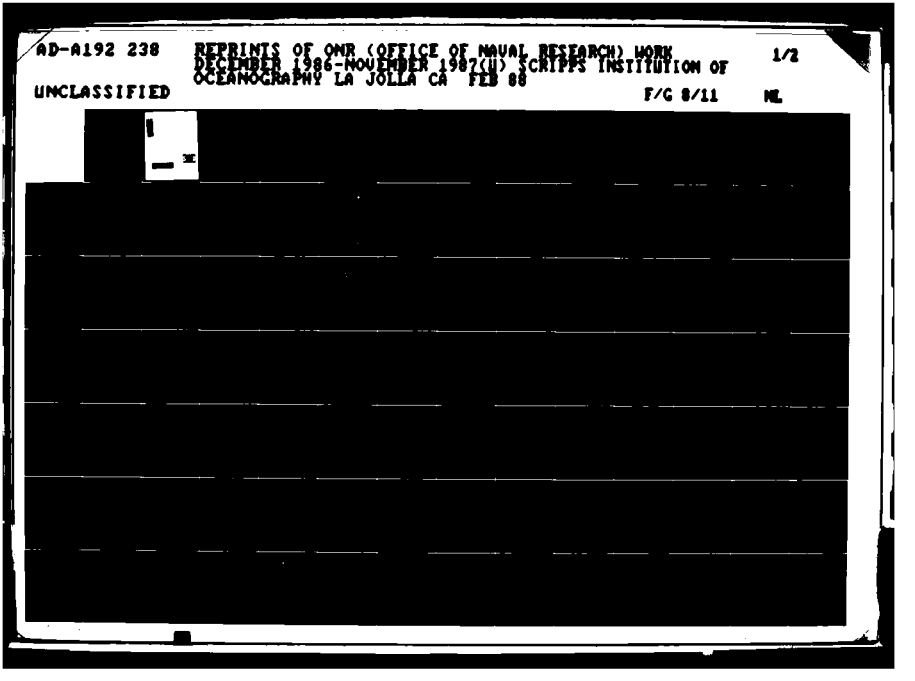
AD-A192 238 REPRINTS OF ONR (OFFICE OF NAVAL RESEARCH) WORK DECEMBER 1986-NOVEMBER 1987(U) SCRIPPS INSTITUTION OF OCEANOGRAPHY LA JOLLA CA FEB 88 1/2

UNCLASSIFIED F/G 8/11 NL



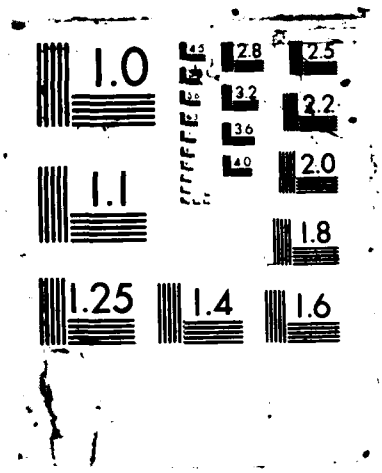

1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

TABLE OF CONTENTS

REVIEWS OF GEOPHYSICS, VOL. 25, NO. 1, PAGES 17 40, FEBRUARY 1987

# The Application of Inverse Theory to Seamount Magnetism

ROBERT L. PARKER

*Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego*

LOREN SHURE[1]

*Woods Hole Oceanographic Institute, Woods Hole, Massachusetts*

JOHN A. HILDEBRAND

Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego

The traditional least squares method for modeling seamount magnetism is often unsatisfactory because the models fail to reproduce the observations accurately. We describe an alternative approach permitting a more complex internal structure, guaranteed to generate an external field in close agreement with the observed anomaly. Potential field inverse problems like this one are fundamentally incapable of a unique solution, and some criterion is mandatory for picking a plausible representative from the infinite-dimensional space of models all satisfying the data. Most of the candidates are unacceptable geologically because they contain huge magnetic intensities or rapid variations of magnetization on fine scales. To avoid such undesirable attributes, we construct the simplest type of model: the one closest to a uniform solution as measured by the norm in a specially chosen Hilbert space of magnetization functions found by a procedure called seminorm minimization. Because our solution is the most nearly uniform one we can say with certainty that any other magnetization satisfying the data must be at least as complex as ours. The theory accounts for the complicated shape of seamounts, representing the body by a covering of triangular facets. We show that the special choice of Hilbert space allows the necessary volume integrals to be reduced to surface integrals over the seamount surface, and we present numerical techniques for their evaluation. Exact agreement with the magnetic data cannot be expected because of the error of approximating the shape and because the measured fields contain noise of crustal, ionospheric, and magnetospheric origin. We examine the potential size of the various error terms and find that those caused by approximation of the shape are generally much smaller than the rest. The mean magnetization is a vector that can in principle be discovered from exact knowledge of the external field of the seamount; this vector is of primary importance for paleomagnetic work. We study the question of how large the uncertainty in the mean vector may be, based on actual noise, as opposed to exact, data; the uncertainty can be limited only by further assumptions about the internal magnetization. We choose to bound the rms intensity. In an application to a young seamount in the Louisville Ridge chain we find that remarkably little nonuniformity is required to obtain excellent agreement with the observed anomaly while the uniform magnetization gives a poor fit. The paleopole position of ordinary least squares solution lies over 30° away from the geographic north, but the pole derived from our seminorm minimizing model is very near the north pole as it should be. A calculation of the sensitivity of the mean magnetization vector to the location of the magnetic observations shows that the data on the perimeter of the survey were given the greatest weight and suggests that enlargement of the survey area might further improve the reliability of the results.

CONTENTS



## 1. INTRODUCTION

Ever since C. Darwin deduced the general subsidence of the seafloor from his observations of coral atolls, seamounts have been valuable sources of information about marine geology. The existence of chains of seamounts is the best evidence for deep hot spots stationary in the mantle [*Morgan*, 1971], and the regular lateral spacing of these rows may indicate the presence of longitudinal convection cells [*Richter and Parsons*, 1975]. The statistics of seamount distribution has been used to provide information on the variability of tectonic activity throughout recent geologic past [*Batiza*, 1982]. The weight of the seamount is a load that deforms the ocean crust, and analysis of the bathymetry of the seafloor around a seamount yields estimates of the strength of the oceanic lithosphere [*McNutt and Menard*, 1978] that contributes to our understanding of the thermal evolution of lithospheric plates [*Watts et al.*, 1980].

The first quantitative geophysical studies of seamounts concerned their magnetism. *Vacquier* [1962] developed a method for calculating an average magnetization vector using observations of the magnetic field anomaly and the bathymetric contour of the seamount. This method approximated the seamount body with rectangular prisms

[1]Currently with Information Research Incorporated, Waltham, Massachusetts.



Paper number 6R0757.
8755-1209/87/006R-0757$15.00

and assumed that all the prisms were of uniform magnetization in order to calculate the average magnetization with the minimum least squares error in the fit of the anomaly. *Talwani* [1965] modified the least squares method by representing the body as a collection of horizontal polygonal laminas whose outlines followed the contours of the body, and *Plouff* [1976] refined this method by increasing the accuracy of the integration in the vertical direction. The results from the least squares modeling program were used for paleomagnetic study of seamounts [*Uyeda and Richards*, 1966; *Richards et al.*, 1967; *Vacquier and Uyeda*, 1967], and as soon as the ideas of seafloor spreading became accepted, such analyses were applied to the unraveling of the history of the ocean basins [*Francheteau et al.*, 1970; *Harrison et al.*, 1975; *Gordon and Cox*, 1980; *Sager*, 1983]. Despite widespread use and some suggestive findings, the results of this program have been rather disappointing. The scatter of paleopoles derived from apparently homogeneous groups of seamounts is often large, and the portion of the magnetic anomaly accounted for by the model is often small.

We believe the generally unsatisfactory performance is due in large measure to limitations of the least squares method of analysis of the magnetic data. The interior of the seamount is assumed to be uniformly magnetized down to its base, the plane of the surrounding seafloor, and below this level the magnetization is the same as in the surrounding oceanic crust. In its barest form, the least squares model contains just three unknown parameters: the three components of the magnetization vector. It is generally necessary to include as unknowns the parameters of a local background field varying linearly across the survey region to correct for small errors in the computation of the anomaly from total field measurements; this increases the number of unknowns to six. Linearity of the relation between magnetization and the observed magnetic anomalies permits parameter estimation by a least squares solution of the overdetermined equations connecting the model to the data. The underlying idea is that any discrepancy between the predictions of the model and the observed anomalies originates from uncorrected diurnal variations, crustal magnetic fields, and so on that may be treated as random, independent noise sources. Under these circumstances the Gauss-Markov theorem justifies the application of the least squares formulation by its assertion that the true uniform magnetization will be recovered if enough data are used. Yet when the predicted field is subtracted from the observed one, the residual never has the form of a random, uncorrelated noise signal as it must if the conditions of the theorem are to be satisfied. There are systematic, large-scale residual fields concentrated around the seamount that remain unaccounted for by the uniform model. This may happen even when the seamount exhibits an anomaly of the simplest form, with just one maximum, or when it is situated on crust so young that no reversal can have occurred in the history of the body. It has long been recognized that the pattern in the anomaly residual is indisputable evidence for a more complex internal magnetization.

Several investigators have inferred variations in the strength of internal magnetization using the pattern of magnetic field anomaly residuals. *Richards et al.* [1967] and *Harrison* [1971] observed that short-wavelength residuals centered over the seamount top could be an indication of nonmagnetic rocks capping the seamount. They modeled this by eliminating the uppermost bathymetric layers from the model and were able to increase the correspondence between the model predictions and the observations. *Emilia and Massey* [1974] confirmed this result by allowing their inversion program to vary the magnetization amplitude of the model for each layer in the seamount, although they found their method was unstable if too many independent layers were used. *Schimke and Bufe* [1968] obtained a magnetization for the cap of Chautauqua Seamount by inverting the residual anomaly calculated for the whole seamount. The sum of the cap and the whole seamount magnetization indicated that the cap may be more weakly magnetized than the remainder of the seamount body [*Francheteau et al.*, 1968]. *Blakely and Christiansen* [1978] used the pattern of magnetic residuals at the Mount Shasta Volcano to delineate lateral variations in internal magnetization. They observed that the western portion of Mount Shasta may have greater magnetization than its eastern portion and concluded that nonuniform magnetization could lead to erroneous paleomagnetic poles using the least squares method. Likewise, *Kodama and Uyeda* [1979] used magnetic field inversion to deduced that the eastern portion of Oshima Volcano may have lower magnetization than the rest of the body. To explain this pattern, they presented geological evidence for an older volcanic edifice hidden beneath the eastern part of the volcano.

Other workers have attempted to account for seamount magnetic anomaly residuals by assuming that portions of the seamount contain both normal and reversely magnetized rocks. *Sager et al.* [1982] assumed the top kilometer of Nagata Seamount was of reversed polarity, opposite in direction and equal in magnitude to the remainder of the body. They divided the body into normal and reversed sections by introducing a negative volume for the assumed reversed portion of the body, resulting in an improved goodness of fit between the calculated and the observed anomalies. Likewise, for a collection of seamounts on the Cocos plate, *McNutt* [1986] used a modified least squares method allowing for solution of up to nine distinct regions of magnetization. The number and the location of the magnetically distinct regions were specified before inversion, and in two cases the seamounts appeared to have regions of both normal and reversed polarity. Naturally, including more degrees of freedom in the models improves the fit, but the significance of any conclusions obtained is questionable in view of the arbitrariness involved in the subdivision procedure.

Another approach to removing the effect of nonuniform magnetization is to smooth the magnetic field anomaly before attempting inversion for the magnetization. This approach has been used for complicated magnetic anomalies where short-wavelength components may be imposed on a longer-wavelength anomaly. The justifying assumption for smoothing is that the volume of rocks creating the short wavelength anomalies is small in comparison to the volume creating the long-wavelength anomaly. *Miles and Roberts* [1981] used an orthogonal profile technique to smooth the magnetic anomaly of Rosemary

Bank Seamount before inverting for the least squares magnetization. *Sager* [1984] used an upward continuation algorithm to smooth the magnetic field anomaly of Abbott Seamount. He argued that the advantage of upward continuation is that it is physically equivalent to increasing the height of the plane of the magnetic observations, and therefore it gives a valid estimate of the magnetization amplitude. *Gardner et al.* [1984] used the same procedure to continue upward the magnetic anomaly of Shimada Seamount by 4 km before attempting inversion with the least squares method. A set of field data is not associated with a unique upward continuation, and so numerical upward continuation involves its own (often unstated) assumptions about the true magnetic field. The differences between the various upward continued versions of the data correspond exactly to the differences in the ascribed magnetizations of the various models; the inherent ambiguity cannot be avoided by this kind of preprocessing.

It is clear that the presence of mixed polarity in a seamount may be a source of error in the calculation of its paleomagnetic pole. To demonstrate this, *Lumb et al.* [1973] constructed the synthetic magnetic anomaly for a seamount model with mixed polarity and showed that an inaccurate estimate of the paleopole was obtained if a homogeneous magnetization was assumed. They consequently used mixed polarity magnetization to explain the scatter of paleopoles obtained from inversion of the magnetic anomalies of the Cook Islands. In contrast, *Sager* [1983] argued that seamounts of mixed polarity may yield accurate paleopoles if one polarity clearly dominates the body and the effect of the other polarity is removed by upward continuation. This is probably true for Abbott Seamount [*Sager*, 1984] because of the agreement between its magnetic paleolatitude and the latitude of the Hawaiian hot spot. However, there is a significant discrepancy between the paleopole calculated for Shimada Seamount [*Gardner et al.*, 1984] and the pole position near the north pole implied by its young age. There are several cases of disagreement between the magnetization inferred from least squares magnetic field inversion and the magnetization measured from rock samples. At Cobb Seamount, *Merrill and Burns* [1972] reported difficulty in reconciling the paleopole obtained from magnetic field inversion and that measured from summit rock samples. Similarly, the magnetic field inversion for Suiko Seamount [*Kodama et al.*, 1978] yields a paleolatitude significantly different from that measured for rocks obtained from Suiko during Deep Sea Drilling Project leg 55 [*Kono*, 1977]. Similarly, the magnetic field inversion for the Oshima Volcano [*Kodama and Uyeda*, 1979] provides a magnetization direction that differs from paleomagnetic measurements taken on surface volcanic rocks.

These inconsistencies point to the need for a more general magnetization model. The fundamental difficulty facing anyone who wishes to introduce a more complex structure is the nonuniqueness of the inverse problem. Even when a magnetic field caused by internal magnetization is known exactly at every point outside the seamount, there are infinitely many other magnetizations generating precisely the same exterior field. To get some idea of how large a family of models is compatible with every exterior field, consider $f$, an arbitrary continuously differentiable function, vanishing on $\partial V$, the boundary of the seamount. If a magnetization vector is defined by $\mathbf{M} = \nabla f$, it is easily shown by application of Gauss' theorem that the exterior field associated with $\mathbf{M}$ is identically zero. Thus if $\mathbf{M}_0$ generates a particular exterior field, $\mathbf{M} + \mathbf{M}_0$ will cause an identical one for every $f$ of the specified form. Therefore, on the basis of the magnetic field data alone, it is impossible to distinguish between an enormous variety of different models. To overcome this basic ambiguity, some restriction must be introduced from our knowledge of geology and geophysics to limit the amount of variability. This is just what the uniform magnetization assumption does in a heavy-handed way.

Because the magnetic anomaly data are consistent with magnetizations of infinite complexity, we must try to avoid being misled into believing some accidental feature of a model is truly demanded in the solution. Our approach in this paper is to construct the magnetization model matching the data that is as close as possible to a uniform model. The details of what we mean by "closeness" and the techniques for achieving the desired objective will be discussed in the next section. Having found the model with the minimum amount of nonuniformity, we know the true internal magnetization must possess that degree of nonuniformity, or more. This is useful information because it tells us something about the complexity of the body.

Nonuniform seamount magnetization may be produced by several factors: (1) the duration of seamount volcanism, (2) the variety of seamount rock types with different magnetization characteristics, and (3) the structural complexity of seamounts. The duration of seamount volcanism is not well known, but estimates range from as short as a few hundred thousand years [*Duncan and McDougall*, 1976] to as long as 10 or 20 million years [*Menard*, 1964; *McDougall and Schmincke*, 1976]. These time spans are long in comparison to the time for secular variation of the geomagnetic field, implying that the magnetization of individual seamount lava flows may be deflected by several degrees but that the average magnetization will represent an axial dipole field. The duration of seamount volcanism is short in comparison to the frequency of geomagnetic field reversals during the Cretaceous [*Kent and Gradstein*, 1985] but is long in compaison to the frequency of reversals during the last 5 million years [*Lowrie and Kent*, 1983]. The probability of spanning a field reversal during construction is therefore higher for Tertiary seamounts than for Cretaceous seamounts, and this allows Cretaceous seamounts to be more easily modeled. Seamounts with episodic or post erosional volcanism may be constructed of rocks with imprints from geomagnetic fields of different periods and locations. For example, *Rice et al.* [1980] reported that as much as 32% of Bermuda is made of mid-Tertiary sills that were intruded into a Cretaceous edifice. Additionally, for the southern Line Islands it was reported that both Eocene and Late Cretaceous volcanism are present within the seamount edifices [*Haggerty et al.*, 1982]. Seamount nonuniform magnetization may also result from the variety of rock types involved in their construction. Seamount rocks such as hyaloclastites, pillow lavas, dikes, and gabbros may differ significantly in their

magnetic properties. Hyaloclastites are relatively nonmagnetic rocks composed of ash, sand, and broken pillow rinds associated with explosive underwater volcanism [*McBirney*, 1971; *Lonsdale and Batiza*, 1980]. *Harrison* [1971] proposed that the weakly magnetized top of a seamount may be composed of hyaloclastites. Likewise, *Harrison and Ball* [1975] observed low magnetization at an exposed seamount composed predominantly of hyaloclastite tuff. Higher magnetization will be present in seamount basalts, such as pillow lavas, and dikes as well as in seamount gabbros. It is not known what fractions of the seamount bodies are composed of hyaloclastites relative to basalts or gabbros. However, it has been put forward that large amounts of these materials may be found on the flanks [*Lonsdale and Batiza*, 1980] and summits [*Batiza et al.*, 1984] of seamounts. Structural complexity may be another factor leading to nonuniform seamount magnetization. Seamounts may contain large magma chambers or conduits, and these bodies may require a few hundred thousand years to cool because of the low thermal conductivity typical of basalt [*Grossling*, 1970]. Such a body may partially cool in a polarity interval opposite to the rest of the seamount or may record magnetization changes due to geomagnetic secular variation or grain size variation. It has been proposed that tilting of the flanks of seamounts may occur as a result of inflation of magma chambers [*Staudigel and Schmincke*, 1984] resulting in changes in magnetization inclination. Likewise, dike and rift zones may have distinct magnetic signatures, and they are particularly prominent features on large seamounts and guyots.

The existence of a large family of alternative models each capable of exactly matching the magnetic anomaly certainly means that it is impossible to obtain an exact description of the interior magnetization from these data. Furthermore, the data will usually not allow us to decide unambiguously between competing geological speculations. Therefore it is important to identify any features of a model that can in principle be strictly related to the magnetic anomaly. The dipole moment of the seamount can be computed from its exterior field if this is exactly known. The dipole moment is especially useful geophysically because, after division by the volume of the seamount, it is the vector of average magnetization. This is the vector most diagnostic of the mean geomagnetic field direction during the formation of the seamount, and so it is a most important quantity for paleomagnetic research. As it happens, when we compute the most nearly uniform magnetization, the mean magnetization is automatically separated from the nonuniform remainder; it is then easy to find the paleomagnetic pole associated with the most uniform seamount. The actual dipole moment and our estimate will differ because the magnetic data are incomplete and imprecise. A key question is, How far can the true mean magnetization differ from the vector associated with our model? We develop a theory to provide the answer. We require an upper limit on the intensity of magnetization of the rocks of the seamount; otherwise, the actual difference between the actual mean magnetization and the uniform vector may be arbitrarily large. Preliminary calculations with the theory indicate that more work needs to be done on this topic because the estimated uncertainties remain disappointingly large. The focus of our current theoretical research is the refinement of the bounds on the uncertainty.

The plan of this paper is as follows. Section 2 gives the mathematical details concerning the construction of a most nearly uniform magnetization. For this problem we have chosen a Hilbert space setting in which the norm of the space is proportional to the rms magnetization. In this space the distance between two models is the norm of their difference. We decompose an arbitrary magnetization into two parts: a uniform magnetization (a vector of constant size and magnitude at every interior point of the seamount) and another, nonuniform part that may vary in magnitude and direction. The model we seek is the one that has the smallest nonuniform component and satisfies the measurements of the magnetic field anomaly. The norm of the nonuniform portion is a seminorm of the magnetization in the language of functional analysis, and so we call the modeling process seminorm minimization in contrast with many geophysical inversion techniques which are model norm minimizations. Although it is always possible in principle to obtain exact agreement between the predictions of the model and the measurements, we should allow for misfit because of noise in the measurements and approximations in the theory. Section 3 of the paper deals with the various approximations necessitated by practical calculation and measurement. The shape of the seamount cannot be represented exactly in any actual computation, so we have chosen an approximation for it in terms of an enclosing set of triangular facets on a flat base. We estimate the magnitude of the errors introduced by this approximation and show how they may be kept well below the uncertainties associated with the magnetic observations. To carry out the theory of section 2, a large number of volume integrals must be carried out over the seamount. Even with our simplified body those integrals cannot be performed in closed form, and therefore we adopt a scheme for numerical approximation. Here one of the advantages of our particular Hilbert space formulation becomes evident: the volume integrals can be transformed into surface integrals by means of Gauss' theorem. Despite this the numerical work in obtaining the necessary accuracy is great; we describe efficient numerical processes for computing the surface integrals. Section 4 treats the question of estimating the uncertainty in the uniform part of the magnetization model. We show how a knowledge of the maximum permissible intensity of magnetization can be converted into a bound on the uncertainty in the average magnetization. In section 5 the theory is applied to a seamount in the South Pacific Ocean on the Louisville Ridge seamount chain. Ordinary least squares modeling of this seamount is unsatisfactory in two ways. First, the predicted anomaly has the wrong shape and magnitude, resulting in an rms misfit of 269 nT to an anomaly with rms magnitude of less than 600 nT. Second, the calculated paleomagnetic pole position is more than 30° from the north geographic pole, a displacement most improbable for a young seamount, as this one is by the evidence of radiometric dating and its position in the Louisville chain. Application of our method overcomes both deficiencies: from magnetic field measurements on the approach to the seamount we estimate that the local

crustal fields contribute about 30 nT to an anomaly with a peak magnitude of over 1200 nT; we find the most nearly uniform model with this rms misfit. Its pole position is within 7° of the geographic north pole. The uncertainty ascribed to the pole position by the error theory appears to be much too large, and reasons for this discrepancy are put forward.

Most readers will not require a deep understanding of the mathematical derivations or of the numerical techniques that lie behind a successful application of the theory. To them we suggest the following strategy: first, go quickly through the next section to get an idea of the theoretical framework; then skip to section 5 where an actual magnetic survey is analyzed with the method. In the application we have included signposts to the earlier material in the event that the reader wishes to follow up any particular point in greater detail.

## 2. THEORY I: FINDING A MODEL

This section explains how to find the most nearly uniform magnetization within a seamount consistent with magnetic field data measured in its vicinity. The fundamental geological assumptions are that the seamount was formed by the outpouring of lavas onto a previously existing, relatively level crust and that the new material did not cause the older crust below to become strongly remagnetized. Our model seamount does not have large magnetic "roots"; instead, the significant magnetic sources lie above the level of the surrounding seafloor. Naturally, the magnetic material around the seamount and under it must contribute to the measured fields at the sea surface. These signals are noise as far as we are concerned, and we allow for them by permitting mismatch between the predictions of the model and the observations.

The first task is to solve the forward problem, that is, the calculation of the predicted magnetic fields from a known model of magnetization. Almost all marine magnetic measurements are of the total field intensity $|\mathbf{B}|$, which we shall assume have been reduced to total field anomalies by subtraction of a local, total field computed from a global field model. As we have already noted in the introduction, it may be necessary to include in our model parameters describing the variation of the global field over the survey area. Because the contribution of the seamount to the total observed field is small, the resultant anomaly is well approximated by

$$\Delta|\mathbf{B}| = \hat{\mathbf{B}}_0 \cdot \Delta\mathbf{B}$$

where $\hat{\mathbf{B}}_0$ is a unit vector in the direction of the ambient field at the site, and $\Delta\mathbf{B}$ is the field vector associated with local sources. Suppose for the moment that the entire anomaly is caused by the magnetization of the seamount; then the anomaly found at $\mathbf{r}$, the position of an observer, is just

$$\Delta|\mathbf{B}(\mathbf{r})| = \int_V \mathbf{G}(\mathbf{r}, \mathbf{s}) \cdot \mathbf{M}(\mathbf{s})\, d^3\mathbf{s} \qquad (1)$$

where $\mathbf{M}(\mathbf{s})$ is the magnetization vector at a point $\mathbf{s} \in V$ in the body and $\mathbf{G}(\mathbf{r}, \mathbf{s})$ is Green's function for the problem, namely,

$$\mathbf{G}(\mathbf{r}, \mathbf{s}) = \frac{\mu_0}{4\pi}\hat{\mathbf{B}}_0 \cdot \nabla_{\mathbf{r}}\, \nabla_{\mathbf{r}} \frac{1}{|\mathbf{r}-\mathbf{s}|} = \frac{\mu_0}{4\pi}\left[\frac{3(\mathbf{r}-\mathbf{s})\hat{\mathbf{B}}_0 \cdot (\mathbf{r}-\mathbf{s})}{|\mathbf{r}-\mathbf{s}|^5} - \frac{\hat{\mathbf{B}}_0}{|\mathbf{r}-\mathbf{s}|^3}\right] \qquad (2)$$

This function gives the field component at $\mathbf{r}$ in the direction of $\hat{\mathbf{B}}_0$ owing to an elementary dipole at $\mathbf{s}$. Let us recognize explicitly in the notation the important fact that measurements are obtained at only a finite number of places $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$. We simplify (1) as follows:

$$d_j = \int_V \mathbf{G}_j(\mathbf{s}) \cdot \mathbf{M}(\mathbf{s})\, d^3\mathbf{s} \qquad j = 1, 2, \ldots, N \qquad (3)$$

Here $\mathbf{G}_j(\mathbf{s})$ stands for $\mathbf{G}(\mathbf{r}_j, \mathbf{s})$, and $d_j$ is an abbreviation for the $j$th datum, $\Delta|\mathbf{B}(\mathbf{r}_j)|$. Equations (2) and (3) constitute a complete formal solution to the forward problem; a practical solution requires in addition efficient numerical procedures for the evaluation of the volume integrals over the complicated shape of the seamount. This question is deferred until section 3, because we shall need to evaluate other more involved functions over the same domain to solve the inverse problem.

At this point we introduce the notion of distance between two models of magnetization, so that there is a definite meaning to the idea that some models are closer together than others. A natural mathematical setting for this discussion is a normed linear vector space $X$ containing as elements all the magnetization functions that might occur inside $V$; any magnetization $\mathbf{M}$ is a single element in $X$. The distance between any two elements $\mathbf{M}$ and $\mathbf{N}$ of $X$ is the norm of their difference $\|\mathbf{M} - \mathbf{N}\|$. Equation (3) is interpreted as saying that each observation is given by a linear functional of $\mathbf{M}$. There are several normed vector spaces that might be suitable in this context. In the study of marine magnetic anomalies, it is traditional to reduce the vector-valued magnetization to a scalar function of position times a constant unit vector, in other words, to consider only magnetizations with constant direction. One might at first suppose that the restriction to unidirectional magnetization models might make it impossible to fit the data properly, particularly if the "wrong" direction were chosen, but it can be proved that such models are capable of satisfying any finite data set, no matter what direction is used. The proof follows from the linear independence of the associated representers, something established by the methods used in appendix A. Nonetheless, we believe it is important not to make restrictive assumptions about the magnetization of the seamount, and so we employ a space that allows complete freedom for the magnetization functions that are its elements.

*Parker* [1971] proposed the use in this problem of a Hilbert space, which we shall call $P$ here; elements are vector-valued functions of position $\mathbf{s} \in V$, for example, magnetizations. Technically, an element of $P$ is a certain equivalence class of functions brought into being by the completion of the space; we shall not dwell on these matters here. The inner product of the space is

$$(\mathbf{M}, \mathbf{N}) = \int_V \mathbf{M}(\mathbf{s}) \cdot \mathbf{N}(\mathbf{s})\, d^3\mathbf{s} \qquad (4)$$

The norm of an element is

$$\|\mathbf{M}\| = (\mathbf{M}, \mathbf{M})^{\frac{1}{2}} = \left[\int_V |\mathbf{M}(\mathbf{s})|^2 \, d^3\mathbf{s}\right]^{\frac{1}{2}}$$

If $\|\mathbf{M}\|$ is normalized by the square root of the volume of the seamount, it is just the rms magnetization. There are a number of reasons why a Hilbert space is a convenient choice for the linear vector space; the principal one is that optimization problems (minimization of a norm, for example) are particularly simple in such spaces because they have unique solutions, linearly related to the data in many cases.

Let us first assume that the data $d_1, d_2, \ldots, d_N$ are to be satisfied exactly and that the required uniform magnetization is already known: we call it $\mathbf{U} \in P$. Notice that $\mathbf{U}$ is not simply a vector in ordinary space; it is a vector-valued function throughout the region $V$ that is constant in magnitude and direction at every point. Equation (4) suggests that we can write (3) as an inner product:

$$d_j = (\mathbf{G}_j, \mathbf{M}) \qquad j = 1, 2, \ldots, N \tag{5}$$

This is possible because for observations outside $V$, every $\mathbf{G}_j$ has a bounded norm and is therefore a valid element of $P$. Such elements are called representers in the mathematical literature, a name preferred by the authors to the geophysical term "data kernels." Stated in the context of the Hilbert space $P$, our problem can be posed as follows: we must find $\mathbf{M}$ satisfying (5) such that $\|\mathbf{M} - \mathbf{U}\|$ is as small as possible. This corresponds to making our model as close to a particular uniform model as it can be, in the sense of the norm of $P$. To solve the problem, write the nonuniform part of the magnetization as $\mathbf{R}$:

$$\mathbf{R} = \mathbf{M} - \mathbf{U} \tag{6}$$

and take the inner product with $\mathbf{G}_j$; from (5) we have

$$d_j - (\mathbf{G}_j, \mathbf{U}) = (\mathbf{G}_j, \mathbf{R}) \qquad j = 1, 2, \ldots, N \tag{7}$$

Since everything on the left is known, the problem is to find the element $\mathbf{R}$ of smallest norm obeying a given finite collection of inner product constraints. The solution to this kind of optimization problem has appeared in the geophysical literature many times [e.g., *Backus*, 1970; *Parker*, 1977]. For completeness we sketch the derivation here. We show in Appendix A that the representers $\mathbf{G}_j$, where $j = 1, 2, \ldots, N$, are linearly independent; therefore they form the basis for an $N$-dimensional, and therefore closed, subspace of $P$, which we call $G$. The decomposition theorem for Hilbert spaces [*Luenberger*, 1969] states that any element of $P$ may be written as a sum of two parts, one lying in $G$ and the other in $G^\perp$, with the subspace of elements orthogonal to the elements of $G$ called the orthogonal complement of $G$. We decompose $\mathbf{R}$ in this way:

$$\mathbf{R} = \mathbf{S} + \mathbf{T} \tag{8}$$

where $\mathbf{S} \in G$ and $\mathbf{T} \in G^\perp$; obviously, $(\mathbf{S}, \mathbf{T}) = 0$, and then it follows that

$$\|\mathbf{R}\|^2 = \|\mathbf{S}\|^2 + \|\mathbf{T}\|^2 \tag{9}$$

When (8) is substituted into (7) we see that only the $\mathbf{S}$ component of $\mathbf{R}$ affects the fit to the data, because $(\mathbf{G}_j, \mathbf{T}) = 0$ for all $j$; hence we can choose the part of $\mathbf{R}$ lying in $G^\perp$ at will. Equation (9) shows us that we should make $\mathbf{T}$ the zero element of $P$, for this gives us the smallest norm of all. Now we must adjust $\mathbf{S}$ to obtain agreement with the data. Because the elements $\mathbf{G}_j$ form a basis for $G$,

$$\mathbf{S} = \sum_{j=1}^{N} \alpha_j \mathbf{G}_j$$

We have concluded that $\mathbf{T} = 0$, and we know from (8) that $\mathbf{R} = \mathbf{S}$; from (6), $\mathbf{M} = \mathbf{U} + \mathbf{S}$, and so the magnetization nearest $\mathbf{U}$ satisfying the data is

$$\mathbf{M}_0 = \mathbf{U} + \sum_{j=1}^{N} \alpha_j \mathbf{G}_j \tag{10}$$

The decomposition theorem has reduced the problem of finding an optimum element from a search in an infinite-dimensional space to a problem in a finite number of unknowns. All that remains to be done is to find the expansion coefficients $\alpha_j$, and the model nearest $\mathbf{U}$ has been found. This is accomplished by substituting (10) into (5): we obtain the system of linear equations

$$\sum_{j=1}^{N} \Gamma_{jk} \alpha_j = d_k' \qquad k = 1, 2, \ldots, N$$

where

$$d_j' = d_j - (\mathbf{G}_j, \mathbf{U})$$

and

$$\Gamma_{jk} = (\mathbf{G}_j, \mathbf{G}_k) \tag{11}$$

The matrix $\Gamma$ of all possible pairs of inner products of the representers is fundamental in much of the theory; it is called the Gram matrix. The representers are linearly independent, from which it follows that the Gram matrix is nonsingular [*Luenberger*, 1969], and so there is a unique solution to the linear system for the $\alpha_j$.

In reality we do not know the direction nor the intensity of the uniform magnetization that best approximates the interior magnetization: determination of $\mathbf{U}$ is one of the most important goals of our investigation. Also we must not demand precise agreement between the predictions of the theory and the observations. The complete solution to the problem will be developed in two stages: first we admit $\mathbf{U}$ to be unknown; then we allow misfit.

To determine the unknown $\mathbf{U}$, imagine making a guess for that element, solving (7) for the smallest $\mathbf{R}$, and then repeating the process for a series of different guesses. Clearly the best solution of the series would be the one that causes $\mathbf{R}$ to be smallest, for then it would be the nearest one to some uniform model in the set of guess models. To solve the general problem, we analyze this hypothetical optimization problem over the space of all possible elements $\mathbf{U}$. In fact $\mathbf{U}$ belongs to a three-dimensional subspace of $P$, because any such element can be written

$$\mathbf{U} = \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \beta_3 \mathbf{X}_3$$

where $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$ are fixed elements of $P$ representing uniform magnetizations of unit intensity in three mutually

perpendicular directions. From (10) the most nearly uniform solution takes the form

$$\mathbf{M} = \sum_{n=1}^{3} \beta_n \mathbf{X}_n + \sum_{j=1}^{N} \alpha_j \mathbf{G}_j \tag{12}$$

We now consider both the $\alpha$ and $\beta$ coefficients to be free in constructing a model that fits the data:

$$(\mathbf{G}_j, [\sum_{n=1}^{3} \beta_n \mathbf{X}_n + \sum_{k=1}^{N} \alpha_k \mathbf{G}_k]) = d_j \tag{13}$$

But in addition we want to minimize the distance squared between $\mathbf{U}$ and $\mathbf{M}$:

$$\|\mathbf{M} - \mathbf{U}\|^2 = ([\sum_{j=1}^{N} \alpha_j \mathbf{G}_j], [\sum_{k=1}^{N} \alpha_k \mathbf{G}_k]) \tag{14}$$

These expressions are more easily grasped if we take all the sums outside the inner products and introduce matrix notation:

$$A\,\beta + \Gamma\,\alpha = d \tag{13'}$$

$$\|\mathbf{M} - \mathbf{U}\|^2 = \alpha^T\,\Gamma\,\alpha \tag{14'}$$

Here $\Gamma$ is the Gram matrix and the meaning of the vectors $\alpha \in \mathbf{R}^N$, $\beta \in \mathbf{R}^3$ and $d \in \mathbf{R}^N$ is obvious; the components of the $N$ by 3 matrix $A$ are given by

$$A_{jn} = (\mathbf{G}_j, \mathbf{X}_n)$$

The matrix $A$ contains the solution to the forward problem for uniform seamounts because its elements are the magnetic fields at the observation positions due to unit uniform magnetizations in the three orthogonal axis directions; we call $A$ the Green matrix. Only this matrix is needed in the conventional least squares fitting process.

The simplest way to minimize (14') with (13') as a constraint is to introduce a set of $N$ Lagrange multipliers $\lambda_1, \lambda_2, \ldots, \lambda_N$, which we can collapse into the vector $\lambda \in \mathbf{R}^N$, and minimize the unconstrained functional

$$\alpha^T\,\Gamma\,\alpha + \lambda^T\,[A\,\beta + \Gamma\,\alpha - d]$$

over the vectors $\alpha$, $\beta$, and $\lambda$. The solution requires only elementary calculus; we eliminate $\lambda$ and then solve a linear system of equations for the coefficients, most neatly written:

$$\begin{bmatrix} \Gamma & A \\ A^T & O \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} d \\ 0 \end{bmatrix} \tag{15}$$

where $O$ is an 3 by 3 matrix of zeros and 0 is a 3-vector of zeros. The necessary and sufficient condition that (15) has a unique solution is that the projections of the elements $\mathbf{X}_n$ into the subspace $G$ spanned by the representers should be linearly independent or equivalently that

$$(\mathbf{G}_j, \sum_n \beta_n \mathbf{X}_n) = 0 \qquad j = 1, 2, \ldots, N$$

only if all $\beta_n$ are zero. Unfortunately this is not true in general: for example, if all the observations lie in the same plane as $\hat{\mathbf{B}}_0$ and the plane is a plane of symmetry of the seamount, then the condition does not hold. In such a geometry, the data contain insufficient information to decide between members of a subspace of uniform components each of which yields an equally small $\|\mathbf{R}\|$. Naturally, such highly symmetric situations will never arise in practice, although the reader may have guessed that we stumbled on such a case in tests with artificial models.

After (15) is solved the vectors $\alpha$ and $\beta$ are put into the expansion (equation (12)), and the desired most-uniform model results. We call the most nearly uniform solution $\mathbf{M}_*$ and the associated uniform and nonuniform pieces $\mathbf{U}_*$ and $\mathbf{R}_*$. From the perspective of functional analysis, the quantity we are minimizing is the norm of the projection of the magnetization onto the orthogonal complement of the three-dimensional subspace of uniform magnetizations. Thus we are performing a regularization of the problem in which $\|\mathbf{R}_*\|$ is a seminorm of $\mathbf{M}$ [*Luenberger*, 1969]. A seminorm is a functional possessing all the properties of an ordinary norm save one: $\|\mathbf{R}_*\|$ can vanish when $\mathbf{M}$ is not the zero element of $P$. The optimization problem we have solved is called seminorm minimization. One nice general property is that the part of the solution lying in the subspace penalized by the norm (the nonuniform magnetization) is orthogonal to the other part of the seminorm minimizing solution (the uniform part). Any portion of the solution lying in the subspace spanned by the $\mathbf{X}_n$ is drawn from $\mathbf{U}_*$, and therefore it is not found in $\mathbf{R}_*$ where it would only increase the seminorm unnecessarily.

A final refinement to the theory allows for some disagreement between the predictions of our model and the observations. We determine the model nearest to a uniform magnetization but fitting the data to a precision dictated by the amplitude of the noise in the observations. For computational convenience we turn to the Euclidean distance or two-norm as a measure of the misfit between model predictions and the data. The rms field arising from crustal sources not in the seamount and other extraneous signals can be estimated by examining magnetic data obtained in the survey region but far enough away from the seamount for its influence to be negligible. Another source of uncertainty arises from the approximation of the seamount's shape; we shall treat this factor in detail in the next section.

The discrepancy between observation and model prediction should be no more than the magnitude of the overall estimated uncertainty. Therefore (5) is replaced by

$$\sum_{j=1}^{N} [d_j - (\mathbf{G}_j, \mathbf{M})]^2 \leqslant S^2 \tag{16}$$

where $S/N^{\infty}$ is the estimated rms noise. This condition insures that the seamount we find will be the most uniform of all those in satisfactory accord with the magnetic field observations.

The arguments given earlier using the Decomposition Theorem apply equally well here: $\mathbf{M}$ must take the form given in (12) so that the optimization problem is reduced to finding the vectors $\alpha$ and $\beta$. In terms of these (16) becomes

$$\|A\,\beta + \Gamma\,\alpha - d\|^2 \leqslant S^2 \tag{17}$$

This constraint appears difficult because it is an inequality, but it is not hard to show that equality applies for any normal data set. The John multiplier theorem [see *Smith*,

1974, chap. 3] provides the technique for minimizing a functional subject to inequality constraints. The idea is almost identical to the more familiar method of Lagrange multipliers: the inequality constraints are appended to the functional under variation. The multiplicative factors behave as they do with equality constraints, but the sign of the multipliers is fixed, and for each one there is an additional "complementary slackness condition." In this particular problem there is just one multiplier, $\mu \geq 0$, and we minimize the functional

$$\alpha^T \Gamma \alpha + \mu \|A \beta + \Gamma \alpha - d\|^2 \tag{18}$$

over all vectors $\alpha$ and $\beta$ such that

$$\mu [\|A \beta + \Gamma \alpha - d\|^2 - S^2] = 0 \tag{19}$$

In the complementary slackness condition (equation (19)), either the factor in brackets vanishes; and then equality applies in (17), or $\mu$ is zero. If $\mu$ vanishes it is clear that (18) has its only minimum at $\alpha = 0$, since $\Gamma$ is positive definite. This corresponds to an exactly uniform magnetization. The motivation for our theory is the fact that a totally uniform body cannot fit the magnetic data to the necessary precision, and so normally $\mu$ is nonzero. We conclude, therefore, that equality is achieved in (17) so that (19) may be obeyed. The conditions of the John multiplier theorem apply at any local minimum of the functional, but because the norm is a convex functional, and the constraint (17) constitutes a convex set of points, any local minimum must also be the global minimizer of the functional by a well-known property of convexity [*Luenberger*, 1969].

It may be helpful to interpret the minimization of (18) as an intermediate problem lying between two extremes: the conventional least squares fitting by a uniform body when $\mu$ tends to zero and the construction of an exactly fitting model when $\mu$ becomes large. In the latter limit, the solution is the one that possesses the smallest nonuniform component; it has no particular relation to the smallest norm model.

To find the minimum, we differentiate (18) with respect to $\alpha$, $\beta$, and $\mu$ in the usual way; after some rearrangement the equations derived from variation of $\alpha$ and $\beta$ can be written

$$\begin{bmatrix} I/\mu + \Gamma & A \\ A^T & O \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} d \\ 0 \end{bmatrix} \tag{20}$$

which is a linear system if $\mu$ is fixed. In contrast to the situation with precisely matching data, we find that the multiplier $\mu$ cannot easily be eliminated. To find $\mu$ we must appeal to (17) taking the equality; a little algebra simplifies the condition to

$$\mu^{-2} \|\alpha\|^2 = S^2 \tag{21}$$

In outline the solution of the problem proceeds in this way: the vector $\alpha$ may be regarded as a known function of $\mu$ through the solution of (20) (assuming the symmetry of the problem allows a unique solution); then (21) is a nonlinear equation for $\mu$ which we may write

$$F(\mu) = S^2 \tag{22}$$

After some manipulation we can find an explicit expression for the derivative of $F$:

$$\frac{dF}{d\mu} = -2\mu^{-4} (H\alpha)^T [\mu\Gamma^2 + \Gamma + A (A^T H A)^{-1} A^T] (H\alpha)$$

where the matrix $H$ is defined by

$$H = (I/\mu + \Gamma)^{-1}$$

The positive definiteness of $\Gamma$ insures the same property for the factor in brackets when $\mu > 0$, from which it follows that $dF/d\mu$ is always negative. Thus the solution to (22) is unique if it exists. It may be verified that as $\mu \to \infty$, $F(\mu) \to 0$, which is consistent with the idea that this limit corresponds to the problem of finding a magnetization fitting the data exactly. The maximum $F$ occurs as $\mu \to 0$, corresponding to $\alpha = 0$ and the problem of least squares fitting a uniform model. The value of $\mu$ associated with a particular $S^2 < F(0)$ can be found by iteration with Newton's method, which always converges if the starting approximation is less than the true value of $\mu$; this follows from $d^2F/d\mu^2 > 0$, a result requiring heavy algebra to obtain. In practice, having found a value of $\mu$ that yields plausible misfits to the data, we usually sweep through a range of values in its vicinity to examine the different solutions.

With inexact fitting there is an interpretation of the coefficients $\alpha$ that has no counterpart in the analysis of precise data: the equations obtained upon variation of the functional with $\alpha$ can be expressed as

$$\alpha = \mu [d - (A\beta + \Gamma\alpha)]$$

or

$$\alpha_j = \mu [d_j - (\mathbf{G}_j, \mathbf{M})]$$

Since the term in brackets is the discrepancy between the predictions of the model and the observations, this equation says that the individual misfits to the data are each proportional to an expansion coefficient of $\mathbf{R}$ in the basis of representers.

The last matter to be dealt with in this section is the inclusion of the corrections to the ambient field to allow for inaccuracies of the main field model. We can do no better than the traditional treatment and allow three further unknown parameters that correct for the presence of a linear variation of the ambient field over the survey region. In place of (5) the theoretical prediction from the model takes the form

$$d_j = \gamma_0 + \boldsymbol{\gamma} \cdot \mathbf{r}_j + (\mathbf{G}_j, \mathbf{M}) \qquad j = 1, 2, \ldots, N$$

where $\boldsymbol{\gamma}$ is a horizontal vector (the gradient of the ambient correction) and $\gamma_0$ is the unknown offset of the ambient field from the main field model. Formally this is just an inner product on another Hilbert space $P'$ whose elements consist of ordered triples $[p; \mathbf{q}; \mathbf{F}]$ where $p$ is a real number, $\mathbf{q}$ is a horizontal vector, and $\mathbf{F}$ is an element in $P$. Then the associated inner product for $\mathbf{F}', \mathbf{G}' \in P'$ is

$$(\mathbf{F}', \mathbf{G}')' = ([p; \mathbf{q}; \mathbf{F}], [s; \mathbf{t}; \mathbf{G}])'$$
$$= ps + \mathbf{q} \cdot \mathbf{t} + (\mathbf{F}, \mathbf{G})$$

The development proceeds in exactly the same way in the new space when the representers for the anomaly data are chosen to be

$$\mathbf{G}_j' = [1; \mathbf{r}_j; \mathbf{G}_j]$$

and the origin of coordinates is in the sea surface, so that $\mathbf{r}_j$, the position vector of the $j$th observation point, is horizontal. In $P'$ the completely uniform seamount is one of the form $[\rho, \mathbf{q}, \beta_1\mathbf{X}_1 + \beta_2\mathbf{X}_2 + \beta_3\mathbf{X}_3]$; then all the equations derived earlier apply without change. Now the coefficient vector in the matrix equations $\beta$ lies in $\mathbf{R}^6$ while $\alpha$ is unchanged in dimension and meaning.

## 3. Approximations

In this section we explain the nature of the approximations and show that the errors introduced by them can be kept to acceptably low levels in comparison to other sources of uncertainty in the data. The first question we address is the calculation of the Gram matrix elements. The definition of the representers from (2), together with the inner product of $P$ from (4) and the Gram matrix from (11), gives an explicit expression for an element of $\Gamma$:

$$\Gamma_{jk} = \left(\frac{\mu_0}{4\pi}\right)^2 \int_V \left[\hat{\mathbf{B}}_0 \cdot \nabla \nabla \frac{1}{|\mathbf{r}_j - \mathbf{s}|}\right] \cdot \left[\hat{\mathbf{B}}_0 \cdot \nabla \nabla \frac{1}{|\mathbf{r}_k - \mathbf{s}|}\right] d^3\mathbf{s} \qquad (23)$$

Here $\nabla$ means $\nabla_s$, differentiation with respect to $\mathbf{s}$; we have used the fact that

$$\nabla_s \frac{1}{|\mathbf{r} - \mathbf{s}|} = -\nabla_r \frac{1}{|\mathbf{r} - \mathbf{s}|}$$

After we have taken advantage of the symmetry of $\Gamma$ there are $N(N+1)/2$ integrals like this to be evaluated; a similar set of $3N$ integrals is required to find the elements of the Green matrix $A$. Because $N$ is typically 100 or more, many thousands of integrals must be carried out. Several approximations are required in order to calculate these numbers.

The region $V$ in (23) is the set of points defining the seamount. We cannot know the exact shape of the bottom boundary of $V$ where the newer lavas of the volcanic body lie in contact with the original crust. We approximate this surface by a horizontal plane at the mean level of the surrounding terrain. The upper surface of the seamount is known in considerable detail, but it too must be approximated in our calculations because it is so complex. It is important that the approximation can be formed from samples of the bathymetry not disposed in any regular manner because we shall show that the computationally optimal spacing of bathymetry samples depends directly on the water depth: topography in shallow water should be sampled more densely than that in deep water. Our approach has been to represent the surface of the seamount by a tessellation of triangular facets: the facets are the plane interpolations of sample points drawn from the bathymetric data. As suggested by Figure 1, there are many ways in which a given arbitrary collection of points in the plane may be connected together to form nonoverlapping triangles, and each of these yields a different interpolation. In appendix B we show that the rms error of the interpolation can be deduced from the power spectrum of the topography. For young seafloor, *Fox and Hayes* [1985] find that a power law is a good description of the spectrum over a large range of wave numbers, and as proved in the appendix, this leads to the following expression for the rms interpolation error averaged over a triangular facet $T$:

$$\delta^2_T = \eta^2 (\Delta / l_0)^{1-\eta} \Theta(T) \qquad (24)$$

where $\Delta$ is the area of $T$, $\Theta(T)$ is a factor depending on the triangle's shape, and $\eta$, $\eta'$, and $l_0$ are constants, values computed from the analysis of Fox and Hayes are $\eta = 2.48$ and $\eta' = 15.5$ m when $l_0$, which is an arbitrary length scale, is set to 1 km. The shape factor $\Theta(T)$, given by equation (B10), is contoured in Figure 2. For a fixed area, the equilateral triangle produces the least error with $\Theta_{min} = 0.320$, but any triangle whose angles all exceed 20° is associated with an error only slightly larger, for then $\Theta \leq 0.575$. Table 1 gives the rms error for equilateral triangles of various sizes. Although serious errors will not be incurred unless a triangle is severely elongated, the analysis indicates that the cells of the tessellation should be chosen to be as nearly equilateral as possible. An automatic procedure for doing this was given by *Watson* [1982], the method is based upon a theorem of G. F. Delaunay stating that a triangular tessellation may be arranged so that the circumcircle of every triangle contains no vertex of any other triangle of the set. A slightly modified version of Watson's program has proved to be highly satisfactory: it is efficient, is reliable, and yields sensible tessellations, for example, the one in Figure 1*b*.

The error in interpolation of the topography discussed above is secondary to the consequent error introduced in the computed magnetic anomaly by the approximation of

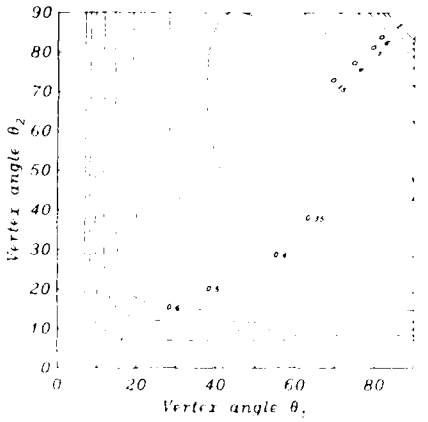



Fig. 2. Shape factor $\Theta(T)$ of equation (24) for triangles two of whose vertex angles are $\theta_1$ and $\theta_2$ in degrees. All possible triangles are covered by this diagram.

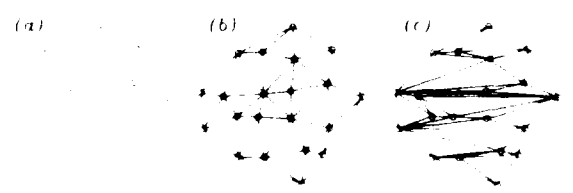



Fig. 1. (*a*) A collection of 21 points placed at random in the plane. (*b*) A triangular tessellation of them based upon *Watson*'s [1982] algorithm for the most nearly equilateral configuration. (*c*) A quite different tessellation of the same points.

TABLE 1. Rms Error of Interpolation by Equilateral Triangles

| Side, m | $\delta^2_{\ I}$, m |
|---|---|
| 50 | 0.70 |
| 100 | 1.2 |
| 200 | 2.0 |
| 300 | 2.6 |
| 500 | 3.9 |
| 1000 | 6.4 |
| 1500 | 8.7 |
| 2000 | 10.7 |
| 2500 | 12.6 |
| 3000 | 14.4 |

Root-mean-square error in topography after interpolation by an equilateral triangle with the given side length. The values are calculated from equation (24) and the constants of *Fox and Hayes* [1985].

the true surface of the seamount by an artificially smooth one. This factor will be treated as a noise term, that is, a part of the measured field not fitted by the model. We have no control over the contribution from approximation of the base of the seamount by a level plane, but we shall show that this error is not large in comparison with effects of diurnal variations and residual crustal fields. The approximation of the upper surface can be made as accurate as we desire by choosing the triangle size sufficiently small. It would be wasteful of computing resources, however, to reduce this error far below those from other sources.

The magnetic effect of the lower surface will be modeled by a thin horizontal layer located at the level of the surrounding crust with variable thickness $\delta$, the difference between the true topography and that of the model. The same kind of model is used for the upper boundary, but we choose the level to be that of the shallowest part of the seamount; this safely overestimates the noise signal. We simplify the magnetization in the layer by using a constant vector $\mathbf{M}$. The expected squared magnetic signal is found by integrating the power spectrum of the magnetic field at the sea surface:

$$\begin{aligned} \Delta B^2 &= \mathrm{E}[(\hat{\mathbf{B}}_0 \cdot \Delta\mathbf{B})]^2 \\ &= \int_{\mathbf{R}^2} S_\delta(\vec{k})|U(\vec{k})|^2\, d^2\vec{k} \end{aligned} \tag{25}$$

where $S_\delta$ is the power spectral density of the layer thickness and where $U(\vec{k})$ is the (approximately linear) transfer function between topography and surface magnetic field given by *Parker* [1973],

$$U(\vec{k}) = \pi\mu_0 e^{2\pi i(\vec{k}\cdot\vec{r}_0 - kz_0)}\hat{\mathbf{B}}_0 \cdot (|\vec{k}|\hat{\mathbf{z}} - i\vec{k})\mathbf{M} \cdot (|\vec{k}|\hat{\mathbf{z}} - i\vec{k})/|\vec{k}|$$

where $z_0$ is the depth to the thin layer and $\hat{\mathbf{z}}$ is a vertical unit vector, here we have taken only the first term of Parker's series and converted to the Fourier transform conventions set out in appendix B. The complex wave number terms in $U(\vec{k})$ achieve their largest magnitudes when $\hat{\mathbf{B}}_0$ and $\mathbf{M}$ are vertical; thus (25) gives the following bound:

$$\Delta B^2 \leqslant \int_0^\infty \pi^2\mu_0^2 e^{4\pi k z_0}|\mathbf{M}|^2 S_\delta[k]\, 2\pi k\, dk \tag{26}$$

where it has been assumed that the spectrum is isotropic (that is, depends only upon $|\vec{k}|$) and we have substituted $k$ for $|\vec{k}|$. From this equation we can estimate the error introduced by giving the seamount a flat base. In this case the effective area of the triangles is very large; thus we may replace $S_\delta[k]$ by the original isotropic spectrum of the bottom topography, $S[k]$, and choose $z_0$ to be the local mean depth of the ocean in the absence of the seamount. In appendix B we give an expression for $S[k]$ derived from the data of *Fox and Hayes* [1985], using the power law model:

$$S[k] = c_3(kl_0)^{-(\eta+1)}$$

where $c_3 = 27{,}400\ \mathrm{m}^4$ for young volcanic terrain and $\eta$ and $l_0$ are as before. Substituting into (26) we find

$$\begin{aligned} \Delta B^2 &\leqslant 2\pi^3\mu_0^2|\mathbf{M}|^2 c_3 l_0^{-4}\Gamma(3-\eta)\left(\frac{l_0}{4\pi z_0}\right)^{3-\eta} \\ &= 121\left(\frac{l_0}{z_0}\right)^{0.52}\ \mathrm{nT}^2 \end{aligned} \tag{27}$$

for $|\mathbf{M}| = 10\ \mathrm{A\ m}^{-1}$, a reasonable figure for young oceanic basalts [*Vacquier*, 1971]. Thus we estimate the rms error from neglecting the roughness of the base of the seamount ranges from about 9 nT in shallow water, say $z_0 \approx 1500$ m, to about 5 nT in water 4 km deep.

The major difficulty in using (26) for the upper surface of the seamount is in calculating the proper spectral density $S_\delta[k]$. We simplify the interpolation process by approximating it as a low-pass filter that never magnifies the original topographic signal and whose gain falls with increasing $k$. Thus we treat the residual $\delta$ as a high-passed version of the topography. From the analysis of appendix B we have one other property of the filter: the total variance of the residual topography, $\delta^2_{\ I}$, gives us the power gain. Of all the filters obeying these constraints we find the one acting on $S[k]$ that gives the largest possible variance in $\Delta B$. Thus even though we may not know the exact form of $S_\delta[k]$ we can still set an upper limit on the interpolation error. It can be shown by the application of the principles of linear programming [see *Luenberger*, 1969] that the optimal filter is a pure high-pass filter that rejects all energy below some critical wave number and passes a signal above that with constant gain. Using this result and the expression for the power spectrum of the field, we can maximize the variance with respect to the two free parameters of the filter: its gain and the cutoff wave number. After some lengthy algebra we obtain the results about to be summarized. Define a dimensionless quantity $\sigma$ by

$$\sigma = \frac{(\eta-1)\,\delta^2_{\ I} l_0^2}{2\pi c_3}\left(\frac{l_0}{z_0}\right)^{\eta-1}$$

The largest possible mean square magnetic field error due to neglect of the surface roughness is estimated to be

$$\Delta B^2_{\ \max} = \mu_0^2|\mathbf{M}|^2(\eta-1)Q(z_0)\,\frac{\delta^2_{\ I}}{16z_0^2} \qquad \sigma \leqslant \left(\frac{4\pi}{z_0}\right)^{\eta-1}$$

$$\Delta B^2_{\ \max} = \mu_0^2|\mathbf{M}|^2(\eta-1)Q(4\pi\sigma^{1/(\eta-1)})\,\frac{\delta^2_{\ I}}{16z_0^2} \qquad \sigma \geqslant \left(\frac{4\pi}{z_0}\right)^{\eta-1}$$

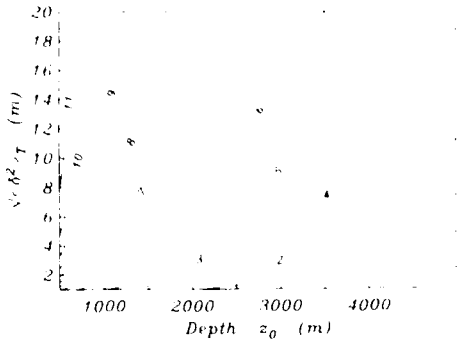



Fig. 3. Upper bound on the rms error in the magnetic anomaly due to smoothing of the surface of the seamount by interpolation. $z_0$ is the water depth over the shallowest point of the seamount. $\delta^2_I$ is the rms interpolation error. The magnetization is taken to be 10 A m$^{-1}$, and the error varies in proportion to this figure. Contour values are in nanoteslas.

where $Q(q) = q^{\eta-1}\Gamma(q, 3-\eta)$ and (in this definition only!) $\Gamma(q, a)$ is the incomplete gamma function [*Abramowitz and Stegun*, 1965, chap. 6]; $q_c$ is defined to be the positive $q$ that makes $Q(q)$ maximum, and with the value of $\eta$ that we have adopted, $q_c = 1.163$ and $Q(q_c) = 0.286$. In Figure 3 we have contoured the rms magnetic field in nanoteslas from these expressions, using the value of 10 A m$^{-1}$ for $|\mathbf{M}|$. Combining the information in Table 1 with that in Figure 3 we may conclude that depending on the water depth, triangles with sides between 500 m and 2000 m can be used without introducing errors larger than 5 to 10 nT.

The numerical evaluation of the integrals in (23) represents the greatest computational burden in practical calculations. A large computational saving is achieved by reducing the volume integrals over $V$ to integrals over the surface $\partial V$: following *Parker* [1971], we write (23) in the form

$$\Gamma_{jk} = \left(\frac{\mu_0}{4\pi}\right)^2 \int_V \nabla F_j \cdot \nabla F_k \, d^3\mathbf{s} \qquad (28)$$

where

$$F_j = \hat{\mathbf{B}}_0 \cdot \nabla \frac{1}{|\mathbf{r}_j - \mathbf{s}|}$$

which holds because $\hat{\mathbf{B}}_0$ is a constant vector. Next consider the following identity, which is valid for sufficiently smooth functions:

$$\nabla F_j \cdot \nabla F_k = \nabla \cdot (F_j \nabla F_k) - F_j \nabla^2 F_k$$

All the measurement positions lie outside the seamount; therefore $|\mathbf{r}_k - \mathbf{s}| > 0$, and so $\nabla^2 |\mathbf{r}_k - \mathbf{s}|^{-1} = 0$ for all $\mathbf{s} \in V$. Since $\nabla^2$ commutes with $\hat{\mathbf{B}}_0 \cdot \nabla$, this implies $\nabla^2 F_k = 0$, and thus

$$\nabla F_j \cdot \nabla F_k = \nabla \cdot (F_j \nabla F_k)$$

Substituting this into (28) and applying Gauss' theorem gives

$$\Gamma_{jk} = \left(\frac{\mu_0}{4\pi}\right)^2 \int_{\partial V} F_j \nabla F_k \cdot \hat{\mathbf{n}}(\mathbf{s}) \, d^2\mathbf{s} \qquad (29)$$

where $\hat{\mathbf{n}}(\mathbf{s})$ is the unit normal to the surface at the point $\mathbf{s} \in \partial V$. We remark that there is no guarantee that $\hat{\mathbf{n}}(\mathbf{s})$ exists for any surface point nor do we know if Gauss' theorem is valid for a surface whose power spectrum is like that in equation (B7); since we apply this formula to the simplified body with triangular facets, no difficulty arises, for the theorem is undoubtedly correct for regions with a piecewise plane boundary [*Kellogg*, 1953, chap. 4].

The computation of the Gram matrix has been reduced to the evaluation of (29) over the set of plane triangles by which we have approximated the surface of the body. For programming simplicity we also tessellate the lower, plane boundary with the projections of the triangles defining the upper surface. The elements of the matrix $A$ can be converted to surface integrals by an analogous process. The surface integrals for the Gram matrix cannot be performed in terms of elementary functions except in certain special cases; therefore we use a numerical integration formula, technically known as a "cubature" rule, designed for triangular regions. *Stroud* [1971] gives a complete review of this question and provides several examples of the required type. The integral of a smooth function defined for points in the triangle $T$ is approximated by a weighted sum of samples of the function:

$$\int_T f(\bar{x}) \, d^2\bar{x} \approx \sum_j w_j f(\bar{x}_j)$$

By an appropriate choice of $w_j$ and $\bar{x}_j$ it is possible to make the cubature formula exact for all polynomial functions in the plane with degree less than some upper limit, $p$; these are called degree-$p$ formulas. This is equivalent to the familiar Gaussian quadrature method on the real line. The theory for functions of more than one variable is far from complete: for example, the smallest number of sampling points that will yield a degree-$p$ rule is not known in general. It is nonetheless possible to generate a suboptimal formula by taking combinations of one-dimensional Gaussian rules in a so-called "conical product." The optimal degree-5 rule for triangular regions is known (formula $^*T_2$ of *Stroud* [1971, p. 314]): it uses seven sampling points. In contrast, the degree-5 conical product formula requires nine sampling points. With a complete theory, we would be able to choose a cubature rule with guaranteed accuracy for every surface integral but with the presently available methods this is impractical. Even for simple functions bounding the error of this kind of approximation is difficult. Furthermore, *Stroud*'s [1971, chap. 5] examples give the impression that the bounds yielded by the available methods are of the crudest kind, overestimating the true error by several orders of magnitude in almost every case. *Sard*'s [1963] theory, summarized by Stroud, provides a useful result for purposes of comparison: for an integration rule of degree $p$, the error depends principally upon the magnitude of the largest derivative of the order of $p + 1$; this rests on the assumption that the integrand possesses all these derivatives which is true for our functions because they are analytic. Roughly speaking, the integrand of (29) behaves like $\partial^2 R_j^{-1} \partial^3 R_k^{-1}$ where $\partial$ is a first-order differential operator acting in the plane of the triangle and $R_j$ is the distance from the observer $j$ to a point in the triangle (and similarly for $R_k$). Combined with Sard's result, this suggests that the largest errors occur in contributions to the

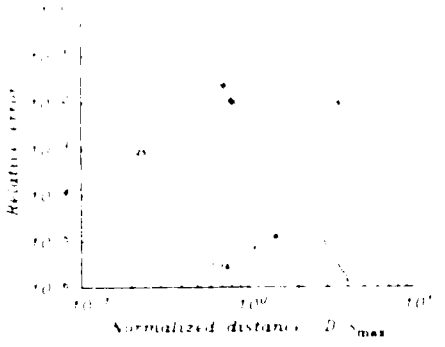


Fig. 4. Relative error as a function of distance of four cubature rules $C_7$, $C_9$, $C_{16}$, and $C_{25}$. $C_7$ is a seven point optimal degree-5 rule. The others are conical product rules of degrees 5, 7, and 9. The integrals are the contributions to the diagonal element of the Gram matrix of an equilateral facet, side $s$, and the observer over the center of the triangle at a distance $D$.

diagonal elements of $\Gamma$ by facets close to the measurement position because these are the conditions under which the highest derivatives arise.

As was noted earlier, some of the integrals in (29) can be evaluated in elementary functions; this happens when $\hat{\boldsymbol{\imath}} = \hat{\mathbf{B}}_0 = \hat{\mathbf{n}}$ and $j = k$, that is, on the diagonal of the Gram matrix. The expressions are complicated and will not be set out here. By comparing the exact and numerical values for these diagonal elements we can assess the performance of the approximation when it is likely to be at its worst. In Figure 4, we illustrate the performance of four numerical cubature rules over a range of observer-triangle distances. The integration is carried out over an equilateral triangle, and the observer position moves on a line normal to the facet passing through the centroid; all the distances, $D$, shown on the horizontal axis are measured to the closest point in the facet normalized by the side. The top line, marked $C_7$, gives the relative error as a function of distance when the optimal seven-point rule is used. The most important point to note is the strong dependence of the error on distance between the observation point and the facet. As is expected, the error is generally smaller the more distant the observer is. We can be more precise by using Sard's result: the formula is exact for fifth-degree polynomials, so the absolute error should be be bounded by some constant times the maximum magnitude of the sixth derivative of the integrand found in the region of integration. A short calculation involving spherical harmonics shows that for large distances the sixth derivative of the integrand should decrease as the inverse eleventh power of distance while the value itself falls as the inverse fifth power; thus the relative error ought to decrease with the inverse sixth power of distance. Our numerical calculations confirm that the error falls at this rate asymptotically. The line labeled $C_9$ shows the error resulting from the nine-point, degree-5 conical product rule; the error here also falls like the inverse sixth power. Although the error is less at every distance for $C_9$ than for $C_7$, the improvement is slight in view of the need for two more integrand evaluations. The behavior of the 16-point, conical product rule is shown by the line $C_{16}$; the rule is exact for polynomials of degree 7, and so for large $D$ the relative error drops as the inverse eighth power. Similarly, the 25-point, conical product formula is a degree-9 rule, although the behavior of the relative error curve is not a simple monotone decrease with $D$. We have experimented with a variety of triangle shapes in the comparison with the exact integration; we conclude that if the distance is normalized by the maximum length of a side, Figure 4 gives an upper bound on the relative error and so it may be used as a guide to the expected accuracy of the different cubature rules.

Based upon the foregoing discussion, we can develop a strategy for efficient numerical integration. Because the error exhibits such a strong functional dependence on the ratio of observer distance to triangle side, we should keep this ratio carefully under control. In the approximation of the seamount by triangular facets, the sides of every triangle are arranged to be shorter than the water depth over the shallowest corner times a factor; in practice we have insured that the longest side of any triangle never exceeds twice the water depth. This arrangement guarantees that the ratio $D/s_{\max}$ of Figure 4 is always greater than one half. Using the equilateral triangle data as a basis for an error model, we conclude that with the 16-point conical product rule the relative error should remain below one part in a hundred for the diagonal elements and better than that for the off-diagonal ones. The coordinates and weights for this formula are given in Table 2. To check the actual accuracy of the integration, we have tested the scheme in a few cases by applying $C_{16}$ and $C_{25}$ to the same body; this test indicates that the error estimate is conservative and that we always achieve accuracies of a few parts in a thousand.

We must understand how precise the approximation to the elements of the various matrices needs to be. We briefly consider the perturbation theory for the solution to (20) which we abbreviate by

$$B\xi = b$$

The right side represents data (magnetic anomaly values),

TABLE 2. Degree-Seven Cubature Rule $C_{16}$

| $x$ | $y$ | $w$ |
|---|---|---|
| 0.860240136 | 0.009703785 | 0.005423226 |
| 0.583590432 | 0.028912084 | 0.022584049 |
| 0.276843014 | 0.050210123 | 0.035388068 |
| 0.057104196 | 0.065466995 | 0.023568368 |
| 0.860240136 | 0.046122080 | 0.010167260 |
| 0.583590432 | 0.137419104 | 0.042339725 |
| 0.276843014 | 0.238648660 | 0.066344216 |
| 0.057104196 | 0.311164552 | 0.044185089 |
| 0.860240136 | 0.093637784 | 0.010167260 |
| 0.583590432 | 0.278990463 | 0.042339725 |
| 0.276843014 | 0.484508327 | 0.066344216 |
| 0.057104196 | 0.631731252 | 0.044185089 |
| 0.860240136 | 0.130056079 | 0.005423226 |
| 0.583590432 | 0.387497483 | 0.022584049 |
| 0.276843014 | 0.672946863 | 0.035388068 |
| 0.057104196 | 0.877428809 | 0.023568368 |

Sampling points and weights of the sixteen-point, degree-seven, conical product cubature rule on the right-angled triangle with corners (0, 0), (0, 1), and (1, 0) constructed as described by *Stroud* [1971]. The corresponding values for any other triangle may be found by a simple linear transformation.

and these are known at best to about a few parts in $10^2$. Let $\xi + \epsilon$ be the solution to a slightly perturbed data vector $b + e$, and for the moment, take $B$ to be exact. Then

$$B(\xi + \epsilon) = b + e$$

and it is well known [e.g., *Golub and Van Loan*, 1983] that the perturbation in the solution is bounded as follows:

$$\frac{\|\epsilon\|}{\|\xi\|} \leqslant \kappa(B)\frac{\|e\|}{\|b\|} \tag{30}$$

where $\kappa$ is the condition number of $B$ given by

$$\kappa(B) = \|B\|\,\|B^{-1}\|$$

Here we take the vector norm to be the ordinary Euclidean length, and we take the Frobenius norm for the matrices, the square root of the sum of elements squared. The condition number, which is never less than unity, governs the way in which perturbations in the data are magnified in the solution vector. It is impossible to predict the value of $\kappa$ from general principles although we expect it to increase with $\mu$ in (20). If the answers we obtain are to be useful, they must not be sensitive to small errors in $b$, and therefore the condition numbers of our matrices should be less than $10^2$, and we have checked this in actual examples. Now let us examine how errors in the matrix $B$ alter the solution; let the numerical approximation to the true $B$ be $B + E$ and the correspondingly corrupted solution vector be $\xi + e'$; obviously

$$(B + E)(\xi + e') = b$$

and then there is a companion result of (30) that states

$$\frac{\|e'\|}{\|\xi\|} \leqslant \kappa(B)\frac{\|E\|}{\|B\|}\left[1 - \kappa(B)\frac{\|E\|}{\|B\|}\right]^{-1}$$

provided $\kappa(B)\|E\|/\|B\| < 1$. We show that with an appropriately small relative error in the numerical integration, the right side of the above inequality is small in comparison with the right side of (30). If the relative error in cubature is always less than $\nu$,

$$|E_{ij}| \leqslant \nu|B_{ij}|$$

because in (20) the diagonal of $\Gamma$ is positive. From the definition of the Frobenius norm $\|E\|/\|B\| \leqslant \nu$. It follows that the cubature error introduces effects in the solution small in comparison to those due to data uncertainty if the relative precision is an order of magnitude smaller than the relative error in the measurements. This is why we have set the target for the level of accuracy in the cubature rule at a few parts in a thousand.

## 4. Theory II: Appraising the Solution

A complete analysis of any inverse problem includes an assessment of the reliability of the solution. This is commonly provided by an estimate of the "resolution" associated with the given set of data. In our problem that kind of study is inappropriate for two reasons. First, it requires a theory capable of dealing with a vector-valued function in three dimensions, making the display and computation of the results extremely complicated. Second, even exact values of the magnetic field measured at every exterior point are incapable of yielding a unique magnetization model which implies that the resolution of any conceivable set of data will be poor.

Recall that the volume of the seamount times the mean magnetization is just the total dipole moment of the body, and if magnetic anomaly data were available everywhere (or just on a sphere enclosing the seamount), the dipole moment could be deduced exactly from the field data. Therefore some aspects of the model can be obtained unambiguously in ideal circumstances, and we shall focus on the average magnetization, $\mathbf{U}$. The direction of the associated vector allows the calculation of the pole position of the magnetic field at the time of formation of the seamount if we assume the seamount formed rapidly enough that the motion of the tectonic plate may be neglected, and if we accept the axial dipole hypothesis [*McElhinny*, 1973, chap. 6]. The fluctuations of the secular variation are expected to average to zero over a period of formation of the seamount. The validity of the paleopole calculation might hold even if the seamount captured a polarity reversal, provided the opposite polarities are not present with exactly equal magnetic moments. Thus, instead of assessing the quality of our model at every interior point with an analysis of resolution, we seek the uncertainty in the single important property, the average magnetization vector. Since this is a linear functional of the unknown model we take up an idea of *Backus* [1970] and *Parker* [1977] on bounding collections of linear functionals in Hilbert space; our approach differs from these, however, because of the way in which we handle uncertainty in the measurements.

Even though a complete knowledge of the external magnetic field does uniquely determine $\mathbf{U}$, practical magnetic anomaly data cannot. This is shown as follows: choose $\mathbf{U}$ arbitrarily, and demand that a model magnetization simultaneously possess this average and satisfy the given data. The requirement of a particular average magnetization may be written

$$(\mathbf{X}_n, \mathbf{M}) = (\mathbf{X}_n, \mathbf{U}) \qquad n = 1, 2, 3$$

Since the right side can be calculated, we have three additional equations to be included with the $N$ in (5) provided by observation. Because the enlarged set of representers, including the three artificial ones $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$, remains linearly independent, the associated Gram matrix is nonsingular, and therefore it is always possible to find a magnetization $\mathbf{M}$ exactly matching the observations for any choice of $\mathbf{U}$ whatever. The proof of linear independence is an easy extension of the one given in appendix A. This negative result is not as upsetting as it might at first appear: when unreasonable $\mathbf{U}$ are chosen, the models generated may be unacceptable because fitting the data together with an "unnatural" mean magnetization may cause extremely large intensities and rapid variations of the solution on a small scale. For example, any model with rms magnetization of 100 A m$^{-1}$ could be rejected even if it did explain the magnetic anomaly precisely: such an intensity is considerably outside the range of observed values for marine basalts established by extensive direct sampling in the Deep Sea Drilling Project of the crust [*Bleil and Petersen*, 1983] and the sparser sampling of seamounts themselves [*Kono*, 1977]. Since there is a generally agreed upon upper limit of magnetization for a plau-

sible model, this factor may be included as part of the information to be used in determining the reliability of the mean magnetization.

The Hilbert space $P$ is apparently suited to this problem because its norm is, after a scaling, just the rms magnetization. An acceptable model is one obeying the two constraints:

$$(\mathbf{M}, \mathbf{M}) \leqslant VM_{\max}^2 \tag{31}$$

$$\sum_{i=1}^{N} [d_i - (\mathbf{G}_i, \mathbf{M})]^2 \leqslant S^2 \tag{32}$$

where $V$ is the volume of the seamount. (Strictly we should write $m(V)$, that is, the measure of the set $V$.) The average magnetization of an arbitrary $\mathbf{M} \in P$ is given by

$$\mathbf{A}[\mathbf{M}] = V^{-1} \sum_{n=1}^{3} \mathbf{X}_n (\mathbf{X}_n, \mathbf{M})$$

This follows from the definition of the orthogonal uniform elements $\mathbf{X}_n$ and

$$(\mathbf{X}_m, \mathbf{X}_n) = V\delta_{mn}$$

We ask how far in the sense of the norm the mean magnetization can be from $\mathbf{U}_*$ and still satisfy both the data and a constraint that the rms intensity be bounded by $M_{\max}$. By solving this maximization problem we find that limits on the uncertainty in the mean magnetization: if we find quite large values of $M_{\max}$ yield a small range for $\mathbf{A}[\mathbf{M}]$, the data may be said to determine the mean magnetization well; if, however, the range in $\mathbf{A}[\mathbf{M}]$ is large even when $M_{\max}$ approaches values within observational experience, say 10 A m$^{-1}$, it must be concluded that the true average magnetization is not strongly constrained by the data. As we shall see, for the seamounts we have studied so far the second alternative seems to apply: the rigorous analysis of uncertainty is disappointing and leads to the unwelcome conclusion that a very wide range of average magnetizations are compatible with the data and the imposition of a prior upper limit on rms magnetization. We have reasons for believing the true uncertainty is much less than indicated by the bound given in the theory; we will take up this matter again. But first, we sketch the theory for obtaining a strict bound.

We write the mathematical problem as a minimization, rather than a maximization: we seek the element of $P$ obeying (31) and (32) that minimizes $-\|\mathbf{A}[\mathbf{M}] - \mathbf{U}_*\|^2$. As in the previous problem containing inequalities, we appeal to the John multiplier theorem, constructing the functional

$$F[\mathbf{M}] = -\|\mathbf{A}[\mathbf{M}] - \mathbf{U}_*\|^2 + \mu\{\sum_{i=1}^{N} [d_i - (\mathbf{G}_i, \mathbf{M})]^2\} + \lambda(\mathbf{M}, \mathbf{M}) \tag{33}$$

where $\lambda, \mu \geqslant 0$ are two John multipliers associated with complementary slackness conditions

$$\mu\{\sum_{i=1}^{N} [d_i - (\mathbf{G}_i, \mathbf{M})]^2 - S^2\} = 0$$

$$\lambda\{(\mathbf{M}, \mathbf{M}) - VM_{\max}^2\} = 0$$

Once more, we expect that neither $\mu$ nor $\lambda$ will vanish and that (31) and (32) will be equalities in all cases of practical importance. For this problem we use a variational approach: the stationary points of (33) are located by taking the Gateaux derivative of $F$ [see *Smith*, 1974, chap. 2] and setting it to zero. The Gateaux derivative is

$$\Delta F[\mathbf{M}] = -2\sum_{n=1}^{3} [V(\mathbf{X}_n, \mathbf{M}) - (\mathbf{X}_n, \mathbf{U}_*)]\mathbf{X}_n + 2\lambda\mathbf{M} + 2\mu\sum_{i=1}^{N} [(\mathbf{G}_i, \mathbf{M}) - d_i]\mathbf{G}_i \tag{34}$$

Suppose this vanishes when $\mathbf{M} = \mathbf{M}_1$. Then, because $\lambda$ is positive, (34) can be rearranged thus

$$\mathbf{M}_1 = \sum_{1}^{N} a_i\mathbf{G}_i + \sum_{n=1}^{3} b_n\mathbf{X}_n \tag{35}$$

where the coefficients $a_i$ and $b_i$ are various combinations of inner products in the representers and uniform elements. In other words, the stationary value of $F$ occurs at an element of $P$ that is a linear combination of elements $\mathbf{G}_i$ and $\mathbf{X}_n$, which puts $\mathbf{M}_1$ in the same subspace of $P$ as $\mathbf{U}_*$. Equation (35) is substituted into (33), and the minimization problem is reduced to one in a finite-dimensional space, finding the appropriate coefficients $a_i$ and $b_n$. The equations are expressed in matrix form if all the inner products are performed. We find

$$F = -\|[A^t \; VI]c - \beta\|^2 + \lambda c^t \begin{bmatrix} \Gamma & A \\ A^t & VI \end{bmatrix} c + \mu\|[\Gamma \; A]c - d\|^2$$

$$= -\|B_0 c - \beta\|^2 + \lambda c^t B_1 c + \mu\|B_2 c - d\|^2 \tag{36}$$

where the vector $c \in \mathbf{R}^{N+3}$ is a composite of the coefficients $a_i$ and $b_n$,

$$c = (a_1, a_2, \ldots, a_N, b_1, b_2, b_3)^t$$

and $\beta \in \mathbf{R}^3$ is the vector of coefficients defining $\mathbf{U}_*$ in the uniform basis. In (36) the norms are ordinary Euclidean lengths of vectors and $I$ is the $3\times3$ unit matrix; the other matrices and vectors are the same as those appearing in section 2; the correspondence between the matrices $B_0$, $B_2$, $B_3$, and the ones in (36) is straightforward.

Viewed as a quadratic form in $c$, the functional $F$ has only one stationary point which can be found by differentiation: for any given John multipliers $\mu$ and $\lambda$, $F$ is stationary when

$$c = (-B_0^t B_0 + \lambda B_1 + \mu B_2^t B_2)^{-1}(\mu B_2^t d - B_0^t \beta) \tag{37}$$

The next problem is to find the appropriate multipliers so that the data misfit and the maximum allowable norm are reproduced, the classical problem of unknown multipliers. We consider (37) to define the vector $c$ as a function of $\lambda$ and $\mu$, that is, $c = c(\lambda, \mu)$. In these terms the two conditions to be obeyed are that $f_1(\lambda, \mu) = f_2(\lambda, \mu) = 0$, where

$$f_1(\lambda, \mu) = c(\lambda, \mu)^t B_1 c(\lambda, \mu) - M_{\max}^2 \tag{38}$$

$$f_2(\lambda, \mu) = \|B_2 c(\lambda, \mu) - d\|^2 - S^2 \tag{39}$$

In principle we can apply Newton's method to solve this pair of equations once the derivatives are known:

$$\frac{\partial f_1}{\partial \lambda} = 2B_1\frac{\partial c}{\partial \lambda} \qquad \frac{\partial f_2}{\partial \lambda} = 2[B_2c - d]^T\frac{\partial c}{\partial \lambda}$$

$$\frac{\partial f_1}{\partial \mu} = 2B_1\frac{\partial c}{\partial \mu} \qquad \frac{\partial f_2}{\partial \mu} = 2[B_2c - d]^T\frac{\partial c}{\partial \mu}$$

and from (37)

$$\frac{\partial c}{\partial \lambda} = -(-B_0^TB_0 + \lambda B_1 + \mu B_2^TB_2)^{-1}B_1c$$

$$\frac{\partial c}{\partial \mu} = (-B_0^TB_0 + \lambda B_1 + \mu B_2^TB_2)^{-1}B_2^T(d - B_2c)$$

Here the functional under minimization is concave, not convex, and so uniqueness of the solution cannot be guaranteed. In particular, there may be more than one $(\lambda, \mu)$ pair giving rise to the root $f_1(\lambda, \mu) = f_2(\lambda, \mu) = 0$, that is, the desired values of norm and misfit. If there are several solutions, we must choose the one that gives the largest value of $\|B_0c - \beta\|^2$. In the absence of any theoretical results on this matter, the only safe procedure is to explore the positive quarter plane $\lambda, \mu > 0$, evaluating a measure of misfit to the root such as $f_1^2 + f_2^2$. Where this indicator function is small, the iterative process can be invoked to produce a precise root.

If (37) is used together with (38) and (39) the search process would be very expensive computationally, since for every $\lambda, \mu$ point we must solve (37), a linear system of the order of $N_1 = N + 3$, where $N$ is the number of magnetic anomaly observations, typically of the order of 150. After advantage is taken of the symmetry of the matrix, each solution requires of the order of $N_1^3/6$ operations, or "flops" [*Golub and Van Loan*, 1983, p. 90]. We can entirely avoid these costly solutions at the expense of two initial spectral factorizations; the details are set out in appendix C. In most cases, the rearrangement described in appendix C can achieve reductions in computational cost of a factor of 200 or more.

Having found the vector $c$ that minimizes the functional $F$ of (36) we compute the corresponding error cone in the mean magnetization vector as follows. The vector $u \in \mathbf{R}^3$ given by

$$u = B_0c = [A^T \quad VI]c$$

is the set of expansion coefficients of $\mathbf{A}[\mathbf{M}]$, the average magnetization of the solution $\mathbf{M}$, in the basis of uniform elements $\mathbf{X}_n$. We may treat $u$ exactly as the the uniform part of the magnetization of the seamount; the vector $\beta$ plays the same role for $\mathbf{U}_*$. Thus the angle between $u$ and $\beta$, namely, $\cos^{-1}(u^T\beta/\|u\|\,\|\beta\|)$ is an upper limit on the angular uncertainty in our determination of the direction of average magnetization, which can easily be converted into the error of the paleomagnetic pole position. With considerably more work one could obtain the cone of uncertainty, by maximizing the angle in all the planes containing the vector $\beta$; the figure would be entirely contained within the cone we have just found and would certainly not have a circular section. The complexity of the maximization problem and the rather poor performance of the present approach with field data deters us from pursuing the question further.

## 5. Application to Field Data

We now apply the theory to a seamount survey performed in the South Pacific. P. F. Lonsdale collected the data on the R/V *Thomas Washington* of the Scripps Institution of Oceanography during leg 6 of the Marathon Expedition in September of 1984. The seamount chosen for study, at 48.2°S, 148.8°W, designated LR148.8W on account of its longitude, is the youngest of a long series to be found on the Louisville Ridge. Total field magnetic values were measured, and precise bathymetric information (contoured in Figure 5) was available through a Sea Beam system. Global Positioning System Navigation was available for a part of the survey and for the remainder, navigation relied upon dead reckoning based upon the Doppler log.

Morphological and petrological evidence [*Hawkins et al.*, 1985] supports the idea that the Louisville Ridge is a hot spot chain similar to the Hawaii-Emperor chain in the North Pacific but on a somewhat smaller scale; it appears to have been active for a comparable period of time. LR148.8W is a large edifice at the southern and therefore younger end of the chain. It is believed to be quite recent, less than 10 m. y. in age. We expect that the virtual geomagnetic pole (VGP) of the uniform part of the magnetization should lie close to the present-day rotation axis of the earth, assuming the effects of secular variation have been averaged out during the time of construction. From Figure 6 we see that the magnetic anomaly is complex and does not resemble that of a uniformly magnetized body (Figure 7), perhaps suggesting the presence of both normal and reversed magnetization, although normal material clearly predominates.

The original records were prepared for inversion as follows. The Sea Beam bathymetric data were sampled with a spacing approximately proportional to the local water depth in order to ensure accurate numerical cubature in the computation of the Gram matrix elements as discussed in section 3. A total of 295 samples went into this

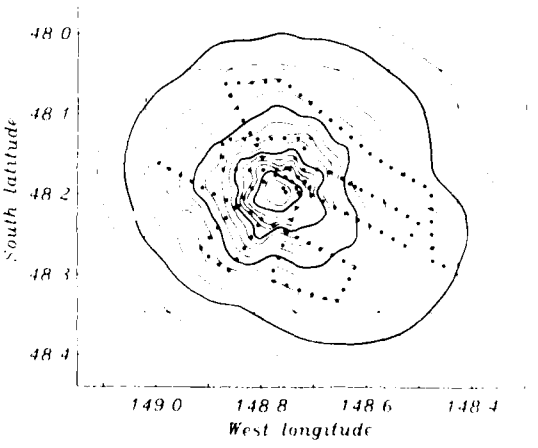



Fig. 5. Bathymetry of LR148.8W contoured at 250-m intervals with 1000-m levels plotted as heavy solid lines. The solid circles give the locations of the magnetic field intensity observations used in the inversion procedure. The interior dashed box is the boundary of Figures 6, 7, and 12. This box measures approximately 44 km by 34 km.

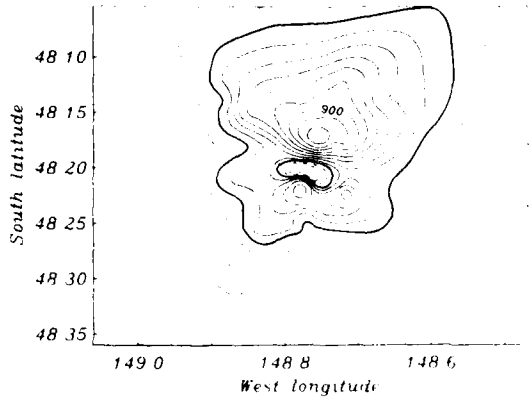



Fig. 6. Magnetic anomaly associated with LR148.8W, contoured at an interval of 150 nT. Negative contours are dashed lines, and the zero level is shown as a heavy solid line. The maximum anomaly contour is 1200 nT. Notice that the scale of this map is slightly larger than that of Figure 5.

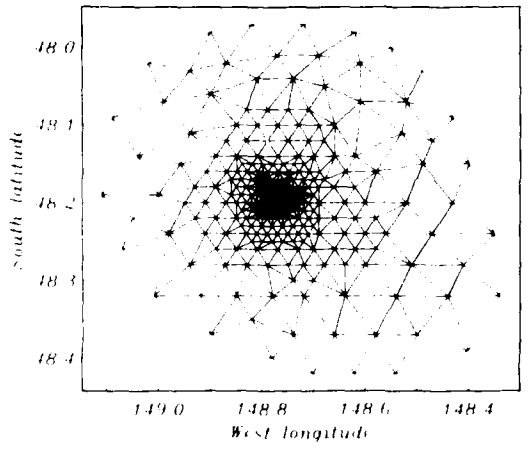



Fig. 8. Plan view of the triangular facet model of LR148.8W used in the numerical calculations. There are 565 triangular facets in the upper surface of the model; the lower boundary is a plane located on the 4250-m contour.

description. The coordinates in the horizontal plane were organized into a tessellation best approximating equilateral triangles by *Watson*'s [1982] algorithm, yielding 565 triangles (Figure 8). When the appropriate vertical coordinate is assigned to each vertex, the result is a model for the upper surface of the seamount in terms of triangular facets; the lower boundary of the model is the horizontal plane bounded by the 4250-m contour. We showed in section 3 that replacement of the true bottom surface by a plane and of the upper surface by triangular facets introduces quite negligible error into the calculations. The model seamount is shown in Figure 9 as it would be seen by an observer in the south looking in a direction 20° west of north and downward 10° below horizontal; in this figure there is a factor of 5 vertical exaggeration. Notice the almost flat top and nearly pseudo sphere appearance of the flanks of the volcano. Magnetic anomaly values were computed from the original intensity measurements by subtracting a standard main field model, International Geomagnetic Reference Field 80 [*Peddie and Fabiano*, 1982]. At the latitude of the survey, diurnal variations are of sufficiently small amplitude that they may safely be ignored. A subset of 131 anomaly values as shown in Figure 5 was chosen to be the data for inversion purposes. Notice that no interpolation or estimation from contour maps was required; only magnetic values from the original survey were used along with corrected positions from the Sea Beam map. Also analyzed were magnetic readings taken during the ship's passage between seamounts of the chain. Seafloor spreading anomalies were prominent, but LR148.8W does not straddle a reversal boundary; it lies on reversely magnetized crust between anomalies 26 and 24 [*Lonsdale*, 1986]. The magnetic field fluctuations in the vicinity were found to have an rms amplitude of 32 nT; the power spectrum had the expected $\exp(-4\pi kz_0)$ form for a field of crustal origin, but the amplitude is much higher than can be attributed to bottom roughness (see section 3) and a reasonable uniform magnetization, and we infer that the observed field is due primarily to local variations in the intensity of the reversely magnetized crust. We chose 30 nT as the level of misfit for model predictions; notice that the peak magnetic anomaly of LR148.8W is over 1200 nT.

The calculations of the Gram matrix $\Gamma$ and the Green matrix $A$ were carried out as described using the $C_{16}$ and $C_{25}$ cubature rules. The construction of the Gram matrix is by far the most time-consuming part of all the computations; on the Crav XM/P-48 at the National Science Foundation San Diego Supercomputer Center, the $C_{25}$ calculation took 10 min, which would translate into several days

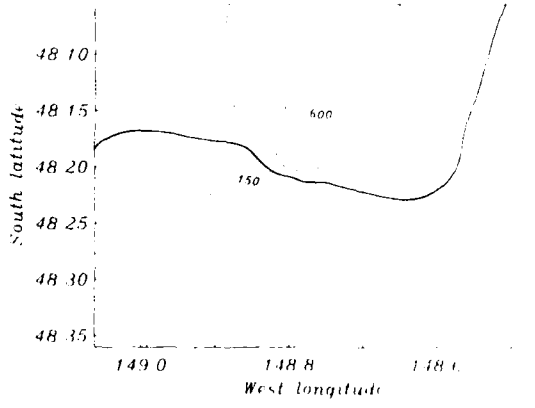



Fig. 7. Best approximation to the observed anomaly of Figure 6 by means of an uniform internal magnetization, contoured at an interval of 150 nT. Negative contours are dashed lines, and the zero level is shown as a heavy solid line. The maximum contour is only 750 nT. Three additional parameters describing a linearly varying background field have been fitted, to allow for small errors in the main field model.

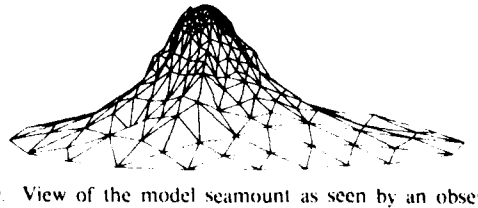


Fig. 9. View of the model seamount as seen by an observer to the south looking 20° west of north and downward 10°. The vertical scale has been exaggerated by a factor of 5.

on a lesser machine. To check the accuracy of the numerical cubature, many of the subsequent calculations were performed with the matrices found by the two rules; no significant differences were observed, from which we conclude that the $C_{16}$ rule was satisfactory.

The first model calculation to be performed was the traditional least squares fitting of a uniform model to the anomaly (with the additional free parameters for slope and offset of the ambient field); in our notation this is just the least squares problem

$$A\beta \sim d$$

which we solved in the usual way by QR factorization [*Golub and Van Loan*, 1983, chap. 6]. As we have already remarked, the field predicted by a uniform magnetization (Figure 7) is a very poor approximation to the measured one. The rms misfit between the observed anomaly and the model field is 269 nT, nearly 10 times the crustal background noise of 30 nT. Therefore little confidence is to be placed in the associated VGP location of 57.8°N, 118.9°W, which is far from the north geographic pole. The intensity of magnetization of the uniform model is 3.7 A m$^{-1}$, a perfectly satisfactory figure.

We turn next to the most nearly uniform solution obtained by seminorm minimization. The calculations involve the solution of the linear system (equation (20)) in section 2; recall that the multiplier $\mu$ is unknown and must be chosen to yield a model with the desired misfit. Rather than single-mindedly exhibiting one solution with the desired 30-nT misfit, we swept through a large range of multipliers, calculating the VGP position and misfit for each. This is not a computationally expensive procedure, and it is instructive. We saw in section 2 that very small values of $\mu$ cause the solution to approach the least squares solution with its large misfit of 269 nT, while large values of the multiplier lead to misfits approaching zero; this behavior is apparent in Figure 10*a*. At the preferred misfit of 30 nT the VGP location for the uniform component of the solution, **U**, is 83.0°N, 47.2°W, remarkably close to the geographic north pole. The position of the VGP is quite insensitive to misfit over a considerable range: it moves less than 4° as the misfit varies from 5 nT to 100 nT. This is illustrated in Figure 10*b* where the trajectory of the pole is plotted on a map of the polar region. Such stability is highly desirable because our estimate of the uncertainty in the magnetic measurements is not precise. The most nearly uniform model corresponding to the 30-nT misfit has an overall rms intensity of magnetization of 6.22 A m$^{-1}$. The nonuniform component of the solution accounts for only 0.78 A m$^{-1}$ in the sense of the norm. (Recall that the uniform part **U** and the nonuniform portion **R** of the model are orthogonal elements of $P$.) Thus only about 13% nonuniform magnetization is needed to reduce the misfit from the 269 nT of the best-fitting purely uniform model to 30 nT. Although the nonuniform part of the model is small in its contribution to the overall magnetization, its presence is a decisive factor in obtaining good agreement with the data. In Figure 11 we attempt to display the internal magnetization vector of our preferred solution: the seamount is cut in three horizontal planes, and in each section we draw arrows, whose sizes and orientations represent the magnetization distribution. We see that the bulk of the model is normally magnetized and resembles **U** with its dip of 65.2° and declination of −4.8°; there is no reversed material in these sections. It is the nature of our solution that the magnetization is more irregular at those points nearest the observations, that is, the points on the upper surface of the seamount, and this is evident in Figure 11. Therefore near the peak there probably will be small pockets of reversed magnetization because of the greater fluctuation of the solution on the boundary. Little credence should be given to such features even though a seamount formed during the Cenozoic might be expected to contain normal and reversed material. It is not possible to prove rigorously from the external field data that reversed material is present; this follows from the fact that any finite magnetic data set can be satisfied by a body normally magnetized in its entirety, but we shall not go into the proof here.

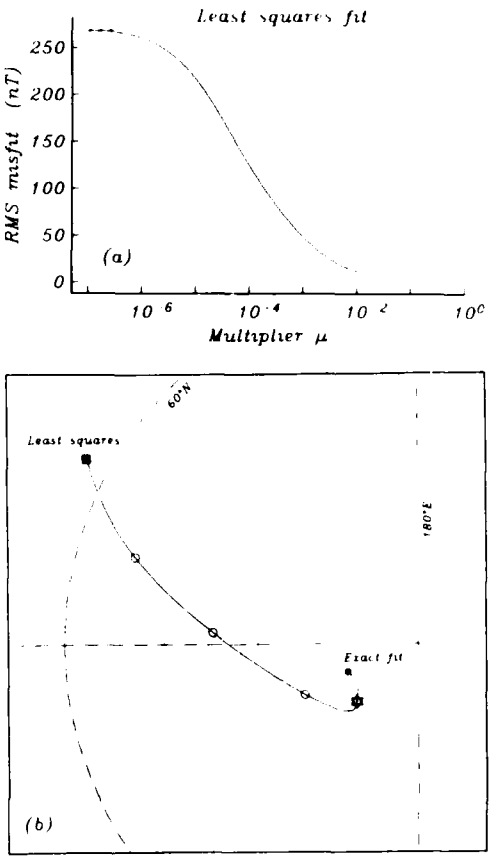



Fig. 10. Behavior of the solutions as the parameter $\mu$ varies. (*a*) Rms misfit of computed magnetic anomaly to observed values as a function of John multiplier $\mu$. A misfit of 30 nT is considered appropriate. (*b*) Pole path of the different solutions shown in a Lambert equal-area projection of the polar region. Open circles are poles with rms misfits of 200, 150, and 100 nT from left to right. The star is the pole of the 30-nT misfit solution.

How accurate is the paleopole of the mean magnetization? In the previous section we described a theory that was capable in principle of providing the answer if we are prepared to supply an upper limit on the rms magnetization to be found in marine basalts. The principle of the method is to construct the worst possible case: the model

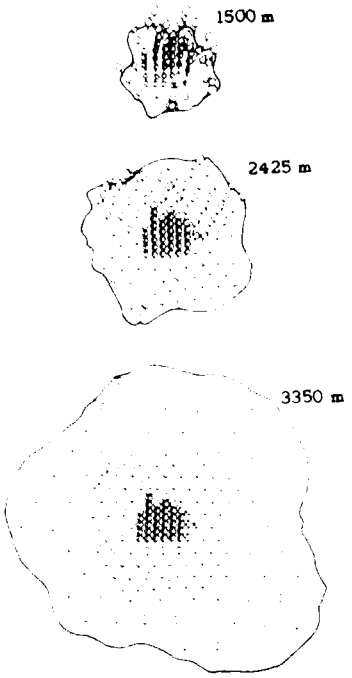



Fig. 11. Magnetization vectors in three horizontal sections of the seamount at bathymetry contour levels 1000 m, 2000 m, and 3000 m. The solution illustrated is the most uniform model with rms misfit to the data of 30 nT. The length of the arrow depicts the strength of magnetization and the orientation can be inferred from the aspect of the conical head.

that lies as far as possible from the preferred solution while lying within the constraints set by the the data and an upper limit on the norm of **M**, or equivalently, the rms magnetization. We could embark upon the search over the $\lambda$, $\mu$ plane seeking stationary values of $F$ in (33), but in the present example this is unnecessary because we can find a model obeying the constraints with a completely different VGP position; in other words we need not seek the worst case. In section 2 we discussed the problem of finding an element in $P$ nearest to a fixed element $\mathbf{U}$; if the fixed element is 0, the zero element, we will find the smallest model of $P$ fitting the data. The magnetization of this smallest model is, from (10), a linear combination of the representers $\mathbf{G}_j$. When this theory is adjusted to include misfit and the three parameters for the ambient field, it becomes an example of seminorm minimization on $P'$, and the necessary matrix elements are already available. Arranging $\mu$ to give the expected 30 nT misfit, we find an rms magnetization averaged over the seamount of only 1.70 A m$^{-1}$, and the pole position of the projection of **M** onto the subspace of uniform elements is 60.1°N, 34.4°E. Obviously this model has a small enough norm and a good fit but its pole lies 30° from that of the most uniform model; there is no doubt that by allowing a larger norm, solutions with poles even further away must exist.

Is it really true that the uncertainty in the uniform VGP is so large? The fact that the direction is very close to what one might expect on geological grounds encourages us to believe otherwise, it must be remembered that we have approached the problem of uncertainty by estimating an upper bound and that the true error could be far smaller than that bound. An explanation for the generosity of the error bound in the theory of this paper may be traced to a property of the norm of $P$: although the rms magnetization of the minimum norm model is only 1.70 A m$^{-1}$, we calculate that at some points within the model the local magnetization intensity rises to nearly 1500 A m$^{-1}$, a totally unacceptable value. Thus a reasonable rms magnetization is by itself no guarantee of a plausible magnetization model. Perhaps a more suitable restriction would be that the magnetization at any point in $V$ should not exceed a prescribed limit, say 20 A m$^{-1}$. This is a kind of uniform norm, but uncertainty estimates based upon it would be much more complicated than those undertaken in this paper; the idea is probably worth pursuing, but we do not take it up here.

Far from the seamount its magnetic field closely resembles that of a point dipole, the influence of higher multipole terms having fallen away more rapidly than that of the leading term. An analysis of three suitable field measurements can yield the vector moment of a dipole. But the dipole moment divided by the volume is just the mean magnetization of the seamount. Another way of looking at this question was given by *Parker* [1971]: we imagine trying to construct the linear functional for $\mathbf{X}_1$, say from a linear combination of the representers $\mathbf{G}_j$. The representers for distant observers are much smoother and more nearly constant, so that they are much more valuable to the approximation. This discussion ignores the problem of noise in the observations; while the field values far from the seamount have a more direct connection with the average magnetization, their relative accuracy is much less than those nearer the body because the noise fields do not drop off with distance from the seamount while the signal does. There must be an optimal distance from the seamount at which the field values yield a maximum amount of information about the U; did the survey of LR148.8W extend to that optimum distance? We have reason to believe that it did not. A certain amount of manipulation of (20) yields the following informative expression for $\beta$, the vector of coefficients for $\mathbf{U}$ in the basis $\mathbf{X}_k$:

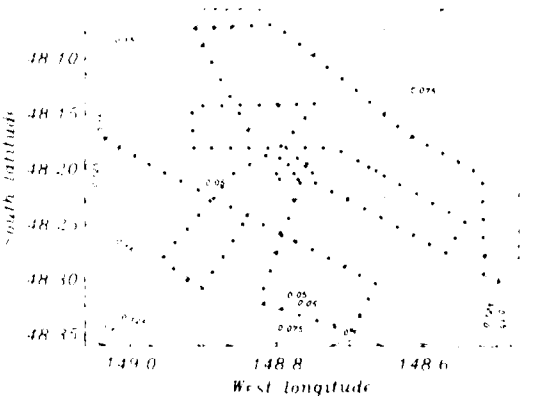



Fig. 12. Dimensionless weight function expressing the influence of each field measurement upon the estimation of the average magnetization. The small levels in the central region indicate that magnetic measurements there have little effect on U.

$$\beta = (A^T(I/\mu + \Gamma)^{-1}A)^{-1}(I/\mu + \Gamma)^{-1}d$$
$$= (\mu_0/4\pi)w^T d$$

where $w$ is an $N \times 6$ matrix (remember the inclusion of the three additional free parameters describing the background field); the factor $\mu_0/4\pi$ renders $w$ dimensionless. This equation shows how every magnetic field measurement contributes to each coefficient of the uniform component of the solution. Column 1 of $w$ can be interpreted as a weight vector in an equation of the form

$$\beta_1 = (\mu_0/4\pi)w_1^T d$$

When $w_1$ is mapped we see which regions of the survey affect our determination of $\beta_1$ by the size of the weight function. The average magnetization is insensitive to the magnetic anomaly in places where $w$ is small, and conversely, it depends heavily on data where $w$ is large. (This interpretation is not strictly valid because if a different data value appeared at any point, the size of $\mu$ would have to be adjusted to retain the desired misfit, thereby changing $w$; if the putative changes are small, so will the perturbation to $\mu$.) As a synopsis of the influence on all of U we have contoured $(w_1^2 + w_2^2 + w_3^2)$ in Figure 12. We discover that our knowledge of U is determined for the most part by values at the edge of the survey. If 36 data are removed in a 15-km disk at the center of the magnetic anomaly, thus obliterating the central maximum and the negative patch to the south of it, the uniform part of the new solution has a pole position only 6° away from the one obtained with all the data. This confirms the relative lack of importance of the central region for the purposes of obtaining an estimate of the uniform component. The way in which the weight function grows toward the edge of the map strongly suggests that the optimum distance for recovering information about the average magnetization probably lies beyond the boundary of the present survey, and it must be concluded that to obtain the best estimate of the average magnetization, the area should be extended considerably. We anticipate that with a more nearly optimal distribution of magnetic observations the error will bound significantly smaller.

LR148.8W is a challenging seamount for analysis: the magnetic anomaly is complex and obviously incompatible with the assumption of a uniform interior. The fact that we are able to recover a reasonable VGP and that the minimum departures from perfect uniformity are very mild gives us encouragement as we contemplate a wider application of the technique. Our seamount is an unusually large body, nearly 1600 km$^3$ in volume, and it comes within 500 m of the ocean surface. These two properties are major factors in determining the computing time needed for our approach, particularly when it is recalled from section 3 that the size of each triangular facet is governed by the local water depth. The majority of seamounts would make much more modest computational demands.

## 6. Discussion

We have presented a method using linear inverse theory to construct an internal magnetization function for seamounts based upon their magnetic anomalies and shapes. We have shown why it is fundamentally impossible to deduce the true magnetization from external field data no matter how precise or complete it is. Therefore our model has been selected to correspond as closely as possible to the simplest structure, the uniformly magnetized body. Approximations to this ideal are expected on geological grounds if the seamount formed rapidly or during a period of single polarity of the main geomagnetic field. Nonetheless, extensive modeling of actual seamounts has shown that the uniform model by itself rarely gives an accurate account of the magnetic anomaly so that significant nonuniformity is certainly present. Application of the method to a young seamount in the Louisville Ridge chain yields a magnetization accounting accurately for the magnetic anomaly; we find that in the measure of the norm, only 13% nonuniformity is required to obtain the good agreement, even though the observed anomaly is complex and poorly approximated by the field of a uniform body. This result suggests that the picture of an essentially uniform seamount may not be so inaccurate after all, but since even quite small amounts of heterogeneity have a disproportionately large influence on the form of the magnetic anomaly, this has been impossible to appreciate until now. We predict that relatively small proportions of nonuniformity will be needed in all the seamounts with simple magnetic anomalies.

The mean magnetization is a property of a seamount that can in principle be obtained from the magnetic anomaly alone. The direction of the mean magnetization vector is diagnostic of the paleomagnetic field averaged over the period of formation of the body, and it is the most valuable information about the seamount for tectonic studies. In our example we find that the mean magnetization vector predicts a paleopole very close to the north geographic pole in agreement with our expectations; this is in contrast to the pole position of the best-fitting uniform model, which lies 30° away. We have developed a theory for the uncertainty in the estimate of the mean magnetization requiring an upper limit on the overall rms magnetization allowed in the volcanic rocks; sampling and knowledge of rock magnetism puts such estimates on a secure footing. Unfortunately, the results for the Louisville Ridge seamount are disappointing: the uncertainty is so large as to give the impression that the calculated paleopole position is unreliable. The excellent location of the model pole leads us to believe that the true error is much smaller and that a better theory is needed for its estimation. We show that significant improvements in the uncertainty estimates can be expected to follow from a more extensive survey, since it is apparent from our calculations that data taken at a surprisingly great distance from the magnetic sources contain substantially more information about the mean magnetization than those closer to them. We would not recommend dispensing with coverage over the central region in future surveys, however, because the shape of the seamount must still be known in detail if its magnetic field is to be properly analyzed. It is to be hoped of course that progress on the theoretical front will obviate the necessity of resurveying every seamount. In fact we may anticipate here a refinement in the theory of error estimation: recent work [*Parker*, 1987] relying upon a plausible statistical characterization of the magnetic nonuniformities

promises to provide the basis for an alternative and more powerful theory for the uncertainties.

## APPENDIX A: LINEAR INDEPENDENCE OF THE REPRESENTERS

The linear independence of the representers in (2) guarantees the positive definiteness of the Gram matrix, an essential property for a number of results in the paper. For any particular seamount and set of observer positions one could in principle verify that $\Gamma$ had a positive determinant, but it is more satisfactory if this fact can be shown in general as we shall now proceed to do. We follow the usual path of investigating the consequences of assuming that the functions are linearly dependent. Linear dependence would imply the existence of a set of constants $\gamma_j$ not all zero, such that

$$0 = \sum_{j=1}^{N} \gamma_j G(\mathbf{r}_j, \mathbf{s})$$

for every point $\mathbf{s} \in V$. From (2) we see this is equivalent to

$$0 = \hat{\mathbf{B}}_0 \cdot \nabla \nabla F(\mathbf{s}) \tag{A1}$$

where

$$F(\mathbf{s}) = \sum_{j=1}^{N} \frac{\gamma_j}{|\mathbf{r}_j - \mathbf{s}|} \tag{A2}$$

For this appendix only, let us erect a local Cartesian axis system with the $z$ direction aligned with $\hat{\mathbf{B}}_0$. Then $\hat{\mathbf{B}}_0 \cdot \nabla = \partial/\partial z$. It is elementary that the only solutions to (A1) are of the form

$$F(x, y, z) = c_1 z + c_2 + f(x, y)$$

where $c_1$ and $c_2$ are constants and $f$ is an arbitrary continuously differentiable function of $x$ and $y$. Thus linear dependence implies the existence of coefficients $\gamma_j$ such that

$$c_1 z + c_2 + f(x, y) = \sum_{j=1}^{N} \frac{\gamma_j}{|\mathbf{r}_j - \mathbf{s}|} \tag{A3}$$

for $(x, y, z) = \mathbf{s} \in V$. The functions $1/|\mathbf{r}_j - \mathbf{s}|$ are all analytic in $\mathbf{s}$ inside $V$, so that if those coefficients exist, (A1) remains valid by analytic continuation of the individual components of the vector everywhere outside $V$, except right at the singularities $\mathbf{s} = \mathbf{r}_j$. Therefore the property (A3) may be extended outside $V$ in the same way.

Consider a sphere centered at the origin of coordinates and enclosing all the observation positions; its radius $R_1$ is greater than $\max|\mathbf{r}_j|$. We evaluate $F$ using (A2) at a position vector $\mathbf{s}$ outside this sphere such that $|\mathbf{s}| > 2R_1$; it follows that

$$|F(\mathbf{s})| \leqslant \sum_{j=1}^{N} \frac{|\gamma_j|}{|\mathbf{r}_j - \mathbf{s}|}$$

$$\leqslant \frac{N\gamma_{\max}}{|\mathbf{s}| - R_1}$$

$$< \frac{2N\gamma_{\max}}{|\mathbf{s}|}$$

where $\gamma_{\max}$ is the largest of all the magnitudes of the $\gamma_j$. Thus by choosing $|\mathbf{s}|$ large enough we can make $|F|$ as small as we please. If $\mathbf{s} = (x, y, z)$ and $z$ increases without bound while $x$ and $y$ are fixed, $F$ must tend to zero. Thus $c_1$ must be zero, or otherwise the left side of (A3) would grow in magnitude in the limit of large $z$. But if $c_1$ vanishes, the left side of (A3) no longer depends on $z$. Therefore the $c_2 + f(x, y)$ must vanish identically for every $x$ and $y$, if $F$ is to tend to zero when $z$ tends to infinity. In other words, the left side of (A3) is zero.

Thus equation (A3) has become

$$0 = \sum_{j=1}^{N} \frac{\gamma_j}{|\mathbf{r}_j - \mathbf{s}|} \tag{A4}$$

From this we show that all the $\gamma$ coefficients vanish too. Renumber the coordinates and coefficients so that $\gamma_1$ is the coefficient of largest magnitude; also consider a position $\mathbf{s}$ so that $|\mathbf{r}_1 - \mathbf{s}| = \epsilon$. Then from (A4)

$$\frac{\gamma_{\max}}{\epsilon} \leqslant \sum_{j=2}^{N} \frac{|\gamma_j|}{|\mathbf{r}_j - \mathbf{s}|}$$

Now let $R_2$ be the smallest of all the interobserver distances $|\mathbf{r}_j - \mathbf{r}_k|$; if $\epsilon$ is chosen to be smaller than $R_2/2$ it is easily verified that

$$\frac{\gamma_{\max}}{\epsilon} \leqslant \sum_{j=2}^{N} \frac{2|\gamma_j|}{R_2} \tag{A5}$$

or

$$\gamma_{\max} \leqslant \epsilon C$$

where $C$ is some constant independent of $\epsilon$. Since we may choose $\epsilon$ to be as small as we please, this means that the $\gamma_j$ largest in magnitude must vanish; so then they all must. This contradicts the original assertion that not all of them could be zero, and therefore we must conclude that the representers are not linearly dependent.

Notice that the proof fails, as it should, if two of the observer positions are in fact identical; then $R_2$ would vanish, and (A5) would not be legitimate.

## APPENDIX B: APPROXIMATION OF TERRAIN BY TRIANGULAR FACETS

When the power spectral density of the bathymetry is known, it is possible to estimate the probable error committed by replacing the true surface by a plane triangle that interpolates the bathymetric values at its corners. Although to our knowledge no spectral studies exist for the surface of an actual seamount, it seems plausible to assume that on scales much less than the diameter of the seamount, the surface roughness is approximately the same as that of very young oceanic seafloor or of terrestrial lava flows, and for both of these, quantitative analysis is available. We begin with data presented by *Fox and Hayes* [1985] on the spectrum of bathymetric profiles; one of their spectra describes the Gorda Rise, and we shall use the parameters estimated for this region.

We find it most convenient to develop the theory using the autocorrelation function of the topography. The (two-dimensional) autocorrelation function, $R(\vec{x})$, is defined by the expectation of the product of two samples of the topography taken at positions $\vec{y}$ and $\vec{y} + \vec{x}$ (we use $\vec{x}$ rather than $\mathbf{x}$ to stress the fact that the vectors give positions in the plane), and this is related to the power spectral density through the Fourier transform:

$$R(\vec{x}) = \mathrm{E}[h(\vec{y})h(\vec{y}+\vec{x})] \tag{B1}$$

$$= \int_{\mathbf{R}^2} S(\vec{k}) \exp(2\pi i \vec{k}\cdot\vec{x})\, d^2\vec{k}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(k_x, k_y) \exp(2\pi i[k_x x + k_y y])\, dk_x\, dk_y \tag{B2}$$

where $S(\vec{k}) = S(k_x, k_y)$ is the wave number power spectral density of the topography. In practice it is found that the spectra are fairly isotropic, so that $S$ depends only on $|\vec{k}|$ and not on the direction of $\vec{k}$; it follows that $R$ is a function independent of the direction of its argument vector. We denote the isotropic spectral density by $S[|\vec{k}|]$ and the isotropic autocorrelation function by $R[|\vec{r}|]$, admitting the slight risk of confusion from this notation. Fox and Hayes do not provide power spectral densities $S(\vec{k})$ or isotropic spectral functions $S[|\vec{k}|]$, but give instead the one-dimensional spectral amplitudes of profiles, functions we denote by $A_1(k_x)$. Now

$$A_1(k_x)^2 = \int_{\mathbf{R}} S(\vec{k})\, dk_y \tag{B3}$$

which is just the two-dimensional power spectrum integrated in a wave number direction normal to that of the profile [see *Shure and Parker*, 1981]; since all directions are equivalent, any constant direction may be chosen. If we evaluate (B2) for $y = 0$ we see from (B3) that

$$R(x, 0) = \int_{-\infty}^{\infty} A_1(k_x)^2 \exp(2\pi i k_x x)\, dk_x$$

$$= R[x] \tag{B4}$$

Thus the isotropic autocorrelation function is just the one-dimensional inverse Fourier transform of the profile power spectral density.

Fox and Hayes find for many kinds of terrain a power law holds quite well; we rewrite their result:

$$A_1(k_x)^2 = c_1 |l_0 k_x|^{-\eta} \tag{B5}$$

where $l_0$ is an arbitrary constant with dimensions of length. After a careful interpretation of the unusual units used in the paper, we calculate that for the young volcanic terrain of the Gorda Rise, $c_1 = 4800$ m$^3$ and $\eta = 2.48$ when $l_0$ is set to 1 km. Equation (B5) cannot be substituted into (B4) in a classical way because the Fourier integral diverges owing to the singularity at $k_x = 0$; in fact this shows us that the extrapolation of the spectrum to all wave numbers is not strictly legal since it possesses infinite energy caused by contributions at long wavelengths. However, it is only the short wavelengths that concern us here and so we use distribution theory to evaluate the Fourier transform [*Gel'fand and Shilov*, 1964, p. 359]:

$$R[r] = 2(2\pi)^{\eta-1} c_1 l_0^{1-\eta} \sin(\pi\eta/2)\Gamma(1-\eta)(r/l_0)^{\eta-1}$$

$$= c_2 (r/l_0)^{\eta-1} \tag{B6}$$

In this appendix, $\Gamma$ is the familiar special function, not the Gram matrix. We find $c_2 = -239$ m$^2$. It is at first astonishing that $R[0] = 0$ and that it is negative elsewhere; the explanation is that the generalized Fourier transform has suppressed an infinite constant term. The true autocorrelation function has a large but unknown value at zero, but the result we shall obtain (equation (B8)) is invariant under addition of any constant to $R$ so that our ignorance of the constant is unimportant.

From (B6) we may calculate the isotropic spectrum $S[k]$ by means of a Hankel transform which is just the inverse transform of (B2) expressed in polar form [*Bracewell*, 1978]:

$$S[k] = \int_0^{\infty} J_0(2\pi k r) R[r]\, 2\pi r\, dr$$

$$= c_3 (k l_0)^{-(\eta+1)} \tag{B7}$$

where

$$c_3 = \frac{c_2 l_0^2}{\pi\eta} \frac{\Gamma((1+\eta)/2)}{\Gamma((1-\eta)/2)}$$

For the constants we are using $c_3 = 27{,}400$ m$^4$. This result will be useful in calculating the magnetic effect of the roughness in the terrain.

Consider now a triangular region $T$ defined by its corners at position vectors $\vec{x}_1$, $\vec{x}_2$, and $\vec{x}_3$ which lie in a horizontal plane. The plane-interpolated topography for $\vec{x} \in T$ is found from the values of the corners via a formula of the kind

$$\tilde{h}(\vec{x}) = \sum_{i=1}^{3} h(\vec{x}_i)\phi_i(\vec{x})$$

where the basis functions, $\phi_i$ are each 1 at $\vec{x}_i$ and zero along the opposite edge of the triangle, varying linearly over $T$. We can find the variance of the interpolation via the expectation, assuming as usual that $h$ has zero mean. Define

$$\delta^2 = \mathrm{E}[(\tilde{h}(\vec{x}) - h(\vec{x}))^2]$$

Then, using the definition of the autocorrelation function (B1) and performing a certain amount of algebra, we obtain

$$\delta^2 = R[0](1 + \sum \phi_i(\vec{x})^2) + \sum_{i,k} 2R[|\vec{x}_i - \vec{x}_k|]\phi_i(\vec{x})\phi_k(\vec{x})$$

$$- \sum 2R[|\vec{x} - \vec{x}_i|]\phi_i(\vec{x})$$

Next we average over the triangle $T$

$$\delta^2_T = \int_T \delta^2\, d^2\vec{x}/\Delta$$

where $\Delta$ is the area of the triangle. Then, performing the integrals involving the basis functions we find

$$\delta^2_T = \frac{3}{2}R[0] + \frac{1}{6}\sum_{i,k} R[|\vec{x}_i - \vec{x}_k|]$$

$$- \frac{1}{\Delta}\sum_i \int_T 2R[|\vec{x} - \vec{x}_i|]\phi_i(\vec{x})\, d^2\vec{x} \tag{B8}$$

This is a general result, not depending on any particular choice of autocorrelation function, except that isotropy has been assumed. In substituting (B6) into (B8) it is convenient to define a cyclic extension of the corner vectors so that $\vec{x}_{i+3} = \vec{x}_i$ for $j = 1, 2$ and similarly for the angles at the corners associated with each $\vec{x}_i$, which will be called

$\theta_j$. We wish to separate the effects of size from those of triangle shape, and so we introduce normalized side lengths, referred to the area of the triangle: $\hat{s}_j = |\bar{x}_{j-1} - \bar{x}_{j-2}|/\Delta^{1/2}$. With all this we find after some effort that (B8) becomes

$$\{\delta^2\}_j I = -c_2 (\Delta^{1/2}/I_0)^{\eta-1}\Theta(T) \tag{B9}$$

where $\Theta(T)$ is a shape factor for the triangle given by

$$\Theta(T) = \sum_{j=1}^{3} \left[ \frac{2^{\eta-2}|\Psi(\theta_{j-2}) - \Psi(-\theta_{j-1})|}{(\eta+1)(\eta+2)(\hat{s}_j)^{\eta-1}} - \frac{(\hat{s}_j)^{\eta-1}}{6} \right] \tag{B10}$$

and here the function $\Psi$ is the integral

$$\Psi(\theta) = \int_{\pi/2}^{\pi/2} \frac{d\phi}{\sin^{\eta-1}\phi}$$

$$= \cos\theta\, F(1+\eta/2,\ 1/2;\ 3/2;\ \cos^2\theta)$$

where $F$ is Gauss' hypergeometric function; power series expansions given by *Abramowitz and Stegun* [1965, chap. 15] are quite convenient for evaluating $\Psi$. The function $\Theta(T)$ is plotted in Figure 2 for the value of $\eta = 2.48$ given for the Gorda Rise by *Fox and Hayes* [1985]. The smallest value is attained by equilateral triangles with $\Theta(T) = 0.3198$ and only extremely scalene triangles achieve values above unity.

The Fox and Hayes study gives spectral estimates up to $k_x = 5$ km$^{-1}$, or a wavelength of 200 m; it would be helpful in confirming our analysis if we could show that the same power law extended to smaller scales. We have compared their spectrum with one found for a subaerial lava flow in Bonito Arizona [*Jaeger and Schuring*, 1966]. The largest wavelength estimated in this spectrum is 10 m, so that there is, unfortunately, no overlap. We find that the marine spectrum, extrapolated to the shorter scale is consistently a factor of 10 smaller in power. It is likely this discrepancy results from the fact that the Bonito flow consists mostly of aa, which is almost certainly much rougher than the expected surface of a seamount on the small scales.

## Appendix C: Efficient Exploration of the $\lambda, \mu$ Plane

In this appendix we describe a practical means for solution of (38) and (39). Certain roots of these equations give rise to magnetizations of prescribed norm and data misfit that are as far as possible from the uniform state. It is possible that other stationary points exist to the optimization problem that do not correspond to the desired maximum, so that a fairly complete exploration of the $\lambda, \mu$ plane should be performed to uncover all the roots. For this to be a practical proposition, an economical method must be found for solving (37), the large system of linear equations associated with each point in the plane.

The fundamental idea is the possibility of finding a single similarity transform simultaneously mapping two symmetric matrices into diagonal form [*Golub and Van Loan*, 1983, chap. 8]. First we solve the eigenvalue problem for the matrix $B_1$ and write the solution as the spectral factorization:

$$B_1 = Q_1 \Lambda Q_1^t$$

where $Q_1$ is an $N_1 \times N_1$ orthogonal matrix, whose columns are the eigenvectors of $B_1$, and $\Lambda$ is the diagonal matrix of eigenvalues. Notice that all the eigenvalues should be positive because $B_1$ is positive definite; in practice, small errors in the cubature may cause this to be untrue for the actual numerical array and so it is necessary to regularize the problem by adding a small positive constant to the diagonal of $B_1$ or by taking absolute values for the eigenvalues. When this is done, the square root of $\Lambda$ is defined as well as its inverse; we denote this in the obvious way. Next we compute a singular value decomposition [*Golub and Van Loan*, 1983]: let

$$B_2 Q_1 \Lambda^{-1/2} = Z_2 \Sigma Q_2^t$$

where the matrix on the left has been computed from its known factors and those on the right are the singular value decomposition factors. $Z_2$ and $Q_2$ are orthogonal arrays of the appropriate sizes, $\Sigma$ is an $N \times N_1$ matrix,

$$\Sigma = [\Sigma' \quad O_3]$$

where $\Sigma'$ is the $N \times N$ diagonal array of singular values, and $O_3$ is a $N \times 3$ array of zeros. Notice that these two computationally expensive factorizations need be done only once per seamount. Now define the $3 \times N_1$ matrix $\tilde{B}_0$ by

$$\tilde{B}_0 = B_0 Q_1 \Lambda \quad Q_2$$

When these factorizations are substituted into (37) we obtain

$$c = Q_1 \Lambda \quad Q_2 \tilde{c}$$

where

$$\begin{aligned}\tilde{c} &= (-\tilde{B}_0^t \tilde{B}_0 + \lambda I + \mu \Sigma^t \Sigma)^{-1} Q_1^t \Lambda \quad Q_1^t (\mu B_1^t d - B_0^t \beta) \\ &= (-\tilde{B}_0^t \tilde{B}_0 + \lambda I + \mu \Sigma^t \Sigma)^{-1} \\ &\qquad \cdot (\mu Q_1^t \Lambda \quad Q_1^t B_1^t d + Q_1^t \Lambda \quad Q_1^t B_0^t \beta) \\ &= (-\tilde{B}_0^t \tilde{B}_0 + \lambda I + \mu \Sigma^t \Sigma)^{-1} (\mu v_1 - v_2)\end{aligned} \tag{C1}$$

Observe that the vectors $v_1$ and $v_2$ do not depend on $\lambda$ or $\mu$. Equation (C1) is a transformed version of (37). The functions $f_1$ and $f_2$ are simply expressed in terms of the vector $\tilde{c}$, which becomes the working variable during the heavy calculations:

$$f_1 = \|\tilde{c}\|^2 - M_{\max}^2$$

$$f_2 = \|\Sigma \tilde{c} - Z_2^t d\|^2 - S^2$$

In (C1) we still must solve an $N_1 \times N_1$ linear system for each $\lambda, \mu$ pair, the key to the efficient realization of this process is the fact that $\tilde{B}_0$ is not a square array but only $3 \times N_1$ and that $\lambda I + \mu \Sigma^t \Sigma$ is diagonal. We denote this diagonal matrix by $D(\lambda, \mu)$ and apply the following matrix identity, called the Sherman-Morrison-Woodbury formula [*Golub and Van Loan*, 1983, p. 3]:

$$(A + UW^t)^{-1} \equiv A^{-1} - A^{-1} U (I + W^t A^{-1} U)^{-1} W^t A^{-1}$$

where $A$ is any square invertible matrix, $U$ and $W$ are of the proper sizes, but not necessarily square, and the left side must exist. If we choose $A = D(\lambda, \mu)$ and $U = -W = \tilde{B}_0^t$ then

$$(-\tilde{B}_0^I\tilde{B}_0 + \lambda I + \mu\Sigma^I\Sigma)^{-1} = D^{-1} + D^{-1}\tilde{B}_0^I(I - \tilde{B}_0D^{-1}\tilde{B}_0^I)^{-1}\tilde{B}_0D^{-1} \quad \text{(C2)}$$

The reason that this rearrangement represents a massive computational saving is that the matrix inversions in the expression are either of the diagonal matrix $D(\lambda, \mu)$ and hence trivial or of the matrix $(I - \tilde{B}_0D^{-1}\tilde{B}_0^I)^{-1}$, which is only a 3×3 symmetric array. When $N_1$ is large, the number of computer operations to evaluate $\tilde{c}$ using (C2) and (C1) is approximately $13N_1$; with $N_1 = 150$, this represents a factor of nearly 300 improvement over the equivalent calculation using (37).

*Acknowledgments.* The authors wish to express their appreciation for the support of the Office of Naval Research through contract N00014-85-C-0104. The large-scale numerical work was carried out for the most part on the Cray XM/P-48 computer at the San Diego Supercomputer Center; we thank the National Science Foundation for grants to the Scripps Institution of Oceanography for use of this facility and to its staff for their assistance. In addition we are especially grateful to Peter Lonsdale for collecting the magnetic and bathymetric data for the seamount LR148.8W as part of the Marathon Expedition leg 6, funded by the Office of Naval Research under contract USN N00014-80-0440. Finally we thank Alan Chave for his help in installing an early version of our programs on a Cray computer at Los Alamos National Laboratory.

## REFERENCES

Abramowitz, M., and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1965.

Backus, G. E., Inference from inadequate and inaccurate data, I, *Proc. Natl. Acad. Sci. U. S. A.*, *65*, 1–7, 1970.

Batiza, R., Abundances, distribution and sizes of volcanos in the Pacific Ocean and implications for the origin of non-hotspot volcanos, *Earth Planet. Sci. Lett.*, *60*, 195–206, 1982.

Batiza, R., D. J. Fornari, D. A. Vanko, and P. Lonsdale, Craters, calderas and hyaloclastites on young Pacific seamounts, *J. Geophys. Res.*, *89*, 8371–8390, 1984.

Blakely, R. J., and R. L. Christiansen, The magnetization of Mount Shasta and implications for virtual geomagnetic poles determined from seamounts, *J. Geophys. Res.*, *83*, 5971–5978, 1978.

Bleil, U., and N. Petersen, Variations in magnetization intensity and low-temperature oxidation of ocean floor basalts, *Nature*, *301*, 384–387, 1983.

Bracewell, R., *The Fourier Transform and Its Applications*, McGraw-Hill, New York, 1978.

Duncan, R. A., and I. McDougall, Linear volcanism in French Polynesia, *J. Volcanol. Geotherm. Res.*, *1*, 197–227, 1976.

Emilia, D. A., and R. L. Massey, Magnetization estimation for nonuniformly magnetized seamounts, *Geophysics*, *39*, 223–231, 1974.

Fox, G. F., and D. E. Hayes, Quantitative methods for analyzing the roughness of the seafloor, *Rev. Geophys.*, *23*, 1–48, 1985.

Francheteau, J., J. G. Sclater, M. L. Richards, and C. G. A. Harrison, Comments on paper by G. R. Schimke and C. G. Bufe, 'Geophysical description of a Pacific Ocean Seamount,' *J. Geophys. Res.*, *73*, 7713–7715, 1968.

Francheteau, J., C. G. A. Harrison, J. G. Sclater, and M. L. Richards, Magnetization of Pacific seamounts: A preliminary polar curve for the northeastern Pacific, *J. Geophys. Res.*, *75*, 2035–2061, 1970.

Gardner, J. V., W. E. Dean, and R. J. Blakely, Shimada Seamount: An example of recent mid-plate volcanism, *Geol. Soc. Am. Bull.*, *95*, 855–862, 1984.

Gel'fand, I. M., and G. E. Shilov, *Generalized Functions, vol. 1*, Academic, New York, 1964.

Golub, G. H., and C. F. Van Loan, *Matrix Computations*, Johns Hopkins Press, Baltimore, Md., 1983.

Gordon, R. G., and A. Cox, Calculating paleomagnetic poles for oceanic plates, *Geophys. J. R. Astron. Soc.*, *63*, 619–640, 1980.

Grossling, B. F., Seamount magnetism, in *The Sea*, vol. 4, pp. 129–156, John Wiley, New York, 1970.

Haggerty, J. A., S. O. Schlanger, and I. Primoli-Silva, Late Cretaceous and Eocene volcanism in the southern Line Islands, and implications for hot spot theory, *Geology*, *10*, 433–437, 1982.

Harrison, C. G. A., A seamount with a non-magnetic top, *Geophysics*, *36*, 349–357, 1971.

Harrison, C. G. A., and M. M. Ball, Geophysical observations on an exposed seamount in the Afar depression, *Bull. Volcanol.*, *38*, 26–43, 1975.

Harrison, C. G. A., R. D. Jarrard, V. Vacquier, and L. Larson, Paleomagnetism of Cretaceous Pacific seamounts, *Geophys. J. R. Astron. Soc.*, *42*, 859–882, 1975.

Hawkins, J., P. Lonsdale, and R. Batiza, Petrological evolution of the Louisville "Ridge" (LVR), *Eos Trans. AGU*, *66*, 405, 1985.

Jaeger, R. M., and D. J. Schuring, Spectrum analysis of terrain of Mare Cognitum, *J. Geophys. Res.*, *71*, 2023–2028, 1966.

Kellogg, O. D., *Foundations of Potential Theory*, Dover, New York, 1953.

Kent, D. V., and F. M. Gradstein, A Cretaceous and Jurassic geochronology, *Geol. Soc. Am. Bull.*, *96*, 1419–1427, 1985.

Kodama, K., and S. Uyeda, Magnetization of Izu Islands with special reference to Oshima Volcano, *J. Volcanol. Geotherm. Res.*, *6*, 353–373, 1979.

Kodama, K., S. Uyeda, and N. Isezaki, Paleomagnetism of Suiko Seamount, Emperor seamount chain, *Geophys. Res. Lett.*, *5*, 165–168, 1978.

Kono, M., Paleomagnetism of DSDP Leg 55 basalts and implications for the tectonics of the Pacific plate, *Initial Rep. Deep Sea Drill Proj.*, *55*, 737–758, 1977.

Lonsdale, P., Tectonic and magmatic ridges in the Eltanin fault system, South Pacific, *Mar. Geophys. Res.*, *8*, 203–242, 1986.

Lonsdale, P., and R. Batiza, Hyaloclastite and lava flows on young seamounts examined with a submersible, *Geol. Soc. Am. Bull.*, *91*, 545–554, 1980.

Lowrie, W., and D. V. Kent, Magnetic reversal frequency since the late Cretaceous, *Earth Planet. Sci. Lett.*, *62*, 305–313, 1983.

Luenberger, D. G., *Optimization by Vector Space Methods*, John Wiley, New York, 1969.

Lumb, J. T., M. P. Hochstein, and D. J. Woodward, Interpretation of magnetic measurements in the Cook Islands, South-west Pacific Ocean, in *The Western Pacific: Island Arcs, Marginal Seas and Geochemistry*, pp. 79–101, Crane, Russak and Co., New York, 1973.

McBirney, A. R., Oceanic volcanism: A review, *Rev. Geophys.*, *9*, 523–556, 1971.

McDougall, I., and H. U. Schmincke, Geochronology of Gran Canaria, Canary Islands: Age of shield building volcanism and other magmatic phases, *Bull. Volcanol.*, *40*, 1–21, 1976.

McElhinny, M. W., *Palaeomagnetism and Plate Tectonics*, Cambridge University Press, New York, 1973.

McNutt, M., Nonuniform magnetization of seamounts: A least squares approach, *J. Geophys. Res.*, *91*, 3686–3700, 1986.

McNutt, M. K., and H. W. Menard, Lithospheric flexure and uplifted atolls, *J. Geophys. Res.*, *83*, 1206–1212, 1978.

Menard, H. W., *Marine Geology of the Pacific*, McGraw-Hill, New York, 1964.

Merrill, R. T., and R. E. Burns, A detailed magnetic study of Cobb Seamount, *Earth Planet. Sci. Lett.*, *14*, 413–418, 1972.

Miles, P. R., and D. G. Roberts, The magnetization of Rosemary Bank Seamount, Rockall Trough northeast Atlantic, *Earth Planet. Sci. Lett.*, *54*, 442–448, 1981.

Morgan, W. J., Convection plumes in the lower mantle, *Nature*, *203*, 42–43, 1971.

Parker, R. L., The determination of seamount magnetism, *Geophys. J. R. Astron. Soc.*, *24*, 321–324, 1971.

Parker, R. L., The rapid calculation of potential anomalies, *Geophys. J. R. Astron. Soc.*, *31*, 447–455, 1973.

Parker, R. L., Linear inference and underparameterized models, *Rev. Geophys.*, *15*, 446–456, 1977.

Parker, R. L., A statistical theory of seamount magnetism, *J. Geophys. Res.*, in press, 1987.

Peddie, N. W., and E. B. Fabiano, A proposed international geomagnetic reference field for 1965–1985, *J. Geomagn. Geoelectr.*, *34*, 357–364, 1982.

Plouff, D., Gravity and magnetic fields of polygonal prisms and

applications to magnetic terrain corrections, *Geophysics*, *41*, 727–741, 1976.

Rice, P. D., J. M. Hall, and N. D. Opdyke, Deep drill 1972: A paleomagnetic study of the Bermuda seamount, *Can. J. Earth Sci.*, *17*, 232–243, 1980.

Richards, M. L., V. Vacquier, and G. D. Vanoorhis, Calculation of the magnetization of uplifts from combining topographic and magnetic surveys, *Geophysics*, *32*, 678–707, 1967.

Richter, F. M., and B. Parsons, On the interaction of two scales of convection in the mantle, *J. Geophys. Res.*, *80*, 2529–2541, 1975.

Sager, W., Seamount paleomagnetism and Pacific plate tectonics, Ph.D. thesis, University of Honolulu, Honolulu, Hawaii, 1983.

Sager, W., Paleomagnetism of Abbott Seamount and implications for the latitudinal drift of the Hawaiian hot spot, *J. Geophys. Res.*, *89*, 6271–6284, 1984.

Sager, W., G. T. Davis, B. H. Keating, and J. A. Philpotts, A geophysical and geologic study of Nagata seamount, northern Line Islands, *J. Geomagn. Geoelectr.*, *34*, 283–305, 1982.

Sard, A., *Linear Approximation*, American Mathematical Society, Providence, R. I., 1963.

Schmnke, G. R., and C. G. Bufe, Geophysical description of a Pacific Ocean seamount, *J. Geophys. Res.*, *73*, 559–569, 1968.

Shure, L., and R. L. Parker, An alternative explanation for intermediate-wavelength magnetic anomalies, *J. Geophys. Res.*, *86*, 11,600–11,608, 1981.

Smith, D. R., *Variational Methods in Optimization*, Prentice-Hall, Englewood Cliffs, N. J., 1974.

Staudigel, H., and H. U. Schmincke, The Pliocene Seamount series of La Palma, Canary Islands, *J. Geophys. Res.*, *89*, 11,195–11,215, 1984.

Stroud, A. H., *Approximate Calculation of Multiple Integrals*, Prentice-Hall, Englewood Cliffs, N. J., 1971.

Talwani, M., Computation with the help of a digital computer of magnetic anomalies caused by bodies of arbitrary shape, *Geophysics*, *30*, 797–817, 1965.

Uyeda, S., and M. L. Richards, Magnetization of four Pacific seamounts near the Japanese Islands, *Bull. Earthquake Res. Inst., Univ. Tokyo*, *44*, 179–213, 1966.

Vacquier, V., A machine method for computing the magnetization of a uniformly magnetized body from its shape and a magnetic survey, pp. 123–137, *Benedum Earth Magnetism Symposium*, University of Pittsburgh Press, Pittsburgh, Pa., 1962.

Vacquier, V., *Geomagnetism in Marine Geology*, Elsevier, New York, 1971.

Vacquier, V., and S. Uyeda, Paleomagnetism of nine seamounts in the western Pacific and of three volcanos in Japan, *Bull. Earthquake Res. Inst., Univ. Tokyo*, *45*, 815–848, 1967.

Watson, D. F., Acord: automatic contouring of raw data, *Comput. and Geol.*, *8*, 97–101, 1982.

Watts, A. B., J. H. Bodine, and N. M. Ribe, Observations of flexure and the geological evolution of the Pacific Ocean basin, *Nature*, *283*, 532–536, 1980.

J. A. Hildebrand, Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093.

R. L. Parker, Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093.

L. Shure, Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, MA 02543.

(Received July 2, 1986;
accepted December 16, 1986.)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>The Application of Inverse Theory to Seamount Magnetism | | 5. TYPE OF REPORT & PERIOD COVERED<br>6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Robert L. Parker, Loren Shure, and John A. Hildebrand | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-85-C-0104 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Scripps Institution of Oceanography<br>La Jolla, California 92093 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research<br>Arlington, Virginia 22217 | | 12. REPORT DATE<br>February 1987<br>13. NUMBER OF PAGES<br>24 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>unclassified<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Reprint from Reviews of Geophysics, Vol. 25, No. 1, 17-40, (1987).

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The traditional least squares method for modeling seamount magnetism is often unsatisfactory because the models fail to reproduce the observations accurately. We describe an alternative approach permitting a more complex internal structure, guaranteed to generate an external field in close agreement with the observed anomaly. Potential field inverse problems like this one are fundamentally incapable of a unique solution, and some criterion is mandatory for picking a plausible representative from the infinite-dimensional space of models all satisfying the data. Most of the candidates are unacceptable . . .

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102 LF 014 6601

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 92, NO. B12, PAGES 12,664–12,674, NOVEMBER 10, 1987

# Frequency Dependent Polarization Analysis of High-Frequency Seismograms

JEFFREY PARK[1]

*Geophysical Fluid Dynamics Program, Princeton University, Princeton, New Jersey*

FRANK L. VERNON III AND CRAIG R. LINDBERG

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

We present a multitaper algorithm to estimate the polarization of particle motion as a function of frequency from three-component seismic data. This algorithm is based on a singular value decomposition of a matrix of eigenspectra at a given frequency. The right complex eigenvector $\hat{z}$ corresponding to the largest singular value of the matrix has the same direction as the dominant polarization of seismic motion at that frequency. The elements of the polarization vector $\hat{z}$ specify the relative amplitudes and phases of motion measured along the recorded components within a chosen frequency band. The width of this frequency band is determined by the time-bandwidth product of the prolate spheroidal tapers used in the analysis. We manipulate the components of $\hat{z}$ to determine the apparent azimuth and angle of incidence of seismic motion as a function of frequency. The orthogonality of the eigentapers allows one to calculate easily uncertainties in the estimated azimuth and angle of incidence. We apply this algorithm to data from the Anza Seismic Telemetered Array in the frequency band $0 \leq f \leq 30$ Hz. The polarization is not always a smooth function of frequency and can exhibit sharp jumps, suggesting the existence of scattered modes within the crustal waveguide and/or receiver site resonances.

## 1. INTRODUCTION

The polarization of particle motion as measured by a three-component seismometer has been studied by a number of straightforward methods, most simply by tracing the projection of the motion as a function of time onto a chosen plane of reference. Although useful to illustrate the particle motion of simple arrivals, this practice is qualitative and less useful with complicated signals.

The problem of extracting a particular type of wave (e.g., *P*, *SH*, Rayleigh) from a noisy background has been studied by correlation techniques and special filters [e.g., *Kanasewich*, 1981; *Archambeau and Flinn*, 1965; *Vidale*, 1986]. Most of these techniques are designed for time domain analysis and implicitly assume that the waveform has essentially the same polarization over all or most frequencies. *Samson* [1977, 1983*a*,*b*,*c*] describes a method of estimating the polarization as a function of frequency. This is important for the analysis of seismic records. The seismic waveforms of local and regional distance events are often superpositions of direct, refracted, reflected, and scattered waves, with no guarantee that the polarization or phase are constant in frequency. In the presence of strong scattering, one might not expect a respectable "pure state" polarization at any frequency. Alternatively, coherent addition of scattered waves within the crustal waveguide will produce traveling modes whose signature in extended body wave codas may be a well-defined polarization and phase that varies with frequency. The distinct spectral peaks seen by *Park et al.* [this issue] in seismic spectra observed on the Anza Seismic Telemetered Array [*Berger et al.*, 1984] suggest that waveguide modes may be evident in the complex waveforms of events at epicentral distances of 100–250 km. Inhomogeneities in the crustal waveguide can lead to scattering and coupling of these propagating modes (see, e.g., *Kennett* [1986] and *Odom* [1986] for a description of these effects) which will, in general, cause frequency dependent scattering. In such cases, it is more useful to determine the type of seismic motion from its polarization signature, as in the study of *Vidale* [1986], than to attempt to isolate phases.

In this paper we develop and demonstrate another algorithm for determining the frequency dependence of the polarization of high-frequency seismic records. We have used multitaper spectral analysis [*Thomson*, 1982] to estimate the spectral density matrix $\mathbf{S}(f)$ of *Samson* [1983*a*]. This has several advantages. By employing prolate spheroidal wave functions as tapers (instead of cosine or boxcar tapers) to obtain direct spectral estimates, the elements of the estimated spectral density matrix will be less biased [*Lindberg*, 1986; *Park et al.*, 1987]. It is also not necessary to apply a moving average to the density matrix estimate to smooth it; smoothing is obtained by summing the eigenspectra of each component of motion (see equation (3)). Using multitapers to estimate the spectral density matrix is more suitable for very short records, such as those which include a single seismic phase. This is because data are not discarded by applying a single bell-shaped taper to the record. (A similar method has been independently developed and applied to magnetometer data by *Lanzerotti et al.* [1986].)

We analyze a number of three-component records of seismic codas. In these observations the source pulse has

[1]Now at Department of Geology and Geophysics, Yale University, New Haven, Connecticut.



Paper number 6B6235
0148-0227/87/0068-6235$05.00

been dispersed and scattered within the crust. In an idealized picture the shape of the source spectrum is retained in the shape of the coda spectrum, but the spectral phase is randomized by scattering effects. Despite this randomized phase, one might expect the particle motion to retain the polarization behavior of the type of wave motion dominant within a selected frequency band. Polarization analysis in the frequency domain offers an opportunity to characterize the signal better. With three-component data we have potentially three independent polarizations. If scattering is not great, a single polarization will predominate. This assumption is often true for the $P$ wave coda. If, for instance, interaction with crustal structure decouples $SH$ and $SV$ motion, there may be two principal polarizations in the $S$ wave coda. The algorithm we describe in this paper offers a quantitative criterion for identifying the single dominant polarization.

In section 2, our multitaper polarization analysis method is described. We apply the algorithm to a synthetic pulse example in section 3. In section 4 we show examples from the $P$ wave codas of data observed on the Anza Seismic Telemetered Network. Section 5 summarizes our findings. Uncertainty estimates for polarization angles and phases are derived in the appendix.

## 2. Polarization Analysis With the Multitaper Algorithm

Polarization analysis involves determining the eigenstructure of the spectral density matrix $\mathbf{S}(f)$. Suppose one has three-component data recorded in the time domain of the form

$$\mathbf{x}(t) = (x^1(t), x^2(t), x^3(t)) \qquad t = n\tau;\ n = 0, 1, \ldots, N-1$$

where $\tau$ is the sampling interval, $N\tau$ is the length of the time series, the coordinate system is right-handed, and $x^1(t)$ is the vertical component. If the $j$th record $x^j(t)$ has the frequency domain representation $z^j(f)$, the spectral density matrix $\mathbf{S}(f)$ has components

$$S_{jk}(f) = E\{z^j(f)^* z^k(f)\}$$

where $E$ denotes the expectation operator. *Samson* [1983*a*] forms an estimate of the spectral density matrix, $\hat{\mathbf{S}}(f)$, with components

$$\hat{S}_{jk}(f) = (y^j(f))^* y^k(f) \qquad i, j = 1, 2, 3$$

where

$$y^j(f) = \frac{1}{N\tau} \sum_{n=0}^{N-1} w_n x^j(n\tau) e^{-2\pi i f n\tau} \tag{1}$$

is a discrete Fourier transform of the $j$th component of $\mathbf{x}(t)$ and $\{w_n\}_0^{N-1}$ is a chosen data taper. The matrix $\hat{\mathbf{S}}(f)$ is then smoothed in the frequency domain by applying a moving average, and the eigenvectors and eigenvalues of the smoothed matrix are found.

To apply the multitaper algorithm to the estimation of $\mathbf{S}(f)$, one employs a set of $K$ prolate spheroidal wave function "eigentapers" $v_n^{(k)}(N, W)$: $k = 0, 1, \ldots, K-1$, which are optimally resistant to spectral leakage from outside a chosen frequency band of width $2W$ [*Thomson*, 1982; *Lindberg*, 1986; *Park et al.*, 1987]. For $k = 0, 1, \ldots, K-1$ the spectral estimates

$$y_k^j(f) = \frac{1}{N\tau} \sum_{n=0}^{N-1} v_n^{(k)}(N, W) x^j(n\tau) e^{-2\pi i f n\tau} \tag{2}$$

of each component of $\mathbf{x}(t)$ can be made. Then a multitaper estimate of the spectral density matrix is

$$\frac{1}{K} \mathbf{M}^H(f) \cdot \mathbf{M}(f) \tag{3}$$

where superscript $H$ denotes conjugate transpose and

$$\mathbf{M}(f) = \begin{bmatrix} y_0^{(1)}(f) & y_0^{(2)}(f) & y_0^{(3)}(f) \\ y_1^{(1)}(f) & y_1^{(2)}(f) & y_1^{(3)}(f) \\ \vdots & \vdots & \vdots \\ y_K^{(1)}(f) & y_K^{(2)}(f) & y_K^{(3)}(f) \end{bmatrix}$$

The value of $K$, the number of eigenspectra used, depends on $2W$, the width of the frequency band in which the spectral energy at frequency $f$ is concentrated. The $K = 2NW\tau - 1$ lowest order eigentapers possess sufficient spectral leakage resistance to be useful [*Slepian*, 1983].

To investigate the eigenstructure of $\mathbf{S}(f)$, we perform a singular value decomposition $\mathbf{M}(f) = \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{V}^H$, where $\mathbf{U}$ is a $K \times K$ unitary matrix of left eigenvectors of $\mathbf{M}$, $\mathbf{V}$ is a $3 \times 3$ unitary matrix of right eigenvectors $\mathbf{v}_j$ of $\mathbf{M}$, and $\mathbf{D}$ is a $K \times 3$ matrix with $D_{jj} = d_j$, $j = 1, 2, 3$, the singular values of $\mathbf{M}$, and $D_{ij} = 0$ for $i \neq j$ [*Golub and Van Loan*, 1983]. The polarization vector $\hat{\mathbf{z}}$ is the right eigenvector corresponding to the largest singular value of the matrix $\mathbf{M}$. It specifies the direction of particle motion at frequency $f$ which contains the largest fraction of seismic energy [*Samson*, 1983*b*]. The components of $\hat{\mathbf{z}}$ can be complex, allowing for phase lags between components. Phase lags between components represent elliptical particle motion. Our ability to identify $\hat{\mathbf{z}}$ with the principal polarization of motion at $f$ can be qualitatively assessed by comparing the singular values $d_1 \geq d_2 \geq d_3$. If $d_1 >> d_2, d_3$, the polarization $\hat{\mathbf{z}} = \hat{\mathbf{v}}_1$ is well determined. We can use the ratio of the singular values to estimate the uncertainty in $\hat{\mathbf{z}}$ and any quantities we calculate from it. The estimation of the polarization uncertainty follows the derivation of *Park and Chave* [1984] and is outlined in the appendix. If $d_1 \approx d_2 >> d_3$, there is a strong possibility that coherent seismic motion at $f_0$ exists at two separate polarizations. The dot product $\hat{\mathbf{v}}_1^* \cdot \hat{\mathbf{v}}_2 = 0$ by virtue of the singular value decomposition, but this orthogonal relationship need not carry over into the seismic polarizations. In an $S$ wave arrival, one expects $SV$ and $SH$ motion to be orthogonal to first order in most situations, but the superposition of other signals (e.g., reflected $P$ arrivals) need not have orthogonal polarizations.

If $d_1 >> d_2, d_3$, the three-component particle motion $\mathbf{x}(t)$ in the neighborhood of frequency $f$ can be represented by the real part of $R\hat{\mathbf{z}}e^{-2\pi i f t}$, where $R$ is the amplitude of motion. We can adjust the phase of $\hat{\mathbf{z}}$ so that $R$ is real. If there exists a phase $\phi$ such that $\hat{\mathbf{z}}e^{i\phi}$ is purely real, then all motion described by $\hat{\mathbf{z}}$ lies along a single line in three space. More generally, particle motion will follow an ellipse confined to the plane spanned by the two real vectors $\mathrm{Re}(\hat{\mathbf{z}})$ and $\mathrm{Im}(\hat{\mathbf{z}})$. If this ellipse is strongly elongated along its major axis, reasonable horizontal and vertical azimuths can be found. If the wave type is known, such as a $P$ wave, then the propagation direction can be determined. Strongly elliptical polarization suggests modelike particle motion (for example, a Rayleigh wave) with a poorly defined angle of incidence.

We can project the particle motion described by the complex unit vector $\hat{\mathbf{z}}$ onto an ellipse in the horizontal

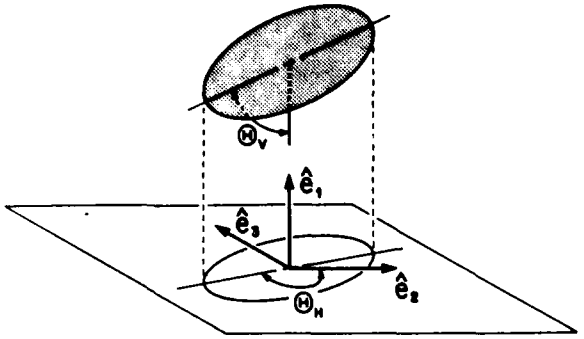



Fig. 1. Diagram to illustrate the definitions of the polarization angles $\Theta_H$ and $\Theta_V$. The azimuth $\Theta_H$ is restricted to [−180°, 180°] and is measured counterclockwise from $\hat{e}_2$. The angle $\Theta_H$ is chosen by determining the maximum horizontal displacement of the particle motion for which $\Theta_1$ will fall in the range $0 \leq \Theta_1 \leq 90°$. The ellipticity of the particle motion is defined by the amplitudes $|z_1|$, $|z_2|$, $|z_3|$ and the phase angles $\phi_{HH}$ and $\phi_{1H}$ (defined in text).

plane which is defined by $\mathbf{z}_H = \hat{\mathbf{z}} - (\hat{\mathbf{e}}_1 \cdot \hat{\mathbf{z}})\hat{\mathbf{e}}_1$, where $\hat{\mathbf{e}}_1 = (1,0,0)$. The major axis of this horizontal ellipse is taken to be the principal direction of horizontally polarized motion. To find the azimuth of the major axis, we determine the point of greatest displacement for the projection $\mathbf{z}_H$ in the horizontal plane by finding the maximum value of

$$|\mathrm{Re}\,(\mathbf{z}_H e^{i2\pi ft})|^2 \tag{4}$$

If the components $(z_1,z_2,z_3)$ of $\hat{\mathbf{z}}$ are expressed in the form $z_j = |z_j| e^{i\phi_j}$, this is equivalent to finding the maxima of

$$|z_2|^2\cos^2(2\pi ft+\phi_2) + |z_3|^2\cos^2(2\pi ft+\phi_3) \tag{5}$$

The extremes of this expression, remembering

$$|z_2|^2\sin 2\phi_2 + |z_3|^2\sin 2\phi_3 = \mathrm{Im}[z_2^* z_3] \tag{6}$$

are found when the phase angle $\theta$ defined as $\theta = 2\pi ft$ takes the values

$$\theta_l = -\tfrac{1}{2}\arg[z_2^* z_3] + \frac{l\pi}{2} \tag{7}$$

where $l$ is an integer. Let $l$ be the integer closest to zero which minimizes (5), the horizontal displacement, and for which $\mathrm{Re}\,(z_1) < 0$. Define the phase angle $\theta_H$ to be the value of $\theta_l$ for this $l$. Once $\theta_H$ has been determined, the horizontal azimuth of the major axis $\Theta_H$ measured counterclockwise from $\hat{\mathbf{e}}_2 = (0,1,0)$ can be defined as

$$\Theta_H = \tan^{-1}\left[\frac{\mathrm{Re}(z_3 e^{-i\theta_H})}{\mathrm{Re}(z_2 e^{-i\theta_H})}\right] = \mathrm{Re}(\tan^{-1}(z_3/z_2)) \tag{8}$$

The range of the arctangent function is $0° < \Theta_H \leq 180°$ if $\mathrm{Re}(z_1 z_3^*) < 0$ and $-180° < \Theta_H \leq 0°$ if $\mathrm{Re}(z_1 z_3^*) \geq 0$. If the particle motion is $P$ like, $\Theta_H$ can be interpreted as pointing in the direction of the wave source. A representation of an elliptical motion for which $\Theta_H < 0$ is shown in Figure 1.

Another useful quantity is $\phi_3 - \phi_2 = \phi_{HH}$, the phase difference between the horizontal components of particle motion. If $\phi_2 - \phi_3 \approx 0°$ or 180°, the particle motion is predominantly linear. The value $\phi_2 - \phi_3 \approx 90°$ represents elliptical motion with the major and minor axes oriented along the axes of the instruments. If $z_2 = \pm i z_3$, the particle motion is circular, with no definable azimuth. In this case, the uncertainty in $\Theta_H$, given in the appendix, goes to infinity as it is proportional to $|z_2^2 + z_3^2|^{-1}$.

The expressions relating horizontal to vertical motion are similar. We want to find the angle $\Theta_V$ made with the vertical by the major axis of the ellipse defined by $\mathrm{Re}(\tilde{\mathbf{z}} e^{-2\pi ft})$. Define the phase angles

$$\theta_m = 2\pi ft = -\tfrac{1}{2}\arg[z_1^2 + z_H^2] + \frac{m\pi}{2} \tag{9}$$

where $m$ is an integer and $z_H^2 = z_2^2 + z_3^2$. The phase angle $\theta_1$ is the value of $\theta_m$ at an $m$ for which the particle motion displacement is maximized. The angle of incidence is

$$\Theta_1 = \tan^{-1}\left|\frac{\mathrm{Re}[z_1 e^{-i\theta_1}]}{\mathrm{Re}[z_H e^{-i\theta_1}]}\right| \tag{10}$$

where $\mathrm{Im}\, z_H \geq 0$. The absolute value is taken to restrict $\Theta_1$ to lie between 0° and 90°, the usual convention for the angle of incidence (Figure 1). The phase lag between vertical and horizontal motion can also be defined. Define $\phi_{1H} = \theta_H - \phi_1$. Since the end points of the major axis of the horizontal motion ellipse correspond to $\theta_H$ and $\theta_H \pm \pi$, we can restrict the range of $\phi_{1H}$ to (−90°, 90°).

## 3. A Synthetic Example

We first illustrate the definitions of $\Theta_H$, $\Theta_1$, $\phi_{HH}$, and $\phi_{1H}$ in a synthetic example. We constructed a three-component record (Figure 2) from a sum of cosinusoids:

$$\begin{aligned}
x^1(n\tau) &= \sum_{f=0}^{100} \cos\left[\frac{\pi f}{80}\right] \cos\left(2\pi f n\tau - \frac{\pi f}{50}\right) \\
x^2(n\tau) &= \sum_{f=0}^{100} \cos\left[\frac{\pi f}{20}\right] \sin\left[\frac{\pi f}{80}\right] \cos(2\pi f n\tau) \\
x^3(n\tau) &= \sum_{f=0}^{100} \sin\left[\frac{\pi f}{20}\right] \sin\left[\frac{\pi f}{80}\right] \cos(2\pi f n\tau)
\end{aligned} \tag{11}$$

where $n = 0, 1, \ldots, N-1$ and the sampling interval is $\tau = 0.004$ s. The polarization vector of this signal can be written immediately as

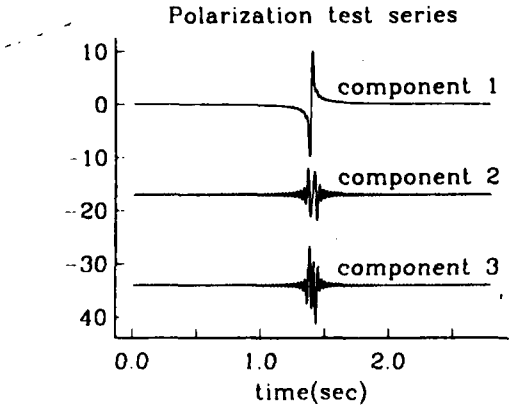



Fig. 2. Polarization test series.

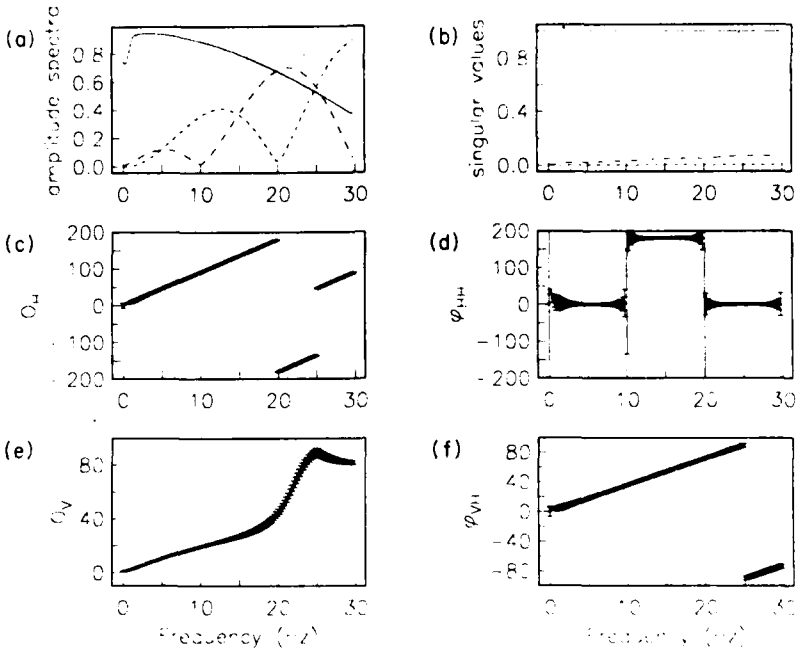



Fig. 3. (*a*) Amplitude spectra and polarization angles calculated from the test series. Spectra for components 1 (solid line), 2 (coarse dashed line), and 3 (fine dashed line). (*b*) The singular value associated with principal polarization is plotted against frequency (solid line), and the secondary singular values (dashed lines). (*c*) Horizontal azimuth of particle motion. (*d*) Phase angle defined by the major and minor axis of the horizontal particle motion ellipse. (*e*) Angle of incidence of particle motion measured from nadir. (*f*) Phase angle defined by major and minor axis of the vertical particle motion ellipse.

$$\hat{z}(f) = \left[\cos\left(\frac{\pi f}{80}\right)e^{i\phi_{1H}}, \cos\left(\frac{\pi f}{20}\right)\sin\left(\frac{\pi f}{80}\right), \sin\left(\frac{\pi f}{20}\right)\cdot\sin\left(\frac{\pi f}{80}\right)\right] \quad (12)$$

where $\phi_{HH} = 0$ and $\phi_{1H} = (-\pi f)/50$. Figure 3 shows the results of a multitaper polarization analysis for frequencies $0 \leqslant f \leqslant 30$ Hz. The uncertainties are plotted as one standard deviation error bars in this and succeeding figures. Figure 3*b* shows the three scaled singular values as a function of frequency. The principal polarization appears well determined. The amplitude spectra for the three components are plotted in Figure 3*a*. The angles $\Theta_H$ and $\phi_{HH}$ are plotted in Figures 3*c* and 3*d*. The angle $\phi_{HH}$ is not well determined near zero frequency, as the horizontal signal amplitude is dwarfed by vertical component energy. The apparent horizontal azimuth $\Theta_H$ "wraps around" from 180° to −180° at 20 Hz and jumps 180° at 25 Hz. The former jump is obvious; the latter is an artifact of $\phi_{1H}$ passing through 90°. The phase angle $\phi_{HH}$, estimated from the synthetic record, has a value of 0° or ±180°, to observational accuracies. These values correspond to rectilinear motion and are dependent on the quadrant where the horizontal azimuth is directed. The phase lag $\phi_{1H}$ between vertical and horizontal components is well determined everywhere except very near zero frequency where the horizontal component amplitude vanishes. The ellipticity of particle motion disrupts the linear trend in $\Theta_1$, as shown in Figure 3*e*. At 25 Hz, $\phi_{1H} = 90°$ and the particle motion is an ellipse with major and minor axes oriented horizontally and vertically, respectively. Therefore $\Theta_1 = 90°$ at 25 Hz. At higher frequencies, $\phi_{1H} > 90°$, the relative sign of vertical and horizontal motion reverses, and the particle motion ellipse "tips" in an opposite manner relative to its orientation for $\phi_{1H} < 90°$. This causes the observed 180° jump in apparent horizontal azimuth $\Theta_H$. This example suggests that one should use caution in interpreting the angles $\Theta_H$ and $\Theta_1$ wherever the particle motion is nearly fully elliptical, i.e., when $\phi_{HH}$ or $\phi_{1H}$ is within 20° of ±90°.

## 4. Data Examples

We illustrate this method of determining the polarization as a function of frequency with several examples. We analyzed several waveforms which were recorded on the Anza array after an earthquake that occurred at 0521:39.5 UT, September 9, 1982, with hypocenter positioned at 32.93°N, 115.85°W, and depth 4.2 km. The magnitude $M_l$ was determined to be 4.4. The event was located near Superstition Mountain, California, on the western edge of the Imperial Valley. The earthquake was recorded on only four stations in the array (PFO, KNW, FRD, and CRY; see *Berger et al.* [1984] for the definitions of these three-letter acronyms) as the event occurred prior to the completion of the array. The hypocenter was roughly 100 km southeast of the array. The $m = 1$ component is the vertical seismometer output with positive motion defined as up. We choose the $m = 2$ component so that positive motion points 45° east of north. Positive motion along the $m = 3$ axis is directed 45° west of north, forming a right-handed coordinate system. Let the angle $\Theta_H$ be measured counterclockwise from the primary hor-

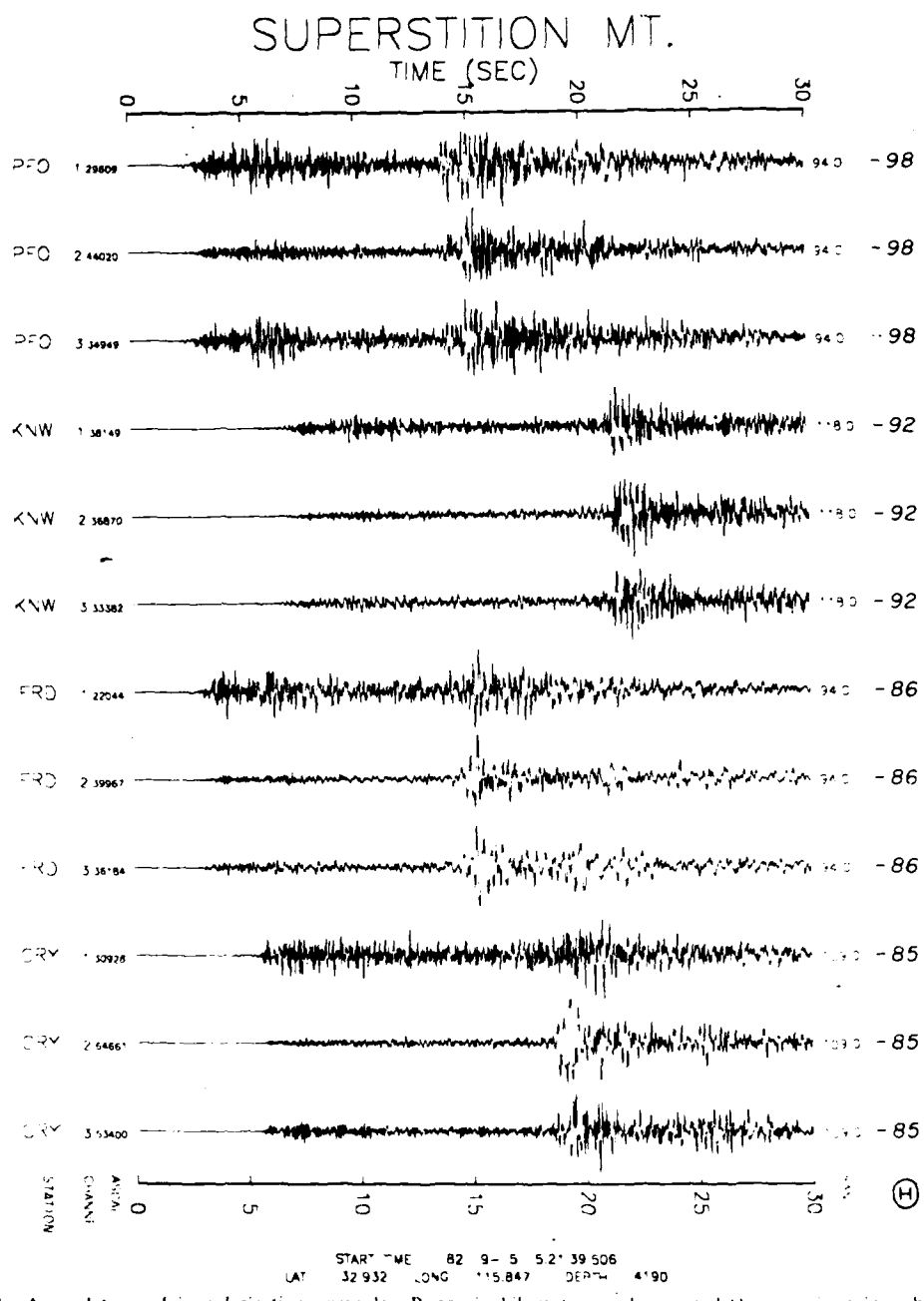



Fig. 4. Anza data used in polarization example. Range in kilometers and expected $\Theta_H$ are given in right-hand columns. Maximum amplitude in counts is given in left-hand column, along with station code and component number.

izontal axis ($m = 2$). If the wave propagation is along a straight line connecting the source and receiver, $-85° > \Theta_H > -98°$ for the four stations. The first 30 s of recorded motion for this event are shown in Figure 4, along with range and azimuth information (azimuth is measured counterclockwise from N45°E). Both *S* and *P* arrivals are extended wave trains, although the *S* energy is more concentrated in time. An interesting feature of this event is the small precursor to the main *P* arrival, shown in the enlarged detail for stations FRD and CRY in Figures 5*a* and 5*b*. This waveform corresponds to a lower crustal phase.

Polarization analysis reveals that the first arrivals have complicated polarization signatures. The time window taken is short (1.6 s), corresponding to a Rayleigh frequency $1/(N\tau)$ of 0.625 Hz. Analysis using seven $4\pi$ prolate tapers averages energy over a band of width $8/(N\tau)$, so that all of the estimates shown represent an average over a 5-Hz frequency band. If the true polarization varied significantly over this bandwidth, one would expect $\Theta_V$, $\Theta_H$, $\phi_{VH}$, and $\phi_{HH}$ to be relatively poorly determined. The results for FRD are shown in Figure 6. The singular values $d_2$ and $d_3$ displayed in Figure 6*b* show local maxima at several places in the spectrum from 0 to 30 Hz. Max-

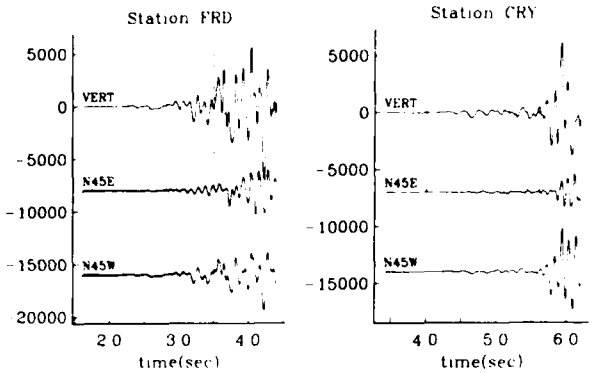



Fig. 5. (*a*) Plots of precursory waveform observed on station FRD. (*b*) Plots of precursory waveform observed on station CRY. The portion used for spectrum analysis is bounded by dashed lines. Both horizontal components at station FRD exhibit visible 60-Hz power line noise. The spectral leakage resistance of the $4\pi$ prolate eigentapers used in the analysis guards against bias in the frequency band of interest.

ima at 2.5, 7.5, and 14 Hz correspond to boundaries between distinct spectral features (Figure 6*a*). All the maxima below 25 Hz correspond to frequencies at which one or more of the polarization angles change rapidly. Horizontal motion is roughly rectilinear below 13 Hz, but its azimuth is variable and significantly different from the nominal azimuth of −87°. In fact, the largest amplitude signal, from 8 to 13 Hz, is oriented clockwise 125° from the primary component, a deflection of nearly 40° from the nominal *P* wave arrival azimuth. The phase lag between horizontal and vertical motion is alternately positive and negative in adjacent frequency bands but is never more than partially elliptical. The angle between vertical and horizontal motion, which can be interpreted in this case as the nominal angle of incidence, varies smoothly with frequency in Figure 6*e*, with $\Theta_1 \approx -25°$ 30° for $f \leq 10$ Hz, and $\Theta_1 \approx 15°$ above 13 Hz.

Figure 7 shows an analysis of the small amplitude *P* precursor observed at station CRY. The variation of the largest singular value $d_1$ with frequency shows four frequencies (2.5, 7, 12, and 16 Hz) at which the principal polarization vector is poorly determined, and there is a peak in $d_2$. Each of these peaks in $d_2$ occurs where there is an abrupt change in the three-component spectra and in one or more of the polarization angles. Although the estimated uncertainties are larger than those in the last example, the variability among frequency bands is clearly visible in Figures 7*c*–*f*. Motion in the horizontal plane is dominantly elliptical below 14 Hz, but particle rotation proceeds in opposite senses in the two frequency bands 2.5 Hz $\leq f \leq$ 7 Hz and 7 Hz $\leq f \leq$ 14 Hz. The azimuthal angle $\Theta_H$ hovers near the value expected for the epicenter (−85°), but our synthetic example in Figure 3 suggests that this may be due to the −90° phase lag between component motions. At higher frequencies, including the substantial spectral peak at 18–20 Hz, the observed horizontal azimuth of particle motion is roughly transverse to the arrival azimuth, as though the energy at these frequencies were *SH* in character. A better interpretation is in terms of side-scattered *P* energy, as the vertical azimuth of particle motion $\Theta_1$ remains in the 20°–40° range across all frequencies in Figure 7*e*.

Similar behavior is observed on stations PFO and KNW. The nature of this polarization behavior is quite puzzling. It is unlikely that instrument calibrations are at fault. A timing error among components would result in a linear drift in the relative phase angles, similar to that shown in Figure 3*f*. There are no poles or zeroes in the instrument response over the frequency region shown. A perturba-

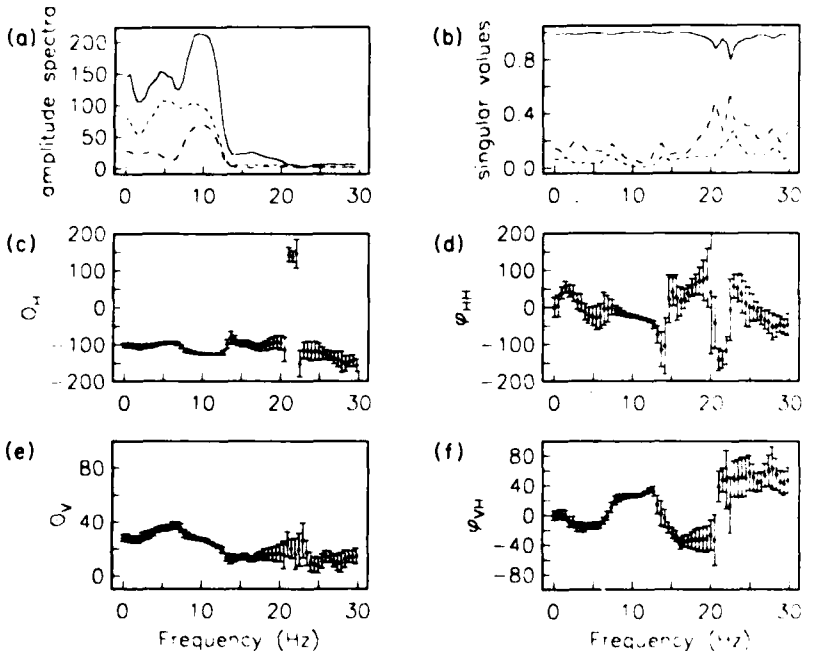



Fig. 6. Amplitude spectra and polarization angles for precursory waveform observed at station FRD. Solid/dashed line conventions are identical to those of Figure 3.

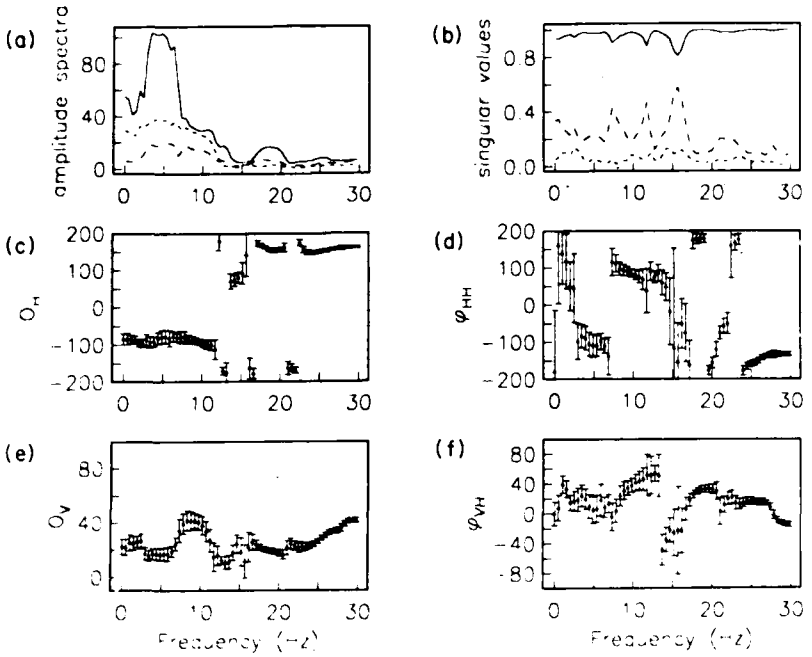



Fig. 7. Amplitude spectra and polarization angles for precursory waveform observed at station CRY. Solid/dashed line conventions are identical to those of Figure 3.

tion in the response filter characteristics would have difficulty mimicking the apparent boundaries between spectral processes. Moreover, we show below that the relative polarization shift from frequency band to frequency band varies greatly within the *P* coda. This argues for a signal-generated effect rather than an instrument effect. This behavior may reflect the modal structure of an intercrustal head wave in a stratified crust. Another interpretation is in terms of resonant vibrational modes in the earth structure near the receiver. Structure of scale lengths 100–200 m could account for the higher-frequency resonances observed in Figures 6 and 7.

We performed experiments to see if such resonant behavior could be found in the *P* codas for this event. When the entire coda was used for polarization analysis, the results were poor. The three-component seismogram recorded at station KNW is shown in Figure 8. Figure 9 presents polarization data from the 14-s *P* wave coda. There appear to be competing signals at nearby frequencies, creating either rapid variations in the polarization, which are difficult to interpret, or else large uncertainties in the polarization. Likewise, the presence of both *SV*- and *SH*-polarized energy in the *S* arrivals made the identification of a "principal" polarization uncertain.

We chose, therefore, to analyze the *P* codas of these records in successive 2-s (500 sample) segments. We observed what appear to be resonances over 4–6 Hz frequency bands and variations in polarization over time that suggest the arrival of *P* energy which has been scattered within the crust. The results of a polarization analysis of the first, fourth, and sixth 2-s time segments of the *P* wave coda recorded at station KNW are shown in Figures 10–12. The growth of the "noise" singular values $d_2$ and $d_3$ as the time window moves through the coda suggests an increase in scattered energy. The most prominent features in the spectra of the principal polarization components are the spectral peaks near 5 and 14 Hz. Comparison of the values of $\Theta_H$ in the time windows indicates that there is a boundary between two distinct spectral processes at 7–7.5 Hz. The 7–14 Hz process is characterized by dominantly rectilinear horizontal motion and steeply vertical particle motion. The relative phase angles $\phi_{VH}$ and $\phi_{HH}$ for the lower-frequency process exhibit more variability. Within a 2-s time window the horizontal azimuth varies only slightly within the 0–7 Hz frequency band, with more shallow vertical angles. Figure 12*e* shows that $\Theta_V \geq 60°$ in this frequency band, which may indicate *SV*-converted motion. Particle motion at frequencies greater than 15 Hz bears little relation to the higher-amplitude low-frequency signal and often cannot be interpreted in terms of *P*-, *SV*- or *SH*-polarized motion traveling directly from source to receiver.

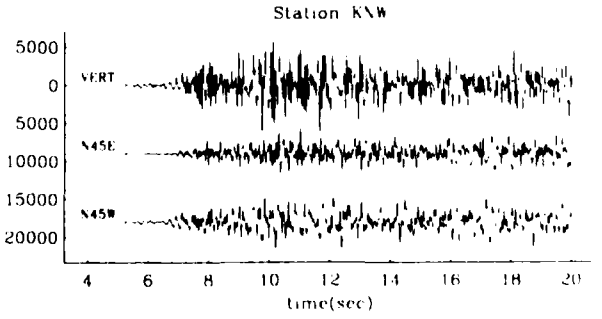



Fig. 8. Three-component seismogram for Superstition Mountain event observed at Anza station KNW. The 14-s segment chosen for polarization analysis is within the dashed lines.

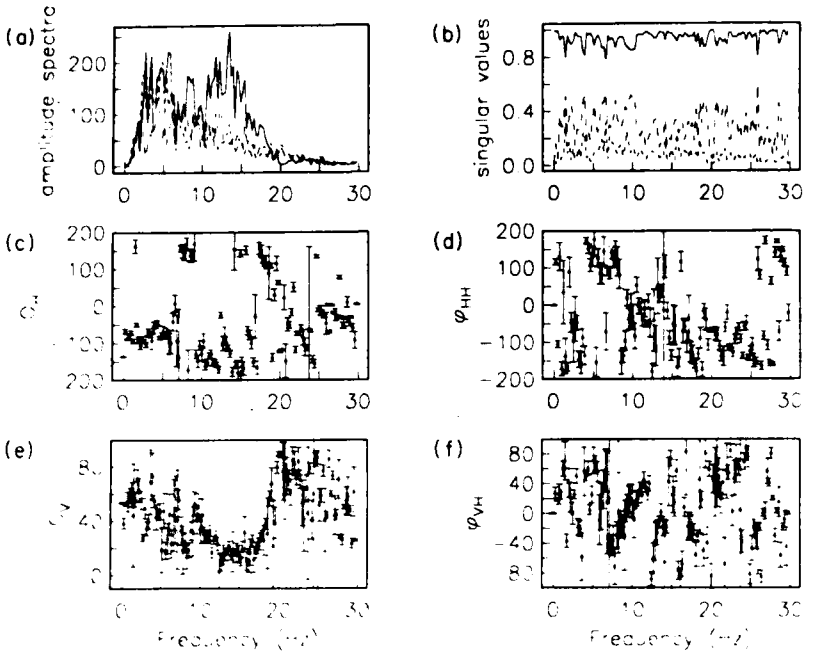



Fig. 9. Amplitude spectra and polarization angles for the 14-s $P$ coda segment shown in Figure 8. Solid/dashed line conventions are identical to those of Figure 3.

The similar frequency dependences of $\Theta_H$ and $\Theta_1$ in these 2-s time windows contrasts with the absence of a clear pattern in the larger time analysis shown in Figure 9. Similar effects are found when records from the other three stations for this event are analyzed. This is not surprising when one notes the large variation of polarization among the three time windows shown in Figures 10–12. The azimuth of the epicenter has $\Theta_H = -92°$ (i.e., clockwise) from the second component. The horizontal azimuth $\Theta_H$ of particle motion is, for 7.5 Hz $< f <$ 14 Hz, always oriented more to the south, with values that vary among time windows by 40° or more. At $f \leq 7$ Hz, several of the time windows tested were consistent with −92° relative azimuth, but the fourth and

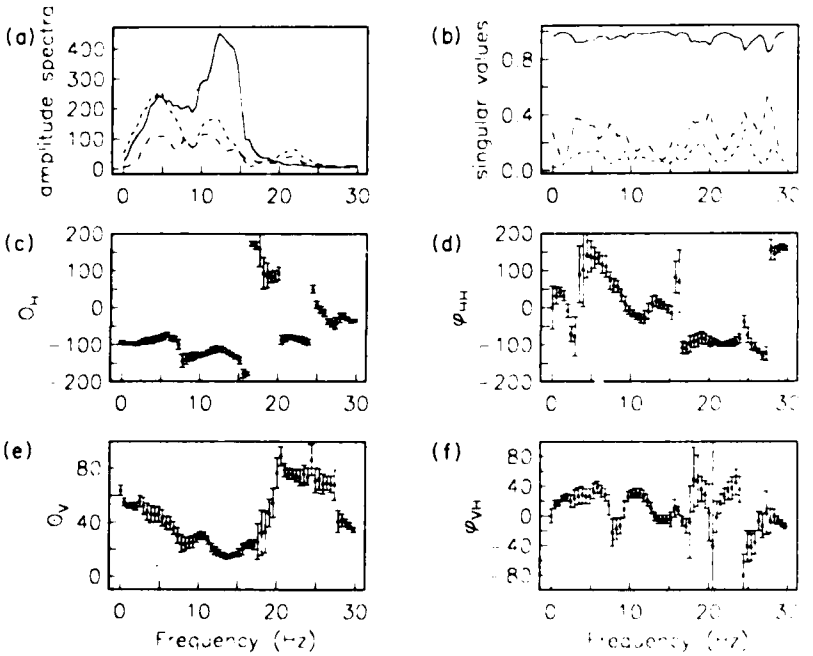



Fig. 10. Amplitude spectra and polarizaton angles for the first 2 s of the 14-s $P$ coda segment shown in Figure 8. Solid/dashed line conventions are identical to those of Figure 3.

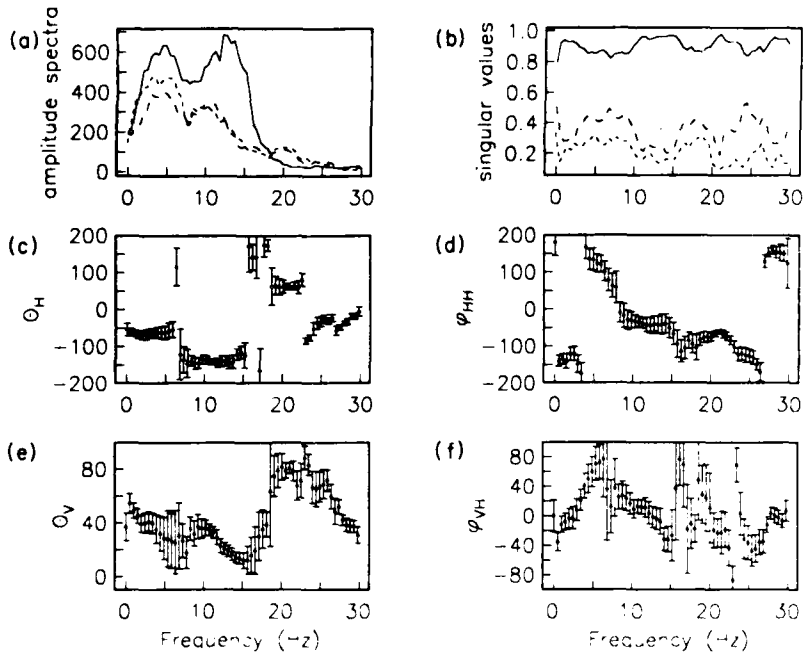



Fig. 11. Amplitude spectra and polarization angles for the seventh and eighth seconds of the 14-s *P* coda segment shown in Figure 8. Solid/dashed line conventions are identical to those of Figure 3.

sixth segments, shown in Figures 11 and 12, show particle motion whose horizontal orientation is nearly pure east-west. We take this variation as evidence for the arrival of scattered off-azimuth *P* energy.

A detailed interpretation of these results is beyond the scope of this paper, but we can draw parallels with recent studies of high frequency seismic spectra. *Sereno and Orcutt* [1985] have shown that the extended $P_n$ wave train observed in ocean bottom seismic data can be modeled by reverberations in the oceanic sediment layer and overlying water column, buttressing their comparison by demonstrating a simple pattern of spectral peaks corresponding to leaky vibrational modes. *Bard and Bouchon* [1985] have shown spectra from seismic events for which the retrieval

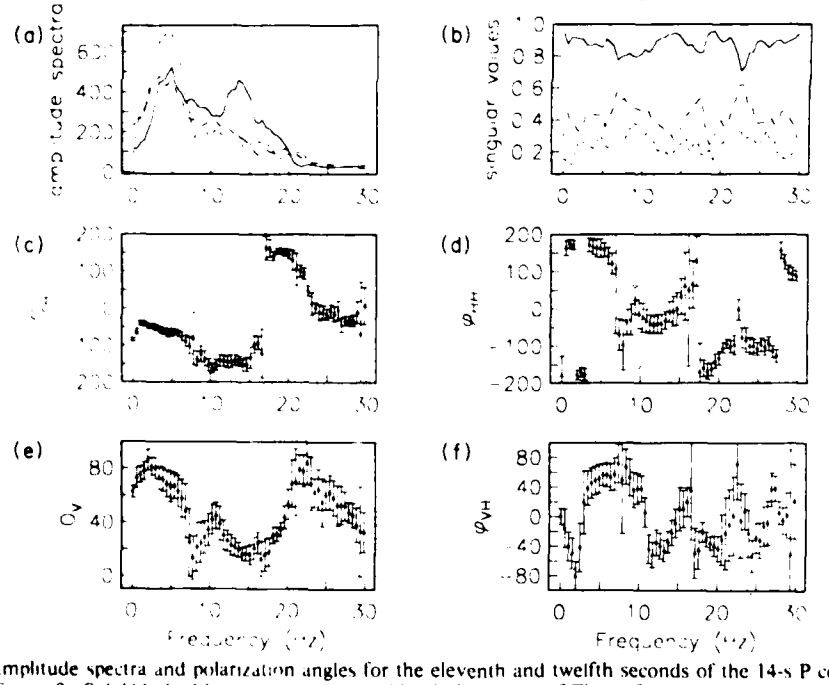



Fig. 12. Amplitude spectra and polarization angles for the eleventh and twelfth seconds of the 14-s P coda segment shown in Figure 8. Solid/dashed line conventions are identical to those of Figure 3.

of simple source parameters like corner frequency and high-frequency roll-off is contaminated by a high-frequency resonance which they model as a reverberation in the low-velocity surface layer. The apparent polarization resonances observed in the $P$ wave codas of the September 1982 Superstition Mountain event probably argue for an even more complex structure than was postulated in these studies.

The interpretation of the coda using resonance models may offer a more direct method for characterizing near-receiver structure than time-domain models of scattered waves [e.g., *Sato*, 1984]. If the resonances of the structure beneath one's receivers are known, we can hope to determine better the spectral shape of the original seismic source. If we model the response $\mathbf{R}(f,\hat{\mathbf{z}})$ of the crustal structure local to a receiver to waves traveling in the lithospheric wave guide with frequency $f$ and polarization $\hat{\mathbf{z}}$, we expect observed three-component amplitude spectra $\mathbf{U}(f)$ to be found by integrating

$$\mathbf{U}(f) = \int_{\Omega} \mathbf{R}(f,\hat{\mathbf{z}}) s(f,\hat{\mathbf{z}}) d\Omega \tag{13}$$

where $s(f,\hat{\mathbf{z}})$ is the amplitude of the impinging signal. We integrate $\hat{\mathbf{z}}$ over the lower half of the unit sphere in order to account for energy arriving from all vertical azimuths and out of plane scattering. In the example of *Sereno and Orcutt* [1985], $\mathbf{R}(f,\hat{\mathbf{z}})$ was calculated for a simple layered model. For arrays (such as Anza) positioned atop a heterogeneous medium, constraints on $\mathbf{R}(f,\hat{\mathbf{z}})$ can be found empirically using a number of events at different azimuths. Determination of $\mathbf{R}(f,\hat{\mathbf{z}})$ may be helpful in evaluating the earthquake hazards of a potential building site, especially as polarization analysis specifies both seismic amplitude and particle motion at the recording site. More research is necessary to determine if such a project is feasible. The above examples suggest that $s(f,\hat{\mathbf{z}})$ varies significantly within the coda, complicating the determination of the near-receiver resonant structure.

## 5. Conclusions

We have devised a multitaper algorithm to determine the polarization of particle motion as a function of frequency and applied it to data recorded on the Anza Seismic Telemetered Array [*Berger et al.*, 1984]. We form a matrix of eigenspectra of three-component records and perform a singular value decomposition to estimate the complex-valued unit vector $\hat{\mathbf{z}}$ whose components specify the sense of particle motion in the plane defined by the two real vectors $\mathrm{Re}(\hat{\mathbf{z}})$ and $\mathrm{Im}(\hat{\mathbf{z}})$. We manipulate the components of $\hat{\mathbf{z}}$ in order to specify four angles. The angle $\phi_{HH}$ represents the relative phase between the components of horizontal motion. The angle $\phi_{HH} \cong 0°$ or $\pm 180°$ if the particle motion is rectilinear in the horizontal plane, and $\phi_{HH} \approx \pm 90°$ if the motion is elliptical and oriented along the component axes. The phase angle $\phi_{ZH}$ is the relative phase between horizontal and vertical motion. The apparent azimuth $\Theta_H$ is defined by the maximum displacement of the horizontal projection of the particle motion ellipse. It is measured in the counterclockwise direction from the first horizontal component. Finally, an angle of incidence $\Theta_i$ of the particle motion is estimated. The uncertainties in these polarization angles can be estimated from the singular value decomposition used to obtain $\hat{\mathbf{z}}$ (appendix).

The variability of the spectra and polarization over $0 \leqslant f \leqslant 30$ Hz suggest that the $P$ coda observations can be separated into several distinct varieties of seismic motions, each occupying a separate frequency band. This behavior suggests that in some cases it may be more appropriate to model the $P$ wave coda as a set of resonant modes caused by near-receiver structure rather than a number of randomly scattered compressional pulses. Evidence for scattered energy is not lacking, however, as the principal polarization accounts for a smaller proportion of the total seismic energy late in the $P$ coda, accounting for only 60–65% in some frequency bands. We also observe that the apparent $P$ wave arrival azimuth can vary by up to 50°, both between adjacent frequency bands and in adjacent time windows. Both rectilinear and elliptically polarized signals are found, often coexisting in the same time window in adjacent frequency bands. We find that the apparent modal structure of the signal polarization breaks down if the length of the time window is much greater than 2 s, suggesting incoherent excitation by direct and scattered seismic waves.

We are currently investigating the polarization behavior of the data recorded at each site in the Anza array. We want to use the polarization information to obtain better estimates of the seismic source spectrum. Such an endeavor requres that one be able to identify the factors causing the apparent jumps in polarization, both as a function of frequency and time.

## Appendix
## Formal Uncertainty of Polarization Estimates

We estimate the uncertainties in the angles $\Theta_i$, $\Theta_H$, $\phi_{ZH}$, $\phi_{HH}$ from uncertainties $\delta\hat{\mathbf{z}}(f)$ in the unit eigenvector $\hat{\mathbf{z}}(f)$, which represents the principal polarization of particle motion at frequency $f$. The derivation of the rms expectation of $\delta\hat{\mathbf{z}}$ can be found in the work by *Park and Chave* [1984]. We only define the problem and state the results here. The vector $\hat{\mathbf{z}} = \hat{\mathbf{v}}_1$, the right eigenvector of $\mathbf{M}$ (defined in (3)) associated with largest singular value $d_1$. The uncertainty $\sigma$ is estimated from the two smaller singular values

$$\sigma^2 = \frac{K}{K-1}(d_2^2 + d_3^2)/2 \tag{A1}$$

where $K$ is the number of eigenspectra used in forming $\mathbf{A}$. The covariance matrix for the first-order uncertainty $\delta\hat{\mathbf{z}}$ has expectation value

$$\delta\hat{\mathbf{z}} \otimes (\delta\hat{\mathbf{z}})^* = \frac{\sigma^2}{K d_1} \sum_{j=2}^{3} \hat{\mathbf{v}}_j \otimes (\hat{\mathbf{v}}_j)^* \tag{A2}$$

It is also true that $\delta\hat{\mathbf{z}} \otimes \delta\hat{\mathbf{z}} = 0$. The $\otimes$ symbol denotes the tensor (outer) product of two vectors. Since

$$\delta\hat{\mathbf{z}} \otimes (\delta\hat{\mathbf{z}})^*_{ij} = \delta z_i \delta z_j^*$$

we have complete information on the formal uncertainties of the components of the principal polarization. Note that since $(\delta\hat{\mathbf{z}})^* \cdot \hat{\mathbf{z}} = 0$ as $\hat{\mathbf{z}}$ is a vector of unit length, $\delta\hat{\mathbf{z}}$ is composed of $\hat{\mathbf{v}}_2$ and $\hat{\mathbf{v}}_3$, the right eigenvectors associated with the "noise" singular values $d_2$, $d_3$.

Given (A2), we can determine the formal first-order uncertainty of any well-behaved function of $\hat{z} = (z_1, z_2, z_3)$; $\phi_{HH} = \phi_2 - \phi_3 = \arg(z_2) - \arg(z_3)$ to within an additive constant. Since

$$\delta\phi_j = \frac{i(\delta z_j^* z_j - \delta z_j z_j^*)}{2|z_j|^2} = \mathrm{Im}\left\{\frac{\delta z}{z}\right\} \tag{A3}$$

$$|\delta\phi_{HH}|^2 = \tfrac{1}{2}\left\{\frac{|\delta z_2|^2}{|z_2|^2} - 2\,\mathrm{Re}\left\{\frac{\delta z_2 \delta z_3^*}{z_2 z_3^*}\right\} + \frac{|\delta z_3|^2}{|z_3|^2}\right\} \tag{A4}$$

Since $\Theta_H = \mathrm{Re}(\tan^{-1}(z_y/z_x))$,

$$\delta\Theta_H = \mathrm{Re}\left\{\frac{z_x^2}{z_x^2 + z_y^2} \cdot \delta\left\{\frac{z_3}{z_2}\right\}\right\} = \mathrm{Re}\left\{\frac{z_x \delta z_y - z_y \delta z_x}{z_x^2 + z_y^2}\right\} \tag{A5}$$

so that

$$|\delta\Theta_H|^2 = \tfrac{1}{2}|z_x^2 + z_y^2|^{-2}(|z_3|^2 |\delta z_2|^2 - 2\mathrm{Re}[z_2^* z_3 \delta z_2 \delta z_3^*] + |z_2|^2 |\delta z_3|^2) \tag{A6}$$

Note that $|\delta\Theta_H|^2 \to \infty$ as $z_2 \to iz_3$, i.e., circular polarization.

The uncertainties of the vertical polarization angles $\Theta_1$ and $\theta_{1H}$ are similar. With $\Theta_1$ given by (8), where $\hat{z}^2 = z_x^2 + z_y^2$, we use the relation $\delta\hat{z} = \hat{z}^{-1}(z_x \delta z_x + z_y \delta z_y)$ to find

$$\delta\Theta_1 = \mathrm{Re}\left\{\hat{z}^{-1}\left\{\frac{\hat{z}^2 \delta z_1 - z_1 z_2 \delta z_2 - z_1 z_3 \delta z_3}{z_1^2 + z_2^2 + z_3^2}\right\}\right\} \tag{A7}$$

and

$$|\delta\Theta_1|^2 = \tfrac{1}{2}\left\{\frac{X(\hat{z}, \delta\hat{z})}{|z_x^2 - z_y^2|^2 |z_x^2 + z_y^2 + z_z^2|^2}\right\} \tag{A8}$$

where

$$X(\hat{z}, \delta\hat{z}) = |\hat{z}|^4 |\delta z_1|^2 + |z_1 z_2|^2 |\delta z_2|^2 + |z_1 z_3|^2 |\delta z_3|^2 - 2\,\mathrm{Re}(\hat{z}^2 z_1^* z_2^* \delta z_1 \delta z_2^*) - 2\,\mathrm{Re}(\hat{z}^2 z_1^* z_3^* \delta z_1 \delta z_3^*) + 2\,\mathrm{Re}(|z_1|^2 z_2 z_3^* \delta z_2 \delta z_3^*) \tag{A9}$$

The restriction of the argument of the arctangent to be positive in the definition of $\Theta_1$ does not alter its uncertainty. Following (A3) and (A4), the uncertainty of $\phi_{1H}$ is

$$\delta\phi_{1H} = \mathrm{Im}\left\{\frac{\delta\hat{z}}{\hat{z}} - \frac{\delta z_1}{z_1}\right\} \tag{A10}$$

and

$$|\delta\phi_{1H}|^2 = \tfrac{1}{2}\left\{\frac{X(\hat{z}, \delta\hat{z})}{|z_1|^2 |z_x^2 + z_y^2|^2}\right\} \tag{A11}$$

where $X$ is given in (A9).

*Acknowledgements.* We thank David Thomson, Freeman Gilbert, and Alan Chave for many useful comments in the course of this study. Jeffrey Park thanks Dennis G. Wills for temporary lodging in the Scholar's Lair of the La Jolla Cultural Society. This work was supported by National Science Foundation grants EAR-82-18612 and EAR-84-09612 (J.P.), Office of Naval Research contracts N00014-85-C-0104, N0014-87-K-0005, and N00014-87-K-191 (C.L.), and U.S. Geological Survey contract 14-08-0001-21893 (F.V.). J.P. was supported as a visiting scientist while at Princeton University Geophysical Fluid Dynamics Lab with F.A. Dahlen by NOAA Grant NA84EAD00057. We thank the members of Project Anza for the use of their data.

## REFERENCES

Archambeau, C. B., and E. A. Flinn, Automated analysis of seismic radiations for source characteristics, *Proc. IEEE, 53*, 1876–1884, 1965.

Bard, P., and M. Bouchon, The two-dimensional resonance of sediment-filled valleys, *Bull. Seismol. Soc. Am., 75*, 519–542, 1985.

Berger, J., L. M. Baker, J. N. Brune, J. B. Fletcher, T. C. Hanks, and F. L. Vernon III, The Anza array: A high-dynamic range, broadband, digitally radio telemetered seismic array, *Bull. Seismol. Soc. Am., 74*, 1469–1481, 1984.

Golub, G. H., and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, Md., 1983.

Kanasewich, E. R., *Time Sequence Analysis in Geophysics*, 3rd ed., University of Alberta Press, Edmonton, Canada, 1981.

Kennett, B. L. N., Wavenumber and wavetype coupling in laterally heterogeneous media, *Geophys. J. R. Astron. Soc., 87*, 313–332, 1986.

Lanzerotti, L. J., D. J. Thomson, C. G. Maclennan, and L. V. Medford, Intermediate bandwidth polarization analysis using multiple-window methods, *Eos Trans. AGU, 67* (44), 873, 1986.

Lindberg, C. R., Multiple taper spectral analysis of terrestrial free oscillations, Ph.D. thesis, 182 pp., Univ. of Calif., San Diego, La Jolla, 1986.

Odom, R. I., A coupled-mode examination of irregular waveguides including the continuum spectrum, *Geophys. J. R. Astron. Soc., 86*, 425–454, 1986.

Park, J., and A. D. Chave, On the estimation of magnetotelluric response functions using the singular value decomposition, *Geophys. J. R. Astron. Soc.*, 77, 683–709, 1984.

Park, J. J., C. R. Lindberg, and D. J. Thomson, Multiple taper spectral analysis of terrestrial free oscillations: Part I, *Geophys. J. R. Astron. Soc.*, in press, 1987.

Park, J., F. L. Vernon III, and C. R. Lindberg, Multitaper spectral analysis of high-frequency seismograms, *J. Geophys. Res.*, this issue.

Samson, J. C., Matrix and Stokes vector representations of detectors for polarized waveforms: Theory, with some applications to teleseismic waves, *Geophys. J. R. Astron. Soc., 51*, 583–603, 1977.

Samson, J. C., Pure states, polarized waves, and principal components in the spectra of multiple, geophysical time-series, *Geophys. J. R. Astron. Soc., 72*, 647–664, 1983*a*.

Samson, J. C., The spectral matrix, eigenvalues and principal components in the analysis of multichannel geophysical data, *Ann. Geophysicae, 1* (2), 115–119, 1983*b*.

Samson, J. C., The reduction of sample-bias in polarization estimators for multichannel geophysical data with anisotropic noise, *Geophys. J. R. Astron. Soc., 75*, 289–308, 1983*c*.

Sato, H., Attenuation and envelope formulation of three-component seismograms of small local earthquakes in randomly inhomogeneous lithosphere, *J. Geophys. Res., 89*, 1221–1241, 1984.

Sereno, T. J., and J. A. Orcutt, Synthetic seismogram modelling of the oceanic $P_n$ phase, *Nature, 316*, 246–249, 1985.

Slepian, D., Some comments on Fourier analysis, uncertainty and modelling, *SIAM Rev., 25*, 379–393, 1983.

Thomson, D. J., Spectrum estimation and harmonic analysis, *IEEE Proc., 70*, 1055–1096, 1982.

Vidale, J. E., Complex polarization analysis of particle motion, *Bull. Seismol. Soc. Am., 76*, 1393–1405, 1986.

C. R. Lindberg and F. L. Vernon III, Institute of Geophysics and Planetary Physics, A-025, University of California, San Diego, La Jolla, CA 92093

J. Park, Department of Geology and Geophysics, Yale University, P.O. Box 6666, New Haven, CT 06511

(Received August 7, 1986,
revised July 16, 1987,
accepted July 27, 1987.)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>Frequency Dependent Polarization Analysis of High-Frequency Seismograms | | 5. TYPE OF REPORT & PERIOD COVERED<br>6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Jeffrey Park, Frank L. Vernon III, and Craig R. Lindberg | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-87-K-0005 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Scripps Institution of Oceanography<br>La Jolla, California 92093 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research<br>Arlington, Virginia 22217 | | 12. REPORT DATE<br>November 10, 1987<br>13. NUMBER OF PAGES<br>12 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>unclassified<br>15a. DECLASSIFICATION DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Reprint from Journal of Geophysical Research, Vol. 92, No. B12, 12,664-12,674, (1987)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We present a multitaper algorithm to estimate the polarization of particle motion as a function of frequency from three-component seismic data. This algorithm is based on a singular value decomposition of a matrix of eigenspectra at a given frequency. The right complex eigenvector $\hat{z}$ corresponding to the largest singular value of the matrix has the same direction as the dominant polarization of seismic motion at that frequency. The elements of the polarization vector $\hat{z}$ specify the relative amplitudes and phases of motion measured along the recorded components within a chosen frequency band. The width . . .

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102 LF 014 6601

unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

*Geophys. J. R. astr. Soc.* (1987) **91**, 755–794

# Multiple-taper spectral analysis of terrestrial free oscillations: part I

Jeffrey Park *Department of Geology and Geophysics Yale University, New Haven, CT 06511, USA*

Craig R. Lindberg *Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, CA 92093, USA*

David J. Thomson *Bell Laboratories, Murray Hill, NJ 07974, USA*

Accepted 1987 April 28. Received 1987 April 22; in original form 1986 September 3

**Summary**. We present a new method for estimating the frequencies of the Earth's free oscillations. This method is an extension of the techniques of Thomson (1982) for finding the harmonic components of a time series. Optimal tapers for reducing the spectral leakage of decaying sinusoids immersed in white noise are derived. Multiplying the data by the best $K$ tapers creates $K$ time series. A decaying sinusoid model is fit to the $K$ time series by a least squares procedure. A statistical $F$-test is performed to test the fit of the decaying sinusoid model, and thus determine the probability that there are coherent oscillations in the data. The $F$-test is performed at a number of chosen frequencies, producing a measure of the certainty that there is a decaying sinusoid at each frequency. We compare this method with the conventional technique employing a discrete Fourier transform of a cosine-tapered time-series. The multiple-taper method is found to be a more sensitive detector of decaying sinusoids in a time series contaminated by white noise.

**Key words**: multiple-taper, free oscillations, spectral analysis

## 1 Introduction

The free oscillations of the Earth appear as decaying sinusoids in the records of instruments in the available low-frequency seismic arrays (International Deployment of Accelerometers, hereafter referred to as IDA, and Global Digital Seismic Network, hereafter referred to as GDSN) (Agnew *et al.* 1976; Ingdahl, Peterson & Orsini 1982). Information about the structure of the Earth can be inferred from the frequencies, decay rates and amplitudes of these oscillations.

Conventionally, these characteristics of the decaying sinusoids are estimated from a direct spectral estimate of the data using a cosine taper (Harris 1978; Dahlen 1982; Lindberg 1986), or by producing spherical harmonic-weighted sums of the direct spectral estimates made

from each station's record ('stacking' or 'stripping') (Gilbert & Dziewonski 1975). There are several difficulties with using a single cosine-taper in the harmonic analysis of free oscillations. The time series analysed in free oscillation studies are non-stationary; they are also contaminated with noise. The cosine taper is symmetric and appropriate for stationary time-series; it is not a good taper for minimizing the spectral leakage of decaying sinusoids immersed in noise. The cosine taper also discards much of the data at the ends of the time series, particularly at the beginning where the signal-to-noise ratios of the free oscillation records are large. This is not desirable. In addition, applying a cosine taper to reduce spectral leakage is purchased with greatly increased variance (e.g. figs 7 and 8 of Dahlen 1982). Use of the cosine taper roughly doubles the variance, or equivalently, halves statistical efficiency of the estimate (Jones 1962). Another drawback of a cosine-taper direct spectral estimate is that it does not discriminate between oscillations of constant phase and frequency (harmonic oscillations) and broad distributions of spectral energy caused by other processes.

To overcome these problems, we have developed a method of harmonic analysis for decaying sinusoids immersed in stationary white noise based on the methods developed by Thomson (1982). A set of several 'optimal' tapers is created, each one designed to minimize the spectral leakage of decaying sinusoids immersed in white noise, while maintaining a large value for the ratio of tapered signal energy to tapered noise energy. Multiplying the data by each taper in turn creates several time series. Taking the discrete Fourier transform of these time series yields several complex eigenspectra (called eigencoefficients by Thomson 1982). A decaying sinusoid model is fit by a least-squares procedure to these complex eigenspectra. The least-squares procedure produces an estimate of the initial amplitude of any decaying sinusoids in the data. The fit of the decaying sinusoid model at any given frequency is tested using a statistical $F$-test. This gives a quantitative measure of the confidence that a phase-coherent decaying sinusoid is present in the data at any given frequency.

The multiple-taper method utilizes more of the data than the cosine-taper direct spectral estimate, and, as shown in Section 4 and the appendix, is a more sensitive detector of free oscillations in a seismic record. In one example, the five singlets of ${}_0S_2$ could be detected in a single record of the 1977 Sumbawa event, with measured frequencies in good agreement with those reported by Buland, Berger & Gilbert (1979), who used a six-station global array stack. Only two of the singlet lines are visible in the conventional direct spectral estimate employing a cosine taper.

The multiple-taper technique for free-oscillation analysis is described in the following sections. Section 2 introduces the functionals which are optimized to yield a family of spectral leakage-suppressing eigentapers appropriate for an oscillation with a given attenuation rate. Functionals for decaying sinusoids in time series with and without white noise are discussed. Section 3 introduces the statistical $F$-test for detection of decaying sinusoids. In Section 4 we present a number of frequency measurements of isolated free oscillations using IDA network data. Our conclusions are summarized in Section 5. An error analysis of the method is included in the appendix. Readers interested primarily in the examples are directed to Section 4. To implement the technique on a computer one needs to solve (2.19) to design the tapers, apply (3.15) to estimate the decaying sinusoid amplitudes as a function of frequency, and compute (3.28) to produce an $F$-test plot to test for the existence of decaying sinusoids at any given frequency.

## 2 Optimal data tapers for decaying signals

In this section we adapt the methods described in a series of five papers by Slepian, Landau and Pollak (Slepian & Pollak 1961; Landau & Pollak 1961, 1962; Slepian 1964, 1978). Their

work involved a set of time-limited functions whose spectral energy is optimally concentrated within a given frequency band. These functions have been employed to design optimal tapers for the analysis of stationary processes (Thomson 1982). We have extended Thomson's work to produce tapers for the harmonic analysis of exponentially decaying signals. For signals that decay exponentially with time, we obtain an optimization equation from which one can find the data taper $w_0$ with optimal resistance to spectral leakage from outside a frequency band of chosen width. Solving the optimization equation, one discovers that there exists a family of data tapers $\{w_0(t), w_1(t), \ldots, w_{N-1}(t)\}$ with good spectral-leakage resistance. We refer to the members of this family as eigentapers. These tapers are eigenvectors of a Toeplitz matrix whose elements are values taken by the function $\sin x/x$. In the next section we produce several spectra from a single record multiplied by each of the eigentapers in turn, and we show how these spectra can be combined to provide useful information.

An important factor in the analysis of low-frequency seismic data is the presence of stationary white noise in the records. This was recognized by Dahlen (1982): the presence of stationary noise determined the optimal time-series length for estimation of parameters in Dahlen's analysis. In his work, however, the taper shapes were held fixed. In this study, we extend the methods of Thomson (1982) to derive optimal taper shapes for any length time series, characterized by a parameter depending on the signal-to-noise ratio at the start of the seismic record. These 'noise-cognizant' tapers have less resistance to spectral leakage than those designed using a procedure that ignores stationary noise. In the appendix we show how noise-cognizant tapers improve the sensitivity of the eigentaper analysis if stationary noise is present in the data.

## 2A DECAYING SIGNAL WITH NO NOISE

Consider first a signal $x(t)$ that consists of a sum of decaying sinusoids uncorrupted by noise. Then one can represent

$$x(t) = \sum_j \mu_j \exp(i\omega_j t - \alpha_j t); \qquad t \geq 0,$$

where $\mu_j$ is the complex amplitude of the $j$th decaying sinusoid, which has angular frequency $\omega_j$ and decay rate $\alpha_j$. In practice, one cannot measure $x(t)$, but only the $N$ discrete numbers $x(t_0), x(t_1), \ldots, x(t_{N-1})$. Assume that $t_0 = 0$, and the time between samples $\Delta t = t_{j+1} - t_j$ is a constant, which we scale to be unity. If $\Delta t = 1$, then the Nyquist frequency $f_{\text{Nyquist}} = \frac{1}{2}$, and the angular frequency $\omega = 2\pi f$ is defined on its principal domain $(-\pi, \pi]$. Tapering the time series $\{x(t)\}_{t=0}^{N-1}$ consists of multiplying it by a real valued sequence $\{w(t)\}_{t=0}^{N-1}$ (the taper). Taking the discrete Fourier transform of the tapered signal $\{x(t)w(t)\}_{t=0}^{N-1}$ yields the function

$$y(\omega) = \sum_{t=0}^{N-1} \exp(-i\omega t) w(t) x(t). \tag{2.1}$$

This sum may be quickly computed using the Fast Fourier Transform (FFT) algorithm (Cooley & Tukey 1965; Brigham 1974). A traditional estimate of the energy content of $x(t)$ as a function of frequency is given by $|y(\omega)|^2$, where $\{w(t)\}_{t=0}^{N-1}$ is a conventional taper (Hann, Hamming, Blackman-Harris, Morse no. 2, etc.; Harris 1978 describes many of the popular tapers). The finite length of the time series makes a boxcar taper implicit if $w(t) \equiv 1$

in (2.1). One wishes to choose $\{w(t)\}_{t=0}^{N-1}$ to facilitate determination of the frequency content of $x(t)$.

The primary purpose of a data taper is to minimize spectral leakage. That is, the spectral component of a tapered signal at frequency $\dot{\omega}$ should have minimal energy contribution from outside the interval $(\dot{\omega} - \Omega, \dot{\omega} + \Omega)$, where $0 < 2\Omega < 2\pi$ is a chosen bandwidth. One must also prevent the energy at $\dot{\omega}$ from the leaking out to affect parts of the spectrum at other frequencies. Suppose that $x(t)$ consists of only one decaying sinusoid in $(\dot{\omega} - \Omega, \dot{\omega} + \Omega)$, with frequency $\dot{\omega}$. The tapered signal $\{w(t)\mu \exp(i\dot{\omega}t - \alpha t)\}_{t=0}^{N-1}$ should have as much of its energy as possible in $(\dot{\omega} - \Omega, \dot{\omega} + \Omega)$ relative to its total energy, which covers the entire band $(-\pi, \pi)$. One chooses a taper $\{w(t)\}_{t=0}^{N-1}$ to maximize the functional

$$f = \frac{\int_{\dot{\omega}-\Omega}^{\dot{\omega}+\Omega} |y(\omega)|^2 d\omega}{\int_{-\pi}^{\pi} |y(\omega)|^2 d\omega}, \tag{2.2}$$

where $y(\omega)$ is the discrete Fourier transform of $\{x(t)w(t)\}_{t=0}^{N-1}$:

$$y(\omega) = \mu \sum_{t=0}^{N-1} \exp(-i\omega t) \exp(i\dot{\omega}t) w(t) \exp(-\alpha t)$$

(Slepian 1983 describes how maximizing a similar functional yields solutions to the concentration problem, which is important in electrical engineering.) Since our time signal is limited to $[0, N-1]$, there is no way to confine completely the energy of its frequency transform to $(\dot{\omega} - \Omega, \dot{\omega} + \Omega)$. Therefore, the value $f$ will always be less than unity.

We expand the numerator of (2.2)

$$\begin{aligned}\int_{\dot{\omega}-\Omega}^{\dot{\omega}+\Omega} |y(\omega)|^2 d\omega &= |\mu|^2 \int_{-\Omega}^{\Omega} d\omega \sum_{t=0}^{N-1} \exp(-i\omega t) w(t) \exp(-\alpha t) \\ &\quad \times \sum_{s=0}^{N-1} \exp(i\omega s) w(s) \exp(-\alpha s) \\ &= 2|\mu|^2 \sum_{t=0}^{N-1} \sum_{s=0}^{N-1} w(t) \exp(-\alpha t) \frac{\sin \Omega(s-t)}{(s-t)} \exp(-\alpha s) w(s)\end{aligned} \tag{2.3}$$

and use Parseval's theorem to expand the denominator

$$\int_{-\pi}^{\pi} d\omega |y(\omega)|^2 = 2\pi |\mu|^2 \sum_{t=0}^{N-1} w(t) \exp(-2\alpha t) w(t) \tag{2.4}$$

so that (2.2) becomes dependent entirely on $w(0), w(1), \ldots, w(N-1)$ and simple functions. Define the $N$-vector $\mathbf{w} = [w(0), w(1), \ldots, w(N-1)]$, the matrix $\mathbf{A}$ with elements

$$A_{lm} = \frac{\sin \Omega(l-m)}{\pi(l-m)} \exp(-\alpha(l+m)); \qquad l, m = 0, 1, \ldots, N-1$$

and the diagonal matrix $\mathbf{B}$, where $B_{lm} = \delta_{lm} \exp(-2\alpha l)$; $l, m = 0, 1, \ldots, N-1$. (The symbol $\delta_{ab}$ is the Kronecker delta function; $\delta_{ab} = 1$ if $a = b$, and 0 otherwise.) Then equation (2.2) can be written as

$$f(\mathbf{w}) = \frac{\mathbf{w} \cdot \mathbf{A} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{B} \cdot \mathbf{w}}. \tag{2.5}$$

To find the taper that optimizes the functional $f$, set the variation of $f$ with respect to $\mathbf{w}$ equal to zero

$$\delta f(\mathbf{w};\mathbf{h}) = \frac{d}{d\epsilon} f(\mathbf{w} + \epsilon\mathbf{h})\bigg|_{\epsilon=0} = 0$$

for all $N$-vectors $\mathbf{h}$ (Goldstein 1980, chapter 2; Smith 1974). Some algebra leads to the eigenvalue problem

$$\mathbf{A}\cdot\mathbf{w} - \lambda\mathbf{B}\cdot\mathbf{w} = 0, \tag{2.6}$$

where $\mathbf{A}$, $\mathbf{B}$ are $N \times N$ real symmetric matrices and the eigenvalue

$$\lambda = f(\mathbf{w}).$$

The eigenvalue $\lambda$ is always less than unity, as can be seen from (2.2). The fractional spectral leakage of the signal at $\omega$ outside the frequency band $(\omega - \Omega, \omega + \Omega)$ is $1 - \lambda$. The taper $\mathbf{w}_0 = [w_0(0), w_0(1), \ldots, w_0(N-1)]$ corresponding to the largest eigenvalue $\lambda_0$ is the optimal taper for minimizing spectral leakage. The taper $\mathbf{w}_0$ has roughly the same shape as other popular tapers such as the Hann and Blackman-Harris tapers. (The taper $\mathbf{w}_0$ corresponds to the solid curve labelled '0' in Figs 1 and 2.) The largest eigenvalue $\lambda_0$ is almost 1; one finds that $\lambda_0 \approx 1 - (2.9 \times 10^{-10})$ for $N\Omega = 8\pi$. Moreover, there are several eigenvalues in the descending family $\lambda_0 > \lambda_1 > \lambda_2 > \ldots > \lambda_{N-1}$ that are very close to $\lambda_0$ and hence close to unity. The associated eigenvectors $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{N-1}$ form a sequence of 'eigentapers', the first few of which possess good spectral leakage resistance.

Let the decay rate $\alpha = 0$ in (2.6), noting that $\mathbf{A}$ and $\mathbf{B}$ depend on $\alpha$. Then (2.6) becomes equation (2.9) of Thomson (1982); its solutions are optimal tapers for concentrating the energy of nondecaying sinusoids. As discussed by Slepian (1978) and Thomson (1982), the

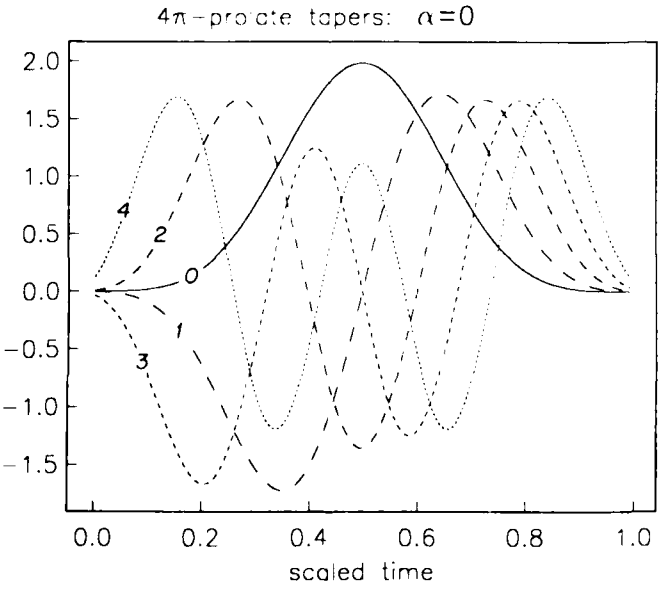



**Figure 1.** The five lowest-order eigentaper solutions to (2.6) when decay rate $\alpha = 0$, and $N\Omega = 8\pi$. The solid black line is the optimal taper. Higher order tapers are successively more oscillatory.

solutions to (2.6) when $\alpha = 0$ are the discrete $N \cdot W\pi$ prolate spheroidal sequences $\{v_t^{(k)}(N, W)\}_{t=0}^{N-1}$, where $W = \Omega/2\pi$ and $k$ is an integer. If $\alpha \neq 0$, the solutions to (2.6) are the eigentapers $w_k(t) = v_t^{(k)}(N, \Omega/2\pi) \exp(\alpha t)$; $t = 0, 1, \ldots, N-1$. A spectral estimate using these tapers is similar to the 'analytic continuation' of the DFT discussed in Buland & Gilbert (1978). In much of the following, the time-bandwidth product $P = N \cdot W = N\Omega/2\pi = 4$. (In Slepian 1978, 1983, the parameter $c = \pi \cdot P$ is used.) $P$ is usually taken to be an integer, but this convention is not required.

The $\{v_t^{(k)}(N, W)\}_{t=0}^{N-1}$ sequences have several properties that are shared with the decaying-signal eigentapers $\{w_k(t)\}_{t=0}^{N-1}$. For example, both possess an orthogonality property:

$$\sum_{t=0}^{N-1} v_t^{(k)}(N, W) v_t^{(k')}(N, W) = \sum_{t=0}^{N-1} \exp(-2\alpha t) w_k(t) w_{k'}(t) = \delta_{kk'}. \tag{2.7}$$

The tapers $\{w_k(t)\}_{t=0}^{N-1}$ sample that part of the signal that decays as $\exp(-\alpha t)$ in an orthogonal manner. Figure 1 shows the five lowest-order eigentapers $w_k(t) = v_t^{(k)}(N, W = 4/N)$; $t = 0, 1, \ldots, N-1$ for a stationary signal ($\alpha = 0$). The zeroth-order taper $\{w_0(t)\}_{t=0}^{N-1}$ is a $4\pi$ prolate taper. Note that the higher-order eigentapers are negative in some places and they weight the data more heavily near the ends of the record. Figure 2 shows eigentapers for a signal that decays by $\exp(-\pi\beta)$, where $\beta = \alpha T/\pi = 1.0$ $Q$-cycles, during the record length $T = N\Delta t$. [One $Q$-cycle refers to the time required for $Q$ oscillations of the harmonic signal. This notation was introduced by Dahlen (1982). One $Q$-cycle is equivalent to an amplitude decay of $\exp(-\pi) \approx 1/23$]. Note the increasing amplitude towards the end of the record, as the tapers try to amplify the decaying signal. The tapers $\{w_k(t)\}_{t=0}^{N-1}$ produce the

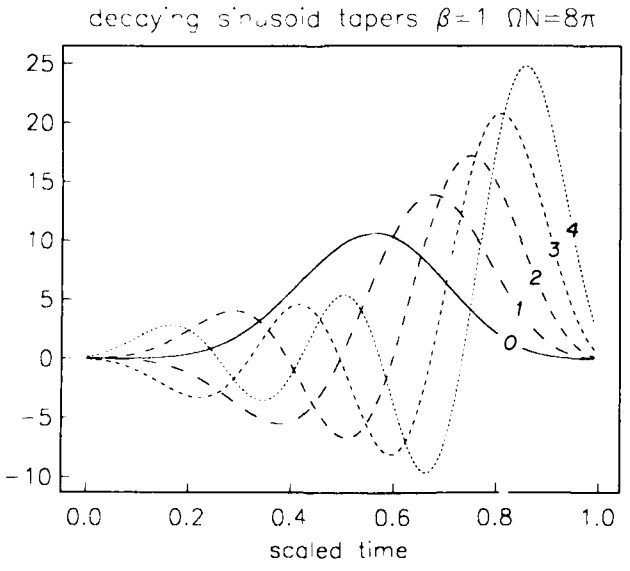



**Figure 2.** The five lowest-order eigentapers for a decaying sinusoid that decays by $\exp(-\pi)$ during the record. Multiplying a decaying sinusoid by these tapers will concentrate its energy in a frequency band of width $2\Omega = 16\pi/N$. The taper amplitudes increase exponentially towards the end of the record to compensate for the signal's decay.

unwelcome result of amplifying the late record noise as well, so that while the signal power remains constant with time in the tapered record, the noise power increases exponentially. In the next subsection we will show how noise-cognizant eigentapers weight the later data more soberly.

Substituting $\{v_t^{(k)}(N, W)\exp(\alpha t)\}_{t=0}^{N-1}$ for $\{w_k(t)\}_{t=0}^{N-1}$ in (2.5) and using the definition of $\lambda$, one can show that the discrete prolate spheroidal sequences and the sequences $\{w_k(t)\}_{t=0}^{N-1}$ have the same eigenvalues $\lambda_k$ for any value of the decay rate $\alpha$. Therefore, the $k$th prolate taper and the $k$th decaying sinusoid eigentaper have the same fractional spectral leakage for a given value of $P = \Omega N/2\pi$. The $2NW$ lowest-order eigenvalues $\lambda_k$ of (2.6) are of order unity, and rapidly drop off thereafter (Slepian 1983). For example, $4\pi$-prolate sequences have eight order-unity eigenvalues, one per Rayleigh frequency spacing ($2\pi/N$) in the central region $(\omega - 8\pi/N,\ \omega + 8\pi/N)$. Values of $\lambda_k$ are given in Table 1 for some examples of $P\pi$ prolate tapers.

The amplitudes of the frequency transforms

$$\tilde{W}_k(\omega) = \sum_{t=0}^{N-1} w_k(t)\exp(\pi\beta t/T)\exp(-i\omega t) \tag{2.8}$$

of the five lowest-order $4\pi$-prolate eigentapers are shown in Fig. 3 over a wide range of frequencies. (Here, record length $T = N$.) Substituting $\{v_t(k)(N, W)\exp(\alpha t)\}_{t=0}^{N-1}$ for $\{w_k(t)\}_{t=0}^{N-1}$ in (2.8), one finds that the functions $\tilde{W}_k(\omega)$ are independent of decay rate. Figure 3 shows the excellent leakage rejection properties of the eigentapers. There is a sharp band-edge at frequency $\omega = 8\pi/T$. Note sidelobe height increases as the order of the taper increases, but remains 30–40 dB below the height of the central region even for the fifth taper. Figure 4 is an expansion of the central peak region displaying both real and imaginary components of the same five eigentaper transforms $\tilde{W}_k(\omega)$. The plots of the central region show that each $\tilde{W}_k(\omega)$ samples the central band $(-\Omega, \Omega)$ in a different manner. The eigentaper transforms $\tilde{W}_k(\omega)$ become increasingly more oscillatory with increasing order. The $\tilde{W}_k(\omega)$ are orthogonal, both within the central band

$$(2\pi)^{-1}\int_{-\Omega}^{\Omega} d\omega\, \tilde{W}_k^*(\omega)\tilde{W}_{k'}(\omega) = \mathbf{w}_k \cdot \mathbf{A} \cdot \mathbf{w}_{k'}$$

$$= \lambda_k \mathbf{w}_k \cdot \mathbf{B} \cdot \mathbf{w}_{k'} = \lambda_k \delta_{kk'} \tag{2.9}$$

**Table 1.** Eigenvalues $\lambda_k$ for lowest-order $P\pi$ prolate tapers.

| $P$ | $k$ | $1-\lambda_k$ | $P$ | $k$ | $1-\lambda_k$ |
|---|---|---|---|---|---|
| 1 | 0 | 0.189 | 3 | 0 | $1.348\times10^{-7}$ |
| | 1 | 0.2504 | | 1 | $9.245\times10^{-6}$ |
| | 2 | 0.7564 | | 2 | $3.850\times10^{-4}$ |
| | | | | 3 | $5.086\times10^{-3}$ |
| | | | | 4 | $5.386\times10^{-2}$ |
| 2 | 0 | $5.725\times10^{-5}$ | 4 | 0 | $2.946\times10^{-10}$ |
| | 1 | $2.438\times10^{-3}$ | | 1 | $2.768\times10^{-8}$ |
| | 2 | $4.061\times10^{-2}$ | | 2 | $1.210\times10^{-6}$ |
| | 3 | 0.2783 | | 3 | $4.245\times10^{-5}$ |
| | 4 | 0.7253 | | 4 | $5.899\times10^{-4}$ |
| | | | | 5 | $7.496\times10^{-3}$ |

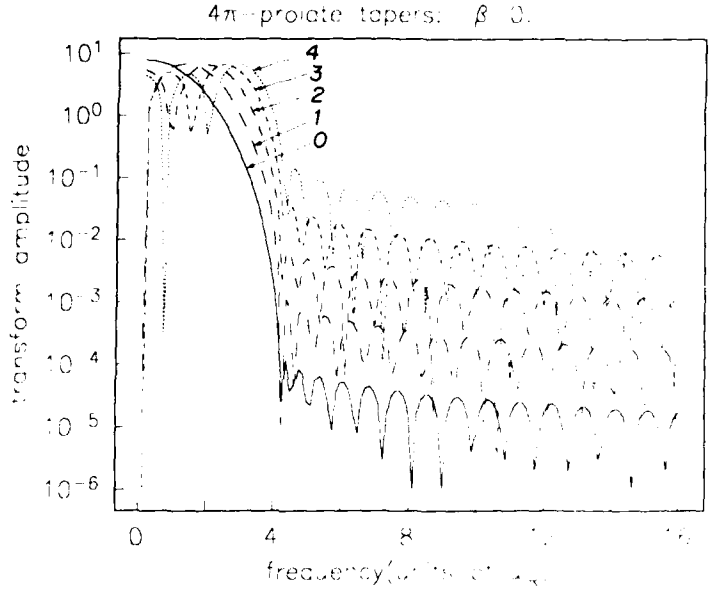



**Figure 3.** Frequency-transform amplitudes of the five lowest-order $4\pi$ prolate eigentapers. The frequency transforms are independent of decay rate by (2.8). The sidelobes are lowest for the optimal eigentaper, and increase in height for higher order eigentapers. The abscissa is in units of $\omega_R = 2\pi/T$, where $T$ is the record length. There is a sharp bandedge at frequency $\omega = 4\omega_R$.

using (2.6) through (2.8), and over the entire discrete Fourier transform frequency band

$$(2\pi)^{-1}\int_{-\pi}^{\pi} d\omega\, \tilde{W}_k^*(\omega)\tilde{W}_{k'}(\omega) = \delta_{kk'} \tag{2.10}$$

by (2.7). In (2.9) and (2.10), the asterisk denotes complex conjugation.

Unfortunately, these tapers are only suitable for the analysis of noise-free records, but low frequency seismic data are noisy. In the next section, tapers designed to analyse noisy records are discussed.

## 2B DECAYING SIGNAL IN WHITE NOISE

Low-frequency seismic records can be modelled as a sum of decaying free oscillations immersed in noise

$$x(t) = \sum_m \mu_m \exp(i\omega_m t - \alpha_m t) + n(t), \qquad t \geq 0, \tag{2.11}$$

where, as before, $\omega_m$, $\alpha_m$ and $\mu_m$ are the frequency, decay rate and complex amplitude of the $m$th free oscillation, with onset at $t = 0$, and $n(t)$ is a realization of a noise process. The sum over $m$ extends in principle over the countably infinite elastic-gravitational free oscillations, but can be taken as finite in a record from a band-limited seismic instrument. We will assume throughout that $n(t)$ is a realization of a stationary, zero-mean, white noise process.

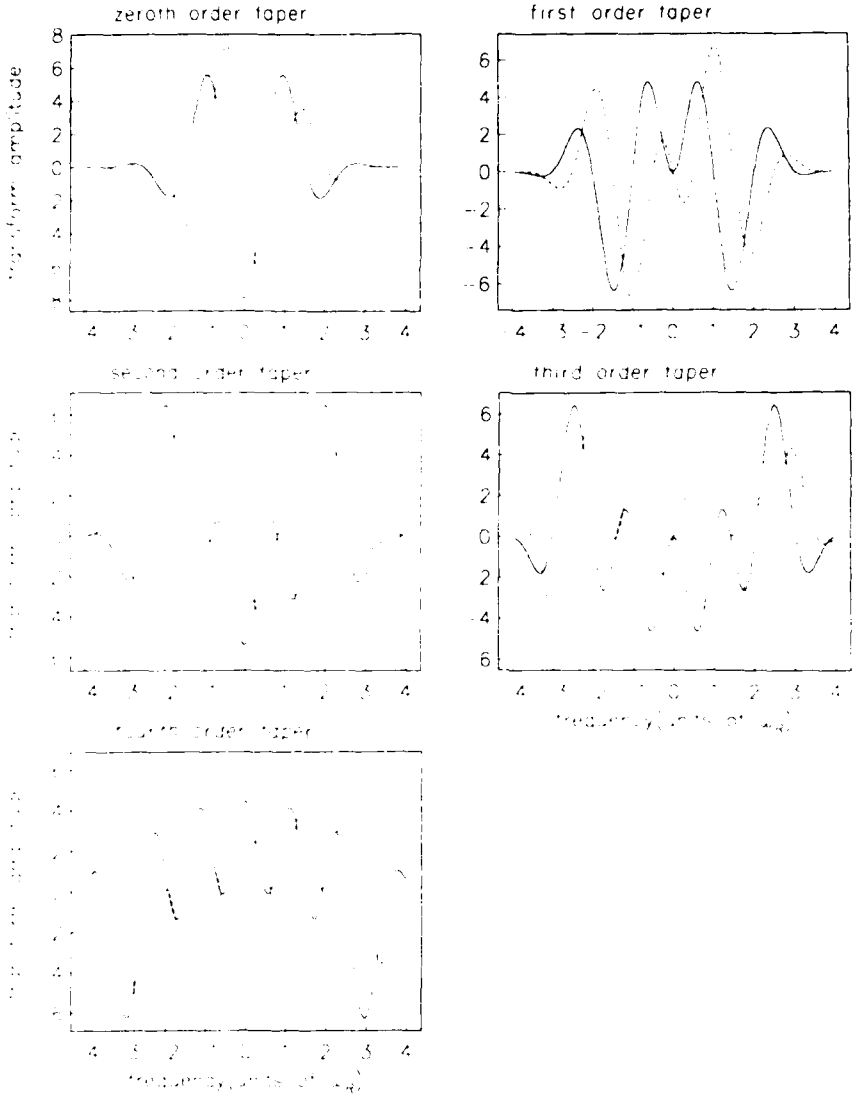



**Figure 4.** Expansion of central peak region of the five eigentaper frequency transforms of Fig. 3. The solid line is the real part; the dotted line is the imaginary part. The central peak region is increasingly more oscillatory for higher-order eigentapers.

In practice, the spectrum of seismic noise does not vary much over the frequency band of interest (Agnew & Berger 1978).

We determine optimal data tapers in this case using an extension of the variational formalism described above. In particular, we wish to balance the need to concentrate as much of the spectral energy of the signal as possible into a region of bandwidth $2W$ against the desire to retain a high ratio of tapered signal power to tapered noise power. The exponential asymmetry of the decaying signal tapers $\{w_k(t)\}_{t=0}^{N-1}$ will increase the

amplitude of stationary noise in the later part of the record. This will degrade the quality of the spectral estimate considerably unless the ratio of tapered signal to tapered noise is constrained to have a reasonable value.

Assume that in the interval $(\dot{\omega} - \Omega, \dot{\omega} + \Omega)$ the record $x(t)$ is composed of the signal, a single decaying sinusoid, plus white noise $n(t)$:

$$x(t) = \mu \exp(i\dot{\omega}t - \alpha t) + n(t). \tag{2.12}$$

Suppose also that we have discrete samples of $x(t)$

$$\{x(t)\}; \qquad t = 0, 1, 2, \ldots, N-1$$

so that the angular frequency $\omega \in (-\pi, \pi]$. We want to choose our taper $\{w(t)\}_{t=0}^{N-1}$ so that the energy of the tapered signal

$$\{w(t)\mu \exp(i\dot{\omega}t - \alpha t)\}_{t=0}^{N-1} \tag{2.13}$$

in $(\dot{\omega} - \Omega, \dot{\omega} + \Omega)$ relative to its total energy is maximized, but now with a constraint: the ratio of the tapered signal power to tapered noise power in $(\dot{\omega} - \Omega, \dot{\omega} + \Omega)$ has a fixed value.

The discrete Fourier transform of the tapered noise is

$$m(\omega) = \sum_{t=0}^{N-1} n(t) \exp(-i\omega t) w(t). \tag{2.14}$$

A measure of the expected energy of the tapered noise at frequency $\omega$ is

$$\langle |m(\omega)|^2 \rangle = \sigma_N^2 \sum_{t=0}^{N-1} |w(t)|^2, \tag{2.15}$$

where $\langle\rangle$ denotes expectation value and $\sigma_N^2$ is the noise variance. The expected power of the noise in the tapered record in $(\dot{\omega} - \Omega, \dot{\omega} + \Omega)$ is

$$\int_{\dot{\omega}-\Omega}^{\dot{\omega}+\Omega} \langle |m(\omega)|^2 \rangle d\omega = 2\Omega\sigma_N^2 \sum_{t=0}^{N-1} |w(t)|^2. \tag{2.16}$$

We generalize (2.5) in order to constrain the ratio of tapered signal to tapered noise within the frequency band $(\dot{\omega} - \Omega, \dot{\omega} + \Omega)$. We now wish to maximize the functional

$$f(\mathbf{w}; \Omega, \eta) = \frac{\mathbf{w} \cdot \mathbf{A} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{B} \cdot \mathbf{w}} + \eta \frac{\mathbf{w} \cdot \mathbf{A} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \tag{2.17}$$

with respect to $\mathbf{w}$, where $\mathbf{w}$, $\mathbf{A}$ and $\mathbf{B}$ are as defined in Section 2A. The second term in equation (2.17) represents the ratio of tapered signal power to tapered noise power; $\eta$ is a Lagrange multiplier. In the limit of very large signal-to-noise ratio, i.e., as $(|\mu|^2)(\sigma_N^2) \to \infty$, one expects $\eta$ to tend to zero. In principle $\eta$ is determined from the constraint equation; in practice we determine its value empirically. The condition $\delta f(\mathbf{w}; \Omega, \eta) = 0$ leads to a non-linear equation for the tapers $\mathbf{w}$ which maximize (2.17). This non-linear equation can be solved approximately (Lindberg 1986).

Alternatively, we can minimize the functional

$$\tilde{f}(\mathbf{w}; \Omega, \nu) = \frac{\mathbf{w} \cdot \mathbf{B} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{A} \cdot \mathbf{w}} + \nu \frac{\mathbf{w} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{A} \cdot \mathbf{w}} \tag{2.18}$$

(F. Gilbert, private communication). Solving $\delta f' = 0$ leads to the equation

$$\mathbf{A} \cdot \mathbf{w} = \lambda' \mathbf{B}' \cdot \mathbf{w}, \tag{2.19}$$

where $\mathbf{B}' = \mathbf{B} + \nu \mathbf{I}$, $\mathbf{I}$ being the $N \times N$ identity matrix and

$$\lambda' = [f'(\mathbf{w}; \Omega, \nu)]^{-1}.$$

The eigenvectors which correspond to the largest eigenvalues $\lambda'$ of (2.19) will *minimize* $f'$.

Given the decay rate $\alpha$ and the noise-weighting parameter $\nu$, (2.19) can be solved for eigentapers $\{w_k(t; \beta, \nu)\}_{t=0}^{N-1}$. When $\nu = 0$, the elements of the $k$th taper $w_k(t; \beta, 0) = w_k(t)$; $t = 0, 2, \ldots, N-1$, and the tapers reduce to those of Section 2A. The fraction of tapered signal power that remains in the frequency band $(\omega - \Omega, \omega + \Omega)$ is

$$(\mathbf{w}_k \cdot \mathbf{A} \cdot \mathbf{w}_k)/(\mathbf{w}_k \cdot \mathbf{B} \cdot \mathbf{w}_k) = \lambda_k \tag{2.20}$$

which can be calculated from the eigenvectors and eigenvalues of (2.19). We have found it helpful to think of the $\lambda_k$ as 'bandwidth retention factors'.

We used EISPACK subroutines (Smith *et al.* 1976) to solve (2.19) for its largest eigenvalues $\lambda'_k$ and associated eigenvectors. We normalized the tapers $\{w_k(t; \beta, \nu)\}_{t=0}^{N-1}$ so that

$$\mathbf{w}_k \cdot \mathbf{B}' \cdot \mathbf{w}_k = \sum_{t=0}^{N-1} (\exp(-2\alpha t) + \nu) w_k(t; \beta, \nu) w_k(t; \beta, \nu) = \delta_{kk}. \tag{2.21}$$

Rather than solve an eigenvalue problem for every data series length, (2.19) was solved for $N = 128$ *and the tapers for* other values of $N$ were found using spline interpolation. This approach takes advantage of the asymptotic relations between the discrete and continuous-time tapers described in Slepian (1978). Tests using these interpolated tapers showed negligible degradation of spectral leakage properties relative to exact solutions for $N > 128$. For $N < 128$, (2.19) should be solved directly (A. Chave, private communication), but such *short time series are rare in free oscillation work. The taper transforms are computed from* the interpolated tapers using an FFT after padding the tapers with zeroes until their lengths were a power of two.

The preceding argument shows that $\nu$ is a complicated function of the signal-to-noise ratio. For large signal-to-noise-ratios $|\mu|^2 \sigma_\xi^2$, $\nu$ will be very small, $\mathbf{B}' \approx \mathbf{B}$, and the solution of (2.19) is not very different from the solution of (2.6). For smaller signal-to-noise ratios, one expects that the optimal tapers will have a $\nu$ of finite size. One could pick an incorrect value of $\nu$ for a particular signal-to-noise ratio, but then the tapers would not perform optimally. Useful values are best determined by experiment. We will show in the appendix that using eigentapers having larger values of $\nu$ results in a marked improvement in the detection capability of the multiple-taper algorithm.

Some examples of noise-cognizant eigentapers are exhibited in Fig. 5 for the case $\nu = 0.01$, $\beta = 0.6$, and $\Omega N = 8\pi$. Note the strong asymmetry of the tapers, with a strong emphasis on data in the earlier section of the record where instantaneous signal-to-noise ratio is greater. The height of the taper's main peak increases with increasing order to compensate for the decay of the signal, as shown in Fig. 2. Figure 6 shows tapers which were designed with $\nu = 0.1$, $\beta = 0.6$, and $\Omega N = 8\pi$. The preference for the early part of the record is more drastic, resulting in significant weighting at the onset of the time series. Here the variational principle minimizing (2.18) has sacrificed resistance to spectral leakage in order to raise the ratio of tapered signal to tapered noise. Figure 7 displays eigentapers designed with $\nu = 0.001$, $\beta = 0.2$, and $\Omega N = 8\pi$. These eigentapers are for a series containing sinusoids that only decay slightly in a more favourable signal-to-noise environment. Asymmetrical weighting remains

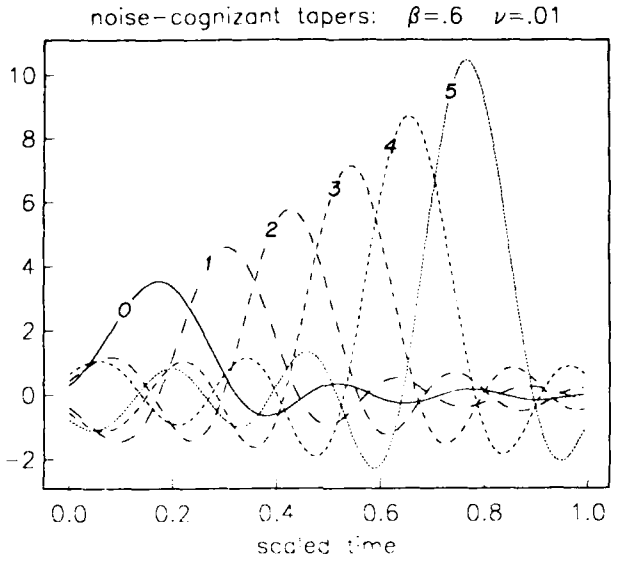



**Figure 5.** Optimal eigentapers for an exponentially decaying sinusoid immersed in white noise. The sinusoid is assumed to decay by $\exp(-0.6\pi)$ during the record. The noise parameter $\nu$ is chosen according to the signal-to-noise ratio of the data. The tapers sample the front of the record where the signal-to-noise ratio is largest, and increase in amplitude towards the end of the record to compensate for the signal's decay.

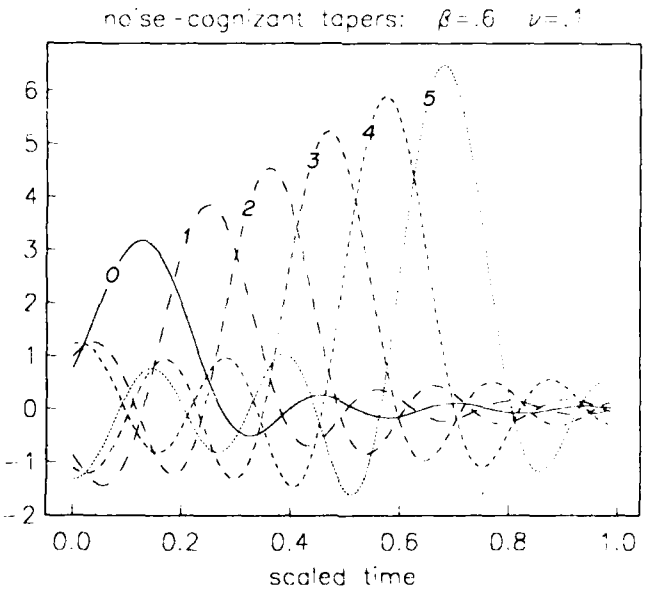



**Figure 6.** Sequence of optimal tapers for an exponentially decaying sinusoid immersed in white noise. These tapers are designed for a lower signal-to-noise ratio than those of Fig. 5, and have a larger noise parameter $\nu$. These tapers sample the data less heavily in the latter part of the record where the signal is obscured by noise, compared to the tapers in Fig. 5.

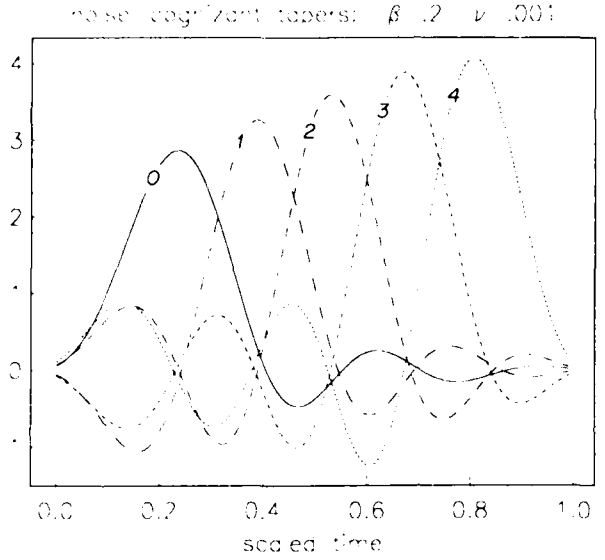



**Figure 7.** Set of optimal eigentapers for a decaying sinusoid contaminated by additive white noise. Compared with the tapers in Figs 5 and 6, these eigentapers are designed for records with a larger signal-to-noise ratio, and a sinusoid that decays less.

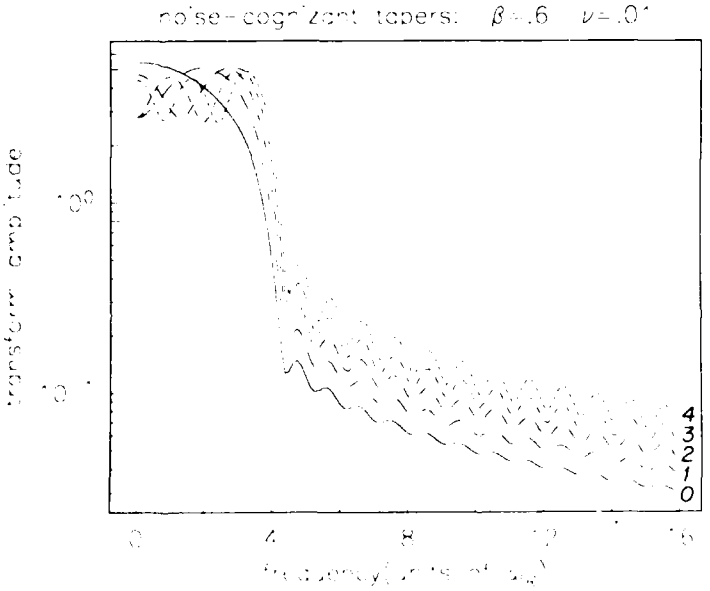



**Figure 8.** Amplitude of the five lowest-order noise-cognizant eigentapers with decay parameter $\beta = 0.6$, noise parameter $\nu = 0.01$ and time-bandwidth product $N\Omega = 8\pi$. The abscissa is in units of $\omega_R = 2\pi/N$. There is a sharp bandedge at frequency $\omega = 4\omega_R = 8\pi/N$. Note that the plot is of absolute magnitude, not power, so the first sidelobes are 20–30 dB down.

**Table 2.** Eigenvalues $\lambda_k'$ and bandwidth retention factors $\lambda_k$ for lowest order noise-cognizant optimal tapers.

| | $\nu = 0.01, \beta = 0.6$ | | $\nu = 0.1, \beta = 0.6$ | | $\nu = 0.001, \beta = 0.2$ | | $\nu = 0.01, \beta = 1$ | |
|---|---|---|---|---|---|---|---|---|
| $k$ | $\lambda_k'$ | $\lambda_k$ | $\lambda_k'$ | $\lambda_k$ | $\lambda_k'$ | $\lambda_k$ | $\lambda_k'$ | $\lambda_k$ |
| 0 | 0.962301 | 0.99869 | 0.73574 | 0.98905 | 0.99729 | 0.99997 | 0.94592 | 0.99682 |
| 1 | 0.940787 | 0.99760 | 0.63985 | 0.98003 | 0.99676 | 0.99995 | 0.89435 | 0.99227 |
| 2 | 0.910363 | 0.99619 | 0.53070 | 0.96888 | 0.99614 | 0.99993 | 0.80953 | 0.98428 |
| 3 | 0.867487 | 0.99361 | 0.41590 | 0.95026 | 0.99542 | 0.99990 | 0.68406 | 0.96699 |
| 4 | 0.808944 | 0.99006 | 0.30618 | 0.92719 | 0.99458 | 0.99975 | 0.52440 | 0.93826 |
| 5 | 0.732283 | 0.98339 | 0.21152 | 0.89053 | 0.98934 | 0.99469 | 0.35619 | 0.88398 |

Values of the eigenvalues $\lambda_k'$ and the bandwidth retention factors $\lambda_k$ for time-bandwidth product $\Omega N = 8\pi$ and various values of the decay parameter $\beta$ and the noise parameter $\nu$. Note that the bandwidth retention parameters $\lambda_k$ are close to 1 for small $k$, and are succeedingly smaller for higher order tapers. The lowest order eigentapers have the smallest fractional leakage $1-\lambda_k$, higher order eigentapers suffer from successively greater spectral leakage.

evident. Table 2 shows values of $\lambda_k'$ and $\lambda_k$ for tapers $\{w_k(t;\beta,\nu)\}_{t=0}^{N-1}$ for a selection of $\beta$ and $\nu$ values. The eigenvalues $\lambda_k'$ drop rapidly from unity with increasing $k$. The bandwidth retention factors $\lambda_k$ remain relatively constant among eigentapers of fixed $\beta$ and $\nu$. This behaviour can be observed qualitatively in the Fig. 8 plots of the amplitudes of the frequency transforms $\tilde{W}_k(w;\beta,\nu)$ of the tapers $\{w_k(t;\beta,\nu)\}_{t=0}^{N-1}$

$$\tilde{W}_k(\omega;\beta,\nu) = \sum_{t=0}^{N-1} w_k(t;\beta,\nu)\exp[(-\pi\beta T)t]\exp(-i\omega t) \tag{2.22}$$

for $\beta = 0.6$, $\nu = 0.01$, and $\Omega N = 8\pi$. The five lowest order eigentapers have sidelobes of comparable height. Enlargements of the central peak regions are shown in Fig. 9.

Because $\mathbf{A}$ and $\mathbf{B}'$ are symmetric, the orthogonality condition (2.9) remains valid for noise-cognizant tapers, using (2.21). However, as the noise-cognizant tapers $\{w_k(t;\beta,\nu)\}_{t=0}^{N-1}$ satisfy (2.21) and not (2.7), the frequency-domain orthogonality relation (2.10) does not

**Table 3.** Elements of matrix $\mathbf{D}$ for $\beta = 0.6$, $\Omega N = 8\pi$ and $\nu = 0.01$.

| | $k'$ | | | | |
|---|---|---|---|---|---|
| $k$ | 0 | 1 | 2 | 3 | 4 |
| 0 | 0.98151 | −0.00183 | 0.00218 | −0.00235 | 0.00227 |
| 1 | 0.00183 | 0.97130 | −0.00352 | 0.00415 | −0.00445 |
| 2 | 0.00218 | −0.00352 | 0.95658 | −0.00620 | 0.00719 |
| 3 | −0.00235 | 0.00415 | −0.00620 | 0.93560 | −0.01043 |
| 4 | 0.00227 | −0.00445 | 0.00719 | −0.01043 | 0.90638 |

for $\beta = 0.6$, $\Omega N = 8\pi$ and $\nu = 0.1$

| | $k'$ | | | | |
|---|---|---|---|---|---|
| $k$ | 0 | 1 | 2 | 3 | 4 |
| 0 | 0.86278 | 0.01271 | −0.01358 | 0.01261 | 0.00995 |
| 1 | −0.01271 | 0.80917 | 0.02210 | 0.02322 | 0.02143 |
| 2 | −0.01358 | 0.02210 | 0.74322 | 0.03364 | 0.03440 |
| 3 | 0.01261 | 0.02322 | 0.03364 | 0.66634 | 0.04693 |
| 4 | 0.00995 | 0.02143 | 0.03440 | 0.04693 | 0.58231 |

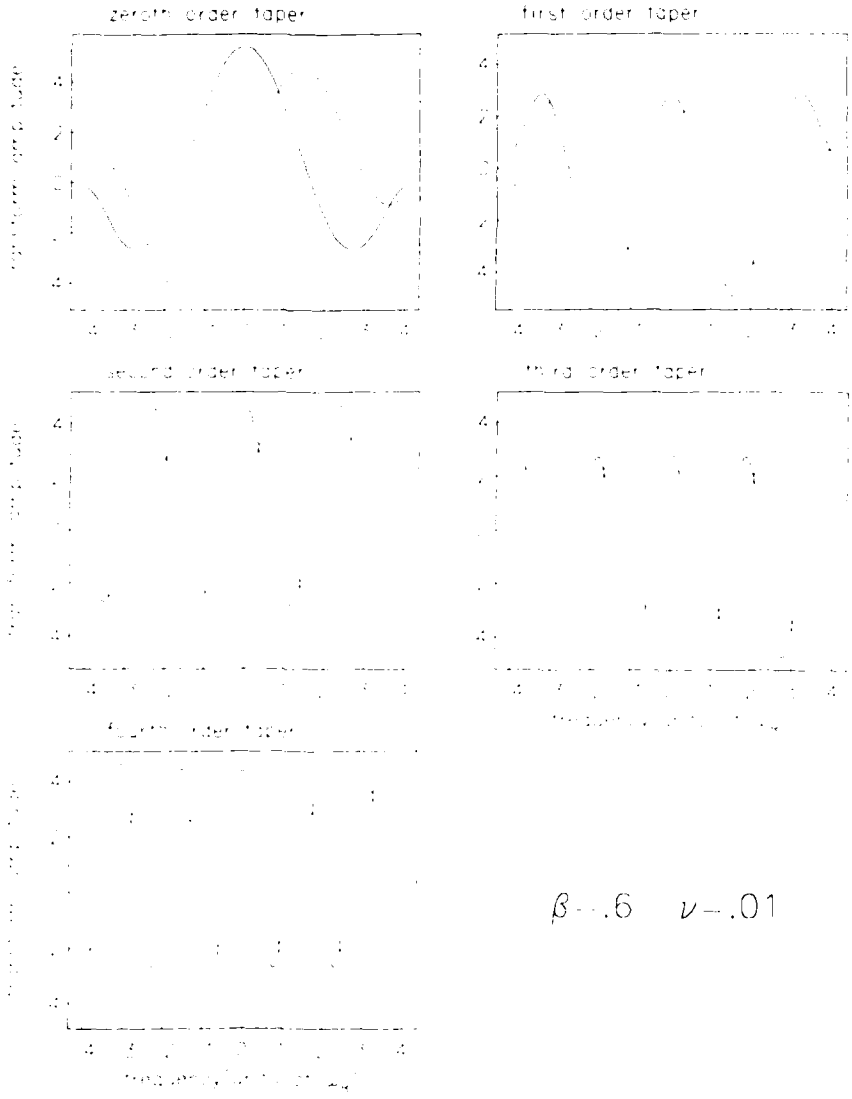



**Figure 9.** Expansion of the central peak region of the frequency transform amplitudes of the five lowest order eigentapers of Fig. 8. The solid line is the real part of the frequency transform; the dashed line is the imaginary part of the transform.

hold. In its place we have

$$(2\pi)^{-1}\int_{-\pi}^{\pi} d\omega \tilde{W}_k^*(\omega;\beta,\nu)\tilde{W}_{k'}(\omega,\beta,\nu) = \sum_{t=0}^{N-1} \exp[(-2\pi\beta_i T)t]\, w_k(t;\beta,\nu) w_{k'}(t;\beta,\nu) = D_{kk'} \tag{2.23}$$

for $k, k' \in \{0, 1, \ldots, K-1\}$. The matrix $\mathbf{D}$ is diagonally dominant for small $\nu$. Table 3 lists the elements of $\mathbf{D}$ for the five lowest order eigentapers with $\Omega N = 8\pi$, $\beta = 0.6$, and $\nu = 0.01$

or $\nu = 0, 1$. The magnitude of the off-diagonal elements of $\mathbf{D}$ indicate the departure from orthogonality of the frequency transforms $\tilde{W}'_k(\omega; \beta, \nu)$ over $(-\pi, \pi]$.

We have required that our data tapers possess certain desirable properties. We want them to have the ability to concentrate most of a decaying sinusoid's energy into a given frequency band, balanced against the capacity to maintain a high signal-to-noise ratio for the tapered data in the frequency domain. This leads to a variational calculus problem, whose solutions are a family of data tapers. These tapers provide a method of orthogonally sampling a decaying sinusoid, in both the time and frequency domains. By sampling a decaying sinusoid repeatedly in different ways, one can obtain superior estimates of its frequency and amplitude. Simple techniques to do this, based on those outlined by Thomson (1982), are the subject of the next section.

## 3 Harmonic analysis

An important part of long-period seismic data analysis is the detection of decaying sinusoids in the data and the measurement of their frequencies and amplitudes. The estimation of decay rate $\alpha$ is also important; we plan to address this problem in later work. In the following it is assumed that the $Q$ of the decaying oscillation is known or has been approximated by some method (e.g. Riedesel *et al.* 1980).

The spectra of low-frequency seismic time-series consist of harmonic 'lines' which have been broadened by decay, and a continuous background spectrum. The decay-broadened 'lines' are treated as signal, whereas the continuous spectrum is considered to be noise. This sets free oscillation data analysis apart from many familiar problems in seismic spectral estimation, e.g. finding the frequency content of body waves, or earthquakes in the near field. The spectra in those cases are predominantly continuous. There are methods of multitaper spectrum analysis that are useful for spectra which do not have harmonic line components (e.g. Thompson 1982; Park, Lindberg & Vernon, in press; Lindberg, Vernon & Park, unpublished manuscript).

The most straightforward method of detecting line components in low-frequency data is to measure obvious spectral peaks in a discrete Fourier transform of the data. If one tapers the time series in a prudent fashion, as indicated in Dahlen (1982), this approach is adequate for well-excited oscillations generated by large earthquakes ($M_L \gtrsim 7$). Unfortunately, most of these well-excited oscillations are surface-wave-equivalent fundamental modes which by themselves allow poor depth resolution. The modes most useful for enhancing the resolution at depth (e.g. the overtone oscillations that correspond to *PKP, PKIKP, SKS* etc. motion) are excited only by very large or very deep earthquake sources. Even then, their spectral peaks may not protrude substantially above the background noise. Masters & Gilbert (1981) show a typical example of this problem in the presumed identification of two inner-core oscillations. The use of spherical harmonic stacking of records from a global array (Gilbert & Backus 1965; Gilbert & Dziewonski 1975; Buland, Berger & Gilbert 1979) can aid mode identification greatly, especially in the case of closely spaced spectral lines caused by splitting of a free oscillation into individual singlets. However, it is difficult to identify decaying sinusoids in a low signal-to-noise environment using conventional methods of spectral estimation.

In the following we propose a method that is designed to yield a quantitative measure of the certainty that there is a decaying sinusoid at any given frequency. The novelty of our algorithm resides in the additional information obtained by sampling the data with more than one taper and the introduction of a statistical theory based on an $F$-test to detect harmonic spectral components and reject continuous, random-phase noise.

### 3.1 REGRESSION ANALYSIS

Suppose that $x(t)$ is a record consisting of noise and a number of decaying sinusoids, one of which has frequency $\hat{\omega}$. Then one can write

$$x(t) = \mu \exp(i\hat{\omega}t - \alpha t) + e(t), \tag{3.1}$$

where $\mu$ is a complex amplitude, $\alpha$ is a decay rate, and $e(t)$ is an error term. The error term consists of other decaying sinusoids and noise. For a sufficiently small value of $\Omega$, $x(t)$ contains only the single decaying sinusoid $\mu \exp(i\omega t - \alpha t)$ in the frequency interval $(\hat{\omega} - \Omega, \hat{\omega} + \Omega)$. This decaying sinusoid represents a deterministic signal in the record $x(t)$, and one can use the method of least squares to estimate its amplitude $\mu$.

We assume in the following that there is not more than one decaying sinusoid in the frequency interval $(\hat{\omega} - \Omega, \hat{\omega} + \Omega)$. This is often not true in practice, but in many applications the various singlets of free oscillation multiplets are observed to combine into a signal that is well approximated by a single resonance. Also, the least-squares procedure can be generalized to the case of two or more decaying sinusoids in a frequency interval of width $2\Omega$ (Thomson 1982).

As we have indicated, it is important to taper the data. Using the optimal tapers of Section 2, we multiply the data by each taper $\{w_k(t;\beta,\nu)\}_{t=0}^{N-1}$, $k = 0, 1, 2, \ldots, K-1$, in turn. We pick only a small number ($K$) of data tapers because higher order tapers have successively poorer leakage resistance. In the case $\nu = 0$ the tapers up to order $K = N \cdot \Omega \pi$ have good spectral leakage resistance; higher-order tapers exhibit vastly poorer performance (Slepian 1978). This is evident from the behaviour of the eigenvalues $\lambda_k$ appearing in Table 1. In the case $\nu \neq 0$, we choose the $K$ noise-cognizant tapers with the largest bandwidth retention factors. Usually $K < N\Omega \pi$ in this case. We show how we choose $K$ in Section 4.

Multiplying the data $\{x(t)\}_{t=0}^{N-1}$ by the $K$ eigentapers $\{w_k(t;\beta,\nu)\}_{t=0}^{N-1}$ one obtains $K$ time series:

$$\{w_k(t;\beta,\nu)x(t)\}_{t=0}^{N-1}; k = 0, 1, \ldots, K-1.$$

From equation (3.1)

$$e(t)w_k(t;\beta,\nu) = x(t)w_k(t;\beta,\nu) - \mu w_k(t;\beta,\nu)\exp(i\omega t - \alpha t); \qquad t = 0, 1, \ldots, N-1. \tag{3.2}$$

Take the discrete Fourier transform of both sides of (3.2):

$$e_k(\omega) = y_k(\omega) - \mu \tilde{W}_k(\omega - \hat{\omega};\beta,\nu), \tag{3.3}$$

where

$$e_k(\omega) = \sum_{t=0}^{N-1} e(t)\omega_k(t;\beta,\nu)\exp(-i\omega t)$$

$$y_k(\omega) = \sum_{t=0}^{N-1} x(t)w_k(t;\beta,\nu)\exp(-i\omega t)$$

and $\tilde{W}_k(\omega;\beta,\nu)$ is defined in equation (2.22). Because of the leakage resistance of the tapers, the $e_k(\omega)$ are approximately the complex eigenspectra of the noise in $(\omega - \Omega, \omega + \Omega)$.

We would like to make an estimate $\mu$ of the amplitude $\mu$ of a decaying sinusoid of frequency $\hat{\omega}$. To do this, a least-squares procedure is performed. At each frequency $\omega$, the complex eigenspectra $y_k(\omega)$, $k = 0, 1, \ldots, K-1$, are taken to be the dependent variables,

$\hat{\mu}$ is the parameter to be estimated, and the $\widetilde{W}_k(\omega - \hat{\omega}; \beta, \nu)$, $k = 0, 1, \ldots, K-1$, are the independent variables. By the Gauss–Markov theorem, to produce a minimum variance estimate of $\mu$ that is unbiased at the decaying sinusoid's true frequency using least squares, the random variables $y_k(\omega)$ must be statistically uncorrelated (Bickel & Doksum 1977, ch. 7; Luenberger 1969, ch. 4; Tukey 1975). However, the $y_k(\omega)$ are not necessarily uncorrelated random variables:

$$\begin{aligned}\mathrm{Cov}\,[y_k(\omega), y_{k'}(\omega)] &= \langle y_k(\omega) y_{k'}^*(\omega)\rangle \quad \langle y_k(\omega)\rangle\langle y_{k'}(\omega)\rangle^* \\ &= \sigma_N^2 \sum_{t=0}^{N-1} w_k(t;\beta,\nu)\, w_{k'}(t;\beta,\nu). \end{aligned} \tag{3.4}$$

The sum

$$H_{kk'} = \sum_{t=0}^{N-1} w_k(t;\beta,\nu)\, w_{k'}(t;\beta,\nu)$$

will not vanish unless $\beta = \nu = 0$ and $k \neq k'$. For $\beta = \nu = 0$, $H_{kk'} = \delta_{kk'}$. Elements of the matrix **H** for $\beta = 0.6$, $\Omega N = 8\pi$, and $\nu = 0.01$ and $\nu = 0.1$ are shown in Table 4.

Since **H** is symmetric and positive definite, it has a Cholesky decomposition. That is, there exists a lower triangular matrix **G** with positive diagonal entries such that

$$\mathbf{H} = \mathbf{G}\mathbf{G}^T \tag{3.5}$$

where the superscript $T$ denotes matrix transpose (Golub & VanLoan 1983).

Transform the complex eigenspectra $y_k(\omega)$ and the independent variables $\widetilde{W}_k(\omega; \beta, \nu)$ using the matrix $\mathbf{G}^{-1}$ as follows:

$$\begin{aligned} v_{k'}(t;\beta,\nu) &= (\mathbf{G}^{-1})_{k'k}\, w_k(t;\beta,\nu) \\ z_{k'}(\omega) &= (\mathbf{G}^{-1})_{k'k}\, y_k(\omega) \\ V_{k'}(\omega_j;\beta,\nu) &= (\mathbf{G}^{-1})_{k'k}\, \widetilde{W}_k(\omega_j;\beta,\nu) \\ g_{k'}(\omega) &= (\mathbf{G}^{-1})_{k'k}\, e_k(\omega). \end{aligned} \tag{3.6}$$

**Table 4.** Elements of matrix **H** for $\beta = 0.6$, $\Omega N = 8\pi$ and $\nu = 0.01$.

| | $k'$ | | | | |
|---|---|---|---|---|---|
| $k$ | 0 | 1 | 2 | 3 | 4 |
| 0 | 1.84870 | 0.18322 | −0.21788 | 0.23523 | −0.22668 |
| 1 | 0.18322 | 2.86999 | 0.35164 | −0.41490 | 0.44470 |
| 2 | −0.21788 | 0.35164 | 4.34203 | 0.61999 | −0.71893 |
| 3 | 0.23523 | −0.41490 | 0.61999 | 6.44037 | 1.04331 |
| 4 | −0.22668 | 0.44470 | −0.71893 | 1.04331 | 9.36201 |

for $\beta = 0.6$, $\Omega N = 8\pi$, and $\nu = 0.1$

| | $k'$ | | | | |
|---|---|---|---|---|---|
| $k$ | 0 | 1 | 2 | 3 | 4 |
| 0 | 1.37223 | 0.12714 | 0.13583 | −0.12606 | −0.09953 |
| 1 | 0.12714 | 1.90834 | −0.22102 | 0.23223 | 0.21434 |
| 2 | 0.13583 | −0.22102 | 2.56776 | 0.33639 | 0.34401 |
| 3 | −0.12606 | 0.23223 | 0.33639 | 3.33664 | −0.46931 |
| 4 | 0.09953 | 0.21434 | 0.34401 | −0.46931 | 4.17692 |

where $v_k(t;\beta,\nu)$, $z_k(\omega_j)$, $V_k(\omega_j;\beta,\nu)$ and $g_{k'}(\omega_j)$ are the transformed tapers, the transformed complex eigenspectra, the transformed independent variables, and the transformed errors respectively. We employ the Einstein summation convention in (3.6) and hereafter, summing repeated indices over the range $0, 1, \ldots, K-1$. From (3.3),

$$g_k(\omega) = z_k(\omega) - \hat{\mu} V_k(\omega - \hat{\omega};\beta,\nu). \tag{3.7}$$

The transformed complex eigenspectra $z_k(\omega)$ are uncorrelated, as

$$\begin{aligned}\mathrm{Cov}\,[z_k(\omega), z_{k'}(\omega)] &= \langle z_k(\omega) z_{k'}^*(\omega)\rangle \quad \langle z_k(\omega)\rangle\langle z_{k'}(\omega)\rangle^* \\ &= (\mathbf{G}^{-1})_{kl}\,\mathrm{Cov}\,[y_l(\omega), y_m(\omega)](\mathbf{G}^{-1})_{k'm} \\ &= \sigma_N^2 \delta_{kk'}\end{aligned} \tag{3.8}$$

by (3.4) and (3.5).

A measure of the error in assuming that the record $x(t)$ consists of a single decaying sinusoid of frequency $\hat{\omega}$ is

$$M(\omega) = \sum_{k=0}^{K-1} |g_k(\omega)|^2 = \sum_{k=0}^{K-1} |z_k(\omega) \quad \hat{\mu} V_k(\omega \quad \hat{\omega};\beta,\nu)|^2. \tag{3.9}$$

Perform a least squares procedure; solve

$$\frac{\partial M(\omega)}{\partial \hat{\mu}^*} = 0 \tag{3.10}$$

for $\hat{\mu}$. Then (3.10) becomes

$$0 = \sum_{k=0}^{K-1} V_k^*(\omega - \hat{\omega};\beta,\nu)\,[z_k(\omega) - \hat{\mu} V_k(\omega \quad \hat{\omega};\beta,\nu)]. \tag{3.11}$$

Note that $\hat{\mu}$ is actually a function of the frequency $\hat{\omega}$:

$$\hat{\mu} = \hat{\mu}(\hat{\omega}) = \frac{\sum_{k=0}^{K-1} V_k^*(\omega - \hat{\omega};\beta,\nu) z_k(\omega)}{\sum_{k=0}^{K-1} |V_k(\omega - \hat{\omega};\beta,\nu)|^2}. \tag{3.12}$$

One can determine $z_k(\omega)$ at a set of discrete frequencies $\omega_j; j = 0, 1, 2, \ldots, J-1$, called bin frequencies, by applying an FFT to the tapered data. The data can be padded with zeroes to interpolate the spectrum. (Note that this 'interpolation' adds no extra information.) To estimate the amplitude $\hat{\mu}$ of the proposed signal at each discrete bin frequency, set $\omega = \hat{\omega} = \omega_j; j = 0, 1, \ldots, J-1$ in (3.12). Then

$$\hat{\mu}(\omega_j) = \frac{\sum_{k=0}^{K-1} V_k^*(0;\beta,\nu) z_k(\omega_j)}{\sum_{k=0}^{K-1} |V_k(0;\beta,\nu)|^2}. \tag{3.13}$$

Substituting for $z_k(\omega)$ in (3.13), it is seen that this 'pointwise regression' formula for $\hat{\mu}$ is equivalent to a Fourier transform of the time series $\{x(t)\}_{t=0}^{N-1}$ with a hybrid taper $\{\hat{w}(t;\beta,\nu)\}_{t=0}^{N-1}$ given by the formula

$$\hat{w}(t;\beta,\nu) = \frac{\sum_{k=0}^{K-1} V_k(0;\beta,\nu)\, v_k(t;\beta,\nu)}{\sum_{k=0}^{K-1} |V_k(0;\beta,\nu)|^2}; \qquad t = 0, 1, \ldots, N-1 \tag{3.14}$$

($V_k^*(0;\beta,\nu) = V_k(0;\beta,\nu)$ since $\{w_k(t;\beta,\nu)\}_{t=0}^{N-1}$ is a real-valued sequence.) Note that $\{\hat{w}(t;\beta,\nu)\}_{t=0}^{N-1}$ is not optimal in the sense of (2.19).

In terms of the complex eigenspectra and taper frequency transforms:

$$\hat{\mu}(\omega_j) = \frac{\sum_{m=0}^{K-1}\sum_{l=0}^{K-1} \widetilde{W}_m^*(0;\beta,\nu)(\mathbf{H}^{-1})_{ml} y_l(\omega_j)}{\sum_{m=0}^{K-1}\sum_{l=0}^{K-1} \widetilde{W}_m^*(0;\beta,\nu)(\mathbf{H}^{-1})_{ml}\widetilde{W}_l(0;\beta,\nu)} \tag{3.15}$$

When $\beta = \nu = 0$, $\mathbf{H} = \mathbf{I}$, and (3.15) reduces to equation (13.5) of Thomson (1982).

If $\nu = 0$ (i.e. tapers designed without provision for stationary background noise) or $\beta = 0$ (tapers designed for non-decaying signals), $\widetilde{W}_k(0;\nu,0) = \widetilde{W}_k(0;0,\beta) = 0$ for odd $k$, since in both cases the $\widetilde{W}_k$ reduce to the transforms of discrete prolate spheroidal sequences. In these instances the pointwise regression technique ignores the odd order tapers completely in constructing $\hat{\mu}$.

By (3.8),

$$\begin{aligned}\mathrm{Var}\,[\hat{\mu}(\omega_j)] &= \frac{\sum_{k=0}^{K-1} |V_k(0;\beta,\nu)|^2 \,\mathrm{Var}\,[z_k(\omega_j)]}{\left(\sum_{k=0}^{K-1} |V_k(0;\beta,\nu)|^2\right)^2} \\ &= \frac{\sigma_N^2}{\sum_{m=0}^{K-1}\sum_{l=0}^{K-1} \widetilde{W}_m^*(0;\beta,\nu)(\mathbf{H}^{-1})_{ml}\widetilde{W}_l(0;\beta,\nu)}\end{aligned} \tag{3.16}$$

The variance of the estimated amplitude increases with increasing noise amplitude.

If there is no decaying sinusoid at frequency $\omega_j$, one would expect $\hat{\mu}$ to be small. However, this is not the best criterion for deciding if there is a decaying sinusoid at frequency $\omega_j$. The sinusoid may be present, but it may have a very small amplitude. Also, the least squares procedure may yield a large value for $\hat{\mu}$ at some frequency, but a decaying sinusoid may not be a good way to characterize the data at that frequency. A method of evaluating the fit of our decaying sinusoid model to the data is needed.

### 3.2 TESTING THE FIT OF THE MODEL TO THE DATA

A common technique for assessing the fit of a least-squares estimate is to perform a statistical

$F$-test (e.g. Wonnacott & Wonnacott 1981). An $F$-statistic is roughly the ratio

$$F = \frac{\text{variance explained by the model}}{\text{unexplained variance}}. \tag{3.17}$$

The random variable $F$ follows the $F$-distribution, which has been tabulated (e.g. Abramowitz & Stegun 1965). We use the $F$-test to compare the fit of the data to a decaying sinusoid model.

Suppose that the record $x(t)$ consists *solely* of zero-mean stationary Gaussian white noise $n(t)$. For free oscillation data, we have found that it is a reasonable approximation to say the background noise is Gaussian white noise and almost stationary. This can be demonstrated by generating ordered value plots of the data, as in Fig. 10 [Wilk & Gnanadesiken (1968) contains details on ordered value, or $P - P$ plotting of data].

As before, one estimates the complex amplitude $\mu$ of a decaying sinusoid of frequency $\omega$ by fitting the model $\mu \tilde{W}_k(\omega \quad \hat{\omega}; \beta, \nu)$ to the random variables

$$y_k(\omega) = \sum_{t=0}^{N-1} w_k(t; \beta, \nu) \exp(-i\omega t) n(t); \qquad k = 0, 1, \ldots, K \quad 1. \tag{3.18}$$

There is a finite probability that a decaying sinusoid model will fit the complex eigenspectra of the noise (3.18) at some frequency. The chance that this will happen is a measure of the confidence that a true decaying sinusoid exists at that frequency.

When no harmonic signal is present, the expected value of each transformed complex eigenspectrum vanishes:

$$\langle z_k(\omega_j) \rangle = 0. \tag{3.19}$$

However, the presence of noise, or signal, may cause any given transformed complex eigenspectrum $z_k(\omega_j)$ to be non-zero at some frequencies. This departure of $z_k(\omega_j)$ from its expected value may be partly 'explained' by the linear regression analysis. Using the estimated value $\hat{\mu}(\omega_j)$ from (3.13), the estimated value of $z_k(\omega_j)$ is

$$\hat{z}_k(\omega_j) = \hat{\mu}(\omega_j) V_k(0; \beta, \nu). \tag{3.20}$$

The deviation of $z_k(\omega_j)$ from $\langle z_k(\omega_j) \rangle$ may be decomposed into an 'explained' deviation, $[\hat{z}_k(\omega_j) \quad \langle z_k(\omega_j) \rangle]$ and an 'unexplained' deviation, $[z_k(\omega_j) \quad \hat{z}_k(\omega_j)]$:

$$[z_k(\omega_j) \quad \langle z_k(\omega_j) \rangle] = [\hat{z}_k(\omega_j) \quad \langle z_k(\omega_j) \rangle] + [z_k(\omega_j) \quad \hat{z}_k(\omega_j)]. \tag{3.21}$$

Or, summing over $k$, and noting that $\langle z_k(\omega_j) \rangle = 0$

$$\sum_{k=0}^{K-1} z_k(\omega_j) = \sum_{k=0}^{K-1} \hat{z}_k(\omega_j) + \sum_{k=0}^{K-1} [z_k(\omega_j) \quad \hat{z}_k(\omega_j)]. \tag{3.22}$$

The same equality holds when one takes the modulus squared of the deviations:

$$\sum_{k=0}^{K-1} |z_k(\omega_j)|^2 = \sum_{k=0}^{K-1} |\hat{z}_k(\omega_j)|^2 + \sum_{k=0}^{K-1} |z_k(\omega_j) \quad \hat{z}_k(\omega_j)|^2 \tag{3.23}$$

by multiplying (3.21) by its complex conjugate, and then summing over $k$. Substituting for $\hat{z}_k(\omega_j)$, (3.23) becomes

$$\sum_{k=0}^{K-1} |z_k(\omega_j)|^2 = |\hat{\mu}(\omega_j)|^2 \sum_{k=0}^{K-1} |V_k(0; \beta, \nu)|^2 + \sum_{k=0}^{K-1} |z_k(\omega_j) \quad \hat{\mu}(\omega_j) V_k(0; \beta, \nu)|^2 \tag{3.24}$$

or

$$\xi(\omega_j) = \theta(\omega_j) + \psi(\omega_j), \tag{3.25}$$

defining

$$\xi(\omega_j) \equiv \sum_{k=0}^{K-1} |z_k(\omega_j)|^2$$

$$\theta(\omega_j) \equiv |\hat{\mu}(\omega_j)|^2 \sum_{k=0}^{K-1} |V_k(0;\beta,\nu)|^2$$

$$\psi(\omega_j) \equiv \sum_{k=0}^{K-1} |z_k(\omega_j) - \hat{\mu}(\omega_j) V_k(0;\beta,\nu)|^2,$$

where $\xi(\omega_j)$ is the total sample variance of the $z_k(\omega_j)$, $\theta(\omega_j)$ is the sample variance explained by the decaying sinusoid hypothesis, and $\psi(\omega_j)$ is the residual, or unexplained sample variance.

We formulate a test to reject the null hypothesis that $\mu = 0$. Consider the random variable $\widetilde{F}$ formed by taking the ratio of the explained sample variance to the unexplained sample variance. Then

$$\widetilde{F}(\omega_j) = \frac{\theta(\omega_j)}{\psi(\omega_j)}$$

$$= \frac{|\hat{\mu}(\omega_j)|^2 \sum_{k=0}^{K-1} |V_k(0;\beta,\nu)|^2}{\sum_{k=0}^{K-1} |z_k(\omega_j) - \hat{\mu}(\omega_j) V_k(0;\beta,\nu)|^2} \tag{3.26}$$

If there is a decaying sinusoid at frequency $\omega_j$, the denominator $\psi(\omega_j)$ will be small, and thus the function $\widetilde{F}(\omega_j)$ will be large. By chance, sometimes a decaying sinusoid model will fit the time series $\{n(t)\}_{t=0}^{N-1}$ reasonably well at some frequency. The probability of this happening can be calculated. Therefore, one can describe quantitatively the confidence that there is a true signal at a given frequency.

We need to know how the random variable $\widetilde{F}(\omega_j)$ is related to the $F$-distribution. In Lindberg (1986) it is shown that

$$F(\omega_j) = (K-1)\widetilde{F}(\omega_j) = \frac{(K-1)\theta(\omega_j)}{\psi(\omega_j)} \tag{3.27}$$

follows an $F$-distribution with 2 and $(2K-2)$ degrees of freedom. Therefore, the chance that the random variable

$$F(\omega_j) = \frac{(K-1)|\hat{\mu}(\omega_j)|^2 \sum_{m=0}^{K-1} \sum_{l=0}^{K-1} \widetilde{W}_m^*(0;\beta,\nu)(\mathbf{H}^{-1})_{ml} \widetilde{W}_l(0;\beta,\nu)}{\sum_{m=0}^{K-1} \sum_{l=0}^{K-1} [y_m(\omega_j) - \hat{\mu}(\omega_j)\widetilde{W}_m(0;\beta,\nu)]^*(\mathbf{H}^{-1})_{ml}[y_l(\omega_j) - \hat{\mu}(\omega_j)\widetilde{W}_l(0;\beta,\nu)]} \tag{3.28}$$

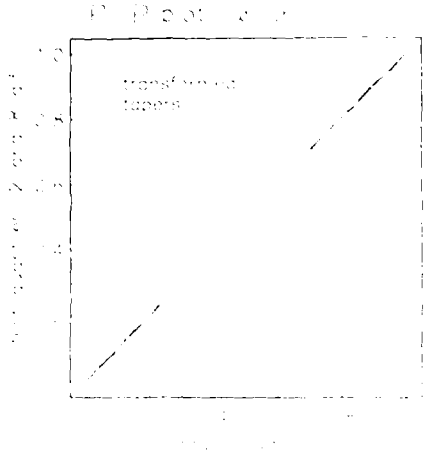


**Figure 10.** Ordered value, or $P$ – $P$ plot of 675 independent values of the ratio $F(\omega_j)$ in (3.28) using synthetic stationary Gaussian white noise as input data. The cumulative probability distribution of the ordered observations $F_{(1)} \leq F_{(2)} \leq \ldots \leq F_{(n-5)}$ is plotted on the ordinant against sample quantiles on the abscissa. The $j$th point is plotted as the ordered pair $[(j-.675), F_{(j)}]$. The graph is almost a straight line, demonstrating that the ratio $F(\omega_j)$ follows an $F$-distribution for Gaussian white noise input data.

takes on a particular value at some frequency due to random noise can be found using standard tables of the $F$-distribution (e.g. Abramowitz & Stegun 1965).

Figure 10 is an ordered value, or $P$ – $P$ plot (Wilk & Gnanadesikan 1968) of 675 independent values of the random variable $F(\omega_j)$ generated from a synthetic record of Gaussian white noise. If the sample followed an $F$-distribution exactly, the ordered value plot would lie on a straight line connecting the points (0, 0) and (1, 1). The departure of the ordered value plot from a straight diagonal line is not significant at the 95 per cent confidence level, using a Kolmogorov–Smirnov test for goodness of fit (Bickel & Doksum 1977). This demonstrates graphically that the ratio $F(\omega_j)$ follows an $F$-distribution when the data consist of Gaussian white noise.

## 4 Data examples

We illustrate the multiple-taper algorithm with two examples of decaying oscillations immersed in white noise. In the first, we analyse a synthetic IDA record in which the signal-to-noise power ratio is known *a priori*. In the second, we study a 340-hr record of the 1977 Sumbawa event from IDA station NNA (Nañia, Peru). Spectral estimates made by taking the DFT of cosine-tapered data are compared to results produced by the multiple taper technique. We find the multiple-eigentaper algorithm is superior for detecting low-amplitude decaying sinusoids in noise.

We have focused our attention on the gravest observed seismic free oscillation, the spheroidal multiplet ${}_0S_2$. (${}_1S_1$ has lower frequency, but this oscillation of the inner core has not yet been conclusively observed.) The multiplet ${}_0S_2$ consists of five decaying sinusoids at distinct frequencies. These 'singlets' are labelled by an azimuthal order number $m \in \{-2, -1, 0, 1, 2\}$. The five singlet frequencies of this oscillation are widely split by the rotation of the Earth, so much so that the magnitude of the quadratic second-order Coriolis

splitting is roughly 60 per cent that of the quadratic splitting caused by the Earth's hydrostatic ellipticity (Dahlen & Sailor 1979). The singlet frequencies have been measured by Buland *et al.* (1979) from spherical harmonic stacks of six 150-hr IDA records of the 1977 Sumbawa event. The multiplet ${}_0S_2$ is difficult to measure as it is excited by only the very largest earthquakes. Even for the Sumbawa event, the signal-to-noise ratio is not large. Also, some singlets have very small amplitudes at some stations because of the dependence of singlet amplitude on latitude. As a result, no more than two or three of the five singlet resonance functions can be seen in any of the conventional amplitude spectra of records from the seven IDA stations existing at that time.

We constructed a 300-hr synthetic IDA record from CMO (College, Alaska) using a source located in Oaxaca, Mexico. The five singlets of ${}_0S_2$, split by rotation and ellipticity, were included in the seismogram (see Park & Gilbert (1986) for an outline of the computation procedure). Gaussian white noise was added to the record with amplitude scaled so that $N|\mu|^2/\sigma_N^2 = 73$ for the $m = 0$ singlet oscillation, $N|\mu|^2/\sigma_N^2 = 22.5$ for the $m = \pm 1$ singlets, and $N|\mu|^2/\sigma_N^2 = 3.6$ for the $m = \pm 2$ singlets. The record was sampled at 160 s intervals to produce a time series of 6750 points. We analysed the record with five eigentapers with $\Omega N = 8\pi$, $\beta = 0.6$, $\nu = 0.01$ to produce amplitude estimates $\hat{\mu}(\omega)$ and an $F$-test of the fit of $\hat{\mu}(\omega)$ to the complex eigenspectra. Five tapers were chosen because the five lowest order eigentapers with $\Omega N = 8\pi$, $\beta = 0.6$ and $\nu = 0.01$ have fractional leakage of 0.01 or less (Table 2). We also produced a spectral estimate using a cosine taper for comparison. According to arguments outlined in the appendix, $\langle F \rangle$ should be near the 99 per cent confidence level for the $m = \pm 1$ lines and considerably greater for the $m = 0$ line. The $m = \pm 2$ lines have $\langle F \rangle \approx 2.25$, but large random fluctuations in $F$ are possible.

The spectral estimate using a cosine taper $|y_c(\omega)|$ is compared with the multitaper amplitude estimate $|\hat{\mu}(\omega)|$ in Fig. 11. We graph the frequency band $280 < f < 340\,\mu$Hz containing the five singlets of ${}_0S_2$ and no other known seismic free oscillation. The ordinate scales of the plots do not match because $y_c(\omega)$ is an estimated amplitude spectrum and $\hat{\mu}(\omega)$ is the amplitude of a presumed harmonic signal at $f = 0$. Many features of the plots are similar, however, because both represent discrete Fourier transforms of tapered data [$\hat{\mu}(\omega)$ corresponds to the DFT of the data times a hybrid taper as shown in (3.13)–(3.14)]. The $m = -2$, 0 and 1 singlets, having frequencies given in Table 5, are readily discernible. The prominence of the $m = -2$ singlet is puzzling in light of its low input-amplitude. The $m = -1$ singlet appears to be obscured somewhat by noise interference.

The $F$-test of the fit of $\hat{\mu}(\omega)$ to the complex eigenspectra is graphed in Fig. 12. All five singlets of ${}_0S_2$ are observable with better than 95 per cent detection confidence. Their measured frequencies are given in Table 5, along with estimates of the expected errors in the

**Table 5.** Frequencies of ${}_0S_2$ in synthetic record

| | Singlet Azimuthal Order $m$ | | | | |
|---|---|---|---|---|---|
| **Input Singlets** | −2 | −1 | 0 | 1 | 2 |
| **Input Frequency (mHz)** | .299800 | .304615 | .309337 | .313874 | .318226 |
| **Input Phase** | −85° | −133° | −2.7° | 126° | 72° |
| ***F*-Test Results** | | | | | |
| **Frequency (mHz)** | .29973 | .30436 | .309356 | .31371 | .31889 |
| **Frequency Uncertainty** | .00022 | .00034 | .000074 | .00018 | .00035 |
| **Phase** | −74° | −118° | −1.5° | 148° | 87° |
| ***F*-value** | 5.5 | 13.7 | 86.0 | 66.6 | 5.9 |

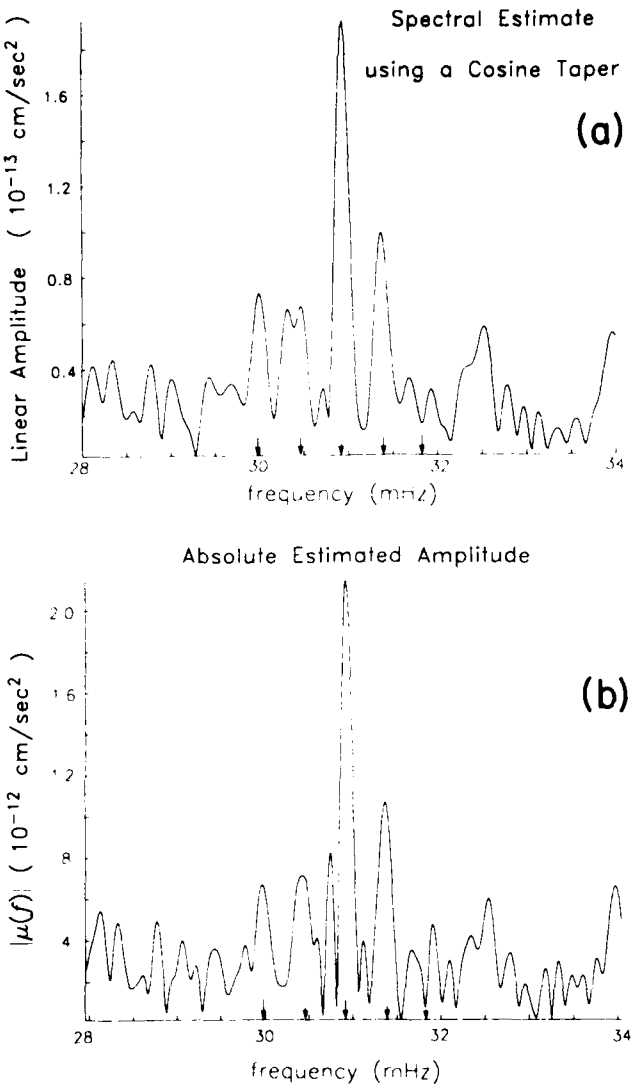



**Figure 11.** (a) Amplitude of a spectral estimate using a cosine taper $|y_c(\omega)|$ for a synthetic record of $_nS_2$. (b) The function $|\hat{\mu}(\omega)|$, where $\hat{\mu}(\omega)$ is the estimated complex amplitude of a decaying sinusoid in a synthetic record of $_nS_2$ using five eigentapers with parameters $\Omega N = 8\pi$, $\beta = 0.6$ and $\nu = 0.01$. In both (a) and (b), three of the five singlets of $_nS_2$ are visible. The true positions of the input singlets are marked

frequencies produced by the method described in the appendix. The most poorly fit frequency observation is within $2\sigma$ of the true value. Note the rough equivalence of the $F$-test values for the $m = \pm 2$ singlet lines. The amplitude of the $m = -2$ singlet in Fig. 11 is enhanced by noise fluctuations, but the noise contribution has incoherent phase, causing the $m = -2$ $F$-test value to fall relative to that of neighbouring oscillation peaks. On the other hand, an apparent noise-minimum at the frequency of $m = +2$ single-line allows its small

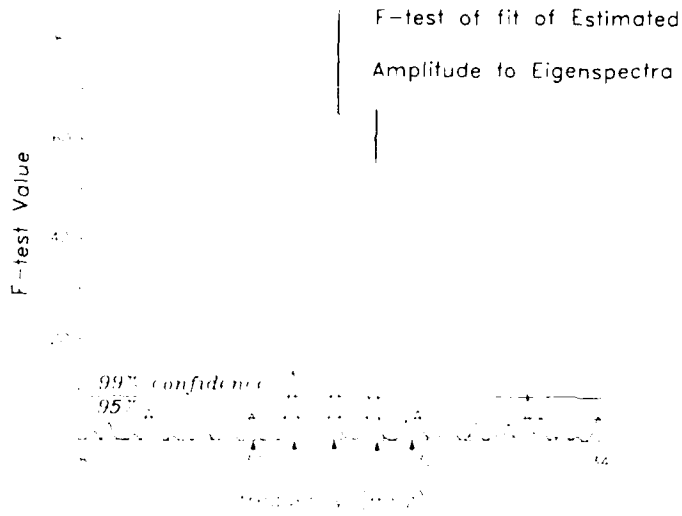



**Figure 12.** $F$-test values resulting from a test of the fit of estimated amplitude $\hat{\mu}(\omega)$ to the eigenspectra obtained using five eigentapers with parameters $\Omega N = 8\pi$, $\beta = 0.6$, and $\nu = 0.01$. The data is a synthetic record of ${}_0S_2$; it consists of five decaying sinusoids whose frequencies are listed in Table 5. All five have peaks above the 95 per cent confidence level. The function $|\hat{\mu}(\omega)|$ is plotted in Fig. 11b. The value $F = 3.11$ corresponds to the 90 per cent confidence level, $F = 4.46$ is the 95 per cent confidence level, and $F = 8.65$ is the 99 per cent level. The true positions of the input singlets are marked.

*amplitude to be detectable in the plot of the F-test.* Note also that the $F$-test has peaks at frequency values not associated with ${}_0S_2$ singlets. These are caused by random statistical fluctuations. The frequency band shown contains 65 independent frequency samples. Therefore, one would expect that due to randomness, roughly three values of the $F$-test in Fig. 12 would protrude above $F \approx 4.5$, the 95 per cent confidence level for the $F$-distribution.

We also took 340 hr of vertical IDA gravimeter data from station NNA, starting 8.5 hr after the onset of the Sumbawa event. This record is relatively complete, with only two data gaps of roughly 2.5 hr each at 95 and 275 hr into the record. Time series points falling in the gaps were assigned the value zero. The data were sampled at 20 s intervals. We low-pass filtered and decimated the record so that it contained 7668 points taken at 160 s intervals. Aftershocks that did not visibly affect the instrument in a non-linear manner were retained, as their effect on the spectrum in the vicinity of ${}_0S_2$ is small. Sections exhibiting non-linear seismometer response contribute significant energy at low frequencies, and so these were removed.

We had to know roughly the $Q$'s of the singlets of ${}_0S_2$ to apply our procedure. The $Q \approx 560$ value for ${}_0S_2$ given by the model of Masters & Gilbert (1983) corresponds to $\beta = 0.68$. Chao & Gilbert (1980) estimate that the $m = 2$ singlet of ${}_0S_2$ has a $Q$ of 415, the $m = 0$ singlet has a $Q$ of 609 and the $m = 2$ singlet has a $Q$ of 509. The $Q$ measurement reported by Hansen & Schnapp (1982) leads to a decay parameter of $\beta = 0.84$. We analysed the record with a set of five eigentapers having parameters $\Omega N = 8\pi$, $\nu = 0.01$ and $\beta = 0.65$.

The function $|\hat{\mu}(\omega)|$ obtained using the eigentapers is plotted in Fig. 13b, and the amplitude of the spectral estimate using the cosine taper $|y_c(\omega)|$ is presented in Fig. 13a. Again, we graph the frequency band $280 < f < 340\,\mu$Hz. Spectra were calculated at frequencies separated by $0.163\,\mu$Hz using the DFT. Table 6 lists the frequency estimates of the five

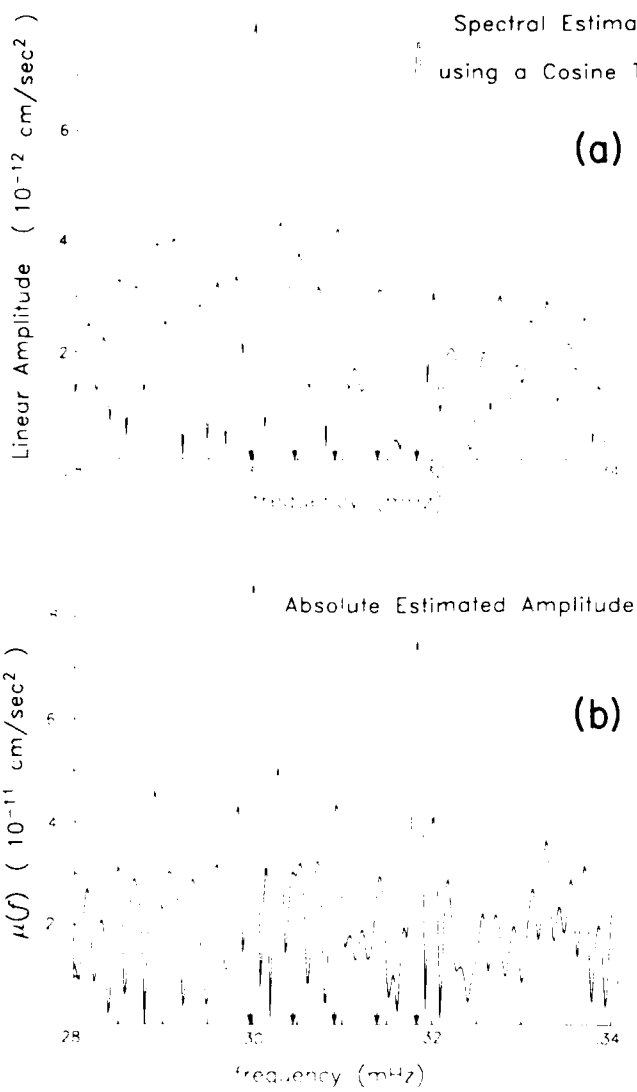



**Figure 13.** (a) Amplitude of a spectral estimate using a cosine taper for a time series of the Sumbawa event recorded at IDA station NNA. We plot the frequency band $280 < f < 340\,\mu$Hz containing the five ${}_0S_2$ singlets. Only two of the singlets are observable. (b) Amplitude of the function $\hat{\mu}(\omega)$ for a time series of the Sumbawa event recorded at IDA station NNA. We plot the frequency band $280 < f < 340\,\mu$Hz containing the five ${}_0S_2$ singlets, but again only two singlets are visible. The positions of the five singlets as determined by stacking are indicated.

singlets of ${}_0S_2$ made by Buland *et al.* (1979); these frequencies are marked in Fig. 13a and b. Only the $m = \pm 2$ lines are clearly visible in Fig. 13a and b. Candidates for the other singlet resonances are evident but do not protrude significantly above the apparent ambient noise level.

Figure 14 is a graph of the $F$-test of the fit of $\hat{\mu}(\omega)$ to the complex eigenspectra. There are

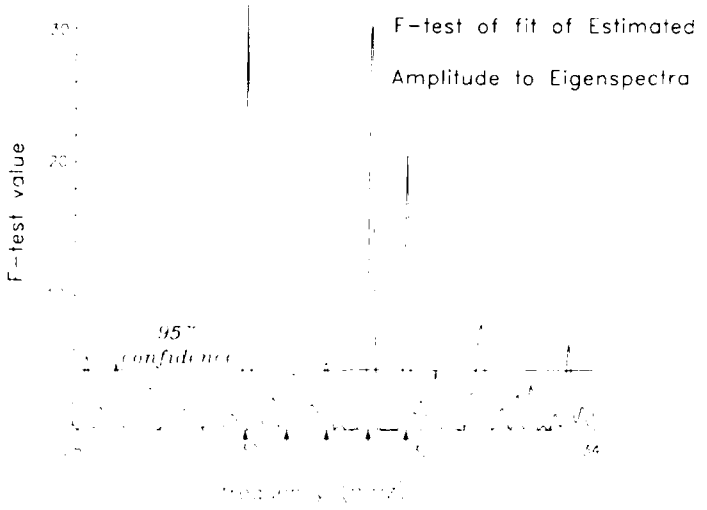



**Figure 14.** $F$-test for the estimated amplitude $\hat{\mu}(\omega)$ plotted in Fig. 13a. The time series being analysed is a record of Sumbawa event from IDA station NNA. We plot the frequency band $280 < f < 340\,\mu$Hz containing the five ${}_0S_2$ singlets. Four of the singlets have $F$-test peaks corresponding to greater than 95 per cent confidence of detection. The positions of the five singlets as determined by stacking are indicated.

four peaks above the 95 per cent detection confidence level in Fig. 14 which correspond to singlets of ${}_0S_2$. The $m = -1$ singlet appears to be contaminated by noise, resulting in a low, asymmetric variance-ratio peak. The estimated frequencies of all five lines, and their associated uncertainties, are listed in Table 6. The discrepancy between the $m = -1$ frequency estimate and that of Buland *et al.* (1979) is another indicator of the noise contamination of the $m = -1$ singlet. The other peaks in Fig. 14 above the 95 per cent confidence level are most likely due to random fluctuations.

In the above examples, we knew (approximately) the frequencies of the decaying oscillations and that they had large enough amplitude to be detectable. To be useful, the multitaper detection algorithm for decaying sinusoids should yield comparable results when either or both of the above conditions are not satisfied. Given the known frequencies of the gravest seismic oscillations, one could use the algorithm to search for so-called 'silent' events (e.g. Kanamori & Cipar 1974), whose existence is still controversial. In the more conservative enterprise of expanding and refining the free-oscillation data set in order to constrain deep Earth structure more reliably, the eigentaper algorithm offers hope of retrieving more

**Table 6.** Frequencies of ${}_0S_2$ in NNA record of Sumbawa event

| | Singlet Azimuthal Order $m$ | | | | |
|---|---|---|---|---|---|
| $F$-Test Results | −2 | −1 | 0 | 1 | 2 |
| Frequency (mHz) | .29988 | .30526 | .30918 | .31423 | .31830 |
| Frequency Uncertainty | .00027 | .00060 | .00041 | .00016 | .00013 |
| Phase | −134° | −48° | −43° | 142° | 10° |
| $F$-Value | 31.7 | 4.2 | 5.0 | 30.0 | 20.5 |
| From Buland *et al.* (1979) | | | | | |
| Frequency (mHz) | .300010 | .304799 | .309490 | .314000 | .318499 |

marginally observable modes than are accessible using single-taper algorithms. Care must be exercised that peaks in the $F$-test due to random noise are not misidentified as seismic free-oscillations. To this end, quantitative comparison of more than one seismic record is essential. This has been done by combining the standard techniques of stacking and stripping of low-frequency seismic records (Gilbert & Dziewonski 1975) with the multitaper algorithm. This is discussed in Part II of this paper.

## 5 Summary

We have described a variational procedure for determining tapers that optimally resist spectral leakage from outside a frequency region of bandwidth $2\Omega$ for exponentially decaying sinusoids contaminated by white noise. Multiplying the data by these tapers creates a number of time series. A decaying sinusoid model is fitted to the discrete Fourier transforms of the tapered data series at each frequency of interest (equation 3.15). The fit of this model to the data is tested at each frequency using a statistical $F$-test (equation 3.28). This gives a quantitative measure of the chance that there is a decaying sinusoid at any given frequency in the data. We have shown that this procedure is a sensitive detector of decaying harmonic lines in free oscillation data.

In Part II of this paper, we shall present a number of extensions to the multiple-taper method of harmonic analysis. We shall explain how the technique has been modified to estimate the harmonic components of records containing gaps. We discuss how sinusoids at frequencies between the discrete FFT bin frequencies can be detected, and how this method can be combined with conventional multi-station stacking procedures. The resolution of closely spaced harmonic lines is treated. Subsequently, we plan to introduce algorithms for finding the decay rates of free oscillations, as well as their frequencies.

## Acknowledgments

The authors would like to thank Drs Freeman Gilbert, Robert L. Parker and Alan D. Chave for their encouragement during this work. Dr John Rice helped us with some of the statistical definitions. Guy Masters provided us with edited IDA records for exploratory data analysis. This work was supported by National Science Fundation Grant EAR-84-09612 and Office of Naval Research contracts N00014-85-C-0104 and N00014-80-C-0440; D. J. Thomson was partially supported by the La Jolla Foundation for Earth Sciences during his stay at the Institute of Geophysics and Planetary Physics as a Green Scholar during 1983.

## References

Abramowitz, M. & Stegun, I. A. (eds), 1965. *Handbook of Mathematical Functions*, Dover, New York.

Agnew, D. C. & Berger, J., 1978. Vertical seismic noise at very low frequencies, *J. geophys. Res.*, **83**, 5420–5424.

Agnew, D. C., Berger, J., Buland, R., Farrell, W. & Gilbert, F., 1976. International deployment of accelerometers: A network for very long period seismology, *Eos, Trans. Am. geophys. Un.*, **57**, 180–188.

Bickel, P. J. & Doksum, K. A., 1977. *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, San Francisco.

Brigham, E. O., 1974. *The Fast Fourier Transform*, Prentice-Hall, Englewood Cliffs, NJ.

Buland, R. & Gilbert, F., 1978. Improved resolution of complex eigenfrequencies in analytically continued seismic spectra, *Geophys. J. R. astr. Soc.*, **52**, 457–470.

Buland, R., Berger, J. & Gilbert, F., 1979. Observations from the IDA network of attenuation and splitting during a recent earthquake, *Nature*, **277**, 358–362.

Chao, B. F. & Gilbert, F., 1980. Autoregressive estimation of complex eigenfrequencies in low-frequency seismic spectra, *Geophys. J. R. astr. Soc.*, **63**, 641–657.

Cooley, J. W. & Tukey, J., 1965. An algorithm for the machine calculation of complex Fourier series, *Maths Comput.*, **19**, 297–301.

Dahlen, F. A., 1982. The effect of data windows on the estimation of free oscillation parameters, *Geophys. J. R. astr. Soc.*, **69**, 537–549.

Dahlen, F. A. & Sailor, R., 1979. Rotational and elliptical splitting of the free oscillations of the Earth, *Geophys. J. R. astr. Soc.*, **58**, 609–623.

Engdahl, E. R., Peterson, J. & Orsini, N. A., 1982. Global digital networks—Current status and future directions, *Bull. seism. Soc. Am.*, **72**, S243–S259.

Gilbert, F. & Backus, G., 1965. The rotational splitting of the free oscillations of the Earth: 2, *Rev. Geophys. Space Phys.*, **3**, 1–9.

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. A.*, **278**, 187–269.

Goldstein, H., 1980. *Classical Mechanics*, 2nd edn, Addison-Wesley, Menlo Park, CA.

Golub, G. H. & VanLoan, C. F., 1983. *Matrix Computations*, Johns Hopkins University Press, Baltimore.

Hansen, R. A. & Schnapp, M. G., 1982. Estimation of the $Q$ of split mode free oscillations (abstract), *Eos, Trans. Am. geophys. Un.*, **63**, 1034.

Harris, F. J., 1978. On the use of windows for harmonic analysis with the discrete Fourier transform, *Proc. IEEE*, **66**, 51–83.

Jones, R. H., 1962. Spectral estimates and their distributions, in *Skandinavisk Aktuarietidskrift*, **45**, Part I: 39–69, Part II: 135–153.

Kanamori, H. & Cipar, J. J., 1974. Focal process of the great Chilean Earthquake May 22, 1960, *Phys. Earth planet. Int.*, **9**, 128–136.

Kendall, M. & Stuart, A., 1979. *The Advanced Theory of Statistics*, vol. 2, 4th edn, Macmillan, New York.

Landau, H. J. & Pollak, H. O., 1961. Prolate spheroidal wave functions, Fourier analysis and uncertainty II, *Bell Systems Tech. J.*, **40**, 65–84.

Landau, H. J. & Pollak, H. O., 1962. Prolate spheroidal wave functions, Fourier analysis and uncertainty III, *Bell Systems Tech. J.*, **40**, 1295–1336.

Lindberg, C., 1986. Multiple taper harmonic analysis of terrestrial free oscillations, *PhD Thesis*, University of California.

Lindberg, C. R., Vernon III, F. L. & Park, J. J., 1988. Comparison of conventional and multitaper spectral estimation methods on synthetic high frequency seismic wavetrains.

Luenberger, D. G., 1969. *Optimization by Vector Space Methods*, John Wiley and Sons, New York.

Masters, T. G. & Gilbert, F., 1981. Structure of the inner core inferred from observations of its spheroidal shear modes, *Geophys. Res. Lett.*, **8**, 569–571.

Masters, G. & Gilbert, F., 1983. Attenuation in the earth at low frequencies, *Phil. Trans. R. Soc. A.*, **308**, 479–522.

Miller, K. S., 1974. *Complex Stochastic Processes: An Introduction to Theory and Application*, Addison-Wesley, Reading, MA.

Park, J. & Gilbert, F., 1986. Coupled free oscillations of an aspherical, dissipative, rotating Earth: Galerkin theory, *J. geophys. Res.*, **91**, 7241–7260.

Park, J., Lindberg, C. R. & Vernon III, F. L., 1987. Multitaper spectral analysis of high frequency seismograms, *J. geophys. Res.*, in press.

Riedesel, M. A., Agnew, D. C., Berger, J. & Gilbert, F., 1980. Stacking for the frequencies and $Q$'s of ${}_0S_0$ and ${}_1S_0$, *Geophys. J. R. astr. Soc.*, **62**, 457–471.

Slepian, D., 1964. Prolate spheroidal wave functions, Fourier analysis and uncertainty—IV, *Bell Systems Tech. J.*, **43**, 3009–3057.

Slepian, D., 1978. Prolate spheroidal wave functions, Fourier analysis and uncertainty—V: The discrete case, *Bell Systems Tech. J.*, **57**, 1371–1429.

Slepian, D., 1983. Some comments on Fourier analysis, uncertainty and modeling, *SIAM Rev.*, **25**, 379–393.

Slepian, D. & Pollak, H. O., 1961. Prolate spheroidal wave functions, Fourier analysis and uncertainty—I, *Bell Systems Tech. J.*, **40**, 43–64.

Smith, B. T., Boyle, J. M., Dongerra, J. J., Garbow, B. S., Ikebe, Y., Klema, V. C. & Moler, C. B., 1976. Matrix Eigensystem Routines—EISPACK Guide, Lecture Notes in *Computer Science*, Springer-Verlag, New York.

Smith, D. R., 1974. *Variational Methods in Optimization*, Prentice-Hall, Englewood Cliffs, NJ.

Thomson, D. J., 1982. Spectrum estimation and harmonic analysis, *Proc. IEEE*, **70**, 1055–1096.
Tukey, J. W., 1975. Instead of Gauss-Markov least squares, what? in *Applied Statistics*, pp. 351–372, ed. Gupta, R. P., North-Holland Publishing House, Amsterdam.
Wilk, M. B. & Gnanadesikan, R., 1968. Probability plotting methods for the analysis of data, *Biometrica*, **55**, 1–17.
Wonnacott, T. H. & Wonnacott, R. J., 1981. *Regression: A Second Course in Statistics*, John Wiley, New York.

## Appendix: error estimation

The methods of Section 3 can be used to obtain estimates of the complex amplitude and the frequency of a decaying sinusoid in a time series. Random noise can cause the estimated amplitude and estimated frequency to deviate from the true values. This appendix outlines methods for calculating the expected size of these deviations.

### A1 Estimated amplitude

First, consider the estimated amplitude $\hat{\mu}(\omega)$. It is a statistical estimator of the true amplitude $\mu$. The utility of $\hat{\mu}$ as an estimator can be gauged by its bias $\langle \hat{\mu} - \mu \rangle$ and its mean square error $\langle |\hat{\mu} - \mu|^2 \rangle$. Let the data $x(t)$ be zero mean white noise $n(t)$ plus a decaying sinusoid with frequency $\omega_T$. Then

$$x(t) = n(t) + \mu \exp(i\omega_T t - \alpha t); \qquad t = 0, 1, \ldots, N-1 \tag{A.1}$$

where $\mu$ is the true complex amplitude, $\alpha$ is the true decay rate, and $\langle n(t) n^*(t') \rangle = \sigma_N^2 \delta_{tt'}$. We also assume that $\langle n(t) n(t') \rangle = 0$. [This is justified as only the real part of the $n(t)$ is actually measured, leaving us free to define its imaginary part. Miller (1974, p. 41) gives further details.] The $k$th transformed complex eigenspectrum of the data is

$$z_k(\omega) = (\mathbf{G}^{-1})_{kl} y_l(\omega)$$
$$= g_k(\omega) + \mu V_k(\omega - \omega_T; \beta, \nu), \tag{A.2}$$

where

$$g_k(\omega) = (\mathbf{G}^{-1})_{kl} \sum_{t=0}^{N-1} \exp(-i\omega t) w_l(t; \beta, \nu) n(t)$$

and $\mathbf{G}$, $y_l(\omega)$ and $V_k$ are as defined in Section 3. It follows that

$$\langle z_k(\omega) \rangle = \mu V_k(\omega - \omega_T; \beta, \nu) \tag{A.3}$$

and

$$\langle z_k(\omega) z_l^*(\omega) \rangle = \sigma_N^2 \delta_{kl} + |\mu|^2 V_k(\omega - \omega_T; \beta, \nu) V_l^*(\omega - \omega_T; \beta, \nu). \tag{A.5}$$

The expected value of $\hat{\mu}$, combining (3.13) and (A.3) is

$$\langle \hat{\mu}_{TW} \rangle = \frac{\mu \sum_{k=0}^{K-1} V_k^*(0; \beta, \nu) V_k(\omega - \omega_T; \beta, \nu)}{\sum_{k=0}^{K-1} |V_k(0; \beta, \nu)|^2} \tag{A.6}$$

When $\omega = \omega_T$, $\langle \hat{\mu}(\omega_T) \rangle = \mu$, so $\hat{\mu}$ is unbiased at the true frequency $\omega_T$. At other frequencies $\hat{\mu}$ is a biased estimator of $\mu$.

The mean square error of $\hat{\mu}$ is constant at all frequencies, and using (A.6), is

$$\langle |\hat{\mu} \quad \mu|^2 \rangle = \frac{\sigma_N^2}{\rho}, \tag{A.7}$$

where

$$\rho = \sum_{k=0}^{K-1} |V_k(0; \beta, \nu)|^2$$

## A2 Frequency estimates

Now consider the estimation of the true frequency $\omega_T$. The true frequency can be estimated from (1) the frequencies of peaks in the modulus of the estimated amplitude $|\hat{\mu}|^2$, (2) minima in the unexplained sample variance $\psi(\omega_j)$ introduced in (3.25), or (3) peaks in the random variable $F(\omega_j) = (K \quad 1)\theta(\omega_j)/\psi(\psi_j)$. These all provide approximately unbiased estimates of the true frequency $\omega_T$, and their mean square errors can be computed, as shown below.

### A2.1 FREQUENCIES ESTIMATED FROM PEAKS IN $|\hat{\mu}|^2$

The function $|\hat{\mu}(\omega)|^2$ achieves a peak at frequency $\omega_\theta$, where

$$0 = \theta'(\omega_\theta) = \left( \frac{d}{d\omega} |\hat{\mu}(\omega)|^2 \right)_{\omega = \omega_\theta}. \tag{A.8}$$

In a neighbourhood of the true frequency $\omega_T$

$$0 = \theta'(\omega_\theta) \approx \theta'(\omega_T) + (\omega_\theta \quad \omega_T)\theta''(\omega_T). \tag{A.9}$$

Taking expectation values of both sides, and assuming that $(\omega_\theta \quad \omega_T)$ and $\theta''(\omega_T)$ are uncorrelated:

$$\langle \omega_\theta \quad \omega_T \rangle \approx - \frac{\langle \theta'(\omega_T) \rangle}{\langle \theta''(\omega_T) \rangle}. \tag{A.10}$$

Define the matrix $\mathbf{M}^{(j)}$ with elements

$$M_{ml}^{(j)} = (m \quad l)^j; \qquad m, l = 0, 1, 2, \ldots, N \quad 1, \tag{A.11}$$

where $j$ is an integer, and the vector $\hat{\mathbf{v}}$ with elements

$$\hat{v}(t) = \frac{\sum_{k=0}^{K-1} V_k^*(0; \beta, \nu) v_k(t; \beta, \nu)}{\rho} \exp(\quad \alpha t); \qquad t = 0, 1, 2, \ldots, N \quad 1. \tag{A.12}$$

Then some algebra shows that

$$\langle \theta'(\omega_T) \rangle = \quad i|\mu|^2 \rho \hat{\mathbf{v}} \cdot \mathbf{M}^{(1)} \cdot \hat{\mathbf{v}} = 0 \tag{A.13}$$

as $\mathbf{M}^{(1)}$ is antisymmetric, and

$$\langle \theta''(\omega_T) \rangle = \quad \rho |\mu|^2 \hat{\mathbf{v}} \cdot \mathbf{M}^{(2)} \cdot \hat{\mathbf{v}}. \tag{A.14}$$

Therefore, from (A.10), $\langle \omega_\theta - \omega_T \rangle = 0$ and $\omega_\theta$ is an unbiased estimator of $\omega_T$. To find the mean square error of $\omega_\theta$, square both sides of (A.9) and take expected values:

$$\langle (\omega_\theta - \omega_T)^2 \rangle \approx \frac{\langle [\theta'(\omega_T)]^2 \rangle}{\langle [\theta''(\omega_T)]^2 \rangle}. \tag{A.15}$$

Using the relation

$$\begin{aligned}\langle x(t_1)x^*(t_2)x(t_3)x^*(t_4)\rangle = |\mu|^4 \exp[i\omega_T(t_1 - t_2 + t_3 - t_4)] \exp[-\alpha(t_1 + t_2 + t_3 + t_4)] \\ + \sigma_N^2 |\mu|^2 \delta_{t_1 t_2} \exp[i\omega_T(t_3 - t_4) - \alpha(t_3 + t_4)] \\ + \delta_{t_3 t_4} \exp[i\omega_T(t_1 - t_2) - \alpha(t_1 + t_2)] \\ + \delta_{t_1 t_4} \exp[i\omega_T(t_3 - t_2) - \alpha(t_3 + t_2)] \\ + \delta_{t_2 t_3} \exp[i\omega_T(t_1 - t_4) - \alpha(t_1 + t_4)] \\ + \sigma_N^4 (1 + \tfrac{1}{2}\delta_{t_1 t_3})(\delta_{t_1 t_2}\delta_{t_3 t_4} + \delta_{t_1 t_4}\delta_{t_2 t_3})\end{aligned} \tag{A.16}$$

one finds that

$$\langle [\theta'(\omega_T)]^2 \rangle = \rho \{\sigma_N^4 \, \mathbf{s} \cdot \mathbf{M}^{(2)} \cdot \mathbf{s} + 2\sigma_N^2 |\mu|^2 \mathbf{s} \cdot \mathbf{r}\} \tag{A.17}$$

and

$$\langle [\theta''(\omega_T)]^2 \rangle = \rho \{\sigma_N^4 \, \mathbf{s} \cdot \mathbf{M}^{(4)} \cdot \mathbf{s} + 2\sigma_N^2 |\mu|^2 \mathbf{s} \cdot \mathbf{r} + |\mu|^4 (\mathbf{v} \cdot \mathbf{M}^{(2)} \cdot \mathbf{v})^2\}, \tag{A.18}$$

where $\mathbf{s}$ has components

$$s_t = [v(t) \exp(\alpha t)]^2; \qquad t = 0, 1, \ldots, N-1 \tag{A.19}$$

and $\mathbf{r}$ has components

$$r_t = [(\mathbf{M}^{(2)} \cdot \mathbf{v})_t]^2; \qquad t = 0, 1, \ldots, N-1. \tag{A.20}$$

For sufficiently large initial signal-to-noise ratios, $N|\mu|^2/\sigma_N^2 \gg 1$, and

$$\langle (\omega_\theta - \omega_T)^2 \rangle \approx \frac{2\sigma_N^2}{|\mu|^2} \frac{\mathbf{s} \cdot \mathbf{r}}{\mathbf{v} \cdot \mathbf{M}^{(2)} \cdot \mathbf{v}}. \tag{A.21}$$

The mean-square error of the estimator $\omega_\theta$ decreases as the signal-to-noise ratio increases.

Figure A1 is a plot of the estimated rms misfit of $\omega_\theta$, defined by

$$\langle \omega_\theta - \omega_T \rangle_{\text{rms}} \equiv \sqrt{\langle (\omega_\theta - \omega_T)^2 \rangle}$$

as a function of initial signal-to-noise ratio for tapers with parameters $\beta = 0.6, 1.05$ and $\nu = 0, 0.01, 0.1, 1$ using (A.15). The misfit is plotted on the ordinate as a fraction of the Rayleigh frequency $\omega_R = 2\pi/T$, where $T = N\Delta t$. The parameter $N|\mu|^2/\sigma_N^2$ is plotted on the abscissa. One expects frequency uncertainty to increase rapidly with decreasing signal-to-noise ratio, but for $N|\mu|^2/\sigma_N^2 < 10$, the estimated frequency uncertainty in Fig. A1 is essentially constant. This is because relation (A.15) ceases to be a good approximation at low signal-to-noise ratios, where the first-order expansion (A.9) fails to hold, and $(\omega_\theta - \omega_T)$ and $\theta''(\omega_T)$ are correlated. The solid curve corresponds to $\nu = 0$; larger values of $\nu$ correspond to succeedingly finer dashed curves. The rms misfit $\langle \omega_\theta - \omega_T \rangle_{\text{rms}}$ tends to decrease with decreasing values of $\nu$ (except for $\beta = 1.05$ and $N|\mu|^2/\sigma_N^2 > 80$).

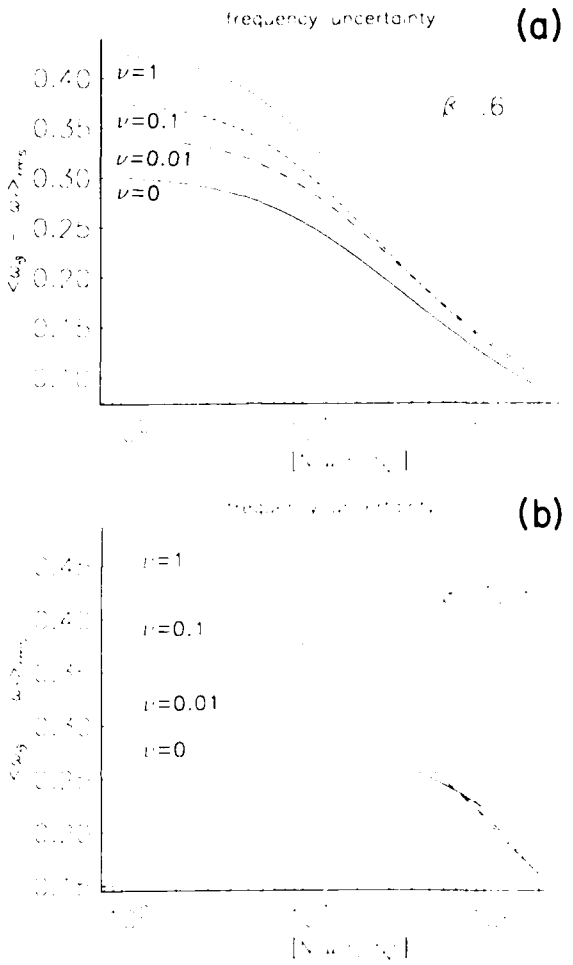



**Figure A1** (a, b). Estimated rms misfit of estimated frequency $\omega_\theta$ to true frequency $\omega_T$ as a function of $N|\mu|^2/\sigma_V^2$ for tapers with $\Omega N = 8\pi$, $\beta = 0.6$, 1.05 and $\nu = 0$, 0.01, 0.1 and 1. The curves are meaningless for $N|\mu|^2/\sigma_V^2 < 10$ because (A.15) fails to be a good approximation. Uncertainty decreases with decreasing values of noise parameter $\nu$ (except for $\beta = 1.05$ and $N|\mu|^2/\sigma_V^2 > 80$). Uncertainty decreases with increasing signal-to-noise ratio.

### A2.2 FREQUENCIES ESTIMATED FROM MINIMA IN $\psi(\omega)$

Another estimator of the true frequency $\omega_T$ is $\omega_\psi$, the frequency of a minimum in the unexplained sample variance defined by

$$\psi'(\omega_\psi) = 0$$

The frequency $\omega_\psi$ is also an unbiased estimator of $\omega_T$, as

$$\langle \omega_\psi - \omega_T \rangle \approx -\frac{\langle \psi'(\omega_T) \rangle}{\langle \psi''(\omega_T) \rangle} = 0. \tag{A.22}$$

The result (A.22) can be obtained using (3.25) and (A.13):

$$\langle \psi'(\omega_T) \rangle = \langle \xi'(\omega_T) \rangle = \sum_{k=0}^{K-1} \left\langle \left( \frac{d}{d\omega} |z_k(\omega)|^2 \right)_{\omega=\omega_T} \right\rangle$$

$$= i|\mu|^2 \sum_{k=0}^{K-1} \sum_{t=0}^{N-1} \sum_{t'=0}^{N-1} (t - t') v_k(t) v_k(t') \exp[-\alpha(t+t')] = 0. \tag{A.23}$$

Also,

$$\langle \psi''(\omega_T) \rangle = -|\mu|^2 tr(\mathbf{M}^{(2)} \cdot \mathbf{\Gamma}^{(1)}) > 0, \tag{A.24}$$

where the matrix $\mathbf{\Gamma}^{(j)}$ has components

$$\Gamma_{tt'}^{(j)} = \sum_{k=0}^{K-1} [v_k(t) v_k(t') \quad |V_k(0;\beta,\nu)|^2 \epsilon_t \epsilon_{t'}]^j, \qquad t, t' = 0, 1, \ldots, N \quad 1 \tag{A.25}$$

and *tr* denotes the trace operation on matrices. Define also the matrix $\Gamma^{[\mu,\nu]}$ with components

$$\Gamma_{tt'}^{[\mu,\nu]} = \Gamma_{tt'}^{(1)} \exp(-\mu t) \exp(\quad \nu t'); \qquad t, t' = 0, 1, \ldots, N-1. \tag{A.26}$$

Then the mean square error of estimator $\omega_\nu$ can be approximated as

$$\langle (\omega_\nu \quad \omega_T)^2 \rangle \approx \frac{\langle [\psi'(\omega_T)]^2 \rangle}{\langle [\psi''(\omega_T)]^2 \rangle}, \tag{A.27}$$

where

$$\langle [\psi'(\omega_T)]^2 \rangle = \sigma_N^4 \operatorname{tr}(\mathbf{M}^{(2)} \cdot \mathbf{\Gamma}^{(2)}) + 2\sigma_N^2 |\mu|^2 \sum_{t=0}^{N-1} [(\mathbf{M}^{(1)} \cdot \mathbf{\Gamma}^{[0,\alpha]})_{tt}]^2 \tag{A.28}$$

and

$$\langle [\psi''(\omega_T)]^2 \rangle = \sigma_N^4 \operatorname{tr}(\mathbf{M}^{(4)} \cdot \mathbf{\Gamma}^{(2)}) + 2\sigma_N^2 |\mu|^2 \sum_{t=0}^{N-1} [(\mathbf{M}^{(2)} \cdot \mathbf{\Gamma}^{[0,\alpha]})_{tt}]^2 + |\mu|^4 [\operatorname{tr}(\mathbf{M}^{(2)} \cdot \mathbf{\Gamma}^{[0,\alpha]})]^2. \tag{A.29}$$

For $N|\mu|^2 \sigma_N^2 \gg 1$:

$$\langle (\omega_\nu \quad \omega_T)^2 \rangle \approx \left( \frac{\sigma_N^2}{|\mu|^2} \right) \cdot \frac{2 \sum_{t=0}^{N-1} [(\mathbf{M}^{(1)} \cdot \mathbf{\Gamma}^{[0,\alpha]})_{tt}]^2}{[\operatorname{tr}(\mathbf{M}^{(2)} \mathbf{\Gamma}^{[0,\alpha]})]^2}. \tag{A.30}$$

As signal-to-noise ratio increases, $\langle (\omega_\nu \quad \omega_T)^2 \rangle$ becomes smaller. Graphs of $\langle \omega_\nu \quad \omega_T \rangle_{rms} = \sqrt{\langle (\omega_\nu \quad \omega_T)^2 \rangle}$ have the same shape as the plots of $\langle \omega_\theta \quad \omega_T \rangle_{rms}$ in Fig. A1, but $\langle \omega_\nu \quad \omega_T \rangle_{rms}$ is 10 25 per cent larger than $\langle \omega_\theta \quad \omega_T \rangle_{rms}$ for a given signal-to-noise ratio. For example, if $\beta = \nu = 0$ and $N|\mu|^2 \sigma_N^2 = 10$, $\langle \omega_\theta - \omega_T \rangle_{rms} = 0.136\,\omega_R$, and $\langle \omega_\nu \quad \omega_T \rangle_{rms} = 0.165\,\omega_R$.

### A2.3 FREQUENCIES ESTIMATED FROM PEAKS IN $F(\omega)$

The true frequency $\omega_T$ can also be estimated from the frequencies $\omega_F$ of peaks in the $F$-test curve defined by $F'(\omega_F) = 0$. As before, to first-order in $(\omega_F \quad \omega_T)$, assuming $(\omega_F \quad \omega_T)$

and $F''(\omega_T)$ are uncorrelated:

$$\langle \omega_F - \omega_T \rangle \approx \frac{\langle F''(\omega_T) \rangle}{\langle F''(\omega_T) \rangle}. \tag{A.31}$$

By (3.27),

$$\psi^2(\omega) F'(\omega) = (K - 1)[\theta'(\omega)\psi(\omega) - \theta(\omega)\psi'(\omega)]. \tag{A.32}$$

Assume that $\psi^2$ and $F'$, $\theta'$ and $\psi$, and $\theta$ and $\psi'$ are uncorrelated at $\omega_T$. Then, the expectation value of the right hand side of (A.32) vanishes at $\omega = \omega_T$. But,

$$\langle \psi^2(\omega_T) \rangle = \sigma_N^4 \left( \sum_{t=0}^{N-1} |v_k(t; \beta, \nu)|^2 \right)^2 \neq 0$$

so $\langle F'(\omega_T) \rangle = 0$. As $\langle F''(\omega_T) \rangle < 0$, $\omega_F$ is an unbiased estimator of $\omega_T$ by (A.31).

The frequency $\omega_F$ can be expressed in terms of $\omega_\theta$ and $\omega_\psi$. Expanding $\theta(\omega_F)$ and $\psi(\omega_F)$ in power series about their extrema to second-order in $(\omega_F - \omega_\theta)$ and $(\omega_F - \omega_\psi)$:

$$F(\omega_F) \approx \frac{(K-1)[\theta(\omega_\theta) + \frac{1}{2}(\omega_F - \omega_\theta)^2 \theta''(\omega_\theta)]}{\psi(\omega_\psi) + \frac{1}{2}(\omega_F - \omega_\psi)^2 \psi''(\omega_\psi)} \tag{A.33}$$

and

$$(\omega_F - \omega_\theta) \frac{\theta''(\omega_\theta)}{\psi''(\omega_\psi)} \approx \frac{(\omega_F - \omega_\psi) F(\omega_F)}{K - 1} \tag{A.34}$$

from (A.33). Substitute (A.33) in (A.34) and let

$$\omega_F = \tilde{\omega} \frac{(\omega_\psi - \omega_\theta)}{2} + \frac{(\omega_\psi + \omega_\theta)}{2} \tag{A.35}$$

so that $\omega_F = \omega_\psi$ when $\tilde{\omega} = 1$, and $\omega_F = \omega_\theta$ when $\tilde{\omega} = -1$. Then $\tilde{\omega}$ satisfies

$$\tilde{\omega}^2 - 2(a + b)\tilde{\omega} - (1 + 2b - 2a) = 0, \tag{A.36}$$

where

$$a = \frac{2\theta(\omega_\theta)}{(\omega_\psi - \omega_\theta)^2 \theta''(\omega_\theta)}$$

$$b = \frac{2\psi(\omega_\psi)}{(\omega_\psi - \omega_\theta)^2 \psi''(\omega_\psi)}.$$

The two solutions of the quadratic equation (A.36) are

$$\tilde{\omega}_\pm = (a + b) \pm \sqrt{(a + b)^2 + 2b + 1 - 2a}. \tag{A.37}$$

The solution $\tilde{\omega}_+$ is spurious because $|\tilde{\omega}_+| \to \infty$ as $a$ or $b \to -\infty$, and truncated Taylor series expansions in (A.33) and (A.34) are invalid for large values of $\tilde{\omega}$. The second solution $\tilde{\omega}_-$ is constrained so that $|\tilde{\omega}_-| < 1$, and corresponds to $\omega_F$ lying between $\omega_\theta$ and $\omega_\psi$.

As $\omega_F$ is a weighted average of $\omega_\theta$ and $\omega_\psi$, one might expect $\omega_F$ to be a more accurate estimator of the true frequency $\omega_T$. This hope is dampened when one realizes that the deviations of $\omega_\theta$ and $\omega_\psi$ from the true frequency $\omega_T$ are strongly positively correlated.

The correlation between $\omega_\theta$ and $\omega_\psi$ can be estimated as

$$\langle(\omega_\theta - \omega_T)(\omega_\psi - \omega_T)\rangle \approx \frac{\langle\psi'(\omega_T)\theta'(\omega_T)\rangle}{\langle\psi''(\psi_T)\theta''(\omega_T)\rangle}. \tag{A.38}$$

where

$$\langle\psi'(\omega_T)\theta'(\omega_T)\rangle = \rho\{\sigma_N^4 \operatorname{tr}(\tilde{\mathbf{M}}^{(2)}\cdot\boldsymbol{\Gamma}^{(1)}) + \sigma_N^2|\mu|^2\sum_{t=0}^{N-1}(\mathbf{M}^{(1)}\cdot\tilde{\mathbf{v}})_t(\mathbf{M}^{(1)}\cdot\boldsymbol{\Gamma}^{[0,\alpha]})_{tt}\tilde{v}(t)\exp(-\alpha t)\} \tag{A.39}$$

and

$$\langle\psi''(\omega_T)\theta''(\omega_T)\rangle = \rho\{\sigma_N^4 \operatorname{tr}(\tilde{\mathbf{M}}^{(3)}\cdot\tilde{\boldsymbol{\Gamma}}^{(1)}) + 2\sigma_N^2|\mu|^2\sum_{t=0}^{N-1}(\mathbf{M}^{(2)}\cdot\tilde{\mathbf{v}})_t(\mathbf{M}^{(2)}\cdot\boldsymbol{\Gamma}^{[0,\alpha]})_{tt}\tilde{v}(t)\exp(-\alpha t) + |\mu|^4(\tilde{\mathbf{v}}\cdot\mathbf{M}^{(2)}\cdot\tilde{\mathbf{v}})\operatorname{tr}(M^{(2)}\cdot\boldsymbol{\Gamma}^{[\alpha,\alpha]})\}, \tag{A.40}$$

where $\tilde{\mathbf{M}}^{(j)}$ has elements

$$\tilde{M}_{kl}^{(j)} = M_{kl}^{(j)}\tilde{v}(k)\exp(\alpha k)\tilde{v}(l)\exp(\alpha l), \qquad k, l = 0, 1, 2, \ldots, N-1$$

Using these equations one finds that the cross-correlation of $\omega_\theta$ and $\omega_\psi$ is almost unity. For example, if $\beta = \nu = 0$ and $N|\mu|^2/\sigma_N^2 = 10$, $\langle(\omega_\theta - \omega_T)(\omega_\psi - \omega_T)\rangle = 0.922$. Any averaging of the two estimators $\omega_\theta$ and $\omega_\psi$ will not result in an estimator which has significantly less error associated with it.

Using (A.34), one can see that for large values of $F$, $\omega_T \approx \omega_\psi$. Therefore, we estimate the errors in the frequencies of the $F$-test peaks using equation (A.27).

### A3 Detection sensitivity

It is useful to know the sensitivity of the $F$-test to the presence of a decaying sinusoid of frequency $\omega_T$. The signal-to-noise ratio required for detection of a sinusoid at a given confidence level can be calculated. Suppose that the time series is given by (A.1), with $\mu$ either purely real or purely imaginary. It can be shown that at frequency $\omega_T$ the random variable $F$ defined in (3.28) follows a noncentral $F$-distribution with noncentrality parameter

$$\gamma = \frac{2|\mu|^2}{\sigma_N^2}\sum_{k=0}^{K-1}|V_k(0;\beta,\nu)|^2 \tag{A.41}$$

(Kendall & Stuart 1979). The expected value of $F(\omega_T)$ is

$$\langle F(\omega_T)\rangle = \frac{(2+\gamma)(K-1)}{2(K-2)} \tag{A.42}$$

(Kendall & Stuart 1979, p. 279).

In Fig. A2, $\langle F(\omega_T)\rangle$ is plotted as a function of signal-to-noise ratio for sets of five tapers with parameters $\Omega N = 8\pi$, $\beta = 0.6$, 1.05 and $\nu = 0$, 0.01, 0.1, and 1. For a given signal-to-noise ratio, the expected value of the $F$-test grows with increasing $\nu$, reaching a maximum

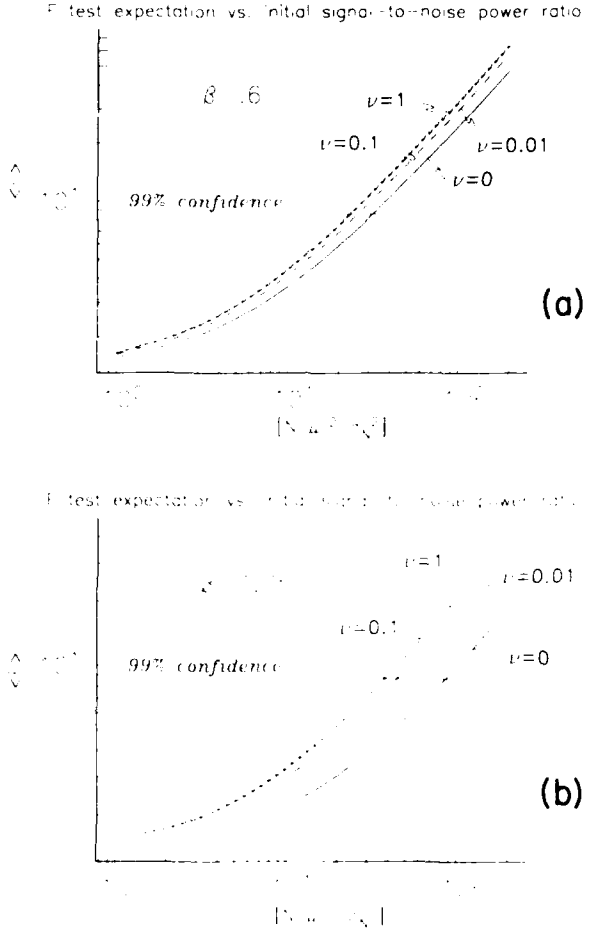



**Figure A2** (a, b). Expected value of the $F$-test at the true frequency $\omega_T$ as a function of $N|\mu|^2/\sigma_N^2$ for five tapers with $\Omega N = 8\pi$, $\beta = 0.6$ and $\beta = 1.05$, and $\nu = 0$, 0.01, 0.1 and 1. Larger values of $\nu$ are plotted with increasingly shorter dashes. The 99 per cent confidence level for an $F$-distributed random variable with 2 and 8 degrees of freedom ($F = 8.65$) is shown. For a given initial signal-to-noise ratio, $\langle F \rangle$ increases as $\nu$ increases. Therefore, it is easier to detect a decaying sinusoid using tapers designed with large values of the noise parameter $\nu$.

when $\nu = 1$. Suppose one wants to detect a decaying sinusoid at the 99 per cent confidence level. To do this using tapers which have $\beta = 0.6$ and $\nu = 0$ requires a 25 per cent higher signal-to-noise ratio than performing the analysis with tapers which have parameters $\beta = 0.6$ and $\nu = 0.1$. Using tapers with $\beta = 1.05$ and $\nu = 0$, a 125 per cent larger value of $|\mu|^2/\sigma_N^2$ is required than employing tapers designed with $\beta = 1.05$ and $\nu = 0.1$. There is a tradeoff between detection capability (Fig. A2) and frequency uncertainty (Fig. A1), but tapers designed with $0.01 \lesssim \nu \lesssim 1$ provide reasonable performance in both areas.

For comparison, consider the spectral estimate obtained by taking the discrete Fourier transform of a time series which has been multiplied by a cosine taper. A cosine taper

$w_c(t)$ is defined by

$$w_c(t) = A\left[1 - \cos\left(\frac{2\pi t}{N - 1}\right)\right]; \qquad t = 0, 1, \ldots, N - 1, \tag{A.43}$$

where $A$ is chosen so that

$$\sum_{t=0}^{N-1} |w_c(t)|^2 = 1.$$

A direct estimate of the spectrum of the data $x(t)$ is $|y_c(w)|^2$, where

$$y_c(\omega) = \sum_{t=0}^{N-1} \exp(-i\omega t) w_c(t) x(t). \tag{A.44}$$

The peak frequency $\omega_c$ defined by

$$\frac{d}{d\omega}|y_c(\omega)|^2\Big|_{\omega=\omega_c} = 0 \tag{A.45}$$

is taken as the estimator of the true frequency $\omega_T$ of a sinusoidal signal in the data. As before, $\omega_c$ is an unbiased estimator of $\omega_T$, and it has mean-square error

$$\langle(\omega_c - \omega_T)^2\rangle \approx \frac{\left\langle\left(\frac{d}{d\omega}|y_c(\omega)|^2\right)^2\right\rangle_{\omega=\omega_T}}{\left\langle\left(\frac{d^2}{d\omega^2}|y_c(\omega)|^2\right)^2\right\rangle_{\omega=\omega_T}} \tag{A.46}$$

Expressions for the expectations on the right-hand side of (A.46) are identical to (A.17) and (A.18) with $w_c(t)\exp(-\alpha t)$ replacing $v(t)$, and $\rho = 1$.

For data consisting only of Gaussian white noise, $2|y_c(\omega)|^2/\sigma_N^2$ is $\chi^2$ distributed with two degrees of freedom, and there is a probability of 0.01 that $2|y_c(\omega)|^2/\sigma_N^2$ will reach or exceed 9.21 (Abramowitz & Stegun 1965). If $|y_c(\omega)|^2$ exceeds the value $9.21\,\sigma_N^2/2$ at some frequency, then one is more than 99 per cent confident that a signal exists at that frequency. It is easy to show that, for the time series (A.1),

$$\langle|y_c(\omega_T)|^2\rangle = |\mu|^2\left(\sum_{t=0}^{N-1} w_c(t)\exp(-\alpha t)\right)^2 + \sigma_N^2 \tag{A.47}$$

so that

$$N|\mu|^2/\sigma_N^2 \approx 3.6N\left(\sum_{t=0}^{N-1} w_c(t)\exp(-\alpha t)\right)^{-2} \tag{A.48}$$

is the value of the initial signal-to-noise power ratio associated with 99 per cent detection confidence at frequency $\omega_T$.

Suppose one wants to detect a decaying sinusoid with decay parameter $\beta = 0.6$ (or decay rate $\alpha = 0.6\pi/T$) at the 99 per cent confidence level. Using the spectral estimate $|y_c(\omega)|^2$, a value of $N|\mu|^2/\sigma_N^2$ of approximately 32 is required, whereas using an $F$-test and five tapers for $\beta = 0.6$, $\nu = 0.1$, $N|\mu|^2/\sigma_N^2 \approx 23.5$ corresponds to detection at the 99 per cent confidence level. If the decaying sinusoid has a decay parameter of $\beta = 1.05$, an initial signal-

to-noise ratio of $N|\mu|^2/\sigma_N^2 \approx 104$ is needed for 99 per cent confidence level detection using the spectral estimate $|y_c(\omega)|^2$, but $N|\mu|^2/\sigma_N^2$ only needs to be 38 when the multitaper method is applied, using five tapers with $\beta = 1.05$ and $\nu = 0.1$. In this case, the multitaper approach is 274 per cent more efficient then the cosine-taper spectral method.

Clearly, the spectral estimate $|y_c(\omega)|^2$ is a less sensitive detector of decaying sinusoids in a time series than the multitaper method. Much of this discrepancy in detection ability is due to the eigentaper's preferential weighting of the start of the record where the signal-to-noise ratio is greater. Also, more information is extracted from a given time-series by applying several tapers; the extra degrees of freedom allow a better-constrained least-squares fit of the decaying sinusoid model to the data. Another advantage of the multiple-taper technique is that it allows one to discriminate between signals which are truly harmonic, and those which have time varying phases; conventional spectral estimates employing single tapers do not.

The variance of the random variable $F(\omega_T)$ can also be expressed in terms of the non-centrality parameter $\gamma$ defined in (A.41):

$$\text{var}\,[F(\omega_T)] = \frac{(K-1)^2\,[4(\gamma+1)(K-1)+\gamma^2]}{8(K-2)^2(K-3)} \tag{A.49}$$

when the data are given by (A.1), and $\mu$ is purely real or purely imaginary. The height of an $F$-test peak is not very well determined; $\sqrt{\text{var}\,[F(\omega_T)]} \geq \frac{1}{3}\langle F(\omega_T)\rangle$ when $K = 5$ for values of $F$ above the 90 per cent detection threshold.

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE *(and Subtitle)*<br>Multiple-Taper Spectral Analysis of Terrestrial Free Oscillations: Part I | | 5. TYPE OF REPORT & PERIOD COVERED<br>6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Jeffrey Park, Craig R. Lindberg, and David J. Thomson | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-85-C-0104 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Scripps Institution of Oceanography<br>La Jolla, California 92093 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research<br>Arlington, Virginia 22217 | | 12. REPORT DATE<br>April 1987<br>13. NUMBER OF PAGES<br>40 |
| 14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)* | | 15. SECURITY CLASS. *(of this report)*<br>unclassified<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

Reprint from Geophysical Journal of the Royal Astronomical Society of London, Vol. 91, 755-794, (1987)

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Summary. We present a new method for estimating the frequencies of the Earth's free oscillations. This method is an extension of the techniques of Thomson (1982) for finding the harmonic components of a time series. Optimal tapers for reducing the spectral leakage of decaying sinusoids immersed in white noise are derived. Multiplying the data by the best $K$ tapers creates $K$ time series. A decaying sinusoid model is fit to the $K$ time series by a least squares procedure. A statistical $F$-test is performed to test the fit of the decaying sinusoid model, and thus determine the probability that there are coherent oscillations in . . .

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102 LF 014 6601

unclassified
SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 92, NO. B12, PAGES 12,675-12,684, NOVEMBER 10, 1987

# Multitaper Spectral Analysis of High-Frequency Seismograms

JEFFREY PARK[1]

*Geophysical Fluid Dynamics Program, Princeton University, Princeton, New Jersey*

CRAIG R. LINDBERG AND FRANK L. VERNON III

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

Spectral estimation procedures which employ several prolate spheroidal sequences as tapers have been shown to yield better results than standard single-taper spectral analysis when used on a variety of engineering data. We apply the adaptive multitaper spectral estimation method of Thomson (1982) to a number of high-resolution digital seismic records and compare the results to those obtained using standard single-taper spectral estimates. Single-taper smoothed-spectrum estimates are plagued by a trade-off between the variance of the estimate and the bias caused by spectral leakage. Applying a taper to reduce bias discards data, increasing the variance of the estimate. Using a taper also unevenly samples the record. Throwing out data from the ends of the record can result in a spectral estimate which does not adequately represent the character of the spectrum of nonstationary processes like seismic waveforms. For example, a discrete Fourier transform of an untapered record (i.e., using a boxcar taper) produces a reasonable spectral estimate of the large-amplitude portion of the seismic source spectrum but cannot be trusted to provide a good estimate of the high-frequency roll-off. A discrete Fourier transform of the record multiplied by a more severe taper (like the Hann taper) which is resistant to spectral leakage leads to a reliable estimate of high-frequency spectral roll-off, but this estimate weights the analyzed data unequally. Therefore single-taper estimators which are less affected by leakage not only have increased variance but also can misrepresent the spectra of nonstationary data. The adaptive multitaper algorithm automatically adjusts between these extremes. We demonstrate its advantages using 16-bit seismic data recorded by instruments in the Anza Telemetered Seismic Network. We also present an analysis demonstrating the superiority of the multitaper algorithm in providing low-variance spectral estimates with good leakage resistance which do not overemphasize the central portion of the record.

## 1. INTRODUCTION

Spectral estimation is a powerful method of data analysis which is often used to study geophysical processes. The estimation of the spectra of background noise, line components, and transient signals is central to the analysis of electric, magnetic, and seismic time series. There have been many techniques developed which are effective for the analysis of long records of stationary processes. Unfortunately, these techniques are not universally applicable to seismic data sets. In many studies it is necessary to estimate a spectrum from a short time series. This situation can occur if some of the data are missing or if the data of interest (e.g., a seismic phase) are contained in a short segment of a longer record.

A new approach for estimating the spectra of short time series, known as multitaper spectral analysis, has been developed recently. We have applied this technique, which was first presented by *Thomson* [1982], to several dozen seismograms; in this paper we analyze two representative records. The spectra estimated using the multitaper technique are compared with several direct spectral estimates employing commonly used single tapers. We show that the multitaper approach can yield superior results when applied to high-frequency seismic data.

Spectral analysis of specific phases within a seismogram, particularly those at regional or local distances, can be difficult. It is often impossible to isolate a particular phase. If one isolates a major phase by discarding the rest of the record and then makes a direct estimate of the waveform's spectrum without first tapering the data (i.e., using a boxcar taper), the high-frequency roll-off of the estimated spectrum can be severely biased by spectral leakage. Therefore it is standard practice to multiply the time series by a taper before performing a discrete Fourier transform (DFT) to reduce spectral leakage (an extensive review of tapering is provided by *Harris* [1978]).

The cosine or Hann taper is popular in seismic analysis, being both effective and easy to calculate. The utility of the Hann taper is bought dearly, however. If one views each data point in a time series as a constraint on the estimated frequency content of the record, the Hann taper discards 5/8 of the statistical information in a given time series. This can be easily seen from the graph of the Hann taper in Figure 1. The data points at the extremes of the record are weighted weakly, while the center of the time series is emphasized. This unequal weighting causes the statistical variance of a direct spectral estimate using a

[1]Now at Department of Geology and Geophysics, Yale University, New Haven, Connecticut



Paper number 7B0660
0148-0227/87/007B-0660$05.00

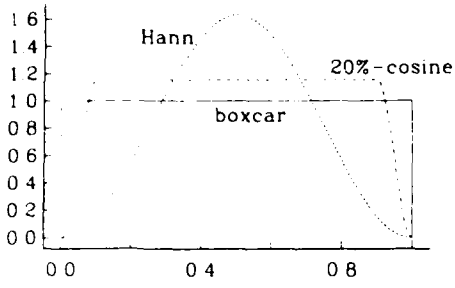



Fig. 1. Comparison plot of boxcar, Hann, and 20% cosine tapers.

Hann taper to be greater than the variance of a periodogram spectral estimate.

Blackman-Tukey tapers are designed to circumvent this loss of information somewhat by applying the cosine weighting to only the extremes of the record. For instance, the 20% cosine taper (Figure 1) discards only 12.5% of the available data variance constraints. However, *Harris* [1978] shows that this taper has less resistance to spectral leakage than a Hann taper. As long as only a single data taper is used, there will be a trade-off between the resistance to spectral leakage and the variance of a spectral estimate.

*Thomson* [1982] introduced the multitaper spectral analysis technique. First, the data are multiplied by not one, but several leakage-resistant tapers. This yields *several tapered time series from one record*. Taking the DFTs of each of these time series yields eigenspectra which are produced which are combined to form a single spectral estimate.

The tapers are constructed so that each taper samples the time series in a different manner while optimizing resistance to spectral leakage. The statistical information discarded by the first taper is partially recovered by the second taper, the information discarded by the first two tapers is partially retrieved by the third taper, and so on. Only a few low-order tapers are employed, as the higher-order tapers allow an unacceptable level of spectral leakage. One can use these tapers to produce an estimate that is not hampered by the trade-off between leakage and variance that plagues single-taper estimates, as we will demonstrate.

Single-taper spectral estimates have relatively large variance (increasing as a larger fraction of the data is discarded and the bias of the estimate is reduced) and are inconsistent estimates (i.e., the variance of the estimate does not drop as one increases the number of data). To counteract this, it is conventional to smooth the single-taper spectral estimate by applying a moving average to the estimate. This reduces the variance of the estimate but results in a short-range loss of frequency resolution and therefore an increase in the bias of the estimate.

The multitaper spectral estimates are formed as a weighted sum of the eigenspectra. Therefore the multitaper spectral estimate is already a smooth estimate; it has less variance than single-taper spectral estimates which have been designed to reduce bias, and it is also a consistent estimator. The comparison between the bias and variance properties of the single taper and multitaper estimates is discussed further in sections 3 and 4.

Another difficulty with seismic data is that the records are nonstationary; that is, the statistical character of the data changes with position in the record. Therefore a spectral estimator which weights the data in the center of the time series more heavily than data at the ends can overemphasize the signal energy in the middle of the record. This can result in a misrepresentation of the spectrum, as we demonstrate in section 3. The multitaper estimate, which discards very little data from the record and weights the data relatively evenly, is not subject to this problem.

Section 2 presents an outline of the basic algorithm. This outline contains sufficient detail to allow the reader to implement the algorithm but avoids derivations that can be found elsewhere. Section 3 describes the seismic data used in this study and presents comparisons of the spectra of seismic time series generated by both regional and local events. We demonstrate the trade-off between spectral leakage resistance and variance of the spectral estimates produced using the boxcar, 20% cosine, and cosine tapers. We compare the bias and variance of these conventional single-taper direct spectral estimates with the bias and variance of the multitaper spectral estimates in section 4. A numerical method for calculating the prolate eigentapers is given in the appendix.

## 2. The Multitaper Algorithm

The multitaper method is based on a family of tapers which are resistant to spectral leakage. We outline the *multitaper method here, and note that more detailed treatments* can be found in the works by *Thomson* [1982] and *Lindberg* [1986].

Suppose that we are given the finite time series $x(t)$, $t = 0, 1, \ldots, N-1$, which is a set of discrete samples of a continuous time process $x(t)$, $0 \le t \le N-1$ (we assume a unit sample interval $\Delta = 1$, without loss of generality). If $x$ has no harmonic components, then it has the Cramér spectral representation [*Doob*, 1953]

$$x = \int e^{i2\pi ft}\, X(f)\, df$$

We wish to estimate the amplitude spectrum $S(f) = E\{|X(f)|^2\}$ (where $E\{\cdot\}$ denotes expected value) of the continuous time process $x(t)$, $0 \le t \le N-1$, from the time series $x(t)$, $t = 0, 1, \ldots, N-1$. A conventional direct spectral estimate $\hat{S}(f)$ of $S(f)$ is found by multiplying the data $x(t)$, $t = 0, 1, \ldots, N-1$, by a sequence $a(t)$, $t = 0, 1, \ldots, N-1$, called a taper, applying a DFT,

$$Y(f) = \sum_{t=0}^{N-1} a_t x_t e^{-i2\pi ft}$$

and finally taking the squared modulus of the resulting function $Y(f)$. Although $t$ is discrete, $f$ is continuous, with $|f| \le \frac{1}{2}$ ($\frac{1}{2}$ is the Nyquist frequency $f_{\text{Nyquist}} = \frac{1}{2\Delta}$). We normalize the taper so that

$$\sum_{t=0}^{N-1} a_t^2 = 1$$

The spectral leakage properties of the data taper $a_t$, $t = 0, 1, 2, \ldots, N-1$ can be deduced from its DFT:

$$A(f) = \sum_{t=0}^{N-1} a_t e^{-i2\pi ft} \qquad (1)$$

For conventional tapers the function $|A|$ has a broad main lobe and a succession of smaller side lobes. For example, for the boxcar taper, $a_t = 1/\sqrt{N}$ and

$$A(f) = \frac{1}{\sqrt{N}} \frac{1-e^{i2\pi fN}}{1-e^{i2\pi f}} = \frac{e^{i\pi(N-1)f}}{\sqrt{N}} \left| \frac{\sin N\pi f}{\sin \pi f} \right|$$

In this case the function $|A(f)|$ is readily observed to have a central lobe flanked by smaller side peaks. (The phase factor $e^{i\pi(N-1)f}$ results from choosing the time origin $t = 0$ to coincide with the first data point. It does not affect the leakage resistance of the taper.) The larger the side lobes, the more spectral leakage is encountered using this taper, biasing the estimate $\hat{X}_a$ away from its desired value. This can be seen by observing that

$$\hat{X}_a(f) = \int A(f-f')X(f')df' \tag{2}$$

which follows from substituting the Cramér spectral representation of $x_t$ in the definition of $\hat{X}_a$, and therefore

$$|\hat{X}_a(f)|^2 = \int |A(f-f')|^2 S(f')df'$$

A good data taper should have a spectral window $A(f-f')$ whose amplitude is large in the central lobe region where $|f-f'|$ is small and has low side lobes at more distant frequencies. This reduces the bias in the estimate by preventing the energy in $X$ at distant frequencies from leaking over to affect the estimate $|\hat{X}_a|^2$ at frequency $f$.

Suppose we wish to minimize the bias at frequency $f$ due to spectral leakage from outside the frequency band $|f'-f| \leq W$, where $2W$ is some chosen bandwidth. We maximize the fraction of energy of $A$ within the chosen band:

$$\lambda(N,W) = \frac{\int_{-W}^{W} |A(f)|^2 df}{\int |A(f)|^2 df} \tag{3}$$

Since no finite time series can be completely band-limited, $\lambda(N,W) < 1$ for finite $N$ and nontrivial $W$. The functional $\lambda$ can be interpreted as follows: in a single-taper direct estimate of the spectrum of a white noise process at frequency $f$, $\lambda$ is the fraction of spectral energy in that estimate that derives from the frequency interval $|f-f'| \leq W$; $1-\lambda$ is the fraction of spectral energy that leaks in from outside that band.

It is convenient from a computational viewpoint to substitute (1) into (3) to express $\lambda$ in terms of the data tapers themselves rather than their transforms. If we seek a taper for an $N$ point time series, the sequence $\{a_t\}_{t=0}^{N-1}$ can be represented as an $N$ vector $\mathbf{a}$. This notation allows us to express (3) in matrix form (following the derivation of equation (2.5) of *Park et al.* [1987], letting the decay rate

$$\lambda(N,W) = \frac{\mathbf{a} \cdot \mathbf{C} \cdot \mathbf{a}}{\mathbf{a} \cdot \mathbf{a}} \tag{4}$$

$\mathbf{C}$ has components

$$C_{t,t'} = \frac{\sin 2\pi W(t-t')}{\pi(t-t')} \qquad t,t' = 0,1,2,\ldots,N-1$$

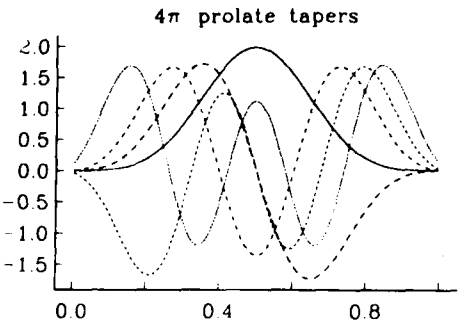



Fig. 2. The five lowest-order $4\pi$ prolate eigentapers. The zeroth-order eigentaper $\mathbf{v}^{(0)}$ is plotted with a solid line, and the higher-order tapers are plotted with dashed lines.

We seek those values of $\mathbf{a}$ for which the functional $\lambda$ is stationary. This leads to the matrix eigenvalue problem

$$\mathbf{C} \cdot \mathbf{a} - \lambda(N,W)\mathbf{a} = 0 \tag{5}$$

which has as its solutions the ordered eigenvalues $1 > \lambda_0 > \lambda_1 > \lambda_2 > \ldots > \lambda_{N-1} > 0$ and associated eigenvectors $\mathbf{v}^{(k)}(N,W)$; $k = 0,1,2,\ldots,N-1$ (which have components $v_t^{(k)}$; $t = 0,1,2,\ldots,N-1$). The $\mathbf{v}^{(k)}(N,W)$ are discrete prolate spheroidal sequences [*Slepian*, 1978], which we also refer to as prolate eigentapers. We will suppress the explicit dependence of $\mathbf{v}^{(k)}$ on $N$ and $W$ in the following. A prolate eigentaper with a time-bandwidth product of $P = NW$ is called a $P\pi$ prolate taper; it concentrates spectral energy in frequency bands of width $2W = 2P/N$. As the Rayleigh frequency $1/N$ is the fast Fourier transform (FFT) frequency bin spacing, a $P\pi$ prolate taper will have a main lobe which is $2P$ "frequency bins" wide. For instance, tapers for which $NW = 4$ minimize the spectral leakage at frequency $f$ from outside the frequency band defined by $|f'-f| \leq 4/N$. For large $N$ ($\gtrsim 100$) one can construct a set of the $\mathbf{v}^{(k)}$ for any value of the time-bandwidth product $NW$. As noted in the appendix, this allows the user to calculate one set of eigentapers $\mathbf{v}^{(k)}$ for a fixed value of $N$ and to interpolate this set to construct tapers for time series of various lengths. We have restricted the following discussion to $4\pi$ prolate tapers, but similar behavior is found for other choices of the time-bandwidth product.

The five lowest-order eigentapers $\mathbf{v}^{(k)}$, $k = 0,1,2,3,4$ shown in Figure 2 have been made for a time series of length $N = 128$ and time-bandwidth product $NW = 4$. The lowest-order taper ($k = 0$) is the familiar $4\pi$ prolate taper advocated by *Thomson* [1971, 1977*a*,*b*] and *Eberhard* [1973] and has a shape similar to conventional tapers such as the cosine taper (Figure 1). The higher-order eigentapers are markedly different from ordinary data tapers. For even values of $k$, the $\mathbf{v}^{(k)}$ are symmetric about the midpoint of the time series. For odd values of $k$, the $\mathbf{v}^{(k)}$ are antisymmetric about the midpoint. All the tapers, except the lowest-order one, have regions of positive and negative data weighting. We normalize the tapers so that

$$\sum_{t=0}^{N-1} (v_t^{(k)})^2 = 1$$

As the eigentapers $\mathbf{v}^{(k)}$ are solutions to (5), they are orthogonal:

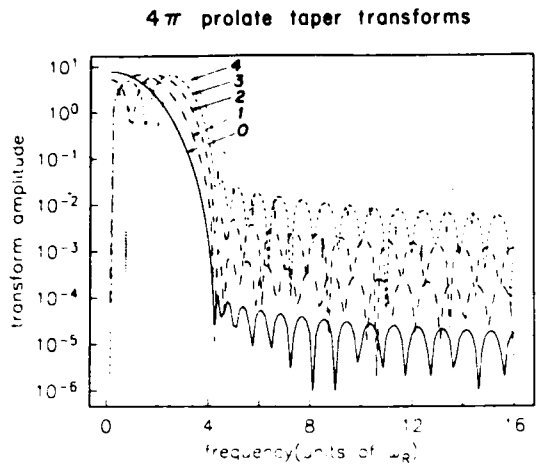



Fig. 3. Fourier transform amplitudes of the five $4\pi$ prolate tapers shown in Figure 2, using the same conventions for dashed and solid lines.

$$\mathbf{v}^{(k)} \cdot \mathbf{v}^{(k')} = \delta_{kk'} \tag{6}$$

(This can be clearly seen in Figure 2.) This relation shows that each $\mathbf{v}^{(k)}$ can be used to provide an orthogonal sample of the data $\{x_t\}_{t=0}^{N-1}$.

Taking discrete Fourier transforms of the prolate eigentapers produces the spectral windows

$$U_k(N,W;f) = \epsilon_k \sum_{t=0}^{N-1} v_t^{(k)}(N,W) e^{-2\pi i f(t-(N-1)/2)} \tag{7}$$

where we have used the time-centered transform to eliminate spurious phase factors in the definition. The function $\epsilon_k = 1$ if $k$ is even; $\epsilon_k = i$ if $k$ is odd. The use of $\epsilon_k$ is a notational convention so that $U_k$ is real-valued. Plots of the $U_k$ for $N = 128$ and $NW = 4$ appear in Figure 3 for $k = 0, 1, \ldots, 4$. Most of the energy of the $U_k$ is concentrated within the specified frequency band as was required by maximizing (3). The spectral windows corresponding to the lowest-order eigentapers have impressively small side lobes, but spectral leakage resistance becomes progressively poorer as the order of the taper increases. The lowest-order $2NW$ eigentapers (e.g., the eight lowest-order $4\pi$ prolate tapers) have eigenvalues $\lambda_k$ close enough to unity that they are useful for minimizing spectral leakage. The eigenvalues $\lambda_k$ of the eight lowest-order eigentapers with time-bandwidth products 4, 3, and 2 are given in Table 1 for $N = 128$. For reference, the value of the functional (3) is given for a boxcar taper which concentrates spectral energy within frequency bands of the same width.

To construct a multitaper spectral estimate, one first calculates the complex "eigencoefficients" $y_k(f)$ by taking a DFT of the product of the data with each $\{v_t^{(k)}\}_{t=0}^{N-1}$

$$y_k(f) = \sum_{t=0}^{N-1} v_t^{(k)} x_t e^{-2\pi i f t} \tag{8}$$

An estimate of the spectrum can be made from weighted sums of the eigenspectra $|y_k|^2$. *Thomson* [1982] formulates the problem of estimating the spectrum of a record as an integral equation. The solution of the integral equation is averaged over $(f-W, f+W)$ to produce the smoothed high-resolution spectral estimate:

$$\bar{S}(f) = K^{-1} \sum_{k=0}^{K-1} (\lambda_k)^{-1} |y_k(f)|^2 \tag{9}$$

where $K$ is the number of tapers used. If $K$ is not large, the smoothed high-resolution estimate (9) differs little from an arithmetic average of the eigenspectra as $\lambda_k \approx 1$ for the lowest-order eigentapers.

Although straightforward, (9) is not the best multitaper spectral estimate to use. An adaptive spectral estimate

$$\hat{S}(f) = \frac{\sum_{k=0}^{K-1} |d_k(f) y_k(f)|^2}{\sum_{k=0}^{K-1} |d_k(f)|^2} \tag{10}$$

can be devised which has frequency-dependent weights $d_k(f)$ chosen to reduce bias from spectral leakage [*Thomson*, 1982]. This technique proves extremely useful in the analysis of highly-colored spectral processes. At frequencies $f$ where the spectrum is reasonably flat, the weights $d_k(f) \approx 1$, reducing the variance of the spectral estimates. At frequencies $f$ where spectrum has a steep slope, the contribution from the higher-order eigentapers, which have poorer leakage resistance, is reduced. The trade-off between spectral leakage and variance of the spectral estimate is balanced at each frequency.

The optimal weights $d_k$ can be found by minimizing the misfit of the estimated spectrum to the true spectrum $S(f)$. This misfit, although unknown, can be estimated statistically. The resulting equation for the weight $d_k(f)$ is

$$d_k(f) = \frac{\sqrt{\lambda_k}\, S(f)}{\lambda_k S(f) + E\{B_k(f)\}} \tag{11}$$

where $S(f)$ is the true value of the spectrum at frequency $f$ and $B_k(f)$ is the spectral energy at frequency $f$ that leaks in from outside the frequency band $(f-W, f+W)$. We replace the unknown value $S(f)$ by its estimate $\hat{S}(f)$. *Thomson* [1982] found it adequate to approximate $E\{B_k(f)\} \cong \sigma^2(1-\lambda_k)$, i.e., as a constant fraction of the total variance of the time series:

$$\sigma^2 = \sum_{t=0}^{N-1} x_t^2 \tag{12}$$

We find the estimate $\hat{S}(f)$ by iteration. We take the arithmetic average of $|y_0(f)|^2$ and $|y_1(f)|^2$ as an initial estimate of $S(f)$, then substitute this value into (11) to

TABLE 1. Fractional Leakage of Eigentapers

| | $P\pi$ Prolate | | |
|---|---|---|---|
| | $P=4$ | $P=3$ | $P=2$ |
| $\lambda_0$ | 0.9999999998 | 0.999999885 | 0.999948125 |
| $\lambda_1$ | 0.999999978 | 0.999992014 | 0.997764652 |
| $\lambda_2$ | 0.999999008 | 0.999750480 | 0.962155175 |
| $\lambda_3$ | 0.999972984 | 0.995477689 | 0.733922358 |
| $\lambda_4$ | 0.999500363 | 0.951033908 | 0.287339619 |
| $\lambda_5$ | 0.993525891 | 0.725208760 | * |
| $\lambda_6$ | 0.943750573 | 0.307789684 | * |
| $\lambda_7$ | 0.721233936 | 0.060764834 | * |
| Boxcar | 0.974748450 | 0.966410435 | 0.949939339 |

* $\lambda_k < .05$

produce first guesses of the weights $d_k(f)$. These weights are then used in (10) to generate a new spectral estimate $\hat{S}(f)$, and the process is repeated. Convergence is usually satisfactory within a few cycles.

Careful examination of the adaptive spectral estimate shows that Parseval's theorem is not explicitly satisfied, i.e., there is no requirement that the energy of the spectrum estimate, integrated over frequency, equal the total variance of the time series. This arises from the way that the multitaper algorithm attempts to compensate for the effects of spectral leakage. If the expected broadband bias $E\{B_k(f)\}$ were to vanish, then (11) would become $d_k(f) = \lambda_k$, and the adaptive estimate (10) would reduce to the smoothed high-resolution estimate (9) (except for a small multiplicative factor due to the departure of the eigenvalues $\lambda_k$ from unity). This would occur if the true spectrum were zero outside the frequency band $|f - f'| \leq W$. As $(1 - \lambda_k)$ of the process variance within the frequency interval $|f - f'| \leq W$ is leaked outside the band, the limiting case $d_k(f) = \lambda_k$ represents an attempt by the estimator to compensate for this spectral energy lost to leakage by boosting the coefficients of the higher-order eigenspectra in the weighted sum. When the spectrum has a steep slope, the higher-order eigenspectra are downweighted and the adaptive spectral estimate tends toward the least biased eigenspectral estimate $|y_0(f)|^2$.

*Thomson* [1982] analyzed two synthetic time series using multitaper methods. Both series had fewer than 100 data points and a numerical precision of roughly 20 bits. In the first example, it was demonstrated that a multitaper approach could accurately estimate a spectrum with a dynamic range of more than seven decades and accurately infer the existence of harmonic lines (i.e., coherent sinusoids) in the data. Thomson also analyzed a 64-point time series used by *Kay and Marple* [1981] in a spectrum analysis "shootout" comparing 11 spectral estimation techniques, including the maximum entropy method as well as a single-taper direct spectral estimate and several other popular spectral estimates. Unlike any of the techniques tested by *Kay and Marple* [1981], a multitaper technique was able to produce a spectral estimate which was similar to the true spectrum of the synthetic time series.

## 3. Spectral Comparisons Using Seismic Data

We compare a number of single-taper direct spectral estimates with the adaptive multitaper spectral estimate techniques on wide dynamic range, high-resolution seismic data. The advent of digital arrays with 16-bit data loggers and the proposed 22- or 24-bit precision instruments demand an improved sophistication in data analysis techniques. We may soon have seismic data which are recorded to the same precision as the synthetic examples of *Thomson* [1982].

The data used in this paper were recorded on seismometers in the Anza Seismic Telemetered Array. The Anza array was designed to record high-frequency seismic signals from local earthquakes. The instruments in this array measure surface velocity, and the data are recorded as 16-bit numbers (this allows a dynamic range of 96 dB). See *Berger et al.* [1984] for a more detailed description of the Anza array.

The multitaper spectral estimate has a smaller variance at each frequency than a single-taper direct spectral estimate. To make a fair comparison between the various direct spectral estimates and the adaptive multitaper method, we will smooth each single-taper estimate using a moving average so that each estimate averages information over roughly the same frequency band as a multitaper estimate using seven $4\pi$ prolate eigentapers.

The effect of smoothing single-taper direct spectral estimates in this way is shown in Figure 4. The section of the seismogram which is analyzed is shown at the top of Figure 4. The unsmoothed spectral estimates are shown below on the left, and the smoothed estimates are displayed on the right below the record. The upper traces are direct estimates using the Hann taper, the middle traces are spectral estimates made with a 20% cosine, and the lower traces are spectrum estimates which employ a boxcar taper. The amplitude is plotted on a logarithmic scale on the vertical axis, and frequency is plotted on a linear scale on the horizontal axis. Each trace is offset by a multiplicative factor of 50 from the adjacent traces. Notice that if one studies the unsmoothed spectral estimates, it is difficult to distinguish any specific features common to each of the estimates except for a general linear trend. In comparison, the smoothed spectral estimates have many of the same features. Each major peak or trough appears at the same frequency in each of the smoothed estimates.

Unfortunately, since we are using real data, it is impossible to know the true spectrum for any of the examples. However, the work of *Thomson* [1982] demonstrates that the multitaper method provides a reasonable spectral estimate. This is confirmed by a study comparing the multitaper estimate with the smoothed direct estimates on a synthetic seismic wave train with a known spectrum (C. Lindberg et al., unpublished manuscript, 1987).

To study how tapering affects the spectra of body wave pulses, we isolate a phase in the middle of a seismogram, produce spectral estimates using each of the four methods, and compare the results. The upper graph in Figure 5 shows the transverse horizontal seismogram of an earthquake which had an epicentral distance of 100 km from the recording station PFO (in Pinyon Flat, California). We extract that section of the seismogram corrresponding to the shear wave arrival and estimate its spectrum by each method. The spectral estimates are plotted on a linear-linear scale in the lower portion of Figure 5 and for clarity are plotted in dimensionless velocity units on the vertical axis. Each of the four spectral estimates have two main peaks in the frequency band from 0 to 20 Hz, near 4 and 14 Hz.

These estimates are interesting to compare. Three of the estimated spectra (those plotted using solid and dashed lines) have almost identical features (except for the offset between them). In these spectral estimates the amplitude of the peak at 14 Hz is about 20% less than the amplitude of the peak at 4 Hz. The other estimated spectrum (curve d, plotted with asterisks) does not resemble the other three estimates closely. The peak at 14 Hz is 10% higher than any other peak in this estimate. This change in the relative amplitude of the two spectral peaks would influence the choice of a corner frequency if these spectra were converted from velocity to displacement or acceleration.

The three spectral estimates which exhibit similar characteristics are the multitaper estimate (curve a, plotted

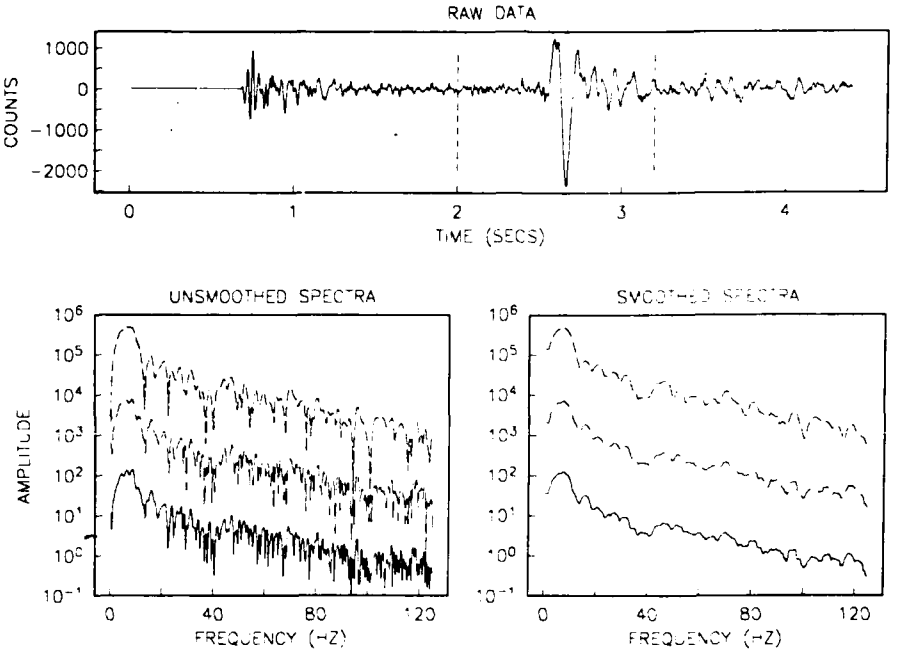



Fig. 4. (Top) Comparison of unsmoothed and smoothed estimates of the spectrum of a high-frequency $S$ wave. The spectra are plotted on a log-linear scale and are offset to facilitate comparison. The boxcar spectral estimates are graphed with a solid line. The dashed lines at the top of each of the lower figures are spectral estimates employing a Hann taper. The middle curves are spectral estimates obtained using a 20% cosine taper.

as a solid line), the 20% cosine direct estimate (curve b, the upper dashed line), and the boxcar direct estimate (curve c, the lower dashed line). The spectrum showing a different distribution of spectral energy was estimated using the Hann taper (curve d). The Hann direct spectral estimate is unlike the other three estimates because it imposes a different emphasis on the time series. Referring back to Figure 1, it is easy to see that the boxcar applies equal weighting to the entire time series and the 20% cosine taper weights 80% of the series equally. Not surprisingly, using either of these two tapers produces essentially the same result. However, the multitaper spectral estimate also gives essentially equal importance to every data point, like the boxcar and 20% cosine estimates (see Figure 2). The Hann taper puts over 80% of its emphasis on the middle 50% of the time series and gives the data in the first and last 25% of the series less weight. This rejection of data near the ends of the series causes the apparent misrepresentation of the distribution of spectra energy shown in Figure 5.

We also compared estimates of the spectrum of a vertical recording of a nuclear explosion. This event had an epicentral range of 412 km and also was recorded at PFO. The section of data which was analyzed is bounded by the vertical dashed lines in the upper trace in Figure 6. The analysis procedures were identical to those used in the previous example except that the log amplitudes of the spectra were plotted on the vertical axis.

The spectrum of the nuclear test has a large dynamic range and has most of its energy concentrated below 20 Hz. By examining the estimated spectra, one can see that some estimates are more effected by spectral leakage than others. The two estimates which are less subject to spectral leakage, the Hann direct estimate (curve d, plotted with asterisks) and the adaptive multitaper estimate (curve a, the solid line), are very similar. Both of these estimated spectra clearly show the spectrum of the signal from 0 to 20 Hz; from 20 to 60 Hz the spectrum of the ground noise is visible. The antialias filters of the recording system are 6 pole Butterworth filters which have a corner frequency of 62.5 Hz. The effect of the filters is visible in the 60–80 Hz band. In the band from 80 to 125 Hz the ground noise is less than the instrument noise. The variance of the adaptive multitaper spectrum is larger in the low-amplitude portion of the spectrum and hence appears unsmoothed. This is because the downweighting of the higher-order eigenspectra minimizes spectral leakage at the cost of reducing the effective number of degrees of freedom of the estimate at each frequency. If smaller variance is desired in the low-amplitude portion of the adaptive multitaper spectrum, then prolate tapers with a larger time-bandwidth product could be used.

The spectra shown in Figure 6 which were obtained using the 20% cosine and boxcar tapers suffer from the effects of spectral leakage. The spectrum estimate employing the 20% cosine (curve c, the lower dashed line) suffers less from spectral leakage than the estimate utilizing the boxcar, as expected. The leakage of spectrum estimated using a 20% cosine taper hides nearly all the features in the ground noise between 20 and 60 Hz. The effect of the antialias filters is completely obscured. The apparent energy in the 20% cosine spectrum estimate is larger than the instrument noise in the 80–125 Hz band by a factor of 10. The performance of the spectral estimate obtained

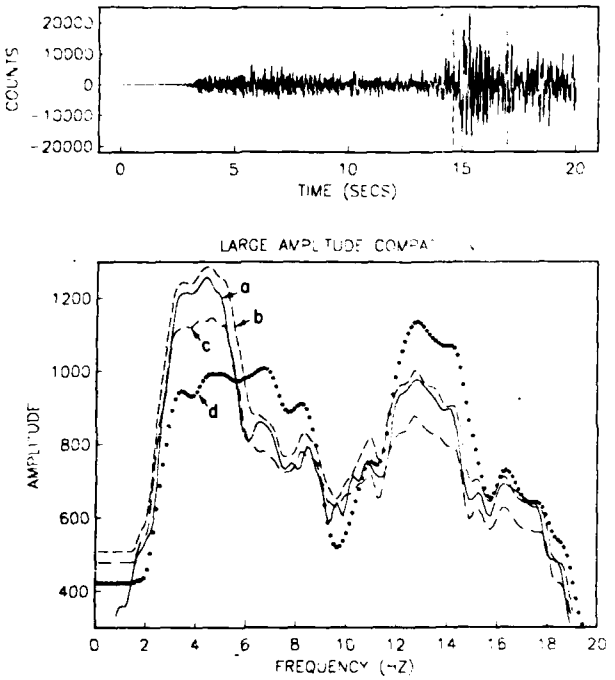



Fig. 5. A multitaper spectral estimate (solid line, labeled a) of the frequency content of an *SH* wave (top) is compared with direct spectral estimates using the boxcar taper (fine dashed line, labeled b), 20% cosine taper (coarse dashed line, labeled c), and Hann taper (asterisks, labeled d). The spectra are plotted using linear scales for the horizontal and vertical axes. The boxcar, 20% cosine, and multitaper estimates of the *S* wave spectrum are almost identical, but the Hann taper estimate is substantially different. This is because the Hann-tapered spectra overemphasize the data in the center of the time series and downweight data points toward the ends of the record. The section of the time series which was analyzed is bracketed by dashed lines in the seismogram at the top.

using a boxcar taper (curve b, the upper dashed line) is even worse, since it does not exhibit any of the features of the true spectrum between 20 and 125 Hz.

These examples show that each of the spectral estimates has different advantages. The smoothed spectrum estimate employing a boxcar taper produces a good estimate of the large-amplitude portions of the spectrum but has very poor spectral leakage properties and is not of much use for spectra which have a large dynamic range. The smoothed spectrum estimate using a Hann taper is less affected by spectral leakage, but this estimate can misrepresent the large-amplitude portion of the spectrum. A smoothed spectral estimate incorporating the 20% cosine taper combines the best properties of the spectral estimates which use the boxcar and the Hann tapers. It retrieves the large-amplitude features almost as well as the boxcar estimate and has spectral leakage properties which are sufficient for many geophysical applications. The adaptive multitaper estimate has even better performance, representing the large-amplitude spectral components as accurately as the boxcar estimate and having excellent spectral leakage properties.

We have also made multitaper estimates of the spectra of more than a dozen events recorded at local and regional distances by the Anza array. Multitaper techniques like the ones presented here and by *Park et al.* [this issue] appear to be useful tools for seismic data analysis.

## 4. Statistical Comparisons

We compare the broadband bias and variance of the smoothed single-taper direct spectral estimates with the smoothed high-resolution and adaptive multitaper estimates. We consider smoothed single-taper and multitaper spectral estimates whose values at some frequency $f$ are formed by averaging seven direct spectral estimates which concentrate the spectral energy at frequency $f$ mainly within the frequency band $(f-W, f+W)$, where $W = 4/N$ (4 times the Rayleigh frequency $1/N$). Therefore we use $4\pi$ prolate eigentapers for the multitaper estimates; the seven lowest-order $4\pi$ eigentapers have good resistance to spectral leakage (see Table 1), but we do not use the seventh-order $4\pi$ prolate eigentaper; it allows excessive spectral leakage, as $\lambda_7 = 0.721233936$. We compare the multitaper estimates with a smoothed single-taper estimate $\bar{S}_a(f)$ which is formed by averaging the seven direct spectral estimates $|\hat{X}_a(f+j/N)|^2$, $j = -3, -2, \ldots, 2, 3$ obtained using a taper $\{a(t)\}_0^N$, i.e.,

$$\bar{S}_a(f) = 1/7 \sum_{j=-3}^{3} |\hat{X}_a(f+j/N)|^2 \qquad (13)$$

This estimate is mostly an average of spectral energy from the band $(f-4/N, f+4/N)$. (The main lobes of tapers

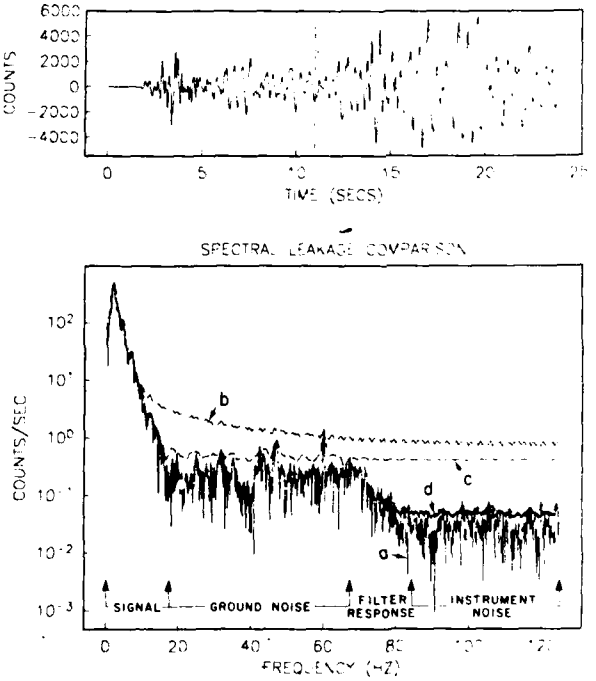



Fig. 6. Comparison of the leakage of various estimates of the spectrum of a vertical seismogram recorded 412 km away from a Nevada Test Site explosion. The spectral estimate using a cosine taper (asterisks, labeled d) and the multitaper spectral estimate (solid line, labeled a) give good representations of the spectra of the seismic signal (0–20 Hz) and ground noise (20–60 Hz). The spectra are plotted using a log-linear scale.

TABLE 2. Statistical Comparison

| Estimate | $7/\sigma^4$ Variance | Fractional leakage $(1-\lambda)$ |
|---|---|---|
| Smoothed boxcar | 1.0000 | 0.0367 |
| Smoothed 20% cosine | 1.0814 | 0.0192 |
| Smoothed Hann | 1.8142 | 0.0093 |
| Smoothed high resolution (seven $4\pi$ prolate eigentapers) | 1.0196 | 0.0094 |
| Adapative multitaper (seven $4\pi$ prolate eigentapers) | 1.0004 | 0.0094 |

other than the boxcar taper are wider than $2/N$, so this is not strictly correct, but we will make this approximation.)

### 4.1. *Variance*

To gain some idea of how smooth the estimators are, we compare the variances of each spectral estimate for a time series composed of Gaussian white noise. For single-taper estimates we define the covariance matrix

$$\Lambda_{ij} = \frac{1}{\sigma^2} E\{\hat{X}_a^*(f + i/N)\hat{X}_a(f + j/N)\} = \int A^*(f + i/N)A(f + j/N)\,df \tag{14}$$

for $i, j = -3, -2, \ldots, 2, 3$ [see *Thomson*, 1982, equation 4.1], where $\sigma^2$ is the process variance defined in (12) and $A(f)$ is the spectral window introduced in (1). If the single-taper estimates $\hat{X}_a(f + i/N)$ and $\hat{X}_a(f + j/N)$ are uncorrelated for $i \neq j$, then $\Lambda$ is a diagonal matrix. If the amount of correlation of the estimates $\hat{X}_a(f + i/N)$ is such that one or more of the $\hat{X}_a(f + i/N)$ can be expressed as a linear combination of the others, then $\Lambda$ is a singular matrix, with at least one zero eigenvalue. In practice, $\Lambda$ has a behavior which lies somewhere between these extremes.

For white noise data the expected value of $\bar{S}_a(f)$ is

$$E\{\bar{S}_a(f)\} = \frac{1}{7}\sigma^2 \sum_{i=-3}^{3} \Lambda_{ii}$$

but as

$$\Lambda_{ii} = \sum_{t=0}^{N-1} a^2(t) = 1$$

for all $i$,

$$\sum_{i=-3}^{3} \Lambda_{ii} = 7$$

and $E\{\bar{S}_a(f)\} = \sigma^2$. The variance of $\bar{S}_a(f)$ is

$$E\{(\bar{S}_a(f))^2\} - (E\{\bar{S}_a(f)\})^2 \tag{15}$$

The first term of (15) is

$$E\{(\bar{S}_a(f))^2\} = \frac{1}{49} \sum_{i=-3}^{3} \sum_{j=-3}^{3} E\{|\hat{X}_a(f + \frac{i}{N})|^2 |\hat{X}_a(f + \frac{j}{N})|^2\}$$

As each function $|\hat{X}_a(f + i/N)|^2$ is the sum of squares of two Gaussian random variables, we can show that

$$E\{(\bar{S}_a(f))^2\} = \frac{\sigma^4}{49} \sum_{i=-3}^{3} \sum_{j=-3}^{3} [\Lambda_{ii}\Lambda_{jj} + |\Lambda_{ij}|^2]$$

[*Papoulis*, 1977, chapter 11], so

$$\mathrm{Var}\{\bar{S}_a(f)\} = \frac{\sigma^4}{49} \sum_{i=-3}^{3} \sum_{j=-3}^{3} |\Lambda_{ij}|^2$$

For the smoothed periodogram estimate (i.e., direct estimate using a boxcar taper), $\Lambda_{ij} = \delta_{ij}$, and $\mathrm{Var}\{\bar{S}_a(f)\} = (\sigma^4/7)$. Values of $(7/\sigma^4)\,\mathrm{Var}\{\bar{S}_a(f)\}$ for the smoothed periodogram, 20% cosine taper and Hann taper direct spectral estimates are tabulated in Table 2. Notice that as more data are discarded by the taper, the variance of the spectral estimate increases.

For the smoothed high-resolution multitaper spectral estimate (9),

$$E\{\bar{S}(f)\} = \frac{1}{K}\sum_{k=0}^{K-1} (\lambda_k)^{-1} E\{|y_k(f)|^2\} = \frac{\sigma^2}{K}\sum_{k=0}^{K-1} (\lambda_k)^{-1}$$

When the $K = 7$ lowest-order $4\pi$ prolate eigentapers are used, then $E\{\bar{S}(f)\} = (1.0095)\sigma^2$, so the estimate $\bar{S}(f)$ is mildly biased for white noise data. Also,

$$\mathrm{Var}\{\bar{S}(f)\} = E\{[\bar{S}(f)]^2\} - (E\{\bar{S}(f)\})^2$$

But

$$E\{|y_k(f)|^2 |y_{k'}(f)|^2\} = \sum_{k=0}^{K-1}\sum_{k'=0}^{K-1} [E\{|y_k(f)|^2\} \cdot E\{|y_{k'}(f)|^2\} + \sigma^2\delta_{kk'}]$$

as the eigentapers are orthonormal (equation (6)) so

$$\mathrm{Var}\{\bar{S}(f)\} = \frac{\sigma^4}{K^2}\sum_{k=0}^{K-1}\sum_{k'=0}^{K-1} \frac{\delta_{kk'}}{\lambda_k\lambda_{k'}} = \frac{\sigma^4}{K^2}\sum_{k=0}^{K-1}\frac{1}{\lambda_k^2}$$

When the $K = 7$ lowest-order $4\pi$ prolate eigentapers are used, then $\mathrm{Var}\{\bar{S}(f)\} = \sigma^4(1.0196)/7$. Therefore the smoothed high-resolution multitaper estimate has only slightly more variance than the smoothed periodogram estimate for Gaussian white noise data.

For the adaptive spectral estimator, $d_k = \sqrt{\lambda_k}$ for white noise data (equation (4.5) of *Thomson* [1982]). Therefore

$$E\{\hat{S}(f)\} = \frac{\sigma^2\sum_{k=0}^{K-1}\lambda_k}{\sum_{k=0}^{K-1}\lambda_k} = \sigma^2$$

and

$$\mathrm{Var}\{\hat{S}(f)\} = \frac{\sigma^4\sum_{k=0}^{K-1}\sum_{k'=0}^{K-1}\lambda_k\lambda_{k'}\delta_{kk'}}{\left[\sum_{k=0}^{K-1}\lambda_k\right]^2} = \frac{\sigma^4\sum_{k=0}^{K-1}(\lambda_k)^2}{\left[\sum_{k=0}^{K-1}\lambda_k\right]^2}$$

For the $K = 7$ lowest-order $4\pi$ prolate eigentapers, $\mathrm{Var}\{\hat{S}(f)\} = \sigma^4(1.00038)/7$. Therefore the adaptive multitaper estimate has slightly less variance than the smoothed high-resolution estimate and slightly more variance than the periodogram estimate.

### 4.2. *Bias*

It is not useful to compare the bias performance of these spectral estimates for white noise data. One is most interested in a measure of broadband bias. Broadband bias is caused when spectral energy at one frequency leaks away to affect the spectral estimate at a distant frequency and is an important factor to consider in the estimation of spectra of colored processes.

We take as our measure of broadband bias the fraction of energy $(1-\lambda)$ in the frequency band $|f-f'| < W$ that leaks out to affect the estimated spectrum at other frequencies. Suppose that the record consists of a single sinusoid, so that the spectrum is highly colored. The $k$th eigenspectrum retains $\lambda = \lambda_k$ of the spectral energy of the sinusoid within a frequency band of width $2W$ centered on the sinusoid frequency. The fractional leakage of the smoothed high-resolution spectral estimate is

$$1-\lambda = 1 - \frac{\sum_{k=0}^{K-1} \frac{1}{\lambda_k} \int_{H}^{H} |U_k(f)|^2 df}{\sum_{k=0}^{K-1} \frac{1}{\lambda_k} \int |U_k(f)|^2 df} = 1 - \frac{K}{\sum_{k=0}^{K-1} \frac{1}{\lambda_k}}$$

If we use the seven lowest-order $4\pi$ prolate eigentapers in the estimate, $\lambda \approx 0.99057$, so $1-\lambda \approx 0.00943$. For the adaptive multitaper spectral estimate, a numerical calculation shows that $1-\lambda \approx 0.00256$.

The smoothed single-taper direct spectral estimates are also biased when the process has a colored spectrum. A single periodogram estimate allows

$$1-\lambda = 1 - \int_{-W}^{W} |A(f)|^2 df \approx 1 - 0.903 = 0.097$$

of the energy of a single sinusoid to leak outside its main lobe. The smoothed periodogram estimate allows

$$1-\lambda = 1 - \frac{1}{7} \sum_{j=-3}^{3} \int_{-4W}^{4W} |A(f + \frac{j}{N})|^2 df \approx 0.0367$$

of the energy in $|f-f'| \leq 4/N$ to leak out. For the smoothed Hann taper estimate, we find $1-\lambda \approx 0.00934$, while the smoothed 20% cosine taper estimate allows $1-\lambda \approx 0.0192$ of the sinusoid's energy to leak out of $|f-f'| \leq W$.

The Hann and 20% cosine tapers do not permit as much spectral leakage as the boxcar taper, but only the smoothed Hann taper estimates exhibit broadband bias characteristics which are as good as the multitaper estimates. Numerical experiments using the $\omega$-square and $\omega$-cube source spectrum models of *Aki* [1967] demonstrate that spectral estimates employing a boxcar taper are inadequate for representing the source spectrum roll-off. The other tapers are sufficient unless the spectrum rolls off more steeply than $f^{-4}$.

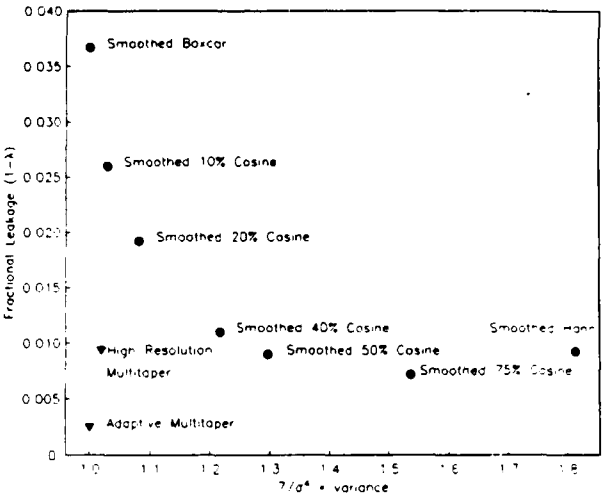



Fig. 7. Comparison of the variance and broadband bias of several single-taper spectral estimates (solid circles) and the multitaper estimates (solid triangles).

Clearly, for smoothed single-taper spectral estimates there is a trade-off. The more severe the taper, the less spectral leakage contaminates the estimate but also the larger the variance of the estimate. The multitaper estimates manage to defeat this trade-off by using several orthogonal leakage-resistant tapers in a single estimate. The relative variances and fractional spectral leakage that are associated with each spectral estimate are listed in Table 2 and are plotted in Figure 7 for comparison.

## 5. CONCLUSIONS

Multitaper spectral analysis techniques offer the seismologist formal and practical advantages over single-taper techniques. Adaptive reweighting of eigenspectra according to the predicted level of spectral leakage enables well-constrained smoothed spectral estimates in portions of the spectrum that have large amplitude, while retaining excellent resistance to spectral leakage in the region where earthquake spectra exhibit a steep roll-off. Comparisons between direct spectral estimates produced using boxcar, Hann, and 20% cosine tapers show that the boxcar taper estimate is contaminated by spectral leakage, that the Hann taper estimates can be misleading in the high-amplitude portion of the spectrum, and that the 20% cosine taper offers a compromise between these two extremes. Therefore a 20% cosine taper may be adequate in many cases but would not be suitable for the analysis of either an unusually dispersive or unusually band-limited seismic signal. However, these pathological situations present no difficulty for the adaptive multitaper estimate.

There are drawbacks to using the multitaper method. The adaptive multitaper algorithm consumes more computer time, since several FFTs must be computed for each time series and one needs to calculate a set of prolate tapers for each time series length. The computational burden is becoming a less serious problem as computer

speeds increase and the cost of computation drops. Also, we have found it adequate to calculate the prolate eigentapers once for a time series of length 128 and generate tapers of other lengths $N > 128$ by interpolating the 128-point tapers using cubic splines (see the appendix).

## APPENDIX: CALCULATING $P\pi$ PROLATE EIGENTAPERS

We follow a procedure similar to that outlined in the appendix of *Thomson* [1982] to calculate eigentapers which have a given time-bandwidth product $NW$ and length $N$. A set of standard tapers of length $N' \gtrsim 100$ is constructed. We use $N' = 128$, as a series whose length is a power of two is convenient for calculating DFTs of the taper. The matrix eigenvalue problem (5) is then solved for the largest $2N'W$ eigenvalues and the associated eigenvectors using EISPACK routines TRED1, BISECT, TINVIT, and TRBAK1 [see *Garbow et al*, 1977]. This procedure determines only the largest eigenvalues and their eigenvectors of a matrix, avoiding the numerical burden of fully decomposing the matrix. In this manner one calculates the prolate eigentapers for a time series of length $N'$. Using the algorithm described by *Thomson* [1982], one approximates the discrete time tapers with the continuous time prolate spheroidal wave functions in order to set up an eigenvalue problem based on Gaussian quadrature. One obtains discrete tapers at nonuniform sample points that can be interpolated to produce tapers with even sampling and of a given length. In our applications we have chosen to use spline interpolation routines to interpolate the evenly spaced tapers of length $N'$ to produce tapers of length $N = N'$.

*Acknowledgments.* We thank David Thomson, Freeman Gilbert, and Alan Chave for many useful comments in the course of this study. Jeffrey Park thanks Dennis G. Wills for temporary lodging in the Scholar's Lair of the La Jolla Cultural Society. This work was supported by National Science Foundation grants EAR-82-18612 and EAR-84-09612 (J.P.), Office of Naval Research contracts N00014-87-K-0005 and N00014-87-K-0191 (C.L.), and U.S. Geological Survey contract 14-08-0001-21893 (F.V.). J.P. was supported as a visiting scientist while at Princeton University Geophysical Fluid Dynamics Lab with F.A. Dahlen by NOAA Grant NA84EAD00057. We thank the members of the Anza Project for the use of their data.

## REFERENCES

Aki, K., Scaling law of seismic spectrum, *J. Geophys. Res.*, *72*, 1217–1231, 1967.

Berger, J., L. M. Baker, J. N. Brune, J. B. Fletcher, T. C. Hanks, and F. L. Vernon III, The Anza array: A high-dynamic range, broadband, digitally radio telemetered seismic array, *Bull. Seismol. Soc. Am.*, *74*, 1469–1481, 1984.

Doob, J. L., *Stochastic Processes*, John Wiley, New York, 1953.

Eberhard, A., An optimum discrete window for the calculation of power spectra, *IEEE Trans. Audio Electroacoust.*, *AU-21*, 37–43, 1973.

Garbow, B., J. Boyle, J. Dongarra, and C. Molar, *Matrix Eigensystem Routines — EISPACK Guide Extension, Lecture Notes in Computer Science*, Springer-Verlag, New York, 1977.

Harris, F., On the use of windows for harmonic analysis with the discrete Fourier transform, *Proc. Instr. Electr. Eng.*, *66*, 51–83, 1978.

Kay, S. M., and S. L. Marple, Jr., Spectrum analysis—A modern perspective, *Proc. IEEE*, *69*, 1380–1419, 1981.

Lindberg, C. R., Multiple taper spectral analysis of terrestrial free oscillations, Ph.D. thesis, 182 pp., Univ. of Calif., San Diego, La Jolla, 1986.

Papoulis, A., *Signal Analysis*, McGraw-Hill, New York, 1977.

Park, J., C. R. Lindberg, and D. J. Thomson, Multiple taper spectral analysis of terrestrial free oscillations: Part I, *Geophys. J. R. Astron. Soc.*, in press, 1987.

Park, J., F. L. Vernon III, and C. R. Lindberg, Frequency dependent polarization analysis of high-frequency seismograms, *J. Geophys. Res.*, this issue.

Slepian, D., Prolate spheroidal wave functions, Fourier analysis, and uncertainty, V, The discrete case, *Bell Syst. Tech. J.*, *57*, 1371–1429, 1978.

Thomson, D. J., Spectral analysis of short series, Ph.D. dissertation, Polytech. Inst. of Brooklyn, New York, 1971.

Thomson, D. J., Spectrum estimation techniques for characterization and development of WT4 waveguide, I, *Bell Syst. Tech. J.*, *56*, 1769–1815, 1977*a*.

Thomson, D. J., Spectrum estimation techniques for characterization and development of WT4 waveguide, II, *Bell Syst. Tech. J.*, *56*, 1983–2005, 1977*b*.

Thomson, D. J., Spectrum estimation and harmonic analysis, *IEEE Proc.*, *70*, 1055–1096, 1982.

C. R. Lindberg and F. L. Vernon III, Institute of Geophysics and Planetary Physics, A-025, University of California, San Diego, La Jolla, CA 92093.

J. Park, Department of Geology and Geophysics, Yale University, P.O. Box 6666, New Haven, CT 06511.

(Received August 7, 1986;
revised July 16, 1987;
accepted July 27, 1987.)

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE *(and Subtitle)*<br>Multitaper Spectral Analysis of High-Frequency Seismograms | | 5. TYPE OF REPORT & PERIOD COVERED<br>6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Jeffrey Park, Craig R. Lindberg, and Frank L. Vernon III | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-87-K-0005 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Scripps Institution of Oceanography<br>La Jolla, California 92093 | | 12. REPORT DATE<br>November 10, 1987<br>13. NUMBER OF PAGES<br>10 |
| 14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*<br>Office of Naval Research<br>Arlington, Virginia 22217 | | 15. SECURITY CLASS. *(of this report)*<br>unclassified<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT *(of this Report)*

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

Approved for public release; distribution unlimited

18. SUPPLEMENTARY NOTES

Reprint from Journal of Geophysical Research, Vol. 92, No. B12, 12,675-12,684, (1987)

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Spectral estimation procedures which employ several prolate spheroidal sequences as tapers have been shown to yield better results than standard single-taper spectral analysis when used on a variety of engineering data. We apply the adaptive multitaper spectral estimation method of Thomson (1982) to a number of high-resolution digital seismic records and compare the results to those obtained using standard single-taper spectral estimates. Single-taper smoothed-spectrum estimates are plagued by a trade-off between the variance of the estimate and the bias caused by spectral leakage. Applying a taper to reduce . . .

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102 LF 014 6601

unclassified
SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 91, NO. B14, PAGES 13,892-13,902, DECEMBER 10, 1986

# Strict Bounds on Seismic Velocity in the Spherical Earth

PHILIP B. STARK, ROBERT L. PARKER, G. MASTERS, AND JOHN A. ORCUTT

*Institute of Geophysics and Planetary Physics, University of California, San Diego, La Jolla*

We address the inverse problem of finding the smallest envelope containing all velocity profiles consistent with a finite set of imprecise $\tau(p)$ data from a spherical earth. Traditionally, the problem has been attacked after mapping the data relations into relations on an equivalent flat earth. Of the two contemporary direct methods for finding bounds on velocities in the flat earth consistent with uncertain $\tau(p)$ data, a nonlinear (NL) approach descended from the Herglotz-Wiechert inversion and a linear programming (LP) approach, only NL has been used to solve the spherical earth problem. On the basis of the finite collection of $\tau(p)$ measurements alone, NL produces an envelope that is too narrow: there are numerous physically acceptable models that satisfy the data and violate the NL bounds, primarily because the NL method requires continuous functions as bounds on $\tau(p)$ and thus data must be fabricated between measured values by some sort of interpolation. We use the alternative LP approach, which does not require interpolation, to place optimal bounds on the velocity in the core. The resulting velocity corridor is disappointingly wide, and we therefore seek reasonable physical assumptions about the earth to reduce the range of permissible models. We argue from thermodynamic relations that $P$ wave velocity decreases with distance from the earth's center within the outer core and quite probably within the inner core and lower mantle. We also show that the second derivative of velocity with respect to radius is probably not positive in the core. The first radial derivative constraint is readily incorporated into LP. The second derivative constraint is nonlinear and can not be implemented exactly with LP; however, geometrical arguments enable us to apply a weak form of the constraint without any additional computation. LP inversions of core $\tau(p)$ data using the first radial derivative constraint give new, extremely tight bounds on the $P$ wave velocity in the core. The weak second derivative constraint improves them slightly.

## INTRODUCTION

There are very few kinds of geophysical data from which we are able to draw sound inferences about deep earth structure. Most of the time, we are of necessity content finding an earth model that adequately accounts for our measurements, disregarding the range of models that predict the data equally well, any of which might resemble the actual earth more closely. Contributing to the nonuniqueness of the solution is the paucity of data available versus the complete description of the earth we seek and the fact that our few data are inexact. Even if we had an infinite amount of noise-free data, deliberate approximations in our assumptions (e.g., that the earth is spherically symmetric) may force us to treat the data as inexact. The issue of nonuniqueness can sometimes be resolved by choosing to optimize some property of the earth model while fitting the data, resulting in a problem with only one solution. In other cases, one can delineate the range of models that satisfy the data and the assumptions of the derivations. One such problem is finding a corridor in the velocity-depth plane within which every velocity model satisfying given seismic travel time data must lie. Since travel time data include triplications and other complications, it is desirable to work with $\tau(p) \equiv T(p) - p\Delta(p)$, the vertical delay time as a function of ray parameter $p$ instead. $T$ is the travel time, and $\Delta$ is epicentral distance in degrees. Ideally, $\tau(p)$ contains the same information as travel time data, $T(\Delta)$, but is a monotonic function, continuous except where there are low-velocity zones. Estimating $\tau(p)$ from the original records is a nontrivial matter, but we shall assume that this step has been taken successfully.



Paper number 6B5937.
0148-0227/86/006B-5937$05.00

The inverse problem of finding the maximum and minimum velocities at a given radius consistent with a set of measurements of $\tau(p)$ on a spherical earth traditionally has been attacked after an exact mapping into a similar problem for a flat earth. ($T(\Delta)$ then becomes $T(X)$, where $X$ is epicentral distance in kilometers.) There are two quite different approaches to the flat earth problem: *Bessonova et al.* [1976] developed a nonlinear scheme (NL), a descendent of the Herglotz-Wiechert integral solution; *Garmany et al.* [1979] transformed the problem so that the data relations were linear and solved it with linear programming (LP). Both NL and LP require the assumption that there are no low-velocity zones in the flat earth, i.e., that $dv/dz \geq 0$, where $v$ is seismic velocity as a function of flat earth depth $z$, or that the effect of low-velocity zones has been removed from the $\tau(p)$ data. *Bessonova et al.* [1974, 1976] discuss how to preprocess the data to remove the traces of low-velocity zones so that NL may be applied; *Orcutt* [1980] shows how the data may be prepared similarly for LP inversion.

The work of *Bessonova et al.* [1976] is the latest in a chain of inversions of travel time data relying upon the Herglotz-Wiechert integral solution [*Aki and Richards*, 1980, vol. II, chapter 12]; other notable papers in the series include *Gerver and Markushevitch* [1966], *Wiggins et al.* [1973] and *Bessonova et al.* [1974]. One of the explicit aims of the $\tau$ method (NL) of *Bessonova et al.* [1974] was to avoid extrapolating $T(X)$ curves from the available finite collection of measurements. Unfortunately, the extrapolation was merely moved from $T(X)$ to $\tau(p)$: NL requires continuous bounds on $\tau$ over a range of $p$. One could imagine using nonparametric estimates of $\tau(p)$ to

construct continuous bounds in a consistent way, although this is not what is done in practice. *Bessonova et al.* [1976] used a statistical technique for estimating confidence intervals for $\tau$ at fixed values of $p$, but their revision of NL still needed continuous bounds on $\tau(p)$ which they constructed by interpolating between the computed points. Interpolation of this sort can violate the maxim that an estimate of a function should not deteriorate if more information becomes available: the interpolated bounds for $\tau$ would be pushed out if an additional wide confidence interval for $\tau$ were computed at a $p$ between two narrow confidence intervals. Interpolation would be acceptable if it did not influence the results unduly, but we show that it can produce radical changes in the velocity-depth envelope. This was not apparent in earlier comparisons of LP and NL [*Garmany et al.*, 1979]. It is very difficult to predict the effect of different data interpolation schemes on the models NL finds because the data and models are nonlinearly related. Interpolation may rule out models that satisfy the finite set of data and yet may allow unphysical models that violate the assumptions of the method. For example, we must insist that candidate velocity models be single-valued. Since NL builds its models from the continuous data bounds, the bounds themselves must be legitimate $\tau(p)$ profiles, which in general they are not. When they are not, NL modifies the continuous data bounds so that they do correspond to realizable velocity models. The rather ad hoc procedure may result in new "data" that violate the original bounds.

These difficulties of NL are not present in the LP formulation of the problem. LP directly incorporates the constraint that candidate models be physically realizable: it avoids the problems of multivalued velocity functions in a straightforward and consistent fashion. LP escapes the need to interpolate by working in the other direction: rather than construct velocity models by transforming the data bounds (which must then be defined over a continuous range of $p$), LP examines all models that satisfy the finite collection of data and chooses those with the greatest and least velocities at some depth. It is possible to use LP to discover the envelope containing all physical $\tau(p)$ profiles that satisfy the finite data set since $\tau$ itself could be used as the penalty functional, but this is not the basis of the inversion.

We would like to be able to incorporate other a priori information about the range of possible earth models into our inversions to tighten the velocity-depth bounds. The assumption that $dv/dz \geqslant 0$ is necessary with LP and NL but translates to an ad hoc proscription in the spherical earth: $dv/dr \leqslant v/r$. This allows low-velocity zones in the spherical earth provided velocity increases more slowly than radius. A preferable and more powerful constraint is that $dv/dr \leqslant 0$: no low-velocity zones in the spherical earth. We support this assumption with thermodynamic arguments applicable to the outer core and less stringently to the inner core and lower mantle (Appendix A). The new constraint on the models is linear and so may be easily incorporated into LP inversions. This restriction would have to be posed in terms of the interpolated $\tau(p)$ bounds to be used with NL. Finding the correct interpolation is practically impossible because it depends nonlinearly on all the data simultaneously.

There are no inherent limitations to the accuracy of LP: the limits are set by the precision of the machine computations and the number of basis functions one uses to represent the earth. This has been proven rigorously by P. B. Stark (unpublished manuscript, 1986), who also proves that within reason, the LP results are insensitive to the particular choice of basis functions. In Appendix B we exploit the linearity of the data relations in the LP formulation and the convexity of the spherical earth to flat earth mapping to prove that the extremal bounds in the spherical earth are just the extremal bounds in the flat earth mapped into the spherical earth. *Gerver and Markushevitch's* [1966] flat earth mapping provides a one-to-one correspondence between spherical and flat earth velocity models predicting exactly the same $\tau(p)$ data for all values of $p$. However, when there is one velocity model that satisfies a finite and inexact data set, there are usually many. It is entirely possible that the velocity model (there may be more than one) that maximizes the flat earth velocity at some depth while satisfying the data might not be the model that maximizes the spherical earth velocity at the corresponding spherical earth depth since depths and velocities in the two domains are nonlinearly related. Thus the coincidence of the extremal models is less than obvious although it has been tacitly assumed heretofore.

Values of $\tau(p)$ are difficult to obtain in some ranges of $p$, while $\Delta(p)$ measurements in that $p$ interval may be more readily available. It is therefore very useful to be able to treat $\Delta(p)$ data jointly with $\tau(p)$ estimates. The LP formulation may employ $\Delta(p)$ and $\tau(p)$ data concurrently [*Orcutt*, 1980].

To test the LP approach in the spherical earth, we have inverted the definitive $\tau(p)$ data set for the core [*Johnson and Lee*, 1985] reduced from 90,000 contemporary International Seismological Centre (ISC) travel times; it is unlikely that better spherically averaged values of $\tau(p)$ for the deep interior of the earth will become available for some time. Like Johnson and Lee, we treat the scatter in the estimates derived from the original $T(\Delta)$ observations as statistical noise disturbing an ideal spherically averaged $\tau(p)$ curve and take the 99.9% confidence intervals as strict bounds on the uncertainties of the $\tau$ values. LP produces generally wider bounds than NL inversion. This might indicate that LP is too conservative, except that the bounds found by LP are achievable: for every velocity-depth point on the bounds there is a model that contains that point and satisfies the finite $\tau(p)$ data exactly. The NL bounds are thus sensitive to the interpolation of the $\tau(p)$ limits, as mentioned earlier.

*Johnson and Lee* [1985] constructed five $\tau(p)$ data using $\Delta(p)$ to constrain the derivative to incorporate additional information in a range of $p$ where $\tau(p)$ data were unavailable. We compare LP inversions of their data with and without these values and also with some $\Delta(p)$ data used directly. We conclude that the five data have a major influence on the inner core boundary and determining the shape of the envelope in that vicinity. When the $\Delta(p)$ data are used directly, a wider and probably more reliable corridor results.

The finite set of $\tau(p)$ and $X(p)$ data is not very restrictive: without additional assumptions the LP bounds are

fairly wide, particularly within the inner core where the flat earth mapping is strong. Forcing the first radial derivative of velocity to be nonpositive substantially narrows the LP bounds. Assuming that the second radial derivative of the $P$ wave velocity in the core is nonpositive enables us to tighten the bounds a bit more using a geometrical construction. Both radial derivative constraints are justified by thermodynamic arguments in Appendix A. The final result, based on nonpositive first and second radial derivatives, is an extremely narrow envelope of velocities in the core consistent with the data. This envelope, roughly comparable to that of *Johnson and Lee* [1985] but tighter particularly in the inner core, is reached by physical arguments.

## METHODS

We denote velocities in the radially symmetric spherical earth model by $\mathrm{v} = \mathrm{v}(r)$ and velocities in the flat earth by $v = v(z)$. Depth $z$ is measured from the surface of the flat earth and radius $r$ is measured from the center of the spherical earth. The velocity at the surface of the sphere, which is the same as at the surface of the half-space, is $w \equiv \mathrm{v}(a) = v(0)$. The variables $\mathrm{v}$, $r$, $v$, and $z$ are related by [*Gerver and Markushevitch*, 1966]

$$v = \frac{\mathrm{v}a}{r} \qquad z = -a \ln\left(\frac{r}{a}\right) \tag{1}$$

Values of the spherical earth ray parameter $dT/d\Delta$ may be converted to their equivalent flat earth values, $dT/dX$, by multiplying by the number of degrees per unit distance at the surface of the sphere.

The forward problems of mapping a one-dimensional monotonic flat earth velocity profile into $\tau(p)$ and $X(p)$ are solved by the familiar transformations [*Aki and Richards*, 1980, vol. II, chapter 12]:

$$\tau(p) = 2\int_0^{z(p)} \left(v(z)^{-2} - p^2\right)^{1/2} dz$$

$$X(p) = 2\int_0^{z(p)} p\left(v(z)^{-2} - p^2\right)^{-1/2} dz$$

where $z(p)$ is the turning depth of the ray with ray parameter $p$, i.e., the depth to velocity $1/p$. Changing the independent variable of integration to $v$, we find

$$\tau(p) = 2\int_w^{1/p} \left(v^{-2} - p^2\right)^{1/2} \frac{dz}{dv}\, dv$$

$$X(p) = 2\int_w^{1/p} p\left(v^{-2} - p^2\right)^{-1/2} \frac{dz}{dv}\, dv$$

For a particular choice of $p = p_i$ these integral relations are linear functionals of $dz/dv$:

$$\tau_i[\zeta] \equiv \tau(p_i) = 2\int_w^{1/p_i} \left(v^{-2} - p_i^2\right)^{1/2} \zeta(v)\, dv \tag{2}$$

$$X_i[\zeta] \equiv X(p_i) = 2\int_w^{1/p_i} p_i\left(v^{-2} - p_i^2\right)^{-1/2} \zeta(v)\, dv \tag{3}$$

where $\zeta(v) \equiv dz/dv$ is the function we will use to represent the earth model. We can find $z(v)$ by integration if we know $\zeta(v)$ and $w$:

$$Z_v[\zeta] \equiv z(v) = \int_w^v \zeta(v)\, dv \tag{4}$$

$Z_v[\zeta]$ is a linear functional of $\zeta(v)$, and knowing $z(v)$ is equivalent to knowing $v(z)$ provided there are no low-velocity zones, that is, provided $\zeta(v) > 0$, so that both $z(v)$ and $v(z)$ are single-valued.

We take certain confidence limits on $\tau(p)$ and $X(p)$ to be strict bounds on $l$ observations of $\tau$ and $m$ observations of $X$; i.e., we assume that we are given two $n$ vectors $\mathbf{d}^-$ and $\mathbf{d}^+$ such that

$$\mathrm{d}_i^- \leqslant \tau_i[\zeta] \leqslant \mathrm{d}_i^+ \qquad i = 1, \ldots, l \tag{5}$$

$$\mathrm{d}_i^- \leqslant X_i[\zeta] \leqslant \mathrm{d}_i^+ \qquad i = l+1, \ldots, n \tag{6}$$

where $n = l+m$ is the total number of data. We assume that the observations are ordered such that $p_i < p_{i+1}$, $i = 1, \ldots, l-1$, and $i = l+1, \ldots, n-1$. The maximum velocity about which we have information is then $y = \max\{1/p_1, 1/p_{l+1}\}$. The data relations expressed in (5) and (6) are a set of $n$ two-sided linear inequalities in $\zeta$. Following *Garmany et al.* [1979], we solve the problem of finding strict limits on the range of velocities by determining the range of depths in which each velocity is allowed: we alternately maximize and minimize $Z_v[\zeta]$ for each target velocity $v$, subject to the collection of $2n$ linear inequality constraints (5), (6), and the positivity constraint $\zeta(v) > 0$. Each of these optimization problems is an infinite-dimensional linear program in the space in which we decide to embed $\zeta$. Appendix B shows that the solutions to these problems in the flat earth are just the extremal bounds in the spherical earth mapped by equations (1), so we may solve the spherical earth extremal bound problem by mapping the data into the flat earth, solving the flat earth problem and mapping the results back into the spherical earth.

## SOLVING THE FLAT EARTH EXTREMAL BOUND PROBLEM

In practice, we must describe the unknown earth model $\zeta(v)$ with a finite collection of numbers — a computer can not store a value of $\zeta(v)$ for every value of $v$. If we write $\zeta$ as a linear combination of a particular finite set of basis functions, the coefficients in the linear combination constitute a finite description of the model a computer can use. The basis set is acceptable if we are able to approximate the data and depth mappings of any $\zeta$ arbitrarily well by using more and more basis functions of the class that we choose (equivalently, if the span of the basis functions is weak-star dense in the limit). The integrals for the mappings may be performed for the basis functions individually, and the resulting numbers, scaled by the coefficients in the expansion, may be added to give the value of the integrals performed on $\zeta$ since the integrals are linear in $\zeta$. We can avoid numerical quadrature and retain the highest accuracy in our computations if the integrals can be performed analytically for the chosen basis functions.

Delta functions are a natural basis set for the flat earth problem because they give rise to homogeneous layers in

$v$. Intuitively, we see that such layers allow changes in velocity to be made as early or as late as possible. We can support this more rigorously: the fundamental theorem of linear programming [*Luenberger*, 1973] states that if there is an optimal linear combination of the basis functions that solves the problem, there is an optimal solution comprised of a linear combination of at most as many basis vectors as there are constraints. If we were to choose boxcar functions to approximate the solution, then in the limit as the number of basis functions goes to infinity and the boxcars become vanishingly narrow, we would still require at most $2n$ basis functions to represent the solution: the solution would tend to a sum of delta functions. (Notice that in using a delta function basis and finite-dimensional linear programming, we must widen the constraints on $\zeta$ from $\zeta > 0$ to $\zeta \geqslant 0$; this extension apparently causes difficulties in the velocity domain because we lose the constraint that $v(z)$ is single-valued. However, if the solution is interpreted in the limit as the lower bound on $\zeta$ approaches zero, merely discontinuous velocities are generated.) The delta function basis has some practical benefits as well: the integrals (2), (3), and (4) are trivial, and we may guarantee positivity of $\zeta(v)$ just by requiring positivity of the coefficients in the basis expansion.

While delta functions are truly optimal only when there is complete freedom to place the layers where needed (i.e., in the limit of an infinite number of basis functions), we can guarantee flexibility near the target velocity $v$ where it is most crucial by inserting basis functions on either side of $v$. This tailoring of the basis expansion to the target velocity has produced substantial improvements in the computational efficiency over earlier realizations of LP [e.g., *Garmany et al.*, 1979]. P. B. Stark (unpublished manuscript, 1986) provides a theoretical account of the improvement.

The choice of a delta function basis for $dz/dv$ suggests that we are approximating a problem where $z(v) \in BV[w,y]$, the Banach space of functions of bounded variation on the interval $[w,y]$ with the variation as the norm. Our delta function expansion of $dz/dv$ leads to a step function expansion of $z(v) \in BV$ upon integration. Spacing the basis functions evenly in velocity is not desirable: most of the sensitivity to the data occurs close to the surface because the integral (2) depends on $1/v$. It is preferable to space the steps evenly in slowness, i.e., reciprocally in velocity. We have found that the numerical solution is much more stable with this spacing, partly because it improves the conditioning of the mapping matrix. It can be proven that the span of a reciprocally spaced set of step functions is weak-star dense in $BV[w,y]$ in the limit, so a basis of this form is acceptable. Our preliminary basis set shall be

$$b_j(v) \equiv \delta(v - v_j) \tag{7}$$

where

$$v_j \equiv \frac{1}{1/y + (L-j-2)h} \qquad j = 1, \ldots, L-2$$

and

$$h \equiv \frac{1/w - 1/y}{L-3}$$

Recall that $w$ is the minimum velocity and $y$ is the maximum one. When we choose a certain target velocity $v$, we insert extra basis functions $b_{L-1}$ and $b_L$ at $1/(1/v + \alpha h)$ and $1/(1/v - \alpha h)$, where $\alpha$ is a small positive constant, about 0.1 typically.

We now write $\zeta$, the unknown earth model, in terms of its basis expansion and perform the integrals (2), (3), and (4):

$$\zeta = \sum_{j=1}^{L} \zeta_j \, b_j = \sum_{j=1}^{L} \zeta_j \, \delta(v - v_j)$$

$$\tau_i[\zeta] = 2 \sum_{J_{1/p_i}} \zeta_j \, (v_j^{-2} - p_i^2)^{1/2} \tag{8}$$

$$X_i[\zeta] = 2 \sum_{J_{1/p_i}} \zeta_j p_i \, (v_j^{-2} - p_i^2)^{-1/2} \tag{9}$$

$$Z_v[\zeta] = \sum_{J_v} \zeta_j \tag{10}$$

where the index set $J_x \equiv \{j \mid v_j < x\}$. Note that (9) is unbounded if there is a basis function at $v = 1/p_i$, $i \in \{1, \ldots, n\}$, and so that particular choice must be avoided. This minor complication has been resolved in a consistent and acceptable fashion by defining the integrals on open intervals, so it is not a fundamental limitation. Equations (8), (9), and (10) may be written as vector dot products:

$$\tau_i[\zeta] = \boldsymbol{\tau}_i \cdot \boldsymbol{\zeta} \tag{11}$$

$$X_i[\zeta] = \mathbf{X}_i \cdot \boldsymbol{\zeta} \tag{12}$$

$$Z_v[\zeta] = \mathbf{Z}_v \cdot \boldsymbol{\zeta} \tag{13}$$

In (11)–(13), $\boldsymbol{\zeta}$ is the vector of coefficients $\zeta_j$ in the basis expansion (7) of $\zeta$,

$$\tau_{ij} \equiv \begin{cases} 2(v_j^{-2} - p_i^2)^{1/2} & v_j < 1/p_i \\ 0 & v_j \geqslant 1/p_i \end{cases}$$

$$\mathrm{X}_{ij} \equiv \begin{cases} 2p_i(v_j^{-2} - p_i^2)^{-1/2} & v_j < 1/p_i \\ 0 & v_j \geqslant 1/p_i \end{cases}$$

$$\mathrm{Z}_{vj} \equiv \begin{cases} 1 & v_j < v \\ 0 & v_j \geqslant v \end{cases}$$

For a given $v$ we wish to minimize

$$\pm \mathbf{Z}_v \cdot \boldsymbol{\zeta} \tag{14}$$

subject to

$$\begin{aligned} &\zeta_j \geqslant 0 && j = 1, \ldots, L \\ &\mathrm{d}_i^- \leqslant \boldsymbol{\tau}_i \cdot \boldsymbol{\zeta} \leqslant \mathrm{d}_i^+ && i = 1, \ldots, l \\ &\mathrm{d}_i^- \leqslant \mathbf{X}_i \cdot \boldsymbol{\zeta} \leqslant \mathrm{d}_i^+ && i = l+1, \ldots, n \end{aligned}$$

Minimizing $+\mathbf{Z}_v \cdot \boldsymbol{\zeta}$ minimizes the depth to $v$, and minimizing $-\mathbf{Z}_v \cdot \boldsymbol{\zeta}$ maximizes the depth. These are standard finite-dimensional linear programs, and software to solve them is widely available. Two moderately large linear programs must be solved for each target velocity so the computational effort is far from trivial.

### CONSTRAINING THE RADIAL DERIVATIVE OF VELOCITY

An essential assumption of the foregoing derivation is that $\zeta(r) \geqslant 0$, i.e., that there are no low-velocity zones in the flat earth. This becomes the ad hoc requirement that $dv/dr \leqslant v/r$ in the spherical earth: there may be low-velocity zones. The low-velocity zones are mild in the crust and upper mantle because velocity increases no faster than radius, but in the core they can be major features of the solution. It is geophysically preferable to assume that there are no low-velocity zones in the spherical earth: $dv/dr \leqslant 0$. In Appendix A we argue that this should be true within homogeneous, adiabatic regions in the earth composed of "normal" materials. These conditions are generally thought to apply approximately throughout the outer core, and are probably not violated within the inner core. The new assumption does not contradict the previous constraint on the flat earth derivative; it is more restrictive:

$$\frac{dv}{dr} \leqslant 0 \Leftrightarrow \zeta(r) \leqslant \frac{a}{r} \tag{15}$$

a result easily obtained from (1). Inequality (15) is linear in $\zeta$, so it can be incorporated into the inversion with the same mathematical machinery. However, we may no longer use a delta function basis for $\zeta$ because delta functions violate the new constraint everywhere they fail to vanish. If (15), which depends on $1/r$, is to apply exactly over the support of a basis function defined on an interval, the basis function must also depend on $1/r$. Logarithmic functions in $z(r)$ correspond to $1/r$ basis functions for $\zeta = dz/dr$. The span of reciprocally spaced pieces of logarithmic functions is also weakly dense in $BV[w,y]$ in the limit as the number of basis functions increases without bound so this basis is acceptable. P. B. Stark (unpublished manuscript, 1986) proves that the solution is not sensitive to the choice of basis functions provided they get closer and closer together as more of them are used and provided enough of them are used. The logarithmic functions are particularly good because fewer of them are required than of other basis functions (e.g., ramps) to get results of the same accuracy. We proceed by expanding $\zeta$ in a new set of basis functions $c_j$ and performing the resulting integrals (2), (3), and (4). Let

$$c_j \equiv \frac{1}{r}\,\Pi\left(\frac{r - r_j}{r_{j+1} - r_j}\right)$$

where

$$\Pi(r) \equiv \{1,\ 0 \leqslant r \leqslant 1;\ 0,\ \text{otherwise}\}$$

The velocities $r_j$ are defined as in equation (7) except that they have been put in increasing order (the two velocities bracketing the target velocity are no longer at the end of the list). Now

$$\tau_i[\zeta] = \sum_{j:\, r_j < p_i} \zeta_j \left[F_i(\hat{\theta}_{ij}) - F_i(\theta_{ij})\right]$$

$$X_i[\zeta] = 2 \sum_{j:\, r_j < p_i} \zeta_j\,(\theta_{ij} - \hat{\theta}_{ij})$$

$$Z_i[\zeta] = \sum_{j:\, r_j < p_i} \zeta_j \ln\left(\frac{\min\{r, r_{j+1}\}}{r_j}\right)$$

where

$$F_i(\theta) \equiv 2p_i(\theta - \tan\theta)$$

$$\theta_{ij} \equiv \cos^{-1}(p_i r_j)$$

$$\hat{\theta}_{ij} \equiv \cos^{-1}(p_i \min\{r_{j+1},\ 1/p_i\})$$

These three expressions may also be written as vector dot products (11), (12), and (13), with the new identifications

$$\tau_{ij} \equiv \begin{cases} F_i(\hat{\theta}_{ij}) - F_i(\theta_{ij}) & r_j < 1/p_i \\ 0 & r_j \geqslant 1/p_i \end{cases}$$

$$\mathbf{X}_{ij} \equiv \begin{cases} 2(\theta_{ij} - \hat{\theta}_{ij}) & r_j < 1/p_i \\ 0 & r_j \geqslant 1/p_i \end{cases}$$

$$\mathbf{Z}_{ij} \equiv \begin{cases} \ln(\min\{r, r_{j+1}\}/r_j) & r_j < 1/p_i \\ 0 & r_j \geqslant 1/p_i \end{cases}$$

The new finite-dimensional problem with radial derivative constraints is to minimize

$$\pm\mathbf{Z}_v \cdot \boldsymbol{\zeta}$$

subject to the constraints

$$a \geqslant \zeta_j \geqslant 0 \qquad j = 1, \ldots, L$$

$$d_i^- \leqslant \boldsymbol{\tau}_i \cdot \boldsymbol{\zeta} \leqslant d_i^+ \qquad i = 1, \ldots, l$$

$$d_i^- \leqslant \mathbf{X}_i \cdot \boldsymbol{\zeta} \leqslant d_i^+ \qquad i = l+1, \ldots, n$$

Minimizing $+\mathbf{Z}_v \cdot \boldsymbol{\zeta}$ minimizes the depth to $r$, and minimizing $-\mathbf{Z}_v \cdot \boldsymbol{\zeta}$ maximizes the depth. These are also standard linear programs and may be solved straightforwardly.

### APPLICATION TO THE EARTH'S CORE

We have applied the theory to $\tau(p)$ data for the core obtained by *Johnson and Lee* [1985] (Figure 1). We follow Johnson and Lee in interpreting the scatter of the data as a noise process distorting an ideal, spherically averaged data curve. The data bounds determined by Johnson and Lee are the 99.9% confidence limits on $\tau(p)$ which both we and they take to be firm bounds within which $\tau$ must lie. P. B. Stark and R. L. Parker (unpublished manuscript, 1986) have developed a method of inverting the confidence interval data without reinterpreting the intervals as strict bounds. The results of inversion with the statistical treatment of the bounds, though different in detail, are surprisingly similar to the results assuming strict data bounds. In reality, the noise components of the observations are probably not statistically independent because they are principally the result of large-scale heterogeneities in the earth and anomalies associated with the sources and receivers; the actual statistics of the $\tau$ values are largely unknown and almost certainly not Gaussian.

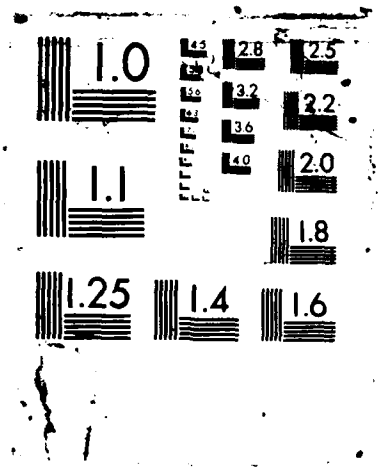
The inversions that follow used 100 basis functions in the preliminary expansion and an additional pair to bound the target velocity. Using 200 did not noticeably change the results; the bounds have converged apparently. The inversions started at a radius of 3480 km, the [illegible] which Johnson and Lee corrected their [illegible] PREM anisotropic earth model [Dziewonski and Anderson, 1981]. We used a minimum [illegible] set about 90 target velocities [illegible] maximize the depth. Figure [illegible]

AD-A192 238 REPRINTS OF ONR (OFFICE OF NAVAL RESEARCH) WORK DECEMBER 1986-NOVEMBER 1987(U) SCRIPPS INSTITUTION OF OCEANOGRAPHY LA JOLLA CA FEB 88 2/2

UNCLASSIFIED F/G 8/11 NL

[redacted]

END
DATE FILMED
8-


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0

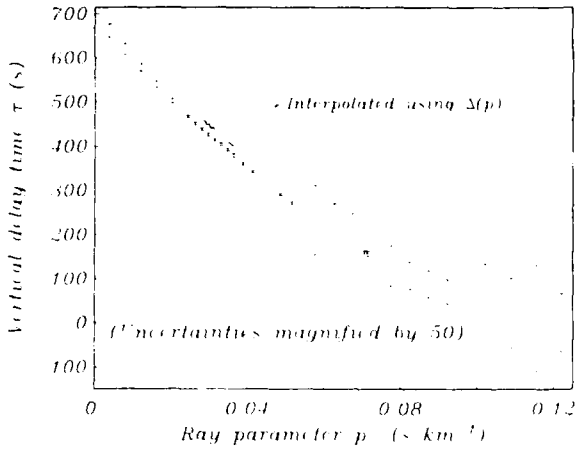



Fig. 1. The $\tau(p)$ data of *Johnson and Lee* [1985] replotted. The ray parameters are the equivalent flat earth values having been scaled by $180/\pi$ and divided by a core radius of 3480 km.

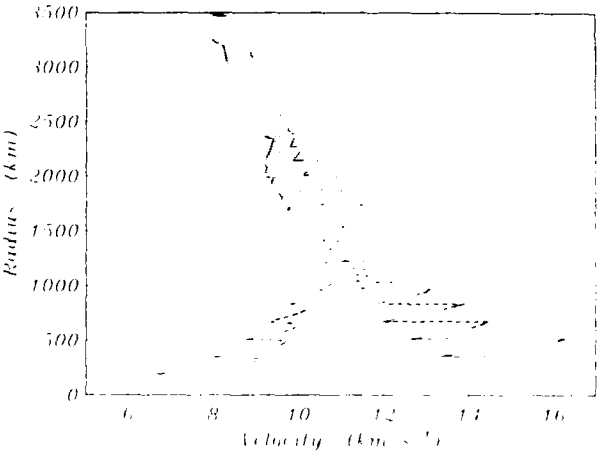



Fig. 3. Velocity bounds calculated without requiring v to decrease with radius. The shaded zone is the solution of *Johnson and Lee* [1985] obtained by NL using all 30 data of Figure 1. The solid line is our LP solution based upon the same data; notice how much wider the bounds found by LP are. The dashed bounds are obtained when the five interpolated values indicated in Figure 1 are omitted; the expansion of the bounds shows how crucial these five data are to the solution near the inner core boundary.

models from an inversion with target velocity $v = 200$ km s$^{-1}$. The effect of the flat earth mapping is seen in the large values of velocity (250 km s$^{-1}$) and depth (10,000 km) as radius tends to zero.

We first found bounds on the velocity in the core assuming that only the flat earth velocity gradient is nonnegative, (i.e., $dv/dr \leqslant v/r$), the identical situation treated by Johnson and Lee. The solid line in Figure 3 is the LP solution using all of *Johnson and Lee's* [1985] $\tau(p)$ values; the dashed line is the LP solution excluding the five values constructed with $\Delta(p)$. The envelopes are determined by about 180 target velocities. The shaded region is Johnson and Lee's NL solution based on all their data. Though both approaches begin with the same $\tau(p_i)$ data, many more solutions are accessible to the LP technique

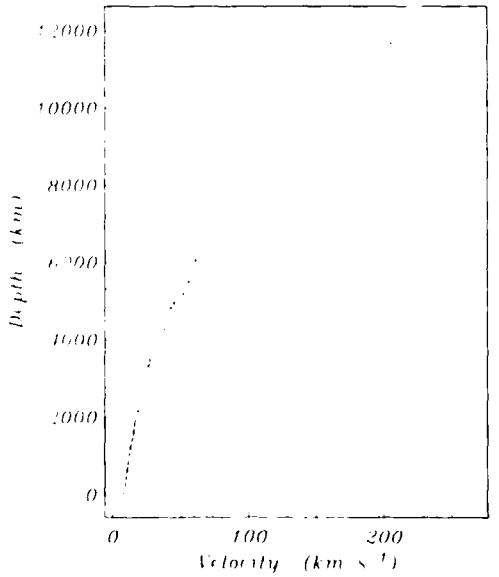



Fig. 2. Two extremal solutions in the equivalent flat earth, one (solid line) minimizing the depth to a velocity of 200 km s$^{-1}$, the other (dashed) maximizing it. Notice the extremely large values of velocity and depth from the exponential mapping.

than to NL. For each point on the LP bounds there is a velocity model that satisfies the measured $\tau(p)$ data and contains that velocity-depth point. We conclude that the NL bounds are narrower than the data require because they exclude models that satisfy the original finite list of $\tau(p)$ data. The jagged excursions in the LP bounds are due to the finite size of the data set: between data, where the constraints on the model are not so demanding, LP is free to make the depth smaller or greater as asked. One should note that even if the inner excursions of the LP bounds were connected, simulating the effect of interpolating the $\tau(p)$ data, the resulting envelope would still lie outside the NL bounds. Neither NL nor LP claims that the bounds themselves are reasonable velocity models, but rather that, in the absence of additional information, each velocity-depth point on the bounds is contained in some velocity model that does satisfy the data. The NL solution does not exhibit the same ragged behavior as LP because the $\tau(p)$ data have been first interpolated to get continuous bounds. It follows from the wide difference between the results of the two approaches that the NL answer is sensitive to the precise interpolation procedure employed. Unfortunately, it is clear that although technically correct, the LP bounds are too wide to be very interesting. This is not a fault of LP; on the contrary, it shows that the present list of $\tau(p)$ data is insufficient to bound velocity very well: if we want tighter bounds, we must either make additional assumptions or introduce more data with the same high accuracy.

*Johnson and Lee* [1985] used $\Delta(p)$ values from the Tonto Forest Seismological Observatory to construct five values of $\tau(p)$ to refine the envelope near the inner core boundary. The sensitivity of the bounds in that vicinity to these five $\tau(p)$ data is clear from comparison of the solid and dashed bounds. There is no need for us to interpolate to incorporate the additional information since LP can exploit $\Delta(p)$ data directly. The $\Delta(p)$ data are shown in

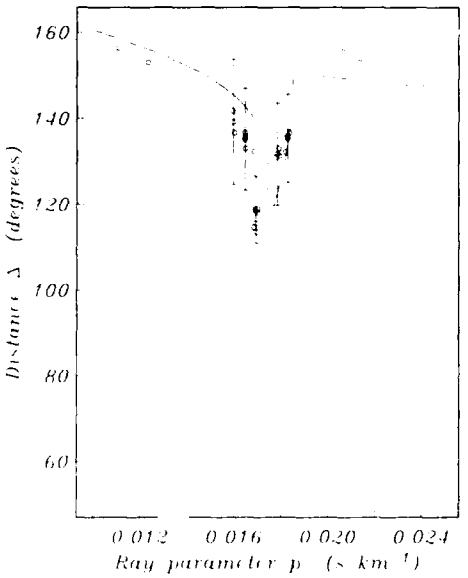



Fig. 4. $\Delta(p)$ data of *Johnson and Lee* [1985] shown as circles. The solid curve shows the predictions of the 1066A earth model. We have assigned uncertainties and adopted values at five ray parameters for our inversion; the numerical values appear in Table 1.

Figure 4 as circles, along with the $\Delta(p)$ predictions of model 1066A [*Gilbert and Dziewonski*, 1975] and our crude assessment of errors in the range of $p$ where $\tau$ was not available. The error bounds are based upon the variance of seven $\Delta(p)$ points at about the same $p$. These conservative values were conferred because so few measurements were available and because the $\Delta(p)$ data, from an array in central Arizona, are not expected to be completely representative of the spherically averaged earth. We adjusted the $\Delta(p)$ values to 3480 km by subtracting the effects of the crust and mantle of the PREM anisotropic earth model to be consistent with Johnson and Lee; the reduced data are given in Table 1.

We can tighten the bounds by making the assumption that $dv/dr \leqslant 0$, as described earlier and justified thermodynamically in Appendix A. The dashed bounds in Figure 5 are determined from the 25 $\tau(p)$ data, while the solid bounds include the five $\Delta(p)$ constraints from Figure 4; the only noticeable difference is in the vicinity of the inner core boundary. The shaded zone is the region obtained from all 30 $\tau(p)$ data of *Johnson and Lee* [1985]. The

TABLE 1. Five Values of Δ for Rays Passing Near the Inner Core Boundary

| $p$, s deg$^{-1}$ | $p$, s km$^{-1}$ | Δ, deg |
|---|---|---|
| 1.775 | 0.02922 | 119.752 |
| 1.830 | 0.03013 | 115.003 |
| 1.880 | 0.03095 | 98.026 |
| 1.983 | 0.03265 | 109.682 |
| 2.033 | 0.03347 | 112.670 |

Each Δ value is assigned an uncertainty of ±16.12 degrees. The values have been reduced to the surface of the core (radius 3480 km) by subtracting the predictions of the PREM mantle from the observations shown in Figure 4.

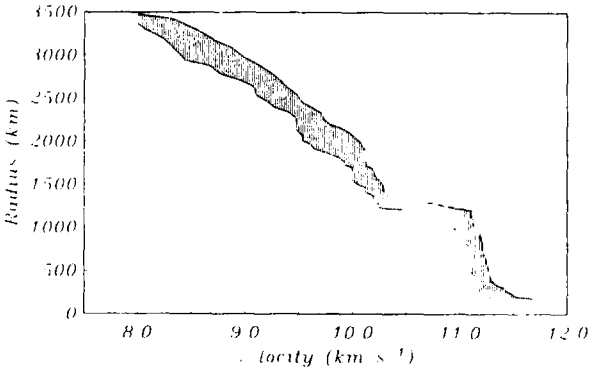



Fig. 5. Velocity bounds in the core when v is required to decrease with increasing radius. The shaded zone is obtained using all thirty data of Figure 1. The region bounded by dashed lines is found when the five interpolated data are omitted; that bounded by the solid lines results from replacing the interpolated $\tau$ values by the Δ values of Figure 4.

solid line represents our preferred solution since we would rather use the $\Delta(p)$ data directly. Each pair of bounds is determined by about 180 points. Figure 6 superposes the bounds from 25 $\tau(p)$ data with and without the radial derivative constraint. The power of the radial derivative constraint is immediately evident: the corridor is extremely narrow, even in the inner core where the exponential mapping takes its greatest toll. An important reason constraints of this sort are so powerful is that they constrain the model at each point, whereas data constrain integrals of pieces of the models.

We would like to require the second radial derivative of velocity in the core to be nonpositive (Appendix A). Manipulation of the flat earth mapping (1) shows that $d^2v/dr^2 \leqslant 0$ is equivalent to $-(dz/dv)^2 \leqslant ad^2z/dv^2$, a nonlinear constraint in this formulation that cannot be imposed exactly in LP. However, we can use a geometrical argument to rule out some of the corridor allowed by the first radial derivative constraint. It must be possible to join any point within bounds incorporating the second derivative constraint to both ends of the envelope with a

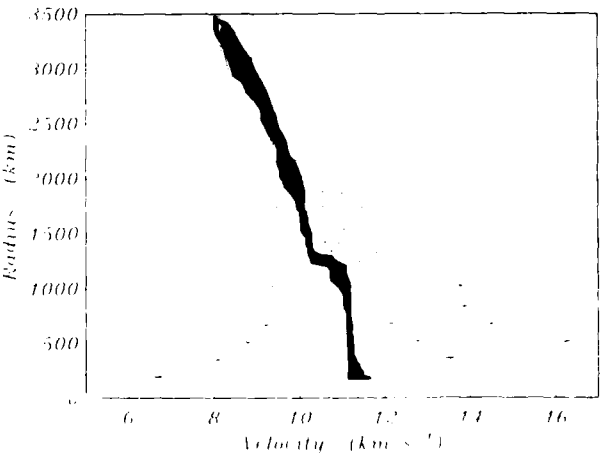



Fig. 6. A comparison of the velocity bounds with and without the constraint that v decrease with radius using the 25 $\tau$ values in Figure 1 not based upon interpolation.

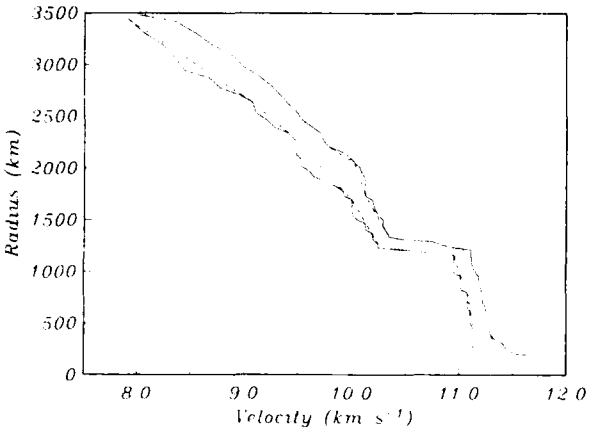



Fig. 7. Most restricted solution discussed in this paper, shown by the large dashes. The velocity must decrease with radius and the weak form of the constraint that $d^2v/dr^2 \leqslant 0$ has been applied separately in the outer and inner core. The data comprise 25 $\tau$ and five $\Delta$ values. The region contained by the solid lines omits the constraint on the second derivative of v. The PREM model is plotted lightly dashed. PREM does not satisfy the $\tau(p)$ data.

curve that is never concave up as our figures are drawn. We can therefore exclude anything to the left of line segments intersecting the lower bound twice. The upper bound is a bit more complicated: we can exclude anything separated from the rest of the envelope by a line that intersects the upper bound twice and the lower bound once. Figure 7 shows the results of this proscription and is our most constrained solution based upon $\tau(p)$ and $\Delta(p)$ observations. (We have plotted the PREM model for reference; it lies within the bounds but, as Johnson and Lee demonstrated, does not satisfy the $\tau(p)$ data.) Although the correct bounds incorporating the second derivative constraint are undoubtedly narrower, the geometrical consideration allows us to rule out some solutions without solving a nonlinear problem. The same type of argument applies to the first radial derivative constraint: it must be possible to join any point in the corridor to both ends of the envelope with a curve that is everywhere nonincreasing with radius. Applying this principle to the bounds based on only the flat earth constraint gives much wider bounds than does proper use of the radial constraint.

## CONCLUSIONS

We have presented a theory and an algorithm for finding the best possible envelope of velocities in a spherical earth consistent with a finite number of $\tau(p)$ and $\Delta(p)$ observations whose uncertainties are expressed as strict intervals. The solution begins by mapping the sphere into a half-space in which velocity varies only with depth, the equivalent flat earth problem. This mapping has been used before, but we show for the first time that maximizing or minimizing the depth functional in the flat earth always leads to a corresponding extremum in the spherical system, even though a nonlinear transformation of sets has taken place.

We use the linear programming approach of *Garmany et al.* [1979] to construct a corridor in the velocity-depth plane that contains all models consistent with the data and the constraint that there are no low-velocity zones in the flat earth. The original formulation has been improved in several ways: we have added the ability to include $\Delta(p)$ data in the manner of *Orcutt* [1980], and we have shown how to select a set of basis functions in the numerical approximation of the problem so that precise bounds are found with a relatively small number of layers.

The requirement that there be no low-velocity zones in the flat earth leads to the ad hoc restriction in the spherical system that $dv/dr \leqslant v/r$. This inequality permits low-velocity zones of increasing intensity as one approaches the center of the earth. We show by thermodynamic reasoning that $P$ wave velocity in the core should increase with depth so that $dv/dr \leqslant 0$. The constraint remains linear when mapped into the flat earth and so may be readily included in the linear programming formulation. We also show that it is quite likely that $d^2v/dr^2 \leqslant 0$ in the core. This condition can not be mapped into a linear condition in the alternative domain and so it has not been fully exploited; a weak form of the condition can be enforced without any additional computation.

We have tested the theory with $\tau(p)$ data for the core prepared by *Johnson and Lee* [1985] and used by them to constrain the P-wave velocity profile. Our technique produces numerous velocity models that satisfy the finite list of $\tau(p)$ data but lie outside Johnson and Lee's bounds; the LP bounds are much wider. The data alone are insufficient to resolve velocity well, so it is desirable to add information to the inversion by making additional assumptions about the earth. Adding the physical constraint on the velocity gradient tightens the LP bounds considerably, bringing them inside the original corridor of Johnson and Lee. Our results suggest that the $P$ wave velocity can be determined to an accuracy of better than $\pm 0.25$ km s$^{-1}$ almost everywhere in the outer core and $\pm 0.1$ km s$^{-1}$ in a large part of the inner core. We can show by the same kind of geometrical argument used to apply the second derivative constraint that the inner core boundary must lie between 1207 and 1242 km if it is a simple discontinuity.

## APPENDIX A: PHYSICAL RESTRICTIONS ON THE VELOCITY IN THE OUTER CORE

It has been accepted that the bulk of the outer core is very nearly adiabatic and homogeneous since the work of *Birch* [1952]. Free oscillation data do not indicate any significant departures from this state [*Masters*, 1979] and so modern earth models tend to build in these properties. We can constrain the first and second radial derivatives of $V_p(r)$, the compressional velocity as a function of radius, if we assume that the outer core is adiabatic and homogeneous and core material is "normal." (In this appendix only we adopt symbols common in geophysical thermodynamics: $p$ is pressure, not ray parameter, $V_p$ will be used for $P$ wave velocity and $T$ for absolute temperature; we believe this is less confusing than using unfamiliar symbols for these variables.) By normal we mean that $(\partial K_S/\partial p)_S$, the isentropic change in the bulk modulus with pressure, is about 3–4, and decreases slowly with isentropic compression. This assumption is supported by the finite-strain fits to the properties of the outer core

done by *Davies and Dziewonski* [1975], who found that $\partial K_S/\partial p$ decreases from about 3.6 at the core-mantle boundary to about 3.45 at the inner core boundary. Assuming hydrostatic equilibrium,

$$\left(\frac{\partial V_p}{\partial r}\right)_{S,C} = \frac{-g}{2V_p}\left[\left(\frac{\partial K_S}{\partial p}\right)_{S,C} - 1\right] \tag{A1}$$

where the subscripts $S,C$ denote constant entropy and composition and $g$ is the acceleration due to gravity. Since ex hypothesi $(\partial K_S/\partial p)_{S,C} \geq 1$, we find immediately that the first radial derivative of $V_p$ is negative. This would be the first result we need if the core had an adiabatic temperature gradient and uniform composition; we examine the effects of departures from these conditions shortly.

Differentiating (A1) gives

$$\left(\frac{\partial^2 V_p}{\partial r^2}\right)_{S,C} = \left(\frac{\partial V_p}{\partial r}\right)_{S,C}\left[\frac{1}{g}\frac{dg}{dr} - \frac{1}{V_p}\left(\frac{\partial V_p}{\partial r}\right)_{S,C}\right] + \frac{\rho g^2}{2V_p}\left(\frac{\partial^2 K_S}{\partial p^2}\right)_{S,C} \tag{A2}$$

Gravitational acceleration $g(r)$ is relatively insensitive to details of the density distribution within the earth and is very well determined. The derivative $dg/dr$ is almost certainly positive in the outer core so the first term on the right side of (A2) is negative. The second term is much smaller than the first because $(\partial K_S/\partial p)_{S,C}$ is relatively insensitive to $p$ but is negative if the material is normal. Thus the second radial derivative of $V_p$ is also negative given these assumptions.

How sensitive are these constraints to the assumption that the temperature gradient is adiabatic? It is extremely unlikely that the outer core can be significantly superadiabatic as the resulting convective instability would relieve the condition through convection [*Masters*, 1979] resulting in an adiabatic interior with thin (seismically unobservable) boundary layers. To examine the effect of a subadiabatic gradient, we write the temperature gradient in the following form:

$$\frac{dT}{dr} = \left(\frac{\partial T}{\partial r}\right)_{S,C}(1-f) \tag{A3}$$

where $f$ is a function of radius. In the isothermal case $f = 1$ and since we have ruled out super-adiabatic gradients, $0 \leq f \leq 1$. The adiabatic temperature gradient is given by

$$\left(\frac{\partial T}{\partial r}\right)_{S,C} = \frac{-gT\gamma}{V_p^2} = \frac{-gT\alpha}{C_p} \tag{A4}$$

where $\gamma$ is Gruneisen's ratio, $\alpha$ is the coefficient of thermal expansion and $C_p$ is the specific heat. With these relations and the assumption of homogeneity one may show

$$\frac{dV_p}{dr} = \left(\frac{\partial V_p}{\partial r}\right)_{S,C} + f\frac{gT\gamma}{V_p^2}\left(\frac{\partial V_p}{\partial T}\right)_{p,C} \tag{A5}$$

The temperature dependence of $V_p$ is unknown but may be estimated using the parameter $\delta_s$ introduced by *Anderson* [1967]. Like $(\partial K_S/\partial p)_{S,C}$, $\delta_s$ is relatively insensitive to pressure and temperature. In a fluid, $\delta_s$ can be written

$$\delta_s = 1 - \frac{2}{\alpha V_p}\left(\frac{\partial V_p}{\partial T}\right)_{p,C}$$

We conclude that $\delta_s \geq 1$ as $(\partial V_p/\partial T)_{p,C}$ is negative in almost all materials. In fact, experiments usually show that $\delta_s \approx (\partial K_S/\partial p)_{S,C}$ so it is reasonable to suppose that $1 \leq \delta_s \leq 4$ in the outer core. Using (A1), (A5) can be written

$$\frac{dV_p}{dr} = \left(\frac{\partial V_p}{\partial r}\right)_{S,C}\left[1 + \alpha T\gamma f X\right] \tag{A6}$$

where

$$X = \frac{\delta_s - 1}{\left(\frac{\partial K_S}{\partial p}\right)_{S,C} - 1} \approx 1$$

i.e., both $f$ and $X$ probably lie between zero and one. Equation (A6) says that $dV_p/dr$ will be slightly more negative in a subadiabatic region. Reasonable estimates of $\alpha, T$, and $\gamma$ [e.g., *Stacey*, 1977] suggest that $\alpha T\gamma$ is about 0.05 in the outer core, so the velocity gradient would be changed by about 5% in an isothermal region.

Differentiating (A6) with respect to radius yields

$$\frac{d^2V_p}{dr^2} = \left(\frac{\partial^2 V_p}{\partial r^2}\right)_{S,C}\left[1 + \alpha T\gamma f X\right] + \left(\frac{\partial V_p}{\partial r}\right)_{S,C}\frac{d}{dr}(\alpha T\gamma f X) \tag{A7}$$

The first term on the right of (A7) is negative, but the second term could cancel it, giving a nonnegative value of $d^2V_p/dr^2$; we can show that this is unlikely.

The only difficulty is the unknown radial variation of $f$. The radial variation of $\alpha T\gamma X$ is dominated by the behavior of $\alpha$, a rapidly decreasing function of pressure. Therefore $d(\alpha T\gamma X)/dr$ is positive. Equation (A7) shows that if $f$ is constant or $df/dr$ is positive (i.e., the core is increasingly stable at larger radius), the second radial derivative of $V_p$ remains negative. It may only become nonnegative if $df/dr$ is large and negative: $d^2V_p/dr^2 \approx 0$ when

$$\frac{df}{dr} \approx -\frac{1}{\alpha T\gamma X}\left[\frac{1}{g}\frac{dg}{dr} + \frac{g}{2V_p^2}\left[\left(\frac{\partial K_S}{\partial p}\right)_{S,C} - 1\right]\right]$$

$$\approx -10^{-5}\ \mathrm{m}^{-1}$$

For this to occur the core must go from an adiabatic state to an isothermal state in a radial distance of about 100 km.

In summary, subadiabatic temperature gradients cause the velocity gradient to steepen slightly (becoming more negative) and leave the second radial derivative of velocity negative provided the core is uniformly stable or becomes more stable toward the core-mantle boundary. The second radial derivative could become nonnegative if the core becomes stable within a very small range of radius as one approaches the inner core boundary, but we consider this case unlikely.

What is the effect of variations in the composition? We make the usual assumption that the outer core is predominantly molten iron with small amounts of light impurities. This model is based partly on shock wave experiments which indicate that the outer core is about 10% less dense than pure iron [e.g., *Jeanloz*, 1979]. The same experiments show that the bulk sound speeds (which equal $V_p$ in a fluid) of the outer core and of iron at core pressures are virtually indistinguishable. It seems that the effect of impurities is to lower $K_S$ as much as they lower the density, resulting in little effect on the velocity. In this case, reasonable radial variations in the composition of the outer core would not affect the negativity of the first and second radial derivatives of velocity. One can use the analysis of *Jeanloz* [1979] to estimate $(\partial V_p/\partial c)_{T,P}$, where $c$ is the concentration of light impurities, and to make an analogous argument to the one for thermal variations given above. The results are similar, and we conclude that the constraints on the radial derivatives are unlikely to be violated for reasonable chemical or thermal departures from the adiabatic and homogeneous state.

## APPENDIX B: EQUIVALENCE OF TWO OPTIMIZATION PROBLEMS

It is not difficult to show that the extrema of $z(v)$ occur for the same models $dz/dv$ as the extrema of $r(v)$. We do this by showing that the variations of $r(v)$ and $z(v)$ differ by a multiplicative constant, so that the perturbations to $dz/dv$ one should make to improve the values of the penalty functionals have the same direction. Since the linear programming solution for $z(v)$ is optimal, that is the functional derivative projected onto a given direction either vanishes or leads outside the constraint set, $r(v)$ can not be improved locally without leaving the region of $dz/dv$ that fit the data. We write $r$ as a functional of $\zeta \equiv dz/dv$:

$$r[\zeta] \equiv a e^{-Z_1[\zeta]/a}$$

where

$$Z_1[\zeta] \equiv \int_w^v \zeta\, dv \quad \text{with} \quad v \equiv vr[\zeta]/a$$

Let $\Delta \equiv Z_1[\zeta+\delta] - Z_1[\zeta]$. Then

$$\begin{aligned} r[\zeta+\delta]-r[\zeta] &= a\left(e^{-(Z_1[\zeta]+\Delta)/a} - e^{-Z_1[\zeta]/a}\right) \\ &= a\sum_{n=0}^{\infty}\frac{(-a)^{-n}}{n!}\left((Z_1[\zeta]+\Delta)^n - Z_1[\zeta]^n\right) \\ &= a\sum_{n=0}^{\infty}\frac{(-a)^{-n}}{n!}\left(Z_1[\zeta]^n + nZ_1[\zeta]^{n-1}\Delta - Z_1[\zeta]^n\right) \\ &= -\Delta\sum_{n=1}^{\infty}\frac{(-a)^{-(n-1)}}{(n-1)!}Z_1[\zeta]^{n-1} \\ &= -\Delta e^{-Z_1[\zeta]/a} = -\Delta\, r[\zeta]/a \end{aligned} \tag{B1}$$

for $|\Delta| << |Z_1[\zeta]|$. Now,

$$\Delta = \int_w^{vr[\zeta+\delta]/a}(\zeta+\delta)\,dv - \int_w^{vr[\zeta]/a}\zeta\,dv$$

$$= \int_w^{vr[\zeta]/a}\delta\,dv + \int_{vr[\zeta]/a}^{vr[\zeta+\delta]/a}\zeta\,dv + \int_{vr[\zeta]/a}^{vr[\zeta+\delta]/a}\delta\,dv$$

Using (B1) and the definition of $v$,

$$\Delta \approx \int_w^v \delta\,dv + \int_v^{v(1-\Delta/a)}\zeta\,dv + \int_v^{v(1-\Delta/a)}\delta\,dv$$

The third term is second order in $\delta$, and the second term may be approximated by $-\zeta(v)\Delta/a$ for $\zeta$ sufficiently smooth and $\Delta/a$ sufficiently small. Note that for the finite-dimensional realization of the problem that we solve, this requires that the perturbation be sufficiently small not to cause the upper limit of integration to move into the support of a different basis function, or that the coefficients of the two match to give a continuous transition. To first order,

$$\Delta \approx \int_w^v \delta\,dv - \zeta(v)\,\Delta/a$$

If we solve for $\Delta$ and substitute the result into equation (B1), we find that

$$r[\zeta+\delta] - r[\zeta] \approx \frac{-r[\zeta]}{a+\zeta(v)}\int_w^v \delta\,dv$$

However,

$$Z_1[\zeta+\delta] - Z_1[\zeta] = \int_w^v \delta\,dv$$

so we may write

$$r[\zeta+\delta] - r[\zeta] \approx \frac{-r[\zeta]}{a+\zeta(v)}\left(Z_1[\zeta+\delta] - Z_1[\zeta]\right)$$

The negative sign is expected since maximizing the depth minimizes the radius. This shows that any extremum of $Z_1$ over the set of $\zeta$ satisfying the constraints is also an extremum of $r$. We must still show that $r$ does not achieve a "better" value for any other $\zeta$ satisfying the constraints.

The functional $r[\zeta]$ is convex since it is the composition of two convex mappings: a linear mapping ($Z_1[\zeta]$) and an exponential. The set of models $\zeta(v)$ that satisfy the data and the radial derivative constraint, if it is used, is convex because it is described by linear inequalities. The familiar theorem concerning the extrema of convex functionals over convex sets applies: the value achieved at a local extremum of $r[\zeta]$ satisfying the data is the global extremal value with respect to the set of feasible solutions. We are therefore justified in solving the problem by finding the extrema of $z(v)$ and mapping the resulting values into values of $r(v)$.

*Acknowledgments.* This work was supported by National Science Foundation Grants EAR-84-03847 and EAR-84-18471, and by the Office of Naval Research Contract N00014-85-C-0140.

## REFERENCES

Aki, K., and P. G. Richards, *Quantitative Seismology: Theory and Methods*. W. H. Freeman, San Fransisco, Calif., 1980.

Anderson, O. L., Equation for thermal expansivity in planetary interiors, *J. Geophys. Res.*, *72*, 3661–3668, 1967.

Bessonova, E. N., V. M. Fishman, V. Z. Ryaboyi, and G. A. Sitnikova, The tau method for inversion of travel times, I, Deep seismic sounding data, *Geophys. J. R. Astron. Soc., 36,* 377–398, 1974.

Bessonova, E. N., V. M. Fishman, M. G. Shnirman, G. A. Sitnikova, and L. R. Johnson, The tau method for inversion of travel times, II, Earthquake data, *Geophys. J. R. Astron. Soc., 46,* 87–108, 1976.

Birch, F., Elasticity and constitution of the earth's interior, *J. Geophys. Res., 57,* 227–286, 1952.

Davies, G. F., and A. M. Dziewonski, Homogeneity and the constitution of the Earth's lower mantle and outer core, *Phys. Earth Planet. Inter., 10,* 336–343, 1975.

Dziewonski, A. M., and D. L. Anderson, Preliminary reference Earth model, *Phys. Earth Planet. Inter., 25,* 297–356, 1981.

Garmany, J., J. A. Orcutt, and R. L. Parker, Travel time inversion: A geometrical approach, *J. Geophys. Res., 84,* 3615–3622, 1979.

Gerver, M., and V. Markushevitch, Determination of a seismic wave velocity from the travel-time curve, *Geophys. J. R. Astron. Soc., 11,* 165–173, 1966.

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

Jeanloz, R., Properties of iron at high pressures and the state of the core, *J. Geophys. Res., 84,* 6059–6069, 1979.

Johnson, L. R., and R. C. Lee, Extremal bounds on the *P* velocity in the earth's core, *Bull. Seismol. Soc. Am., 75,* 115–130, 1985.

Luenberger, D. G., *Introduction to Linear and Nonlinear Programming,* Addison-Wesley, Reading, Mass., 1973.

Masters, G., Observational constraints on the chemical and thermal structure of the earth's deep interior, *Geophys. J. R. Astron. Soc., 57,* 507–534, 1979.

Orcutt, J. A., Joint linear, extremal inversion of seismic kinematic data, *J. Geophys. Res., 85,* 2649–2660, 1980.

Stacey, F. D., A thermal model of the earth, *Phys. Earth Planet. Inter., 15,* 341–348, 1977.

Wiggins, R. A., G. A. McMechan, and M. N. Toksoz, Range of earth structure nonuniqueness implied by body wave observations, *Rev. Geophys., 11,* 87–113, 1973.

G. Masters, J. A. Orcutt, R. L. Parker, and P. B. Stark, Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093.

(Received January 17, 1986;
revised May 5, 1986;
accepted August 6, 1986.)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1 REPORT NUMBER | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4 TITLE (and Subtitle)<br>Strict Bounds on Seismic Velocity in the Spherical Earth | | 5. TYPE OF REPORT & PERIOD COVERED<br>6. PERFORMING ORG. REPORT NUMBER |
| 7 AUTHOR(s)<br>Philip B. Stark, Robert L. Parker, G. Masters, and John A. Orcutt | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-85-C-0140 |
| 9 PERFORMING ORGANIZATION NAME AND ADDRESS<br>Scripps Institution of Oceanography<br>La Jolla, California 92093 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research<br>Arlington, Virginia 22217 | | 12. REPORT DATE<br>December 10, 1986<br>13. NUMBER OF PAGES<br>12 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>unclassified<br>15a. DECLASSIFICATION DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18 SUPPLEMENTARY NOTES

Reprint from Journal of Geophysical Research, Vol. 91, No. B14, 13,892-13,902, (1986)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20 ABSTRACT (Continue on reverse side if necessary and identify by block number)

We address the inverse problem of finding the smallest envelope containing all velocity profiles consistent with a finite set of imprecise $\tau(\rho)$ data from a spherical earth. Traditionally, the problem has been attacked after mapping the data relations into relations on an equivalent flat earth. Of the two contemporary direct methods for finding bounds on velocities in the flat earth consistent with uncertain $\tau(\rho)$ data, a nonlinear (NL) approach descended from the Herglotz-Wiechert inversion and a linear programming (LP) approach, only NL has been used to solve the spherical earth problem. On the basis of the finite. . .

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102 LF 014 6601

unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 92, NO. B3, PAGES 2713-2719, MARCH 10, 1987

# Velocity Bounds from Statistical Estimates of $\tau(p)$ and $X(p)$

PHILIP B. STARK AND ROBERT L. PARKER

*Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego*

We present a new technique for constructing the narrowest corridor containing all velocity profiles consistent with a finite collection of $\tau(p)$ data and their statistical uncertainties. Earlier methods for constructing such bounds treat the confidence interval for each $\tau$ datum as a strict interval within which the true value might lie with equal probability, but this interpretation is incompatible with the estimation procedure used on the original travel time observations. The new approach, based upon quadratic programming (QP), shares the advantages of the linear programming (LP) solution: it can invert $\tau(p)$ and $X(p)$ data concurrently; it permits the incorporation of constraints on the radial derivative of velocity for spherical earth models; and theoretical results about convergence and optimality can be obtained for the method. We compare $P$ velocity bounds for the core obtained by QP and LP. The models produced by LP predict data values at the ends of the confidence intervals; these values are unlikely according to the proper statistical distribution of errors. For this reason the LP velocity bounds can be wider than those given by QP, which takes better account of the statistics. Sometimes, however, the LP bounds are more restrictive because LP never permits the predictions of the models to lie outside the confidence intervals even though occasional excursions are expected. The QP bounds grow narrower at lower levels of confidence, but the corridors at 95% and 99.9% are virtually indistinguishable. The data must be improved substantially to make a significant change in the velocity bounds.

## INTRODUCTION

This paper is a sequel to *Stark et al.* [1986], (hereinafter called SPMO); we assume the reader is familiar with their notation and results. Both papers address the nonlinear inverse problem of ray theoretic seismology on a one-dimensional earth. The earliest approach to the problem, that of Wiechert and Herglotz in the 1900s [*Aki and Richards*, 1980], assumes that an exact travel time curve is available and that the earth does not contain strong low-velocity zones (regions where $dv(r)/dr > v/r$, where $v$ is seismic velocity as a function of $r$, radius). With these assumptions, there is a unique velocity model corresponding to the data. Once stronger low-velocity zones are permitted, many models may satisfy the data [for example, see *Gerver and Markushevich*, 1966]. Even without strong low-velocity zones, usually infinitely many earth models satisfy the available discrete imprecise travel time observations, which do not constitute exact travel time curves. The nonuniqueness introduced by the finite number of data and their contamination by errors is traditionally addressed by trying to delimit the range of models that fit the data adequately.

For reasons stated by SPMO, it is convenient to work with $\tau(p)$, the vertical delay time as a function of ray parameter, rather than $T(X)$, travel time as a function of epicentral distance. *Bessonova et al.* [1976] introduced a method of estimating sample means of $\tau$ and their standard deviations at discrete values of $p$, assuming that the noise contaminating the travel time observations is random and uncorrelated and has zero mean. On these assumptions the sample mean is approximately Student $t$ distributed when $p$ changes very little over the bands of $X$ used in the estimation of $\tau$. The assumption of independent zero-mean random noise is probably not valid for a variety of reasons: Travel times are biased by near-source and near-receiver anomalies and by large-scale heterogeneities within the earth; picking errors are likely to be systematic; and at some point the assumptions of ray theory break down. All these factors tend to correlate travel time measurement errors, some tend to bias the measurements, and without more information one should hesitate to assert that the errors are truly Gaussian. However, the approximation is increasingly reasonable when there are many observations with wide geographic distribution and it enables one to make progress on the problem.

The nonlinear inversion scheme of *Bessonova et al.* [1976] and the linear programming (LP) method of *Garmany et al.* [1979], discussed at length by SPMO, take confidence intervals derived from the means and standard deviations as strict limits within which $\tau$ must lie. *Dorman and Jacobson* [1981] objected that velocity bounds based on this reinterpretation of the statistical data will be erroneous. Some fraction of the time the confidence intervals will not include the $\tau(p)$ values of the real earth; the strict reinterpretation does not allow for this, and so the resulting envelope of models may be too narrow. On the other hand, the models that determine points on the velocity-depth bounds tend to predict values of $\tau(p)$ that lie at the ends of most of the confidence intervals. Since there is really a probability distribution of values within the confidence intervals and the values at the ends are less likely to come from the earth, a model that predicts $\tau(p)$ values consistently at the ends is extremely unlikely to represent the earth and the envelope may well be too wide.

We shall show that it is not necessary to reinterpret the statistical estimates as strict limits on $\tau$: We propose a method, dubbed QP (for quadratic programming), that finds velocity-depth bounds from estimates of the mean values of $\tau(p)$ and $X(p)$ at various $p$ and their standard deviations. QP retains the advantages of LP: interpolation of the data is not necessary, multivalued velocity models are automatically excluded from the inversion, $X(p)$ data



Paper number 6B6146
0148-0227/87/006B-6146$05.00

can be used concurrently with $\tau(p)$, and the radial derivative of velocity can be constrained. We show below that the proof of the equivalence of the spherical and flat earth inverse problems given by SPMO also applies to the inverse problems from statistical data.

For convenience we shall work in the flat earth, following SPMO as closely as possible in our notation. In the application to the earth's core using the $\tau(p)$ data of *Johnson and Lee* [1985], we use the basis functions (pieces of $1/v$) and the radial derivative constraints proposed by SPMO.

## THE QP METHOD

The problems of calculating $\tau(p)$ and $X(p)$ are nonlinear with the customary representation of the earth by $v(z)$, velocity as a function of depth. However, *Garmany et al.* [1979] noted that $\tau(p)$ and $X(p)$ are linear functionals of a one-dimensional flat earth model expressed as the derivative of depth with respect to velocity, $dz/dv$, which we call $\zeta = \zeta(v)$:

$$T_i[\zeta] \equiv \tau(p_i) = 2\int_w^{1/p_i} (v^{-2} - p_i^2)^{-1/2}\,\zeta(v)\,dv$$

$$X_i[\zeta] \equiv X(p_i) = 2\int_w^{1/p_i} p_i\,(v^{-2} - p_i^2)^{-1/2}\,\zeta(v)\,dv$$

where $w$ is the surface velocity. The depth to a fixed target velocity $v_t$ is also a linear functional of the unknown earth model $\zeta$:

$$Z_{v_t}[\zeta] \equiv z(v_t) = \int_w^{v_t} \zeta(v)\,dv$$

Let $y$ denote the reciprocal of the smallest $p_i$ in the data set. Then $y$ is the largest velocity about which the data give us any information, so we will take $\zeta(v)$ to be defined on the interval $[w, y]$. We must insist that $\zeta(v) \geq 0$ to ensure $\zeta$ corresponds to a single-valued velocity model $v(z)$. We may exclude flat earth profiles that correspond to spherical earth profiles with low-velocity zones by requiring $\zeta(v) \leq a/v$, where $a$ is the radius of the spherical earth. SPMO derive this expression for the constraint and justify its use in inversions for core structure.

Our data are sample means $\bar{d}_i$ and their standard deviations $\sigma_i$ of $\tau(p_i)$ for $i = 1,\ldots,n_\tau$ and of $X(p_i)$ for $i = n_\tau+1,\ldots,n$. The measure of misfit to the data we shall use is

$$\mu[\zeta] \equiv \sum_{i=1}^{n_\tau}\left(\frac{\bar{d}_i - T_i[\zeta]}{\sigma_i}\right)^2 + \sum_{i=n_\tau+1}^{n}\left(\frac{\bar{d}_i - X_i[\zeta]}{\sigma_i}\right)^2$$

We will say that a model $\zeta(v)$ fits the data adequateiy if $\mu[\zeta] \leq M^2$, where $M^2$ is some chosen tolerance. We can estimate the probability that the actual $\tau(p)$ and $X(p)$ predictions of the real earth fit the sample means within $M^2$. Let us assume following *Bessonova et al.* [1976] that errors in the travel times are independent and normally distributed and have zero mean; then the sample means are approximately Student t distributed. We will assume further that the estimates of $\bar{d}_i$ have a large number of degrees of freedom and that we have more than a few data ($n >> 1$). Then the weighted misfit of the real earth's predictions to the sample means is approximately $\chi^2_n$ distributed (chi-square with $n$ degrees of freedom). With $M^2$ equal to the $1-\alpha$ percentage point of the $\chi^2_n$ distribution, the requirement that $\mu[\zeta] \leq M^2$ limits our search to a set of models whose predictions include those of the real earth at the $1-\alpha$ confidence level. We shall see later that the precise value of $M^2$ makes very little difference in the bounds we find.

The problem of finding velocity bounds is nonlinear for a host of reasons: First, a nonlinear transformation from the spherical to the flat earth has taken place; second, the data mappings as usually written are nonlinear; third, the measure of misfit to the data is quadratic; and finally, linear inequality constraints are required to ensure that the models are physically reasonable and that the radial derivative constraint is not violated. In general, problems of this kind are not soluble. Here, we show how a finite-dimensional approximation to the problem can be solved and give a numerically stable algorithm to solve it; P. B. Stark (Rigorous velocity bounds from soft $\tau(p)$ and $X(p)$ data, submitted to *Geophysical Journal of the Royal Astronomical Society*, 1986 (hereinafter referred to as Stark (1986)) has demonstrated rigorously that the finite-dimensional approximation converges to the optimal result.

Denote by $U$ the set of flat earth models $\zeta = \zeta(v)$ that satisfy $\mu[\zeta] \leq M^2$, the positivity constraint $\zeta(v) \geq 0$ that ensures that the models are physically realizable and, if we choose to impose it, the radial derivative constraint $\zeta(v) \leq a/v$. We construct velocity bounds by finding the maximum and minimum depths at which a target velocity $v_t$ can occur among the models in $U$. We repeat the procedure with different target velocities until we have a good description of the envelope of acceptable models. This is the same approach used by LP and described by SPMO. The misfit functional $\mu[\zeta]$ is a positive semidefinite quadratic form and thus $\mu[\zeta] \leq M^2$ defines a convex set of models $\zeta$. (This does not mean that any particular model $\zeta(v)$ is a convex function: the set of models satisfying the constraint is a convex set in the space of models from which $\zeta$ is drawn. See *Luenberger* [1969] about convex sets of functions.) The positivity constraint $\zeta(v) \geq 0$ and the radial derivative constraint $\zeta(v) \leq a/v$ are both linear inequality constraints; hence they too describe convex sets. $U$, the intersection of these three sets (the set of models that fit the data adequately, represent single-valued velocity models and have a nonpositive radial derivative in the spherical earth) is also a convex set. The depth to the target velocity $v_t$, $Z_{v_t}[\zeta]$, is a linear functional of the model $\zeta$; linear functionals are convex. Our task is to find the extrema of the convex functional $Z_{v_t}[\zeta]$ over the convex set of models $U$. The familiar theorem that local extrema achieve the global extremal values applies and we conclude that the minimum and maximum depths are unique.

We now know enough to establish the equivalence of the spherical and flat earth inverse problems using statistical data. SPMO proved the equivalence for the strict data problem; their proof relies upon the first-order equivalence of the functional derivatives of radius and

TABLE 1. $X(p)$ Data Used in Some of the Inversions, Abstracted From the $X(p)$ Observations of *Johnson and Lee* [1985]

| $p$, s km$^{-1}$ | $X$, km | $\sigma_1$, km | $\sigma_2$, km |
|---|---|---|---|
| 0.02922 | 7273.4 | 380.216 | 268.8 |
| 0.03013 | 6985.0 | 380.216 | 219.5 |
| 0.03095 | 5953.9 | 380.216 | 143.7 |
| 0.03265 | 6661.8 | 380.216 | 219.5 |
| 0.03347 | 6843.3 | 380.216 | 190.1 |

See Figure 2. The values of ray parameter are given as the flat earth values reduced to the surface of the core (3480 km). The $X(p)$ sample means, labeled $X$, are values reduced to the surface of the core by subtracting the predictions of the PREM anisotropic earth model [*Dziewonski and Anderson*, 1981]. The standard deviations $\sigma_1$ are referred to in the text as the weaker $X(p)$ constraints; $\sigma_2$ are the tighter $X(p)$ data. The values of $\sigma_1$ correspond to the confidence intervals for $X(p)$ used by *Stark et al.* [1986].

depth with respect to $\zeta$ and upon the convexity of the set of solutions that satisfy the strict data, the positivity constraint, and the radial derivative constraint. The depth functional $Z_{v_t}[\zeta]$ and the radius functional are the same for the statistical data problem as for the strict data problem because we are using the same representation of the earth, $\zeta(v) = dz/dv$. We have just seen that the set $U$ of models that satisfy the statistical data within $M^2$ and the positivity and radial derivative constraints is convex, so the proof given by SPMO applies to this problem as well.

How may we find the minimum and maximum depths to $v_t$? A slightly different perspective makes the job fairly straightforward, although it is rather expensive computationally. We shall look at the models that fit the data best subject to the additional linear constraint that they reach a certain depth at the target velocity $v_t$: The penalty functional will be the misfit to the data, not the depth to the target velocity. All the constraints are then linear and the only nonlinearity is in the new penalty functional, which is quadratic in the unknown model $\zeta$.

Consider the earth model $\zeta^*(v)$, where $0 \leqslant \zeta^*(v) \leqslant a/v$, that has the smallest $\mu[\zeta]$, the weighted misfit to the data means $\bar{d}_j$. If we approximate the problem in finite dimensions by writing $\zeta$ as a linear combination of a finite set of basis functions, then an approximation to $\zeta^*$ can be found by quadratic programming with linear inequality constraints. Stark (1986) proves that for any reasonable choice of basis functions, the results obtained by increasing the number of basis functions used in the finite-dimensional approximation converge to the correct answer for the infinite-dimensional problem. The inverse problem is consistent using that set of basis functions provided $\mu[\zeta^*] \leqslant M^2$. The best fitting finite-dimensional model $\zeta^*$ associates the depth $z^* = Z_{v_t}[\zeta^*]$ with the target velocity $v_t$. Provided the problem is consistent, $z^*$ is an upper bound on the least depth to $v_t$ and a lower bound on the greatest depth to $v_t$.

We prove in Appendix A that if we add the linear constraint that the model reach a greater depth than $z^*$ at $v_t$, the best fitting model that we can then find will have a larger misfit $\mu[\zeta]$. If we make the model attain a still greater depth, the misfit will continue to grow. The same thing happens if we require the model to have smaller and smaller depths than $z^*$. Since the constraint that the model arrive at a certain depth at $v_t$ is linear ($Z_{v_t}[\zeta]$ is a linear functional), finding the best fitting model that satisfies the positivity and radial derivative constraints and that achieves a certain depth at $v_t$ is another quadratic programming problem with linear inequality constraints.

The statistical data problem thus can be solved with a single-parameter search: starting with $z^*$, the depth to $v_t$ achieved by the overall best fitting model, we add the constraint that the model attain a slightly larger depth and find the best fitting model; we continue increasing the depth until the best fitting model subject to the constraint has a misfit larger than $M^2$. The depth at which $M^2$ is first exceeded is the maximum depth to $v_t$ in that discretization. Similarly, by decreasing the depth until $M^2$ is passed, we may find the least depth to $v_t$ in the discretization. The monotonicity of the misfit with changes in the depth constraint lets us stop the search as soon as $M^2$ is overrun: the misfit will not fall again.

## APPLICATION TO THE EARTH'S CORE

We implemented QP as described in Appendix B on the National Science Foundation San Diego Supercomputer Center Cray X-MP/48. The computational requirements of the algorithm are fairly heavy: each set of bounds presented required about 30 minutes of central processor time. The inversions that follow use the means and standard deviations of the 25 uninterpolated $\tau(p)$ data tabulated by *Johnson and Lee* [1985]. All our inversions employ the radial derivative constraints and $1/v$ basis functions advocated by SPMO. SPMO used an extra pair of basis functions bracketing the target velocity to enhance the numerical efficiency of LP (Stark (1986) provides a theoretical explanation of this effect); we have followed their practise. We used 100 basis functions in the preliminary expansion and started the inversions at a radius of 3480 km with a minimum surface velocity $w = 7$ km s$^{-1}$, as they did. The smallest number of degrees of freedom in the $\tau(p)$ estimates is 115, and many estimates were derived from thousands of observations, so approximating the distributions of the sample means by Gaussians is reasonable. The approximation is improved further by the summation over the 25 data. The five $X(p)$ means and uncertainties that we use to refine the structure near the inner core boundary in some of the inversions that follow are tabulated in Table 1. We somewhat arbitrarily ascribed two sets of uncertainties to the $X(p)$ observations of *Johnson and Lee* [1985] because the data, from an array study in central Arizona, are relatively few in number and may not be very representative of the spherically averaged earth.

When we refer to statistical bounds at the 99.9% or 95% confidence level, we mean that we have set $M^2$ equal to the appropriate percentage point of the $\chi^2_n$ distribution. Equating $n$ with the number of data is appropriate because we are not estimating a model, nor indeed reducing the number of degrees of freedom at all; we are estimating a bound on a property of the set of models satisfying the data and additional constraints. We used values of $\chi^2_n$ from *Abramowitz and Stegun* [1965]; they range from about 37 to about 60 for 25 and 30 degrees of freedom. A glance at Figure 1, a representative plot of the minimum misfit to the data as a function of the depth that the model

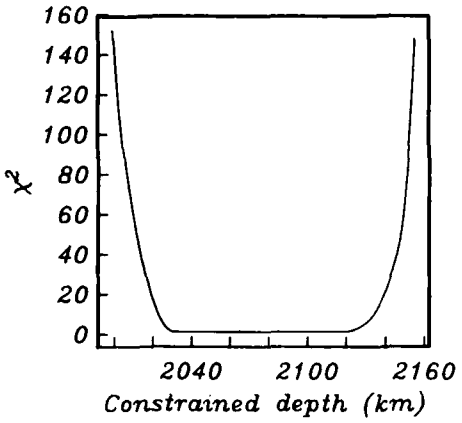



Fig. 1. Representative plot of $\chi^2$ versus depth the model is constrained to achieve, from the inversion of the 25 $\tau(p)$ data with target velocity $v_t = 18$ km $^{-1}$. The minimum misfit of 0.8611 occurs at 2049 km.

is made to attain, shows that the misfit passes rapidly through that span with only a slight change in depth so we do not expect the bounds at those confidence levels to be very different. As we proved, the smallest misfit is monotonic in the depth that the model is constrained to reach. The best fitting model $\zeta^*(v)$ has a misfit of about 0.86 and reaches a depth of 2049 km at the flat earth velocity 18 km $s^{-1}$; it is, however, built unattractively from steps in the spherical earth velocity. It is usually true that models with the smallest misfit to the data are disenchantingly rough. A common alternative to finding bounds on the set of models that satisfy the data, as we do here, is to seek the smoothest model that fits the observations adequately (see *Constable et al.* [1987], for example). Either approach discourages us from attaching too much significance to accidental properties of a particular model.

Figure 2 compares QP and LP bounds based on the 25 data for $\tau(p)$ alone. The dotted bounds are those obtained by SPMO using LP with radial derivative constraints. The solid lines are QP bounds at the 99.9% confidence level ($M^2 = 52.6$) and, as we might have predicted from the steep misfit functional, they are indistinguishable from the QP bounds at the 95% confidence level ($M^2 = 37.7$) in a diagram of this small size. The similarity of these bounds supports our use of the approximation that the individual data errors sum to a $\chi^2_n$ distribution since the results are not sensitive to the precise tolerance $M^2$ we choose. In places the QP bounds lie outside the LP bounds because QP allows the misfit to occur in the most advantageous place while LP limits the misfit at each $p$ independently. In other places the LP bounds are outside; this is because the models produced by LP tend to have data predictions along the ends of the confidence intervals. QP will not permit this since such predictions are jointly extremely improbable. As an interesting note, geometrical constructions of the kind SPMO used with the strict LP bounds can not be applied to the statistical bounds: there is no reason to expect the bounds themselves to meet the same physical constraints as the models in the statistical problem. In the strict problem one knows that bounds using the radial derivative constraint themselves satisfy the radial derivative constraint; here that is clearly false. The slight narrowing within the inner core where the bounds violate the radial derivative constraint is due to the presence of a datum with particularly small standard deviation.

We have used the five looser $X(p)$ data of Table 1 to try to reduce the width of the envelope near the inner core boundary. Figure 3 compares the results at the 99.9% confidence level ($M^2 = 59.7$, solid line) with the LP solution (dashes) using the strict bounds on $X(p)$ given by SPMO. The QP results using the five weaker $X(p)$ constraints are essentially identical to the results in Figure 2 using only the 25 $\tau(p)$ data. The five weaker $X(p)$ constraints make a significant difference to the strict LP bounds near the inner core, but they are too loose to affect the QP results. We assigned more optimistic estimates of the standard deviations of the $X(p)$ observations (Table 1) to improve the bounds and inverted again. Fig-

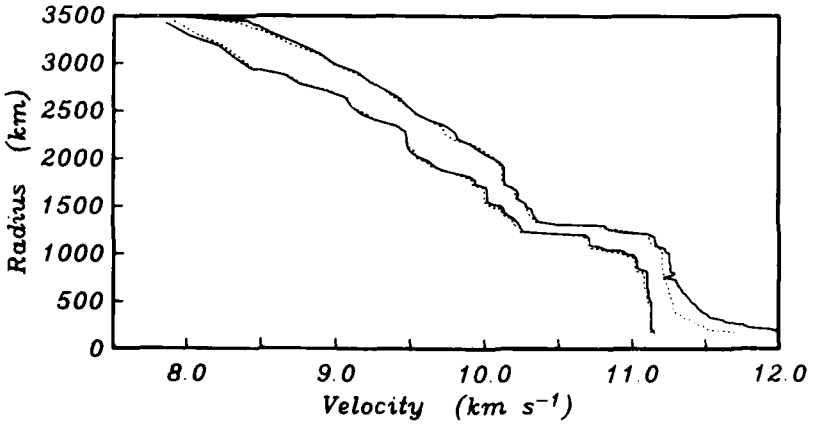



Fig. 2. Bounds based on the 25 $\tau(p)$ data. The solid lines are the statistical bounds at the 95% confidence level; the bounds at the 99.9% confidence level are indistinguishable at this scale. The dotted lines are the strict bounds obtained by SPMO from the 99.9% confidence intervals of *Johnson and Lee* [1985]. The two statistical bounds are so similar because $\chi^2$ changes so abruptly with depth (Figure 1). Also note that while the models are constrained to be monotonic, the statistical bounds need not be monotonic although the strict bounds must.

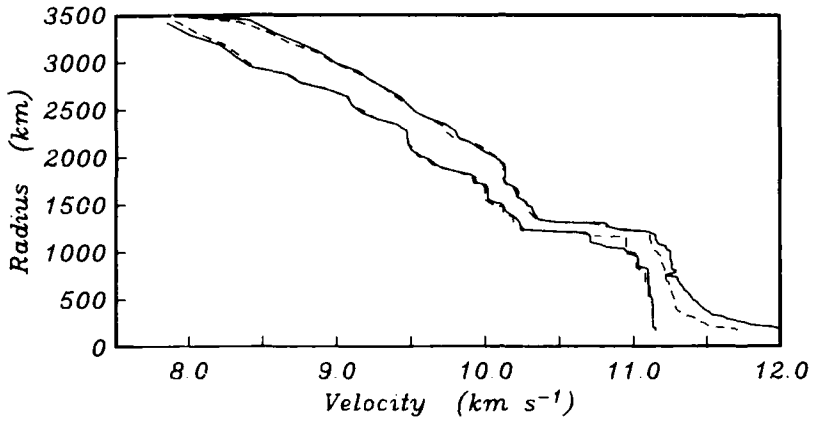



Fig. 3. Statistical bounds at the 99.9% confidence level using the 25 $\tau(p)$ data alone and using in addition the five loose $X(p)$ estimates from SPMO using *Johnson and Lee's* [1985] data. On this scale the results are not distinguishable; both are represented by the solid lines. The dashed bounds are the strict bounds from SPMO using both the $\tau(p)$ and $X(p)$ data. The errors assigned to the estimates of $X(p)$ are too large to change the statistical bounds, although they have a significant effect on the strict bounds near the inner core boundary.

ure 4 plots the results at the 95% confidence level ($M^2 = 43.8$, dashes). The solid lines represent both the results from the 25 $\tau(p)$ data alone and using the weak $X(p)$ constraints. The tighter $X(p)$ data narrow the bounds particularly near the inner core boundary but also generally throughout the core. We tried to invert the corresponding strict data bounds with LP, but the revised data were inconsistent even using 200 basis functions; doubtless this is why they have a strong effect on the QP inversion. (This illustrates, however, that QP is less sensitive than LP to the estimation of data errors.) The smaller $X(p)$ error estimates are probably too optimistic especially with reference to spherically averaged earth structure; we therefore prefer the velocity bounds based on the 25 $\tau(p)$ and five weaker $X(p)$ data (solid lines).

The dotted line in the middle of the bounds is the PREM model of *Dziewonski and Anderson* [1981]. Although PREM lies inside the corridor, its weighted misfit to the sample means of the 25 $\tau(p)$ data is immense. This demonstrates that an arbitrary model within the corridor will not necessarily fit the data. Every velocity-depth point on or within the bounds is consistent with the data: each is contained in some model that fits the data. However, many models in the corridor are invalid. Within the finite-dimensional approximation, the data rule out every point outside the corridor; as the approximation improves, the bounds move slightly outward. We tried unsuccessfully to bring the predictions of PREM into agreement with a $\chi^2$ measure of misfit to the data at the 99.9% level with a baseline shift: there is still some inconsistency between short- and long-period seismic data.

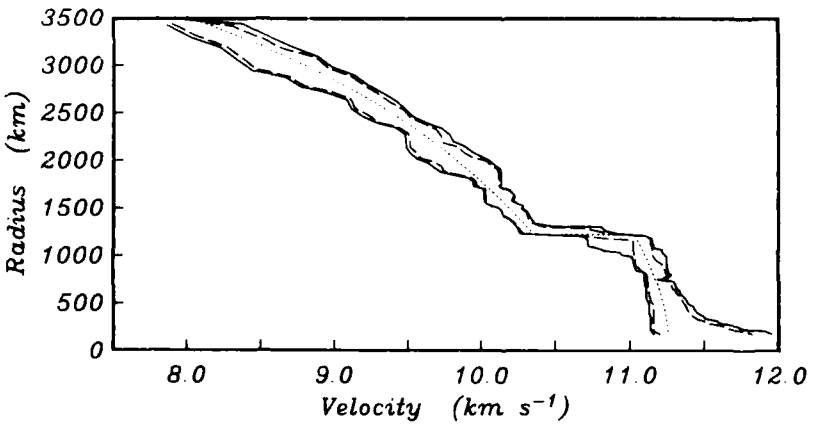



Fig. 4. Statistical bounds at the 95% confidence level. The solid lines are the pair of bounds found using the $\tau(p)$ data alone and using the five weaker $X(p)$ constraints, as in Figure 2. The dashed bounds use instead the five tighter $X(p)$ estimates in Table 1. Note the profound effect near the inner core boundary and the slight general narrowing. In contrast, changes to the $X(p)$ data affect the strict bounds only near the inner core boundary. The PREM earth model (Dziewonski and Anderson, 1981) appears as the dotted line for reference. PREM does not satisfy the $\tau(p)$ data, even at the 99.9% confidence level.

## DISCUSSION

The QP method allows statistical estimates of $\tau(p)$ to be used in a manner that is more consistent with their derivation than the previous methods of *Bessonova et al.* [1976] and *Garmany et al.* [1979]. On the basis of 25 core $\tau(p)$ data from *Johnson and Lee* [1985], QP inversions at the 99.9 and 95% confidence levels find a corridor of velocity models about 0.25 km $s^{-1}$ wide and limit the location of the inner core boundary to approximately 1227—1290 km. The velocity jump at the inner core boundary is about 0.4—0.8 km $s^{-1}$. QP is less sensitive than LP to the estimation of data errors. In places the QP bounds are wider than the corresponding linear programming bounds; in places they are tighter. Overall there is not much difference. This does not indicate that LP and QP will always give comparable results: The nonlinearity of the problem prevents one from predicting what would happen with different data.

While QP brings us closer than any previous method to a completely consistent use of the scattered travel time observations that are available, several steps remain. The sample means of $\tau(p)$ are found by averaging over small bands of $X$ [*Bessonova et al.*, 1976]; through the unknown $X(p)$ function this corresponds to a weighted average in $p$. The averaging should be incorporated into the inversion process: We ought to require that models predict acceptable average values of $\tau(p)$ over appropriate bands of $p$. The present method of estimating $\tau(p)$ will not allow this; it seems likely that another method could be devised whose averaging in $p$ is more easily quantified. The current estimation procedure also requires that the errors in the travel time observations have zero mean and be independent and normally distributed. It might be possible with a great deal more data to estimate the true error distribution of our measurements, but this probably could not be done independently of a reference model, which begs the question.

## APPENDIX A: QUASI-CONVEXITY OF $\mu$

Here we prove that as the depth to $v_i$ is constrained to be further and further from $z^*$, the depth achieved by the model $\zeta^*$ that minimizes the misfit $\mu[\zeta]$, the misfit increases monotonically. This is equivalent to showing that the function of $z$ defined by finding the minimum of $\mu[\zeta]$ among the models $\zeta$ satisfying the positivity and radial derivative constraints and the constraint that $Z_{v_i}[\zeta] = z$, is a quasi-convex functional of $z$ (see *Bazaraa and Shetty* [1979], about quasi-convexity). Let $P$ denote the convex set of models that satisfy the positivity and radial derivative constraints. Define $S^z$ to be the subset of models in $P$ that reach the depth $z$ at velocity $v_i$, that is the elements of $P$ that satisfy the additional constraint that $Z_{v_i}[\zeta] = z$. $S^z$ is obviously convex. By definition $\zeta^* \in S^{z^*}$. Let $\zeta^1$ be a model that minimizes the convex functional $\mu[\zeta]$ over the convex set $S^{z^1}$. We will derive a contradiction from the assumption that there is a model $\zeta^2 \in S^{z^2}$ such that $\mu[\zeta^2] < \mu[\zeta^1]$, where $z^1$ is between $z^*$ and $z^2$, i.e., $z^* < z^1 < z^2$ or $z^* > z^1 > z^2$. All convex linear combinations of $\zeta^*$ and $\zeta^2$ satisfy the convex positivity and radial derivative constraints since $\zeta^*$ and $\zeta^1$ both do. In particular, $\zeta^{1'} \equiv \alpha\zeta^2 + (1-\alpha)\zeta^* \in S^{z^1}$ for $\alpha = (z^1-z^*)/(z^2-z^*) \in [0, 1]$ because then $z^1 = \alpha z^2 + (1-\alpha)z^*$ and $Z_{v_i}[\zeta]$ is linear in $\zeta$. The misfit functional $\mu[\zeta]$ is convex, so by definition

$$\begin{aligned}\mu[\zeta^{1'}] &\equiv \mu[\alpha\zeta^2 + (1-\alpha)\zeta^*] \\ &\leqslant \alpha\mu[\zeta^2] + (1-\alpha)\mu[\zeta^*] \\ &< \alpha\mu[\zeta^1] + (1-\alpha)\mu[\zeta^*] \\ &< \alpha\mu[\zeta^1] + (1-\alpha)\mu[\zeta^1] = \mu[\zeta^1]\end{aligned}$$

$\zeta^1$ was defined to be the model in $S^{z^1}$ that minimized $\mu[\zeta]$ so we have reached a contradiction. Figure 1 is a representative plot of $\mu$ as a function of $z$.

## APPENDIX B: NUMERICAL IMPLEMENTATION OF QP

Our FORTRAN implementation of QP is based on the algorithm NNLS [*Lawson and Hanson*, 1974]. Nonnegative least squares (NNLS) solves the problem

$$\min_{\mathbf{x} \geq 0} \|A \cdot \mathbf{x} - \mathbf{b}\|$$

where $\mathbf{x} \in \mathbf{R}^l$, $\mathbf{b} \in \mathbf{R}^m$ and $A$ is a matrix of $m$ rows and $l$ columns. NNLS is a tremendously robust program even when a large number of variables are used, so much so that SPMO used a weighting scheme similar to the one we will describe to code LP by simulating linear programming with NNLS.

To find the finite-dimensional approximation to the best fitting model $\zeta^*$, we pass NNLS the following matrix $A$:

$$\begin{bmatrix} I & I \\ \gamma W \cdot \begin{bmatrix} \tau \\ \mathbf{X} \end{bmatrix} & 0 \end{bmatrix}$$

and the vector $\mathbf{b}$:

$$\begin{bmatrix} \mathbf{c} \\ \gamma W \cdot \bar{\mathbf{d}} \end{bmatrix}$$

$I$ is the $L$ by $L$ identity matrix, where $L$ is the number of basis functions. The second $I$ matrix is used to introduce a set of $L$ positive slack variables that impose upper bounds on the coefficients in the basis expansion (see *Bazaraa and Shetty* [1979], for a discussion of slack variables). The diagonal matrix of weights, $W$, accounts for the different standard deviations of the data: $W_{ij} \equiv \{1/\sigma_i, i=j;\ 0 \text{ otherwise}\}$. The matrices $\tau$ and $\mathbf{X}$ map the coefficients of the basis expansion for $\zeta$ into their $\tau(p)$ and $X(p)$ predictions (for the exact expressions using the $1/v$ basis functions, see the definitions of $\tau_i$ and $\mathbf{X}_i$ in the fourth section of SPMO). We bound the coefficients of the model to enforce the radial derivative inequality constraint with the vector $\mathbf{c} \equiv (a, a, \ldots, a)^t \in \mathbf{R}^l$, where $a$ is the radius of the body of interest (here 3480 km, the core radius). The small positive constant $\gamma$ downweights fitting the data versus satisfying the radial derivative constraints. (The radial derivative inequalities are then satisfied almost exactly.) The sample means of the data comprise the vector $\bar{\mathbf{d}}$. The first $L$ elements of

the unknown vector $\mathbf{x}$ are the coefficients of the basis expansion we want; the latter $L$ are slack variables mentioned above, which are needed to impose the radial derivative inequality constraints. NNLS automatically forces the unknown to be nonnegative. This in turn ensures that $\zeta(v) \geq 0$ through our choice of basis functions. In general, the radial derivative constraints will be violated slightly since NNLS minimizes the two norm of the misfit to $\mathbf{b}$; however, with $\gamma = 10^3$ they were never violated by more than a part in $10^{11}$ in our applications. Once we have found the best fitting model, we may constrain that the model to attain the depth $z$ at velocity $v_i$ by adding to $A$ a row that is the finite-dimensional representation of $Z_{v_i}[\cdot]$ (see the expression for $Z_{v_j}$ in the same section of SPMO) and adding $z$ as a corresponding element of $\mathbf{b}$. For numerical stability it is important that this new row be inserted above the rows downweighted by $\gamma$ [*Lawson and Hanson*, 1974]. A priori all we know about the misfit as a function of the constraint depth is that it is monotonic about the best fitting depth. We used a bisection method to find $z$ such that the minimum of $\mu[\zeta]$ over $S^2$ is equal to $M^2$ because it is guaranteed to converge. More sophisticated search algorithms would probably not increase the efficiency much since the misfit is so flat near the best fitting depth (Figure 1). The iteration involved in the bisection and the large number of points needed for an accurate description of the bounds make the computational requirements of the method fairly heavy compared with LP.

*Acknowledgments.* We gratefully acknowledge support received from the National Science Foundation under NSF grant EAR-84-03847 and from the Office of Naval Research under ONR contract N00014-87-K-0005. Access to the NSF San Diego Supercomputer Center Cray X-MP/48 was made possible by a grant from the Scripps Institution of Oceanography Computer Policy Committee.

## REFERENCES

Abramowitz, M., and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1965.

Aki, K. and P. G. Richards, *Quantitative Seismology: Theory and Methods*, W. H. Freeman, San Francisco, Calif., 1980.

Bazaraa, M. S., and C. M. Shetty, *Nonlinear Programming*, John Wiley, New York, 1979.

Bessonova, E. N., V. M. Fishman, M. G. Shnirman, G. A. Sitnikova, and L. R. Johnson, The tau method for inversion of travel times, II, Earthquake data, *Geophys. J. R. Astron. Soc.*, *46*, 87–108, 1976.

Constable, S. C., R. L. Parker, and C.G. Constable, Occam's Inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, February, in press, 1987.

Dorman, L. M., and R. S. Jacobson, Linear inversion of body wave data, Part I, Velocity structure from traveltimes and ranges, *Geophysics*, *46*, 138–151, 1981.

Dziewonski, A. M., and D. L. Anderson, Preliminary reference earth model, *Phys. Earth Planet. Inter.*, *25*, 297–356, 1981.

Garmany, J., J. A. Orcutt, and R. L. Parker, Travel time inversion: A geometrical approach, *J. Geophys. Res.*, *84*, 3615-3622, 1979.

Gerver, M., and V. Markushevich, Determination of a seismic wave velocity from the travel-time curve, *Geophys. J. R. Astron. Soc.*, *11*, 165–173, 1966.

Johnson, L. R., and R. C. Lee, Extremal bounds on the *P* velocity in the earth's core, *Bull. Seismol. Soc. Am.*, *75*, 115–130, 1985.

Lawson, C. L., and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, N. J., 1974.

Luenberger, D. G., *Optimization by Vector Space Methods*, John Wiley, New York, 1969.

Stark, P. B., R. L. Parker, G. Masters, and J. A. Orcutt, Strict bounds on seismic velocity in the spherical earth, *J. Geophys. Res.*, *91*, 13,892–13,902, 1986.

R. L. Parker and P. B. Stark, Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, A-025, La Jolla, CA 92093

(Received June 20, 1986,
revised October 13, 1986,
accepted December 17, 1986.)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>Velocity Bounds from Statistical Estimates of $\tau(\rho)$ and $X(\rho)$ | | 5. TYPE OF REPORT & PERIOD COVERED<br>6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Philip B. Stark and Robert L. Parker | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-87-K-0005 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Scripps Institution of Oceanography<br>La Jolla, California 92093 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research<br>Arlington, Virginia 22217 | | 12. REPORT DATE<br>March 10, 1987<br>13. NUMBER OF PAGES<br>7 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>unclassified<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Reprint from Journal of Geophysical Research, Vol. 92, No.B3, 2,713-2,719, (1987)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We present a new technique for constructing the narrowest corridor containing all velocity profiles consistent with a finite collection of $\tau(\rho)$ data and their statistical uncertainties. Earlier methods for constructing such bounds treat the confidence interval for each $\tau$ datum as a strict interval within which the true value might lie with equal probability, but this interpretation is incompatible with the estimation procedure used on the original travel time observations. The new approach, based upon quadratic programming (QP), shares the advantages of the linear programming (LP) solution: it can invert $\tau(\rho)$ and. . .

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102 LF 014 6601

unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

MANDATORY DISTRIBUTION LIST

FOR UNCLASSIFIED TECHNICAL REPORTS, REPRINTS & FINAL REPORTS

PUBLISHED BY OCEANOGRAPHIC CONTRACTORS

OF THE OCEAN SCIENCE AND TECHNOLOGY DIVISION

OF THE OFFICE OF NAVAL RESEARCH

Department of Defense

Office of the Secretary of
Defense (3)
Assistant Director of Defense
Research & Engineering
Washington, D. C. 20301

Navy

Office of Naval Research (3)
Code 460
Arlington, VA 22217

Office of Naval Research
Code 480
Arlington, VA 22217

Office of Naval Research
Code 102 B
Arlington, VA 22217

Office of Naval Research
Commanding Officer
1030 East Green Street
Pasadena, CA 91101

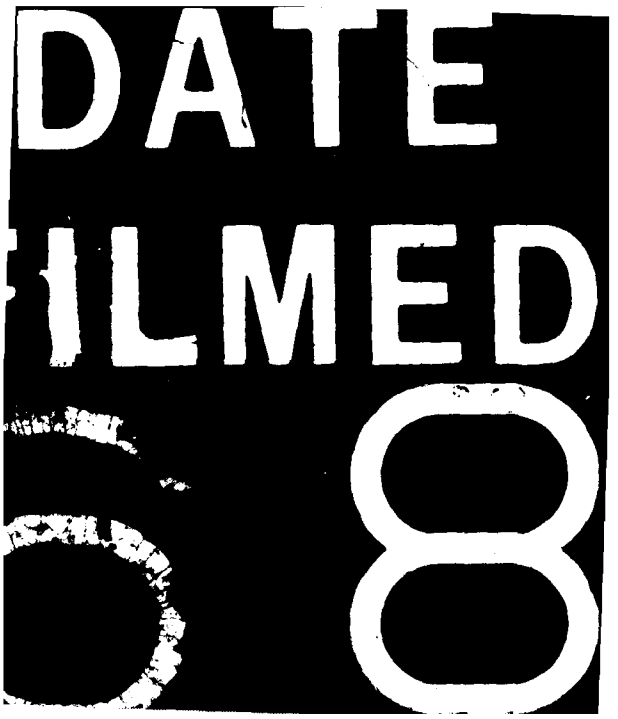
Naval Ocean Research
& Development Activity
NORDA, Code 300
NSTL Station
Bay St. Louis, MS 39529

U.S. Naval Oceanographic Office
Library, Code 8170
NSTL Station
Bay St. Louis, MS 39529

Other Government Agencies

Defehse Documentation Center (12)
Cameron Station
Alexandria, VA 22314

National Oceanic & Atmospheric
Administration
National Oceanographic Data Center
Washington Navy Yard
Rockville, MD 20852


DATE
ILMED
8