. Here very strong geometric constraints are imposed: the two-dimensional system (Figure 2) consists of two quarter-spaces, each horizontally layered; gravity anomaly gradients $\partial \Delta g / \partial x$ are measured on the surface $y = 0$. This is a model for a vertical fault with vertical displacement or the figuration that may be encountered across an oceanic fracture zone where two lithospheric plates of different ages meet. Gravity gradients are no longer commonly measured, but this type of data is assumed here for mathematical convenience. The solution to the
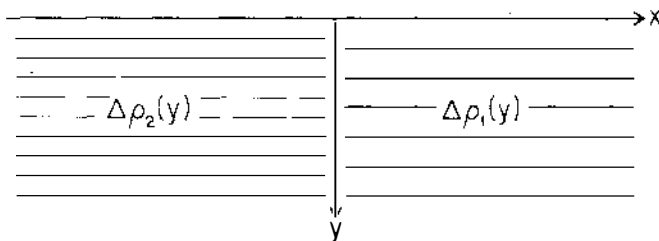

Figure 2   Geometry of gravitational edge effect problem. Each quarter-space is horizontally stratified and they meet along the plane $x = 0$.

forward problem is found to depend only upon the difference in density $\Delta\rho$ between the layers on each side at a given level:

$$\Delta g'(x) = \int_0^\infty \frac{2\mathcal{G}y}{y^2 + x^2} \Delta\rho(y)\, dy, \tag{9}$$

where $\mathcal{G}$ is Newton's gravitation constant and $\Delta g'$ is written for $\partial \Delta g / \partial x$. to recover $\Delta\rho$ from $\Delta g'$.

In order to justify some later steps, it is useful to restrict $\Delta\rho$ to a particular class: $\Delta\rho \in L^1(0, \infty)$. This means that the integral of the magnitude of $\Delta\rho$ is assumed to be bounded. We could reasonably assume also that $\Delta\rho$ itself is everywhere bounded, then $\Delta\rho \in L^2(0, \infty)$ also, but this additional restriction is not particularly valuable in the analysis.

Let us first observe some of the conditions for existence of a solution. From (9) it is obvious that $\Delta g'(x) = \Delta g'($
necessary condition for existence. Indeed, the approximate symmetry of the measured anomalies over a transform fault is a reason for proposing a model of this kind for that system. Consider now complex values of x; we must of course keep y real, but (9) allows us to *calculate* $\Delta g'$ for any x real or complex. It is easy to show that $\Delta g'$ is an analytic function of a complex variable whose only singularities are on the imaginary x axis. This implies that $\Delta g'$ is infinitely differentiable on the real axis (which is accessible to observation!) except possibly at the origin, $x = 0$. The analyticity conditions on $\Delta g'$ are clearly necessary from (9). A little strengthening of these restrictions concerning the decay of $\Delta g'$ at large $|x|$ results in a sufficient condition but the details will not detain us.

Dorman's development proceeds as follows. The Fourier transform of $\Delta g'$ is taken:

$$\begin{aligned}
\overline{\Delta g'}(k) &= \int_{-\infty}^\infty \Delta g'(x)\, e^{ikx}\, dx \\
&= \int_{-\infty}^\infty dx\, e^{ikx} \int_0^\infty \frac{2\mathcal{G}y\Delta\rho(y)}{x^2 + y^2}\, dy \\
&= \int_0^\infty 2\pi\mathcal{G}\Delta\rho(y)\, e^{-|k|y}\, dy
\end{aligned} \tag{10}$$

The validity of interchanging the integrations and the existence of the transform follow from Fubini's theorem and the fact that $\Delta\rho \in L^1$. Restricting ourselves to positive k, we note that (10) is simply a Laplace transform so that it may formally be inverted:

$$\Delta\rho(y) = \mathcal{L}^{-1}[\overline{\Delta g'}(k)]/2\pi\mathcal{G}.$$

This is Dorman's solution. Uniqueness of $\Delta\rho$ is a direct consequence of Lerch's lemma which states that two functions with the same Laplace transform are identical up to a set with measure zero. The analytical machinery for finding $\Delta\rho$ from $\Delta g'$ is an integral in the complex k plane on the Bromwich contour. This procedure is of little account practically because of its inherent instability.

This brings us to the question of stability. Rather than appealing to the well-known instability of the inverse Laplace transform, let us deal directly with (9). We construct an example of noncontinuous dependence of $\Delta\rho$ on $\Delta g'$. First we need a distance measure for the space of data and density functions. In the latter case, a natural choice is based upon the one-norm:

$$\| \Delta\rho \|_1 = \int_0^\infty |\Delta\rho(y)| \, dy;$$

this is because $\Delta\rho \in L^1(0, \infty)$ and therefore all $\Delta\rho$'s of interest possess bounded one-norms. We shall use a one-norm on the space of analytic functions containing $\Delta g'$ just for symmetry. Now a density function $\delta\rho \in L^1$ must be found for which the corresponding $\| \Delta g' \|_1$ can be made as small as desired, while $\| \delta\rho \|_1$ remains bounded away from zero. The smoothing action of the kernel in (9) suggests that an oscillating function might be suitable, and after some experimentation I discovered the function

$$\delta\rho(y) = \frac{\rho_0 a^2 \sin \mu y}{(a^2 + y^2)}$$

with $\rho_0, a, \mu > 0$. Since $|\sin \mu y| \geq \sin^2 \mu y$, it is easily shown that

$$\| \delta\rho \|_1 \geq \pi\rho_0 a(1 - e^{-2\mu a})/4.$$

The form of $\delta\rho$ makes it easy to find $\Delta g'$ from (9) with the calculus of residues: the expression is not important, only the fact that $0 \leq g'(x)$ for all x. Obviously then we have

$$\| \Delta g' \|_1 = \int_0^\infty \Delta g'(x) \, dx$$

$$= \pi\mathscr{G} \int_0^\infty \delta\rho(y) \, dy;$$

this last result comes from (10) with $k = 0$. Thus we have

$$\| \Delta g' \|_1 = \pi\mathscr{G}\rho_0 a^2 \int_0^\infty \frac{\sin \mu y}{a^2 + y^2} \, dy.$$

By the Riemann-Lebesgue theorem, $\| \Delta g' \|_1$ can be made as small as we please by choosing a large enough μ, but from the earlier inequality, $\| \delta\rho \|_1$ does not decrease below some fixed value. Because of the linearity of the problem, this $\delta\rho$ perturbation can be added to any other $\Delta\rho$, and it will be seen that arbitrarily close functions $\Delta g'$ are associated with functions $\Delta\rho$ that never come closer together than some fixed separation. The same test function also serves to demonstrate instability with the two-norm or the uniform norm.

One final analytic point I cannot resist including is the equivalence of (9) to the equation of Weertman (1965) which relates fault displacement to surface strain in a simple system. Weertman found a different way of constructing solutions. His results show that, if $\Delta g'$ is regarded as the real-axis realization of a complex function of

complex x, the corresponding density is simply

$$\Delta\rho(y) = -\text{Im } \Delta g'(iy)/\pi\mathcal{G},$$

that is, $\Delta\rho$ can be obtained by analytic continuation of $\Delta g'$ onto the imaginary x axis. Analytic continuation in the complex plane can be cast into a form that requires the solution of Laplace's equation with over-determined boundary conditions, the unstable problem that Hadamard (1902) first used to illustrate the concept.

We come now to a demonstration of the spectral expansion method. Gravity gradients were computed for a density difference model of 100 kg m^{-3} between 5 and 15 km, zero elsewhere; then zero-mean, normally distributed random errors were added to the 20 exact values. The values are plotted in Figure 3. Note that gradients have been provided on only one side of the density interface; this avoids certain difficulties of exact linear dependence in the Backus-Gilbert analysis, and it would be natural with real observations to average the two half-profiles together to get the best estimate of the symmetric function. The range of these "observations" is from 0–15 μm s^{-2} km^{-1}, and the standard error of the noise is 0.1 μm s^{-2} km^{-1} (there appears to be no definitive statement about gravity units in the SI system; therefore I have picked these. Note 1 μm s^{-2} km$^{-1} \equiv 0.1$ mgal km^{-1}). The scaled data and their kernels are calculated next in order to find the orthogonal kernels . ψ_i. The matrix Γ is easily shown to be

$$\Gamma_{ij} = 2\pi\mathcal{G}^2/(x_i + x_j)\sigma^2,$$

where x_i is the coordinate of the ith measurement and σ^2 the variance of each datum. The eigenvalues of Γ cover an enormous range, as can be seen from Table 1; the number of very small values indicates the very poor conditioning of the matrix. Some of the corresponding orthogonal functions are shown in Figure 4 where the standard pattern, increasingly oscillatory functions with decreasing eigenvalue, is
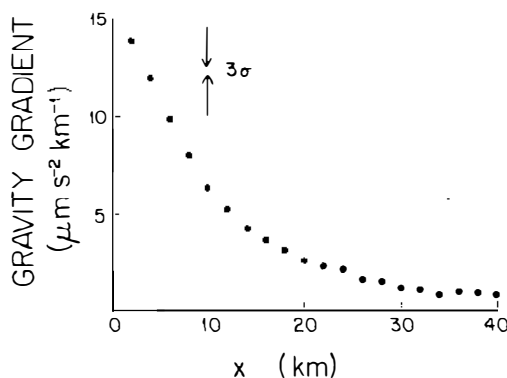

Figure 3   Artificial data values used in numerical examples. There are twenty gravity gradients computed at 2 km intervals from $x = 0$; the value at $x = 0$ is not included because the associated data kernel is singular.

Table 1 Selected values of parameters derived in the numerical solution of the gravitational edge-effect problem: λ_i stands for eigenvalues of matrix Γ; a_i, spectral expansion coefficient in (7); $\lambda_i^{-1/2}$, standard error of a_i; $\chi^2(i)$, normalized squared misfit of expansion when all terms up to and including the ith have been summed in (7).

i	λ_i $(\text{m}^6\,\text{kg}^{-2}\,\text{km}^{-1})$	a_i $(\text{kg}\,\text{m}^{-3}\,\text{km}^{1/2})$	$\lambda_i^{-1/2}$ $(\text{kg}\,\text{m}^{-3}\,\text{km}^{1/2})$	$\chi^2(i)$
1	2.090	166.7	0.692	4441.0
2	0.3721	89.3	1.64	1474.0
3	4.683×10^{-2}	-175.1	4.62	37.8
4	4.855×10^{-3}	-63.1	14.3	18.5
5	4.253×10^{-4}	47.1	48.5	17.6
6	3.179×10^{-5}	302.0	177.4	14.6
7	2.037×10^{-6}	491.0	700.7	14.2
8	1.122×10^{-7}	-4186.0	2986.0	12.2
9	5.312×10^{-9}	-1022.0	13720.0	12.2
10	2.158×10^{-10}	36972.0	68075.0	11.9
20	1.930×10^{-29}	2.976×10^{14}	2.276×10^{14}	0

evident. In Table 1 the column headed $\chi^2(i)$ gives the value of the misfit when all terms in (7) up to i have been included. Since there are 20 data, a value of about 20 is to be expected, a smaller value indicating over-fitting. Thus an expansion with only four functions would seem appropriate; we are led to the same conclusion by examining $\lambda_i^{-1/2}$, the errors in the coefficients, since after $i = 4$ there is a rapid deterioration in the accuracy of a_i. Figure 5 shows the result of taking four terms in (7).

If we had ignored the problems of error and instability and tried to fit the data exactly (recall there are infinitely many such solutions), we would find that the rather small errors in the data are magnified grotesquely in the solution. For example the $\Delta\rho$ with the smallest two-norm is found by summing (7) all the way up to N: this yields a density contrast function with oscillations exceeding $\pm 5 \times 10^{14}$ kg m^{-3} and because that solution minimizes $\| \Delta\rho \|_2$ it represents one of the smallest possible functions that fits the data precisely.

The Backus-Gilbert analysis tells the same story as the spectral expansion. Ignoring the noise in the data, I constructed two delta-function approximations, one for a depth of 15 km, the other for 50 km (Figure 6). There is a little trouble with (6) over the infinite interval, because the integral is divergent for general linear contributions of G_i. This can be remedied in several ways, some more elegant (Backus 1970b) than others; I chose the crudest device of restricting the depth to 100 km, equivalent to the assumption that there are no density variations below that level. As can be seen from the figures, the distribution of observations is capable in principle of resolving rather fine details around 15 km, and the resolution is not too bad even at 50 km. However, when the standard error of the linear estimate is calculated we find that it exceeds 10^{12} kg m^{-3} for both cases. Clearly it is desirable to sacrifice resolving ability for improved statistical properties. Figure 7

is the trade-off diagram for the two depths. This shows that to obtain a standard error of ± 20 kg m^{-3} (recall from Figure 5 typical values are around 50 kg m^{-3} at 15 km) we need a resolving width of about 11 km, which is perhaps geophysically interesting. The curve for $y_0 = 50$ km leads to the discouraging conclusion that little useful information is contained in the data about densities so deep in the Earth.

ANALYSIS OF EXPERIMENTAL DATA: NONLINEAR PROBLEMS

Linearization and the Fréchet Derivative

Many of the most important geophysical problems are nonlinear, which means that the solution to the forward problem cannot be expressed with linear functionals
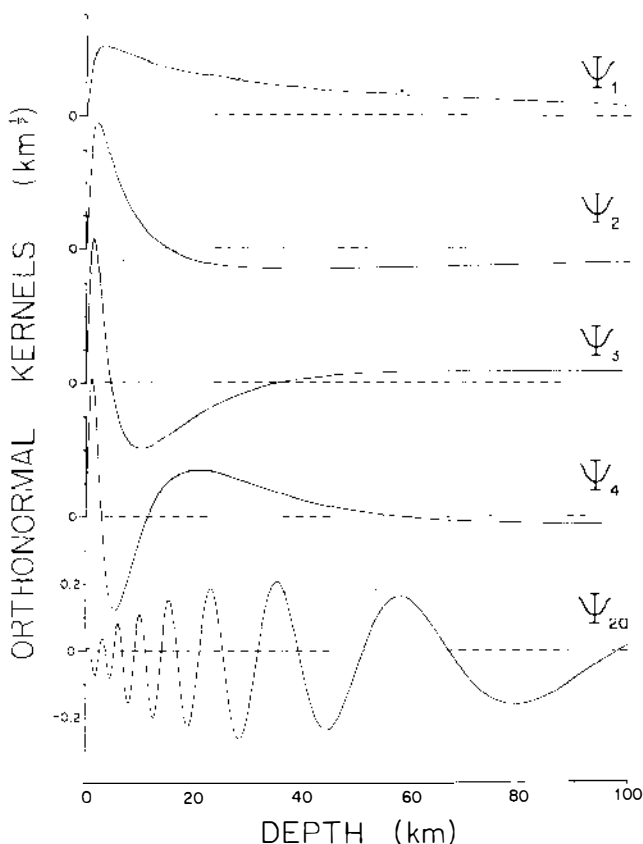
Figure 4   Orthonormal kernels found by diagonalization of Γ. The function ψ_i exhibits i zero-crossings (counting the one at $y = 0$). Thus ψ_{20} has three more zero-crossings than are shown in the figure, all deeper than 100 km.

over the model space. However, the powerful tools of linear theory are so attractive that geophysicists are often willing to make approximations allowing their use. One difficulty is that of establishing the validity of the approximation; cautionary comments (e.g. Anderssen 1975) and examples of the failure of linearization (Wiggins
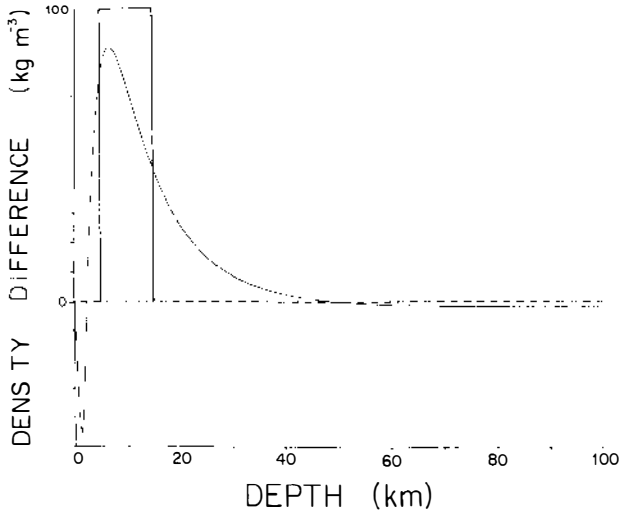
Figure 5 Original density function (box-car) and solution derived from four orthonormal kernels (smooth curve).
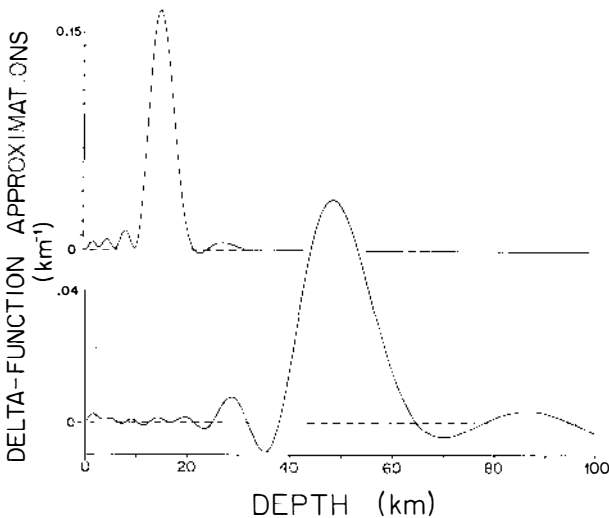
Figure 6 Best delta-function approximations according to the S-criterion for target depths of 15 km and 50 km.

et al 1973, Sabatier 1974) have appeared in the literature warning of the misleading results that can sometimes occur. In some cases, unfortunately, there appears to be no satisfactory alternative for the construction of a solution from real observations or assessment of its significance.

The approximation to which we have been alluding is *linearization*. Suppose the ith observation e_i is related to the model via a (nonlinear) functional F_i, formally:

$$e_i = F_i[m] \qquad i = 1, 2 \ldots N. \tag{11}$$

This deceptively simple equation may hide such complex calculations as the solution of many coupled, ordinary differential equations, as in the case of the inverse problem for the mechanical structure of the Earth from free oscillation data (Backus & Gilbert 1967); it represents the fact that the forward problem is completely solved and that when a particular m is given we can, somehow, find all the appropriate data associated with it. Next, consider a second model $m + \delta m$
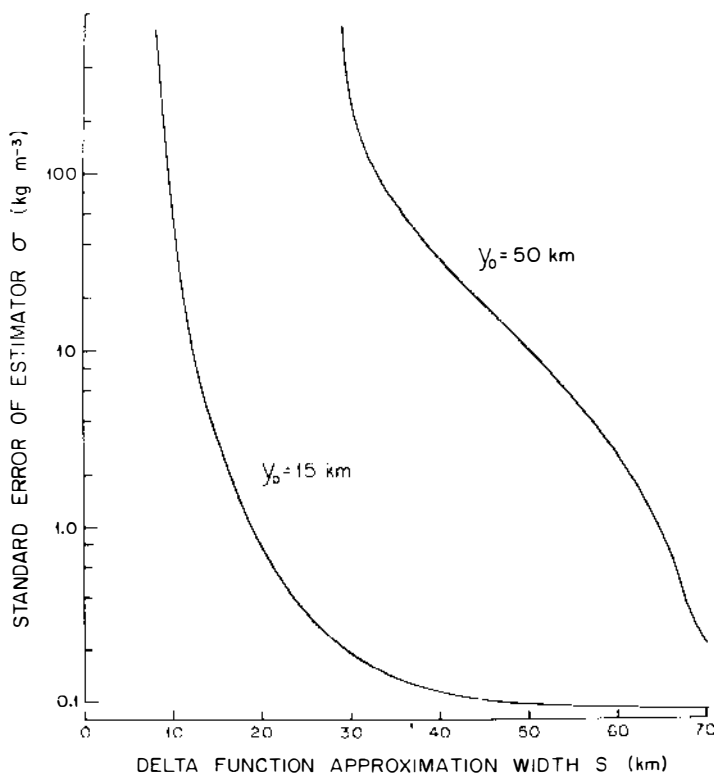
Figure 7   Part of the trade-off curves of standard error σ, and resolution width S according to the S-criterion for target depths of 15 km and 50 km. The complete diagram extends to 10^{14} kg m^{-3} in σ and 200 km in S but such extreme values are meaningless.

close[8] to m and with associated data $e_i + \delta e_i$. To find the new data we might solve a perturbation problem, since m is in a certain sense small; it will be assumed that the solution of this perturbation problem can be written

$$\delta e_i = \int_I D_i(m; y)\, \delta m(y)\, dy + O \int_I [\delta m(y)]^2\, dy \qquad (12)$$

(once more, a simple single-parameter, one-dimensional problem has been chosen for simplicity). Then F_i is said to be *Fréchet differentiable* at m and the function D_i is the *Fréchet derivative* of F_i at m. Another name for D_i, sometimes used in the applied mathematical literature, is the "sensitivity function." Equation (12) can be compared with the more familiar expression obtained when there are a finite number of parameters, μ_j, governing e_i:

$$\delta e_i = \sum_j \frac{\partial F_i}{\partial \mu_j} \delta\mu_j + O\left(\sum_j \delta\mu_j^2\right). \qquad (13)$$

This reveals a correspondence between the partial derivatives $\partial F_i/\partial \mu_j$ and the Fréchet derivative $D_i(m; y)$ if the parameter μ_j is the value of the model m in the jth layer.

Quite surprisingly, perhaps, the solution to perturbations of almost all forward problems in geophysics can be written as (12), and therefore these problems are Fréchet differentiable. One way of arriving at a perturbed solution in the required form is of course through standard perturbation analysis (e.g. Parker 1970 or Johnson & Gilbert 1972); another powerful technique particularly useful in certain seismic problems utilizes the variational formulation of the equations (e.g. Backus & Gilbert 1967 or Gilbert 1976), but there is insufficient space here to deal with these ideas as they deserve.

Having obtained (12) we can see that, if the term $O \int \delta m^2\, dy$ is dropped, it is identical with (2), where $D_i(m; y)$ plays the role of $G_i(y)$. Thus we treat the system as if it were locally linear. Using (12) and the spectral expansion method we could attempt to discover the perturbation δm that would bring an initial guess model into satisfactory agreement with the real observations. If the guess solution and a satisfactory one are not sufficiently close, the neglected term in (12) may not be truly negligible and the new solution will not be satisfactory when substituted into the full nonlinear equation (11). Then the process should be repeated, linearizing about the most recent estimate of m, until improvement ceases or a satisfactory solution is obtained. There is of course no guarantee that this procedure will work at all: it may be that no solution exists matching the observations to an adequate degree, and then it is obvious that the iterative process must fail (recall, however, that in a *linear* problem there are always an infinite number of "good" models, each fitting the observations arbitrarily well, provided the $G_i(y)$ are linearly independent); the other alternative is that, even though satisfactory solutions exist, the initial guess is so far from any of them that divergence ensues. The reader will perhaps recognize this iterative procedure as analogous to Newton's solution of nonlinear equations in several unknowns (Ortega & Rheinboldt 1970).

[8] See footnote 2.

Consider now the situation in which a satisfactory solution has been arrived at; to assess its significance one can simply apply the Backus-Gilbert method to the problem linearized about the best-fitting solution, in the hope that the neglected nonlinear terms do not make an important contribution. One serious problem is that there might be several satisfactory solutions widely separated in the model space, but the linearization affords approximate description of the neighborhood of only one of them. Nothing can be done about this except to attempt a more exhaustive search of the model space for other possible solutions.

There is one final remark I wish to make concerning the use of the Fréchet derivative (12) rather than the corresponding approximate representation (13). Some authors prefer to parameterize their models through layers or cells so that (13) seems a very natural way to express the solution to the perturbation problem. It is apparently not well appreciated that the Fréchet derivative D_i is often far easier to find both algebraically and computationally than the corresponding partial differential coefficients. Even if a finite representation is chosen, there are still great advantages to finding $\partial F_i/\partial \mu_j$ from D_i. A case in point is the following: consider the inverse problem of discovering the shape of crystalline basement that is everywhere buried under sediments. The data are the gravity anomalies caused by the (known) density contrast between the two materials, $\Delta\rho$. If we let $h(\mathbf{x})$ be the depth at a horizontal position $\mathbf{x} = (x, y)$, then the solution to the forward problem is

$$\Delta g(\mathbf{x}_i) = \int_S \frac{\mathscr{G}\Delta\rho \, dS}{\left[|\mathbf{x}_i - \mathbf{x}|^2 + h(\mathbf{x})^2\right]^{1/2}}, \qquad i = 1, 2 \ldots N,$$

where $\Delta g(\mathbf{x}_i)$ is the anomaly observer at $\mathbf{x}_i$ on a horizontal level and $\mathscr{G}$ is Newton's gravitational constant. Clearly this is a nonlinear problem for h. It is easy to show that

$$D_i(h; \mathbf{x}) = -\frac{\mathscr{G}\Delta\rho h(\mathbf{x})}{\left[|\mathbf{x}_i - \mathbf{x}|^2 + h(\mathbf{x})^2\right]^{3/2}}, \qquad i = 1, 2 \ldots N$$

and that these Fréchet derivatives exist for all h with $h(\mathbf{x}) > 0$ for all $\mathbf{x} \in S$. In contrast, the formula for $\partial\Delta g(\mathbf{x}_i)/\partial h_j$, when the plane is divided into square cells and h is then forced to be constant within each cell, apparently extends over many lines (Burkhard & Jackson 1976). It would be very simple to derive an approximation for $\partial\Delta g(\mathbf{x}_i)/\partial h_j$ by rough numerical quadrature of D_i.

Fully Nonlinear Treatment

Linearization is very successful at discovering an acceptable solution to a nonlinear inverse problem, but a finite number of data do not specify a unique solution (except in certain pathological cases). It is not enough to possess one adequate solution, because, in order to draw valid conclusions about the Earth, we need in principle to consider every possible solution, not just one of them. Linearization provides only a partial answer to this question by giving an approximate description of the neighborhood near the preferred model. The difficulty of describing the complete class of viable solutions makes it convenient to consider a simpler problem whose solution is geophysically just as important: the derivation of *properties* that

all solutions have in common. The Backus-Gilbert local averages are an example of a shared property in the linear case. The properties that seem to be most amenable to mathematical analysis in nonlinear systems are upper and lower bounds of various parameters, and these are extremely important geophysical quantities. In the next few paragraphs we discuss three different approaches to the practical calculation of the limits placed by the data on geophysical models.

The method of widest applicability is Monte Carlo modelling (Keilis-Borok & Yanovskaja 1967; Anderssen et al 1972). The principle is simple; one parameterizes the model space with a large but finite number of unknowns and then generates a sequence of structures at random, testing each one against the observations. While most structures will fail the test, some will pass and these form a family of solutions whose common properties are chosen to represent those of the complete class. Normally the structures investigated are profiles of a single parameter, e.g. seismic velocity or density as a function of depth; the property invariably investigated is the "corridor" of upper and lower limits varying with depth within which all solutions are hypothesized to lie. There is unfortunately an objection to the use of a corridor that is seldom mentioned. Consider for example a layer 1 mm thick with density 10^6 kg m^{-3}; should such a layer exist within the Earth it would be undetectable by seismic or gravimetric means because it is so thin. Therefore the addition of such a layer to an acceptable solution would not upset agreement with the observations. Since the layer could be placed at any depth, the upper bound in density cannot be lower than 10^6 kg m^{-3} anywhere, a geophysically uninformative value. There are of course ways of eliminating improbable extreme cases like this. One way is suggested by the Backus-Gilbert trade-off diagram: the corridor is defined to concern the average value of density in a succession of layers with specified thicknesses. Alternatively, gradients larger than a certain amount can be rejected a priori as "geophysically unreasonable." Whatever method is chosen, the fact remains that the corridor is sensitive to additional information not present in the original observations. Practical applications of the Monte Carlo approach usually impose other restrictions as well, such as an a priori corridor of reasonable maximum and minimum parameter values; these assumptions are made to reduce the Herculean effort of conducting a complete search, since they decrease the size of the parameter space. The possibly strong dependence on seemingly unimportant computational details must be kept in mind when one wishes to interpret the results of these studies.

The method has been applied notably to seismic problems, for example the determination of the interior mechanical structure using the dispersion of surface waves (e.g. Jackson 1973, who uses a refinement based upon local linearization to ease the computational labor) or from the frequencies of free oscillation of the Earth (e.g. Press 1968), but the great number and precision of recent observations precludes any further direct application in the latter problem.

The next nonlinear treatment is in some respects the antithesis of the Monte Carlo method: it is an elegant treatment of a specific problem and is based on an analytic construction algorithm. The problem is that of inverting travel-time distance data for seismic velocity, and the underlying analysis is of course the

Herglotz-Wiechert solution discussed earlier. There are different variations (McMechan & Wiggins 1972; Bessonova et al 1974) depending upon what type of data are assumed (array studies can give direct estimates of $dT/d\Delta$, for example), but the objective is again to discover a corridor within which all permissible velocities must fall. The principle of the method is to use the Herglotz-Wiechert formula (in spherical or flat-Earth form) to give the greatest or least depths at which a specified velocity can occur. Considerable complications ensue when the possibility of low-velocity zones is permitted, but only then is the method useful practically, for such layers are commonly encountered in crustal and upper mantle studies. When the velocity decreases (or fails to increase rapidly enough with depth in the sphere), a unique solution no longer is possible even with perfect data. To cope with this difficulty a limit is placed a priori on the permissible magnitude of any velocity reversals, which has the gratifying effect of automatically eliminating troublesome thin layers with high velocities. The size of the bound can be fairly well estimated from a knowledge of the materials likely to be encountered. Unlike Monte Carlo searching the method is cheap computationally, so that different assumed values can be tried if necessary. Clear comparative reviews of this extremal technique and the corresponding linearization have been given by Wiggins et al (1973) and Kennett (1976).

The last approach is one I have personally been working on for several years, and I believe it is the one with the greatest potential. The principle is to choose a scalar property of the model and then to maximize or minimize its value subject to the constraints that the solution fit the data and that any other necessary restrictions be taken into account. If we return briefly to a linear inverse problem and choose a *linear* functional of the model, we discover that there is no upper or lower limit to this value of the functional unless it happens to be made up of a linear combination of data kernals (leading to Backus-Gilbert theory). But when a nonlinear functional of m is used, bounds can often be discovered. This remark suggests that property extremization is a process that must contain some nonlinear element to succeed. The most satisfying application of the method so far has been to the inverse problem of determining the shape of a buried body of known density from the gravity anomalies it produces, a nonlinear problem mentioned several times already. My solution (Parker 1975) depends upon discovering the body with the smallest possible maximum density fitting the observations, which is a nonlinear functional of the model in a linear problem. An additional constraint is frequently needed, namely, the density contrast should be positive. It turns out that the body of least density is uniform and its shape is unique, i.e. there is only one body with that density fitting the data.[9] These two properties enable us to find restrictions on the shape of a uniform body if we now assume the density is known but the shape is not. For example, one can give a level, above which any buried uniform body with the specified density must protrude or another level which it must penetrate from above. Similarly, constraints on lateral extent can be obtained. Very

[9] Here is an example of an inverse problem where a finite number of data are compatible with only one solution, but of course this happens only when the buried body has a very special shape and the density is a very special value.

recently, Sabatier (1976) has shown that my approach, which seemed very specific to this one problem, is in fact an example of a problem that can be dealt with by means of convex analysis and linear programming theory.

The foregoing theory centered upon finding the minimum of a nonlinear functional in a linear forward problem. In principle there is no reason why nonlinear forward problems cannot be treated directly, as I have shown (Parker 1972). Now the Fréchet derivatives D_i enter in place of the linear kernels G_i; when this happened during linearization an approximation was involved, but here none is made. Unfortunately, as with nearly all nonlinear analysis, the resultant equations are difficult to solve even numerically and it is often almost impossible to show rigorously that a global extremum has been found, not merely a local one. Both of these troubles diminish if only a very small number of observations are available.

The place where parameter extremization may make the greatest impact in future is in the testing of hypotheses. Often a physical model of a geological system is arrived at by extrapolation or the synthesis of circumstantial evidence; sometimes (but not as often as one would wish perhaps) the model can be tested against geophysical data. Naturally the test cannot prove the correctness of the model, but it may be able to reject it or choose between two competing hypotheses. One powerful way of making a test is to use the physical model to predict, say, a lower limit on a property such as the mean seismic velocity in a certain depth range or the maximum value of S-wave attenuation. Then the solution to the appropriate inverse problem is found to maximize the chosen parameter; if the greatest value is smaller than that predicted, the model must be rejected; if it is within the predicted range there is no conflict with the geophysical data.

A recent unusual example of this approach has been given by Gubbins (1975), who, using the spherical harmonics of the geomagnetic field, derives a lower limit on the energy requirements of the geomagnetic dynamo to test a mechanism invoking earthquakes to maintain the field. He then compares this with the available energy. Unfortunately the test was inconclusive because of the uncertainty in that figure, but a stronger test involving gravity anomalies (but not using parameter extremization) was more decisive in ruling against the model.

Another ingenious application is that of Jordan (1975): he uses the observed travel-time differences of shear waves between continent and ocean and the Love wave dispersion data over corresponding pure oceanic and continental paths to argue for continental influences as deep as 400 km into the mantle. In this case the parameter minimized is the depth to which velocity differences are permitted between the two regimes. Jordan finds that with limits on only two data (the mean travel time difference in the heterogeneous region and the difference in Love wave dispersion at about 180 s) the smallest allowed layer thickness is 310 km. Additional data force even larger values.

CONCLUSIONS

Indirect measurements will continue to be the major source of information about the Earth's deep interior and about much shallower regions also. The problem of

providing some model consistent with the data has for practical purposes been solved by the use of the spectral expansion (or its finite-dimensional equivalents) and linearization if the equations are nonlinear. However, most geophysicists are now conscious of their obligation to assess the range covered by all satisfactory solutions. In the linear case the elegant concept of resolution and its inverse relation with statistical error has made the Backus-Gilbert analysis a most appealing treatment. With nonlinear equations the two general techniques, linearization and the Monte Carlo search, suffer from drawbacks that, to me at least, make them somewhat unattractive.

One of the most promising alternative strategies is to use the data directly to test a hypothesis. This requires the construction of only one solution, the one violating the hypothesis the least, yet still satisfying the demands of observation. As yet few geophysicists have tried the idea, but it offers a means of bypassing the difficult problem of completely characterizing an infinite set of models.

Literature Cited

Anderssen, R. S., Worthington, M. H., Cleary, J. R. 1972. Density modelling by Monte Carlo inversion—I Methodology. *Geophys. J. R. Astron. Soc.* 29:433–44

Anderssen, R. S. 1975. On the inversion of global electromagnetic induction data. *Phys. Earth Planet. Inter.* 10:292–98

Backus, G. E. 1968. Application of a nonlinear boundary-value problem for Laplace's equation to gravity and geomagnetic intensity surveys. *Q. J. Mech. Appl. Math.* 21:195–221

Backus, G. E. 1970. Non-uniqueness of the external geomagnetic field determined by surface intensity measurements. *J. Geophys. Res.* 75:6339–341

Backus, G. E. 1970a. Interference from inadequate and inaccurate data, I. *Proc. Nat. Acad. Sci. USA* 65:1–7

Backus, G. E. 1970b. Interference from inadequate and inaccurate data, II. *Proc. Nat. Acad. Sci. USA* 65:281–87

Backus, G. E., Gilbert, F. 1967. Numerical application of a formalism for geophysical inverse problems. *Geophys. J. R. Astron. Soc.* 13:247–76

Backus, G. E., Gilbert, F. 1968. The resolving power of gross Earth data. *Geophys. J. R. Astron. Soc.* 16:169–205

Backus, G. E., Gilbert, F. 1970. Uniqueness in the inversion of inaccurate gross Earth data. *Phil. Trans. R. Soc. London Ser. A* 266:123–92

Bailey, R. C. 1970. Inversion of the geomagnetic induction problem. *Proc. R. Soc. London* 315:185–94

Bailey, R. C. 1973. Global geomagnetic sounding—methods and results. *Phys.*

Earth Planet. Inter. 7:234–44

Barcilon, V. 1976. A discrete model of the inverse Love wave problem. *Geophys. J. R. Astron. Soc.* 44:61–76

Bard, Y. 1974. *Nonlinear Parameter Estimation.* New York: Academic. 341 pp.

Bessonova, E. N., Fishman, V. M., Ryaboyi, V. Z., Sitnikova, G. A. 1974. The tau method for the inversion of travel times—I Deep seismic sounding data. *Geophys. J. R. Astron. Soc.* 36:377–98

Bullard, E. C., Cooper, R. I. B. 1948. The determination of the masses necessary to produce a given gravitational field. *Proc. R. Soc. London A* 194:332–47

Bullen, K. E. 1965. *An Introduction to the Theory of Seismology.* Cambridge Univ. Press. 381 pp.

Burkhard, N., Jackson, D. D. 1976. Application of stabilized linear inverse theory to gravity data. *J. Geophys. Res.* 81:1513–32

Cagniard, L. 1953. Basic theory of the magneto-telluric method of geophysical prospecting. *Geophysics* 18:605–35

Cain, J. C. 1971. Geomagnetic models from satellite surveys. *Rev. Geophys. Space Phys.* 9:259–73

Dorman, L. M. 1975. The gravitational edge effect. *J. Geophys. Res.* 80:2949–50

Dorman, L. M., Lewis, B. T. R. 1972. Experimental isostasy (3). *J. Geophys. Res.* 77:3068–71

Franklin, J. N. 1970. Well-posed stochastic extension of ill-posed problems. *J. Math. Anal. Appl.* 31:682–716

Gel'fand, I. M., Levitan, R. M. 1955. On the determination of a differential equation

by its spectral function. *Am. Math. Soc. Transl. Ser.* 2. 1:253–304

Gerver, M., Markusevitch, V. 1966. Determination of seismic wave velocity from the travel-time curve. *Geophys. J. R. Astron. Soc.* 11:165–73

Gilbert, F. 1971. Ranking and winnowing gross Earth data for inversion and resolution. *Geophys. J. R. Astron. Soc.* 23:125–28

Gilbert, F. 1976. Differential kernels for group velocity. *Geophys. J. R. Astron. Soc.* 44:649–60

Gilbert, F., Dziewonski, A. M. 1975. An application of normal mode theory to the retrieval of structural parameters and source mechanisms from seismic spectra. *Phil. Trans. R. Soc. London A* 278:187–269

Gubbins, D. 1975. Can the Earth's magnetic field be sustained by core oscillations? *Geophys. Res. Lett.* 2:409–12

Hadamard, J. 1902. Sur les problèmes aux dérivées partielles et leur significations physiques. *Bull. Univ. Princeton* 13:1–20

Hobbs, B. A. 1973. The inverse problem of the moon's electrical conductivity. *Earth Planet. Sci. Lett.* 17:380–84

Jackson, D. D. 1972. Interpretation of inaccurate, insufficient and inconsistent data. *Geophys. J. R. Astron. Soc.* 28:97–110

Jackson, D. D. 1973. Marginal solutions of quasi-linear inverse problems in geophysics: the edgehog method. *Geophys. J. R. Astron. Soc.* 35:121–36

Johnson, L. E., Gilbert, F. 1972. Inversion and inference for teleseismic ray data. *Methods of Computational Physics,* 12:231–66. New York: Academic

Jordan, T. H. 1975. The continental tectosphere. *Rev. Geophys. Space Phys.* 13:1–12

Jupp, D., Vozoff, K. 1975. Stable iterative methods for the inversion of geophysical data. *Geophys. J. R. Astron. Soc.* 42:957–76

Keilis-Borok, V. I., Yanovskaja, T. B. 1967. Inverse problems of seismology (structural review). *Geophys. J. R. Astron. Soc.* 13:223–34

Kennett, B. L. N. 1976. A comparison of travel-time inversions. *Geophys. J. R. Astron. Soc.* 44:517–36

Lanczos, C. 1961. *Linear Differential Operators.* London: van Nostrand. 564 pp.

Lavrentiev, M. M. 1967. *Some Improperly Posed Problems of Mathematical Physics.* Heidelberg: Springer. 72 pp.

McKenzie, D., Parker, R. L. 1974. Plate tectonics in ω-space. *Earth Planet. Sci.*

Lett. 22:285–93

McMechan, G. A., Wiggins, R. A. 1972. Depth limits in body wave inversions. *Geophys. J. R. Astron. Soc.* 36:239–43

Oldenburg, D. W. 1976. Calculations of Fourier transforms by the Backus-Gilbert method. *Geophys. J. R. Astron. Soc.* 44:413–31

Ortega, J. M., Rheinboldt, W. C. 1970. *Iterative Solution of Nonlinear Equations in Several Variables.* New York: Academic. 572 pp.

Parker, R. L. 1970. The inverse problem of electrical conductivity in the mantle *Geophys. J. R. Astron. Soc.* 22:121–38

Parker, R. L. 1972. Inverse theory with grossly inadequate data. *Geophys. J. R. Astron. Soc.* 29:123–38

Parker, R. L. 1975. The theory of ideal bodies for gravity interpretation. *Geophys. J. R. Astron. Soc.* 42:315–34

Parker, R. L., Huestis, S. P. 1974. The inversion of magnetic anomalies in the presence of topography. *J. Geophys. Res.* 79:1587–93

Press, F. 1968. Earth models obtained by Monte Carlo inversion. *J. Geophys. Res.* 73:5223–34

Rao, C. R. 1973. *Linear Statistical Inference and its Applications.* New York: Wiley. 625 pp.

Richards, P. G. 1975. Theoretical seismology. *Rev. Geophys. Space Phys.* 13:295–98

Sabatier, P. C. 1971. Comparative evolution of inverse problems. *NASA Tech. Memo., X-62, 150*

Sabatier, P. C. 1974. Remarks on approximate methods in geophysical inverse problems. *Proc. R. Soc. London A* 337:49–71

Sabatier, P. C. 1976. Positivity constraints in linear inverse problems. *Geophys. J. R. Astron. Soc.* In press

Siebert, M. 1964. Ein Verfahren zur un Mittelbaren Bestimmung der Vertikalen Leitfähigkeitsverteilung im Rahmen der Erdmagnetischen Tiefensonderung. *Nachr. Akad. Wiss. Göttingen Math. Phys. Kl.* 2

Smith, R. A. 1961. A uniqueness theorem concerning gravity fields. *Proc. Cambridge Phil. Soc.* 57:865–70

Stern, D. P., Bredekamp, J. H. 1975. Error enhancement in geomagnetic models derived from scalar data. *J. Geophys. Res.* 80:1776–82

Twomey, S. 1963. On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature. *J. Assoc. Comput. Mach.* 10:97–101

Weidelt, P. 1972. The inverse problem of geomagnetic induction. *Z. Geophys.* 38: 257–89

Weertman, J. 1965. Relationship between displacements on a free surface and the stress on a fault. *Bull. Seismol. Soc. Am.* 55:945–53

Weston, V. H. 1972. On the inverse problem for a hyperbolic dispersive partial differential equation. *J. Math. Phys.* 12: 1952–56

Wiggins, R. A. 1972. The general linear inverse problem: implication of surface waves and free oscillations for Earth structure. *Rev. Geophys. Space Phys.* 10: 251–85

Wiggins, R. A., McMechan, G. A., Toksöz, M. N. 1973. Range of Earth-structure non-uniqueness implied by body wave observations. *Rev. Geophys. Space Phys.* 11: 87–113

CONTENTS