MULTI-CHANNEL PROCESSING OF DISPERSED SURFACE WAVES

David R. Russell, B.Sc, M.Sc

A Digest Presented to the Faculty of the Graduate School of Saint Louis University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

1987

DIGEST

A multi-channel phase-matched filter is developed for the extraction of phase velocity dispersion from seismic events recorded on multi-channel receivers, under conditions of extremely high incoherent noise levels. Methods of inverting the dispersion for intrinsic earth structure are thoroughly reviewed, with emphasis on determining earth structure where little *a priori* knowledge exists.

To test the theory, data were obtained from a refraction survey of the Saudi Arabian shield, conducted in 1978 by the U.S. Geological Survey. The data are composed of short period (0.1 - 1.0 sec) fundamental mode Rayleigh waves, digitally recorded from 20 (2 hz) seismometers, located between 6 and 45 kilometers from an explosive source. The multi-channel data were processed, and synthetic seismograms were produced to compare with the recorded data. High levels of Gaussian noise were then added to the synthetic seismograms, to demonstrate the viability of the multi-channel technique under low signal-to-noise conditions.

Results of the tests show that the multi-channel phase-matched filter successfully identifies phase-velocity dispersion, even in data sets where the signal-to-noise ratio is so low that surface waves cannot be distinguished from background noise levels. In addition, inversion results indicate that variances predicted by the multi-channel filter map correctly into intrinsic velocity variances of the earth model.

Two potential applications for the multi-channel phase-matched filter are the extraction of Stonely wave dispersion in acoustical well logging, and the identification of shallow structure from Rayleigh-wave ground

roll in seismic reflection exploration.

MULTI-CHANNEL PROCESSING OF DISPERSED SURFACE WAVES

David R. Russell, B.Sc, M.Sc

A Dissertation Presented to the Faculty of the Graduate School of Saint Louis University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

1987

COMMITTEE IN CHARGE OF CANDIDACY

Professor Robert B. Herrmann

Chairperson and Advisor

Professor Brian J. Mitchell

Professor Otto W. Nuttli

This dissertation is dedicated to my wife,

Helen Tong Russell

whose constant support and encouragement made it possible.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Robert B. Herrmann, for his guidance and help on this project, and the others that I have been involved with at St. Louis University. I appreciate the careful review of this manuscript by Dr. Brian Mitchell and Dr. Otto Nuttli. Finally, thanks to Eric Haug, for developing and maintaining an outstanding computer environment for research.

This research was supported by an Amoco Foundation fellowship, and the Air Force Geophysical Laboratory, under contract number F19628-87-K-0049.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT'D)

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

1.1 Formulation of the problem

Normal mode surface wave analysis is a well studied branch of modern seismology, particularly since there are many useful applications associated with the theory. A small listing of the potential uses would include earthquake source mechanisms, magnitude estimates of underground explosions, and shear velocity and Q structure in plane layered media. A significant reason for the value of normal modes is that in the mathematical exposition, eigenvalues (wavenumbers) of the modes are directly measurable in terms of observed phase and group velocities on seismograms. In turn, these eigenvalues can be theoretically related to structural earth parameters such as compressional velocity, shear velocity, and density. However, this relationship is implicit, and variational techniques combined with linear inversion theory are required for finding structural parameters as explicit functions of wavenumber.

This study will concentrate on methods of extracting phase and group velocity dispersion relations from observed seismograms, and the general problem of linear inversion of the relations for structural parameters. In particular, attention will be focused on data sets that are contaminated with high levels of noise, either incoherent or due to such factors as multi-pathing of events. Phase-matched filtering is a suitable method for processing such events, and it will be shown how matched filters can be utilized for both single and multi-channel data sets. The

study will focus on single-station long-period teleseismic events, and multi-channel processing of array data, such as that obtained from refraction surveys. The section on inversion will have a general exposition of inversion theory, and will deal in particular with methods of constraints in the presence of instability. Although linear inversion has been exhaustively studied, little has been published on constraining techniques that allow for simple, unbiased starting models, useful in regions where there is little *a priori* knowledge of the intrinsic shear velocity structure.

1.2 Literature review

Sato (1955) was the first author to use Fourier transform techniques to determine phase velocities of dispersed seismograms. He used two stations, and took the difference in phase spectra to determine phase velocity as a function of frequency, assuming a single mode of propagation. Bloch and Hales (1968) refined the measurement of phase velocities by determining the difference between phases of narrow band pass filtered seismograms.

Landisman *et al.* (1969) studied the "moving window" technique to determine group velocities, and showed how time variable, or frequency variable band pass filtering could be used to improve the method. Moving window analysis involves time windowing data at selected velocities and transforming to the frequency domain to determine the energy content of frequencies at the selected velocities. Dziewonski *et al.* (1969) used a reverse approach. In the "multiple filter" technique, the Fourier spectrum is narrow-band-pass filtered and then transformed to the time domain to determine those velocities at which a given frequency has a

maximum energy content. Herrmann (1973) showed the effects of using different types of filters in multiple filter analysis, and improved the efficiency of the method by using a Hilbert transform in the frequency domain to determine time domain envelopes.

Herrin and Goforth (1977) introduced phase matched filtering of dispersed seismograms to improve estimates of group velocities, by increasing the signal-to-noise ratio of the seismograms. Using an initial estimate of the group velocity, the group delay is integrated to approximate the phase spectrum. This is subtracted from the observed phase, so the signal is approximately zero phase. The signal is then windowed about zero lag in the time domain to increase the signal to noise ratio and remove effects such as multipathing. The phase of the resultant signal is differentiated to correct the initial group velocity estimate.

In a recent paper, McMechan and Yedlin (1981) show how to use multichannel array data to determine phase velocities. Essentially, they use a slant (p-tau) stack of the data which is transformed to the frequency domain. They show that the resultant two-dimensional frequency-wavenumber plot has maxima where the dispersion relation exists, or where phase velocity is a function of frequency. It should be noted that, although the mathematical approach is different, the method of slant stacking followed by a frequency transformation was also reported by Dziewonski and Hales (1972, p. 71).

Two widely used methods of inverting surface wave data for earth structure are direct search techniques and constrained least-squares inversion. Direct searching involves constructing a large set of possible earth models and sorting through these to find which ones fit the observed

dispersed velocities, via the theoretical forward solution. In this analysis, only constrained inversion methods will be considered, mainly due to reasons of computational efficiency. For the same reason, the forward solution will be restricted to the mathematical relation between Rayleigh and Love wave phase velocities and properties of plane, multilayered media. Thomson (1950) and Haskell (1953) were the first to derive this relation, with many subsequent improvements. For a state of the art analysis of the theoretical mathematics of plane layered media, see Wang (1981).

Constrained least squares inversion for surface waves involves transforming the non-linear forward problem into a first order matrix approximation, which is inverted for a perturbation about a starting model. Jeffreys (1961) introduced variational methods for surface-waves to find group velocities in terms of phase velocity perturbations. Takeuchi and Saito (1972) extended the variational technique to find first order perturbations of phase velocities in terms of structural earth parameter perturbations; or, equivalently, phase velocity partial derivatives. Rodi *et al.* (1975) developed an accurate numerical method for calculating group velocity partial derivatives, given phase velocity, phase velocity partial derivatives, and group velocities. The term "constrained" refers to the inverse matrix, which is constructed based on criteria formulated by Backus and Gilbert (1968). They showed that there is an inherent tradeoff between the variance of an inverse solution and its resolution. Inverse matrices can be constrained to produce solutions with small variances, but individual solution elements will no longer be uniquely identified; they will represent an average over adjacent elements. Wiggins (1972) and Jackson (1972) thoroughly studied the properties of con-

strained inversion, utilizing orthogonal decomposition of the matrices based on the theory of Lanczos (1961). Lawson and Hanson (1974) showed how to construct stable, efficient orthogonal decompositions of inversion matrices based on Householder transformations. Twomey (1977) gave an excellent overview of various techniques for constraining inversion matrices. Bache *et al.* (1978), and Russell (1980), used "differential" constraints to demonstrate the feasibility of surface-wave inversion based on simple half-space starting models.

CHAPTER 2

VARIATIONAL THEORY

2.1 General theory

Variational analysis offers a method for calculating first order perturbations in the eigenvalues of normal mode equations without resorting to "brute force" numerical difference techniques. The results can be quite significant in terms of time saving and accuracy, and it gives a clear formulation for setting up inversion problems. Since the theory is central to this study, it will be presented in a general form, in which surface wave analysis can be considered a special case.

Consider the following set of linear differential equations:

$$\mathbf{L}(\omega,\boldsymbol{\beta}_n(z))[\mathbf{u}(\omega,z)] = \sum_{j=0}^{n-1} k^j(\omega)\mathbf{M}_j(\omega,\boldsymbol{\beta}_j(z))[\mathbf{u}(\omega,z)]. \qquad (2.1.1)$$

The following values are understood:

L,M= Linear differential operators, either in scalar or matrix format, allowing the expression to represent coupled differential equations. They are implicitly functions of ω and $\boldsymbol{\beta}$. The brackets indicate that **u** is operated on.

k= eigenvalue found at ω satisfying the equations. The j superscript indicates k raised to the j power.

u= vector of eigenfunctions that solve the coupled differential equations. It is assumed there is an eigenvalue k associated with each vector eigenfunction.

$\boldsymbol{\beta}_j$= vectors of "structural parameters". These can correspond to mass,

density, intrinsic velocity, etc. Notice that for this problem they are considered as one dimensional functions of the variable z.

The above expression can be considered as Fourier transformed wave equations in a depth-dependent medium. Distance and time are transformed into wavenumber k and frequency ω. It is not, however, restricted to wave equations; with suitable variable substitutions, for example, the general Sturm-Liouville operator (Arfken, 1973) can be considered a subset. By choosing this relatively abstract format, several properties of variational analysis can be demonstrated, which may not be clear in special cases.

Rewrite (2.1.1) in the simplified form

$$\mathbf{Lu} = \sum_j k^j \mathbf{M}_j \mathbf{u}. \tag{2.1.2}$$

Multiply by the transpose of $\mathbf{u}$ and integrate over z, giving

$$\int_z \mathbf{u}^{\mathrm{T}} \mathbf{L} \mathbf{u} \, dz = \sum_j k^j \int_z \mathbf{u}^{\mathrm{T}} \mathbf{M}_j \mathbf{u} \, dz.$$

This operation transforms the matrix equations into quadratic scalars, simplifying the system into one equation. The integral bounds are chosen so that a linear combination of the eigenfunctions and their derivatives vanish; this is necessary for the integrals to be self-adjoint, the importance of which will be made clear below.

Introduce a small perturbation in the structural parameters (β_j). Using the above simplified notation, the perturbed system can be written as

$$\int_z (\mathbf{u}+\delta\mathbf{u})^{\mathrm{T}} (\mathbf{L}+\delta\mathbf{L})(\mathbf{u}+\delta\mathbf{u}) \, dz = \sum_j (k+\delta k)^j \int_z (\mathbf{u}+\delta\mathbf{u})^{\mathrm{T}} (\mathbf{M}_j+\delta\mathbf{M}_j)(\mathbf{u}+\delta\mathbf{u}) \, dz.$$

Keeping only first-order terms and rearranging gives

$$\begin{aligned}\int_z (\mathbf{u}+\delta\mathbf{u})^{\mathrm{T}}(\mathbf{L}\mathbf{u}-\sum_j k^j \mathbf{M}_j \mathbf{u})\,dz \\ + \int_z \mathbf{u}^{\mathrm{T}}(\mathbf{L}\delta\mathbf{u}-\sum_j k^j \mathbf{M}_j \delta\mathbf{u})\,dz \\ + \int_z \mathbf{u}^{\mathrm{T}}(\delta\mathbf{L}\mathbf{u}-\sum_j k^j \delta\mathbf{M}_j \mathbf{u})\,dz \\ = \delta k \sum_j j k^{j-1} \int_z \mathbf{u}^{\mathrm{T}} \mathbf{M}_j \mathbf{u}\,dz.\end{aligned} \tag{2.1.3}$$

From (2.1.2), the first integral in (2.1.3) vanishes identically. At this point, the power of variational methods can be shown: if the second integral vanishes identically, perturbations to the eigenvalue (δk) will not require a recalculation of the eigenfunctions ($\mathbf{u}$), since all first order perturbations of the eigenfunctions ($\delta\mathbf{u}$) have vanished. This will be true if the system is self-adjoint, which is defined as follows. Given any two functions $\mathbf{u}_1$ and $\mathbf{u}_2$ (not necessarily solutions to (2.1.2)), if

$$\int_z \mathbf{u}^{\mathrm{T}}_1(\mathbf{L}\mathbf{u}_2-\sum_j k^j \mathbf{M}_j \mathbf{u}_2)\,dz = \int_z \mathbf{u}^{\mathrm{T}}_2(\mathbf{L}\mathbf{u}_1-\sum_j k^j \mathbf{M}_j \mathbf{u}_1)\,dz, \tag{2.1.4}$$

then the system (2.1.2) is considered self-adjoint (Butkov, 1968, p. 341). If this is valid, then the second integral of (2.1.3) can be written as

$$\int_z \delta\mathbf{u}^{\mathrm{T}}(\mathbf{L}\mathbf{u}-\sum_j k^j \mathbf{M}_j \mathbf{u})\,dz$$

which vanishes identically by (2.1.2).

Finally, to solve for δk, define

$$W = \sum_j j k^{j-1} \int_z \mathbf{u}^{\mathrm{T}} \mathbf{M}_j \mathbf{u}\,dz.$$

Solving (2.1.3) for δk gives

$$\delta k = \frac{1}{W} \int_z \mathbf{u}^{\mathrm{T}}(\delta\mathbf{L}\mathbf{u}-\sum_j k^j \delta\mathbf{M}_j \mathbf{u})\,dz. \tag{2.1.5}$$

This expression can be further simplified if it is possible to factor out the structural parameter perturbations from the differential operators $\delta\mathbf{L}$ and $\delta\mathbf{M}$. If so, (2.1.5) can be written as a sum of Fredholm integrals of the first kind (Twomey, 1977). Recalling the explicit parameter dependence in (2.1.1), the result is

$$\delta k(\omega) = \sum_{j=0}^{n} \int_z \mathbf{A}_j(\omega,z)\delta\beta_j(z)dz \tag{2.1.6}$$

where the $\mathbf{A}_j$ are kernel functions calculated from contributions of $\mathbf{L}$ and $\mathbf{M}_j$. The transition may seem obscure at this point, but it will be clarified in practical examples. Equation (2.1.6) forms the basis for linear inversion theory, which will be studied in chapter 5.

2.2 Applications to surface wave theory

The equations of motion for Rayleigh and Love waves can be simply generated for a depth-dependent elastic medium by equating stress and strain relations for plane waves in cartesian coordinates (Takeuchi and Saito (1972). Aki and Richards (1980) give a straightforward derivation of these equations, the result of which is presented below.

For Rayleigh waves:

$$\frac{d}{dz}\begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{pmatrix} = \begin{pmatrix} 0 & -k & \mu^{-1} & 0 \\ k\lambda[\lambda+2\mu]^{-1} & 0 & 0 & [\lambda+2\mu]^{-1} \\ k^2\varsigma-\omega^2\rho & 0 & 0 & -k\lambda[\lambda+2\mu]^{-1} \\ 0 & -\omega^2\rho & k & 0 \end{pmatrix}\begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{pmatrix} \tag{2.2.1}$$

where

$\varsigma = 4\mu[\lambda+\mu] \;/\; [\lambda+2\mu]$

$\lambda,\mu =$ Lame constants

ρ = density

ω = angular frequency k = wavenumber (considered an eigenvalue)

r_1 = radial displacement eigenfunction

r_2 = vertical displacement eigenfunction

r_3 = radial stress eigenfunction

r_4 = vertical stress eigenfunction.

r_1 through r_4 are functions of w , k and z; ρ, λ, and μ are functions of z, and the wavenumber k is only a function of the frequency ω.

For Love waves:

$$\frac{d}{dz}\begin{pmatrix} l_1 \\ l_2 \end{pmatrix} = \begin{pmatrix} 0 & \mu^{-1} \\ k^2\mu - \omega^2\rho & 0 \end{pmatrix}\begin{pmatrix} l_1 \\ l_2 \end{pmatrix} \tag{2.2.2}$$

where k, ω, and μ are the same as above,

l_1 = horizontal displacement eigenfunction

l_2 = horizontal stress eigenfunction.

Boundary conditions for both Rayleigh and Love waves are that the displacement eigenfunctions vanish when $z=\infty$, and the stress eigenfunctions vanish at the free surface $z=0$.

It should be noted that (2.2.1) has a reverse sign in wavenumber (k) from that found in Aki and Richards (1980). This is due to a reverse convention in defining wavenumber and frequency in Fourier transforms; the sign used here is consistent with equivalent expressions for the equations, such as found in Herrmann (1973) and Wang (1981). Aki and Richards define Fourier transforms between distance and wavenumber as

$$F(k) = \int_{-\infty}^{\infty} f(x)\, e^{-ikx}\, dx \ ,$$

whereas the convention used here is

$$F(k) = \int_{-\infty}^{\infty} f(x)\, e^{ikx}\, dx \ .$$

The procedure for finding variational integrals of Love and Rayleigh waves can be simplified if (2.2.1) and (2.2.2) are rearranged into forms explicitly dependent on displacement eigenfunctions only. For Love waves, do the following steps. Multiply the first row of (2.2.2) by μ and differentiate with respect to z. Substitute $\frac{dl_2}{dz}$ of the first row into the second row. The resulting equation is

$$\frac{d}{dz}\left(\mu\frac{dl_1}{dz}\right) = (k^2\mu - \omega^2\rho)l_1 \ . \tag{2.2.3}$$

In Appendix I, this equation is shown to be self-adjoint over the depth interval 0 and ∞. Therefore, the results of the previous section apply, and a variational integral for δk can be found. Define the following:

$$\mathbf{L}=\frac{d}{dz}(\mu\frac{d}{dz}) \ , \quad \mathbf{u}=l_1 \ , \quad \mathbf{M}_0= -\omega^2\rho \ , \quad \mathbf{M}_1=0 \ , \quad \mathbf{M}_2=\mu \ .$$

Equation (2.2.3) is in the same form as (2.1.2), so the variational integral can immediately be written in the form (2.1.5):

$$W = \sum_j jk^{j-1}\int_z \mathbf{u}^{\mathrm{T}}\mathbf{M}_j\,\mathbf{u}\,dz = 2k\int_0^{\infty}\mu {l_1}^2 dz \tag{2.2.4}$$

$$\delta k = \frac{1}{W}\int_0^{\infty}\left\{l_1\frac{d}{dz}(\delta\mu\frac{dl_1}{dz}) + \omega^2\delta\rho {l_1}^2 - k^2\delta\mu {l_1}^2\right\}dz. \tag{2.2.5}$$

This equation can be transformed into a Fredholm integral of the first kind by integrating the first term by parts:

$$\int_0^\infty l_1 \frac{d}{dz}(\delta\mu \frac{dl_1}{dz})dz = l_1(\delta\mu \frac{dl_1}{dz})\Big|_0^\infty - \int_0^\infty \left(\frac{dl_1}{dz}\right)^2 \delta\mu\, dz$$

The first term on the right vanishes by using the boundary conditions for stress and displacement at the free surface and ∞. Equation (2.2.5) can therefore be written

$$\delta k = \frac{1}{W}\int_0^\infty \left\{ -\left[k^2 l_1 + \left(\frac{dl_1}{dz}\right)^2\right]\delta\mu + \omega^2 l_1^2 \delta\rho \right\} dz \ .$$

Relations for shear velocity and phase velocity are

$$\beta^2 = \frac{\mu}{\rho}, \qquad c = \frac{\omega}{k}$$

and first order perturbations of these variables follows for fixed ω:

$$\frac{\delta\mu}{\mu} = 2\frac{\delta\beta}{\beta} + \frac{\delta\rho}{\rho}, \qquad \frac{\delta c}{c} = -\frac{\delta k}{k} \ .$$

Substituting these values into the above integral gives a final relation between perturbations of phase velocity, intrinsic shear velocity, and density:

$$\delta c = \frac{c}{kW}\int_0^\infty \left[k^2 l_1 + \left(\frac{dl_1}{dz}\right)^2\right] 2\mu \frac{\delta\beta}{\beta} dz +$$

$$\frac{c}{kW}\int_0^\infty \left[\left(\mu k^2 - \rho\omega^2\right) l_1 + \mu \left(\frac{dl_1}{dz}\right)^2 \right] \frac{\delta\rho}{\rho}\, dz \ . \qquad (2.2.6)$$

The analysis for Rayleigh waves is similar, but more involved, since it requires coupled differential equations between the radial and vertical displacement eigenfunctions. To set up the equations of motion, solve the first and second rows of (2.2.1) for the stress eigenfunctions r_3 and r_4, giving

$$r_3 = \mu\frac{dr_1}{dz} + k\mu r_2$$

$$r_4 = (\lambda+2\mu)\frac{dr_2}{dz} - k\lambda r_1 \, .$$

Substitute these values into the third and fourth rows for

$$\frac{d}{dz}[\mu\frac{dr_1}{dz} + k\mu r_2] = -k\lambda\frac{dr_2}{dz} + [(\lambda+2\mu)k^2 - \omega^2\rho]r_1$$

$$\frac{d}{dz}[(\lambda+2\mu)\frac{dr_2}{dz} - k\lambda r_1] = k\mu\frac{dr_1}{dz} + [\mu k^2 - \omega^2\rho]r_2 \, .$$

These coupled equations can be put into a single matrix format by defining the following matrices:

$$\mathbf{A} = \begin{bmatrix} \mu & 0 \\ 0 & (\lambda+2\mu) \end{bmatrix} , \quad \mathbf{B} = \begin{bmatrix} 0 & \mu \\ -\lambda & 0 \end{bmatrix} , \quad \mathbf{C} = \begin{bmatrix} (\lambda+2\mu) & 0 \\ 0 & \mu \end{bmatrix} , \quad \mathbf{r} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} .$$

Substitution of the above gives

$$\frac{d}{dz}(\mathbf{A}\frac{d\mathbf{r}}{dz} + k\mathbf{B}\mathbf{r}) = k\mathbf{B}^{\mathrm{T}}\frac{d\mathbf{r}}{dz} + k^2\mathbf{C}\mathbf{r} - \omega^2\rho\mathbf{r} \, . \qquad (2.2.7)$$

By comparing the term in parentheses on the left side to the matrix function (2.2.1), it can be seen that the term corresponds to the stress eigenfunctions r_3 and r_4. Equivalent forms of (2.2.7) are found in Keilis-Borok *et al.* (1965), and Kennett (1983). Appendix I shows that (2.2.7) is self-adjoint over the interval 0 to ∞, so the variational principles of the previous section apply. To show that the equation is equivalent to (2.1.2), let

$$\mathbf{L} = \frac{d}{dz}\mathbf{A}\frac{d}{dz} \, , \quad \mathbf{M}_0 = -\omega^2\rho \, , \quad \mathbf{M}_1 = (\mathbf{B}^{\mathrm{T}}\frac{d}{dz} - \frac{d\mathbf{B}}{dz}) \, , \quad \mathbf{M}_2 = \mathbf{C} \, . \qquad (2.2.8)$$

Substitution of these values into (2.2.7) gives an exactly equivalent form to (2.1.2), so that the integral (2.1.5) is a valid expression for eigenvalue perturbation. Substituting (2.2.8) into the integral gives

$$\delta k \int_0^\infty (\mathbf{r}^T \mathbf{B}^T \frac{d\mathbf{r}}{dz} - \mathbf{r}^T \frac{d\mathbf{B}\mathbf{r}}{dz} + 2k\mathbf{r}^T \mathbf{C}\mathbf{r}) dz =$$

$$\int_0^\infty [\mathbf{r}^T \frac{d}{dz}(\delta\mathbf{A} \frac{d\mathbf{r}}{dz}) + \mathbf{r}^T \mathbf{r} \omega^2 \delta\rho - k\mathbf{r}^T (\delta\mathbf{B}^T \frac{d\mathbf{r}}{dz} - \frac{d\,\delta\mathbf{B}\mathbf{r}}{dz}) - k^2 \mathbf{r}^T \delta\mathbf{C}\mathbf{r}] dz.$$

Some of the terms in the integrals involve differentiation of structural parameters $\delta\mathbf{A}$ and $\delta\mathbf{B}$. To remedy this, integrate those terms by parts (the first and fourth terms on the right, and include the second term on the left side):

$$\delta k \int_0^\infty (2\mathbf{r}^T \mathbf{B}^T \frac{d\mathbf{r}}{dz} + 2k\mathbf{r}^T \mathbf{C}\mathbf{r}) dz =$$

$$\mathbf{r}^T (\delta\mathbf{A} \frac{d\mathbf{r}}{dz} + k\,\delta\mathbf{B}\mathbf{r} + \delta k \mathbf{B}\mathbf{r}) \Big|_0^\infty + \tag{2.2.9}$$

$$\int_0^\infty [-\frac{d\mathbf{r}^T}{dz} \delta\mathbf{A} \frac{d\mathbf{r}}{dz} + \mathbf{r}^T \mathbf{r} \omega^2 \delta\rho - 2k\mathbf{r}^T \delta\mathbf{B}^T \frac{d\mathbf{r}}{dz} - k^2 \mathbf{r}^T \delta\mathbf{C}\mathbf{r}] dz \ .$$

The term evaluated at 0 and ∞ vanishes due to boundary conditions on the eigenfunctions; the displacement $\mathbf{r}$ is 0 at $z=\infty$, and the term in parenthesis is a first order stress perturbation (see (2.2.7)), which vanishes at the free surface.

Expanding (2.2.9) term by term gives:

$$W = 2k \int_0^\infty [(\lambda+2\mu) r_1^2 + \mu r_2^2] dz + 2 \int_0^\infty (\mu r_2 \frac{dr_1}{dz} - \lambda r_1 \frac{dr_2}{dz}) dz \tag{2.2.10}$$

$$W\,\delta k =$$

$$\omega^2 \int_0^\infty \delta\rho (r_1^2 + r_2^2) dz -$$

$$k^2 \int_0^\infty [\delta(\lambda+2\mu) r_1^2 + \delta\mu r_2^2] dz -$$

$$2k\int_0^\infty(\delta\mu r_2\frac{dr_1}{dz} - \delta\lambda r_1\frac{dr_2}{dz})dz -$$
$$\int_0^\infty\left[\delta\mu\left(\frac{dr_1}{dz}\right)^2 + \delta(\lambda+2\mu)\left(\frac{dr_2}{dz}\right)^2\right]dz \ .$$

This can be further simplified to

$$\delta k = \frac{1}{W}\int_0^\infty\omega^2(r_1^2+r_2^2)\delta\rho dz -$$
$$\frac{1}{W}\int_0^\infty\left\{\left(kr_1-\frac{dr_2}{dz}\right)^2\delta(\lambda+2\mu) + \left[\left(kr_2+\frac{dr_1}{dz}\right)^2+4kr_1\frac{dr_2}{dz}\right]\delta\mu\right\}dz \ . \qquad (2.2.11)$$

Relationships for compressional, shear and phase velocities are

$$\alpha^2 = \frac{(\lambda+2\mu)}{\rho}\ , \qquad \beta^2 = \frac{\mu}{\rho}\ , \qquad c = \frac{\omega}{k}$$

and first order perturbations in the parameters are related by

$$\frac{\delta(\lambda+2\mu)}{(\lambda+2\mu)} = 2\frac{\delta\alpha}{\alpha}+\frac{\delta\rho}{\rho}\ , \qquad \frac{\delta\mu}{\mu} = 2\frac{\delta\beta}{\beta}+\frac{\delta\rho}{\rho}\ , \qquad \frac{\delta c}{c} = -\frac{\delta k}{k}\ .$$

Substituting these values into (2.2.11) gives perturbations of phase velocities in terms of structural velocity variations and density:

$$\delta c = \frac{c}{kW}\int_0^\infty\left(kr_1-\frac{dr_2}{dz}\right)^2 2(\lambda+2\mu)\frac{\delta\alpha}{\alpha}dz +$$

$$\frac{c}{kW}\int_0^\infty\left[\left(kr_2+\frac{dr_1}{dz}\right)^2 + 4kr_1\frac{dr_2}{dz}\right]2\mu\frac{\delta\beta}{\beta}dz +$$

$$\frac{c}{kW}\int_0^\infty\left(kr_1-\frac{dr_2}{dz}\right)^2(\lambda+2\mu)\frac{\delta\rho}{\rho}dz +$$

$$\frac{c}{kW}\int_0^\infty\left[\left(kr_2+\frac{dr_1}{dz}\right)^2 + 4kr_1\frac{dr_2}{dz}\right]\mu\frac{\delta\rho}{\rho}dz \ . \qquad (2.2.12)$$

2.3 Summary

Equations (2.2.6) for Love waves and (2.2.12) for Rayleigh waves are Fredholm integrals of the first kind, which relate perturbations in intrinsic shear velocity, compressional velocity, and density to observed phase velocities at the free surface. By assuming a starting earth model of intrinsic velocities and densities, perturbations in phase velocities (δc) can be found by subtracting the theoretical phase velocities (determined from the eigenvalues of the solution to (2.1.6)), from observed phase velocities. By inverting the above integral equations for $\delta\beta$, $\delta\alpha$, and $\delta\rho$, estimates for earth structure can be determined as functions of δc.

Subsequent chapters in this study will concentrate on methods of measuring observed velocities, and then finding stable methods of inversion for the intrinsic earth structure.

CHAPTER 3

MULTI-CHANNEL DISPERSION MEASUREMENTS

3.1 General

Measurements of phase and group velocity dispersion on observed seismograms usually assume that the wavetrains can be decomposed into a sum of monochromatic, or plane waves. If so, an individual wave can be represented by the Fourier spectrum of the seismogram

$$u(x,t) = A(\omega,x)e^{i(\omega t - k(\omega)x)} \tag{3.1.1}$$

where

A = amplitude spectrum,

ω = angular frequency,

$k(\omega)$ = wavenumber,

x = distance from source,

t = time.

Dziewonski and Hales (1972) define phase velocity as the instantaneous velocity of the plane waves at a given frequency:

$$c(\omega) = \frac{dx}{dt} = \frac{\omega}{k(w)} \tag{3.1.2}$$

and the group velocity as the velocity of energy transmission:

$$u(\omega) = \frac{x}{t} = \frac{d\omega}{dk} \quad . \tag{3.1.3}$$

The assumption of using plane waves as a basis for measuring surface waves on seismograms is generally valid for large distances (several wavelengths) from the source (Aki and Richards, 1980, p. 306). However,

it is incorrect to assume that the phase velocity can be determined directly from the phase component of the Fourier transform of the signal, as (3.1.1) implies, unless it is assumed that only a single mode is propagating. In general, there can be an infinite number of discrete wavenumber (k) solutions, or modes, to the surface wave problem (2.2.1) and (2.2.2). Assuming vertical heterogeneity only, the complete normal mode solution can be written as (Aki and Richards, 1980, p. 316))

$$f(x,t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{m} A_m(\omega,x) e^{i(\omega t - k_m(\omega)x)} \, d\omega. \tag{3.1.4}$$

where

$$A_m(\omega,x) = I(\omega) S_m(\omega) R_m(\omega) \frac{e^{-\gamma_m(\omega)x}}{\sqrt{x}} \tag{3.1.5}$$

and

m = mode number,
$I(\omega)$ = instrument response (complex),
$S_m(\omega)$ = source spectrum (complex),
$R_m(\omega)$ = path response (real),
$\gamma_m(\omega)$ = attenuation coefficient (real).

It should be understood that any constant phase term in (3.1.4) is to be considered as part of the phase of $S_m(\omega)$. Equation (3.1.4) shows that the Fourier spectrum of the seismogram consists of a complex sum of normal modes, and as a result, the phase spectrum will be a complicated function of individual modes, and cannot be used directly to measure the wavenumbers $k_m(\omega)$. Other factors that can distort the phase spectrum are multipathing, where the seismogram consists of the original signal plus reflections; multiple events, where signals from different sources

overlap; and ambient site noise levels. In addition, the phase response of $A_m(\omega,x)$ must either be known or removable, or it will bias the "pure" phase spectrum, defined as the exponent in (3.1.4). The following methods will attempt to isolate, under suitable conditions, individual normal modes of propagation, assuming that (3.1.4) is a valid representation of wave motion.

3.2 P-ω stacking

3.2.1 Theory

P-tau (slant) stacking involves searching multi-channel data sets for phase velocities that will produce constructive interference of monochromatic waves at a given frequency. A direct sum of many recordings at a given frequency will not, in general, constructively interfere, due to phase differences at different stations. However, if it is assumed that (3.1.4) is valid, the phase offset between any two stations will be equal to $k_m(\omega)\Delta x$, where Δx is the interstation distance. If $k_m(\omega)$ is picked correctly, the phases of a propagating mode m can be aligned for constructive interference. From (3.1.2),

$$k_m(\omega) = \frac{\omega}{c_m(\omega)} = \omega p_m(\omega)$$

where $p_m = 1/c_m =$ wave slowness. Therefore, either k, p, or c can be treated as independent variables , and they can be varied for the value (or values) where the phase adjusted sum of the recordings reaches a maximum. This value will determine the phase velocity of a mode at a given frequency.

This technique was presented by Dziewonski and Hales (1972) as "the method of sums and differences." McMechan and Yedlin (1981) derived the method in the context of a Fourier transformed wave slowness - frequency (p-ω) plot, with the assumption of continuous spatial sampling. The analysis below follows McMechan and Yedlin, but discrete spatial sampling will be assumed, and the effect of attenuation will be investigated. The frequency domain analysis will be referred to as a p-ω stack, to distinguish it from the time domain slant stack.

Let n spatially separated stations lie along the same azimuth as a seismic source. Assuming that (3.1.4) describes the wave motion, perform the following slant stack:

$$\sum_n f(x_n, \tau + p x_n) = \sum_n \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_m A_m(\omega, x_n)\, e^{i(\omega\tau + \omega p x_n - k_m(\omega) x_n)}\, d\omega$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega\tau} \sum_m \sum_n A_m(\omega, x_n)\, e^{i(\omega p - k_m(\omega)) x_n}\, d\omega \quad . \qquad (3.2.1)$$

If it assumed that geometrical spreading ($\sqrt{x}$) has been removed prior to stacking, then, from (3.1.5)

$$A_m(\omega, x_n) = A_m(\omega)\, e^{-\gamma_m(\omega) x_n} \quad . \qquad (3.2.2)$$

Taking the Fourier transform of (3.2.1) results in

$$F(\omega, p) = \sum_m A_m(\omega) \left[\sum_n e^{-\gamma_m(\omega) x_n}\, e^{i\omega(p - p_m(\omega)) x_n} \right] ,$$

where $\omega p_m(\omega) = k_m(\omega)$. Let W_m be the function within the brackets. Then,

$$F(\omega, p) = \sum_m A_m(\omega)\, W_m(p - p_m(\omega)) \quad . \qquad (3.2.3)$$

When $p = p_m(\omega)$, the resolving function W_m should reach a maximum value due to constructive interference of the waveforms. In addition, if the amplitude of W_m quickly diminishes to zero when p moves away from $p_m(\omega)$, there will be little interference from adjacent modes. Remembering that $F(\omega,p)$ is a complex function, and *assuming* that W_m is an impulsive function of p, the modulus can be expressed as:

$$|F(\omega,p)| = |\sum_m A_m(\omega)\ W_m(p-p_m(\omega))|$$

$$= \sum_m |A_m(\omega)|\ |W_m(p-p_m(\omega))| \ . \qquad (3.2.4)$$

Notice that the phase term of $A_m(\omega)$ has disappeared from the sum in (3.2.4); the common source and instrument phase are effectively removed from the calculation. Plotting (3.2.4) as a two dimensional function of p and ω and picking the maxima should give the phase velocity of individual modes as a function of frequency, assuming that W_m is sufficiently impulsive and that the modes are sufficiently separated to make (3.2.4) a valid statement.

To clarify the above, rewrite the modulus of the resolving function W_m:

$$|W_m| = |\sum_n e^{z_m x_n}| \ , \qquad (3.2.5)$$

where

$$z_m = -\gamma_m(\omega) + i\omega(p - p_m(\omega)) \ .$$

Consider the following special case: let there be an array of N stations, equally spaced at an interval of Δx, with the first station a distance of x_0 from the source. Then (3.2.5) can be written as

$$|W_m| = e^{-\gamma_m(\omega)x_0} \left| \sum_{n=0}^{N-1} e^{z_m\, n\, \Delta x} \right| . \qquad (3.2.6)$$

The sum in (3.2.6) is a geometric series which can be expressed as

$$|W_m| = e^{-\gamma_m(\omega)x_0} \left| \frac{e^{z_m\, N\, \Delta x} - 1}{e^{z_m\, \Delta x} - 1} \right| . \qquad (3.2.7)$$

The modulus of (3.2.7) can be found explicitly as follows: first factor (3.2.7) for

$$|W_m| = e^{-\gamma_m(\omega)x_0} \left| e^{\left[z_m(N-1)\frac{\Delta x}{2}\right]} \frac{\sin\left(z_m \frac{N\Delta x}{2}\right)}{\sin\left(z_m \frac{\Delta x}{2}\right)} \right| .$$

Then use relations from Churchill (1974, p. 58) for the modulus of the complex sine function to get

$$|W_m| =$$

$$e^{-\gamma_m(\omega)X} \left[\frac{\sin^2\left(\omega[p - p_m(\omega)]\frac{N\Delta x}{2}\right) + \sinh^2\left(\gamma_m(\omega)\frac{N\Delta x}{2}\right)}{\sin^2\left(\omega[p - p_m(\omega)]\frac{\Delta x}{2}\right) + \sinh^2\left(\gamma_m(\omega)\frac{\Delta x}{2}\right)} \right]^{\frac{1}{2}} , \qquad (3.2.8)$$

where

$$X = x_0 + (N-1)\frac{\Delta x}{2} .$$

There are several interesting features about the resolving function (3.2.8) or the equivalent (3.2.7). First, it is a periodic function, with periodicity defined by:

$$|W_m(p)| = \left| W_m\left(p \pm \frac{n}{f\,\Delta x}\right) \right| , \qquad (3.2.9)$$

where

$$2\pi f = \omega , \qquad n = 1,2,3... \ .$$

This can be verified by direct substitution into either (3.2.7) or (3.2.8). The periodicity can be also defined in terms of upper and lower bounds on phase velocity about the m'th mode:

$$c_{per} = \frac{c_m}{1 \pm \frac{n c_m}{f \Delta x}} \tag{3.2.10}$$

Second, the resolving function has a maximum when $p = p_m(\omega)$. This can be shown by differentiating (3.2.8) with respect to p and setting the result to zero, which will occur when $p = p_m(\omega)$. The importance of this is that attenuation will have no effect on the position of the maximum. Notice that when $\gamma_m(\omega)$=0, (3.2.8) reduces to the modulus of the Fourier-series kernel (Papoulis, 1977, p. 72):

$$| W_m | = \left| \frac{\sin\left(\omega[p - p_m(\omega)]\frac{N\Delta x}{2}\right)}{\sin\left(\omega[p - p_m(\omega)]\frac{\Delta x}{2}\right)} \right| . \tag{3.2.11}$$

This has a maximum when $p = p_m(\omega)$ and the value of the maximum is N, which can be shown by a first order expansion of (3.2.11).

Third, the width of the major lobe of (3.2.8) is a function of frequency, attenuation, and the total width of the array ($N\Delta x$). When $\gamma_m(\omega) = 0$, the width can be found by finding the first zero away from the maximum of (3.2.11). This will occur when

$$\omega[p - p_m(\omega)]\frac{N\Delta x}{2} = \pm\pi \ , \tag{3.2.12}$$

or equivalently,

$$\text{width} = 2[p - p_m(\omega)] = \frac{2}{fN\Delta x} \quad . \tag{3.2.13}$$

This can be expressed in terms of velocity bounds about the maximum as

$$c_{\text{lim}} = \frac{c_m}{1 \pm \dfrac{c_m}{fN\Delta x}} \quad . \tag{3.2.14}$$

There is no simple solution to the problem of defining the width of the major lobe when attenuation becomes significant. The derivative of (3.2.8) must be searched numerically for a minimum in the vicinity of (3.2.12). However, if it is assumed that N is large, $\gamma_m(\omega)$ is small, and $\gamma_m(\omega) \ll N\Delta x$, the following first order approximation can substitute for the value under the square root in (3.2.8):

$$|W_m| =$$

$$e^{-\gamma_m(\omega)X} \left(\frac{\left(\omega[p - p_m(\omega)]\dfrac{N\Delta x}{2} - \pi\right)^2 + \left(\gamma_m(\omega)\dfrac{N\Delta x}{2}\right)^2}{\left(\omega[p - p_m(\omega)]\dfrac{\Delta x}{2}\right)^2} \right)^{\frac{1}{2}} \quad . \tag{3.2.15}$$

This assumes that p is close to the value found by (3.2.12). Differentiating (3.2.15) with respect to p and setting to zero will result in the following perturbation to (3.2.13):

$$\text{width} = \frac{2}{fN\Delta x}\left(1 + \left(\frac{\gamma_m(\omega)N\Delta x}{2\pi}\right)^2\right) \quad . \tag{3.2.16}$$

This shows that there is an almost negligible first order contribution from attenuation to the width of the resolving function (3.2.8). However, care must be taken with making approximations to the hyperbolic sine term in

the numerator of (3.2.8). If either $\gamma_m(\omega)$ or the total array width is large enough, the exponential nature of the hyperbolic sine term will dominate the expression, and the major lobe of the function will no longer be a reasonable measure of the width of the resolving function. In this case, an appropriate measure for the width of the resolving function will be the half-power point of the envelope function. Ignoring the sine term in the numerator of (3.2.8), and equating the square of the function with half the maximum value, gives the result

$$\sin^2\left[\omega[p-p_m(\omega)]\frac{\Delta x}{2}\right]+\sinh^2\left[\gamma_m(\omega)\frac{\Delta x}{2}\right]=2\sinh^2\left[\gamma_m(\omega)\frac{\Delta x}{2}\right] .$$

Solving for the width gives

$$\text{width}=\frac{2}{\pi f\,\Delta x}\sin^{-1}\left[\sinh(\gamma_m(\omega)\frac{\Delta x}{2})\right] . \qquad (3.2.17)$$

If it assumed that the attenuation across the entire array is substantial, but that there are enough stations to insure that $\gamma_m(\omega)\Delta x \ll 1$, then (3.2.17) reduces to

$$\text{width}=\frac{\gamma_m}{\pi f} . \qquad (3.2.18)$$

The definition of attenuation in terms of spatial Q (Aki and Richards, 1980, p. 298) is

$$\gamma_m=\frac{\pi f}{c_m Q_m} .$$

Substituting this into (3.2.18) leads to the rather elegant result:

$$\text{width}=\frac{1}{c_m Q_m} . \qquad (3.2.19)$$

In terms of phase velocity limits for the width, this is

$$c_{\lim} = \frac{c_m}{1 \pm \frac{1}{2Q_m}} \quad . \tag{3.2.20}$$

Figure 3.1 illustrates the effect of attenuation on a representative resolving function $|\mathrm{Wm}|$. It should be emphasized that when regions of high attenuation and frequencies are investigated (such as found in very shallow crustal structure: $Q_m < 30$ and $f > 1\ hz$), the approximations leading to (3.2.19) can be quite valid.

3.2.2 Attenuation measurement and velocity errors

P-ω stacking has the distinct theoretical advantage that high resolution of multiple surface wave modes is possible, given enough channels and small inter-channel spacing. This is clear from the periodicity condition (3.2.9) and the resolving function width (3.2.13). However, the effect of attenuation limits the maximum resolution obtainable. If the total array width is large enough, or the attenuation is high enough, the resolving function width approaches a minimum limit given by (3.2.19), which is independent of the number of channels or channel separation. In addition, a fundamental assumption in stacking is that the phase velocity is constant for a given frequency and mode, which is true only if the propagation medium is laterally homogeneous. Therefore, it is expected that velocity error will increase as a function of the total array width when actual data are considered. This will be reflected in the stack as a deviation from the theoretically expected shape of the resolving function (3.2.8). Random changes in velocity structure across a seismic array can cause the resolving function defined in the previous section to "blur", degenerating into incoherent noise if the variations are severe. Another

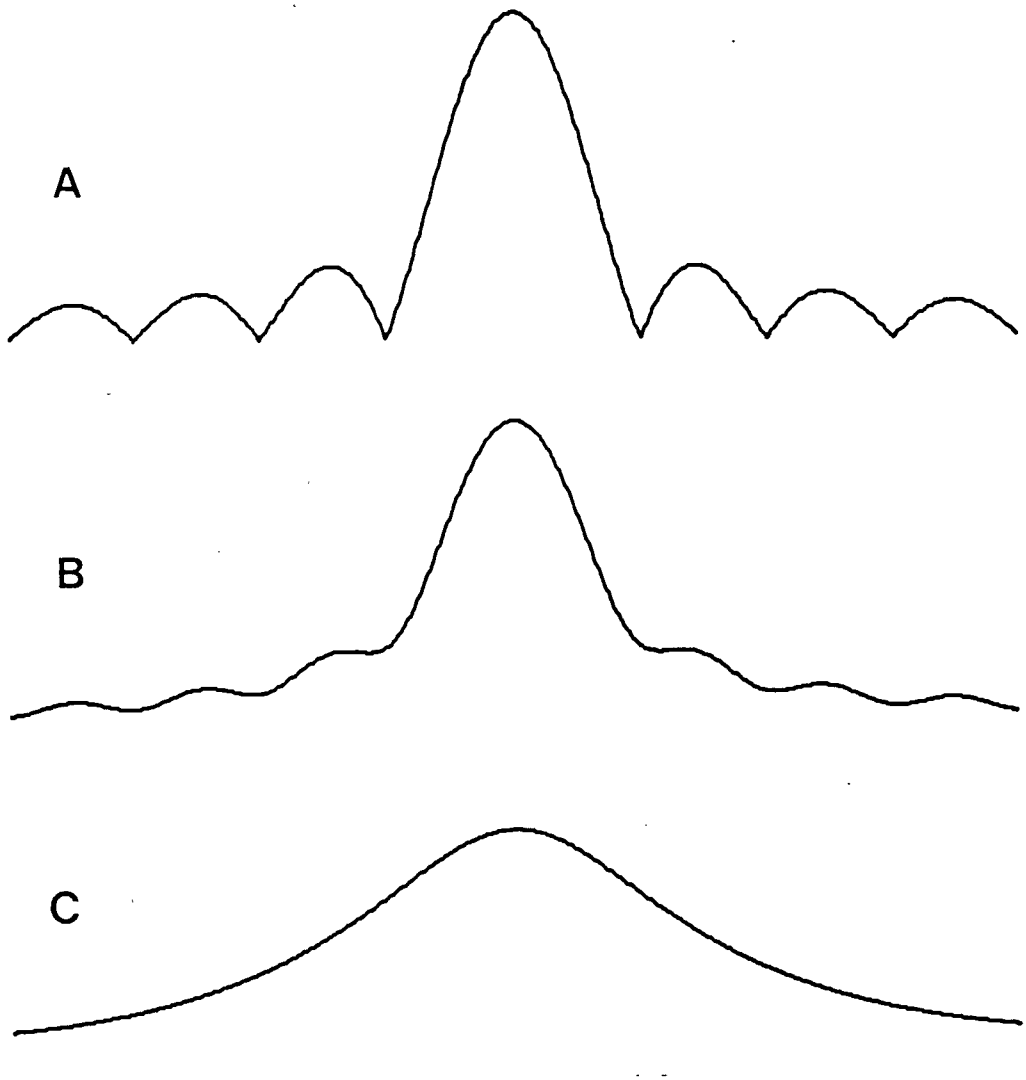

Figure 3.1 P-ω stacking kernel with attenuation. (a) No attenuation. (b) γ=.2. (c) γ=.8.

possibility is the splitting of the resolving function into two or more coherent impulses, due to simple velocity contrasts across the array (normal or reverse faulting could be a typical cause of this). The following analysis will attempt to remove such effects, under the assumption that there is enough coherency in the stack to allow modal separation, and that the envelope maximum of the resolving function defines the average modal wave slowness ($p_m(\omega)$) across the array. The result will be a measure of the error in wave slowness about the maximum, and an estimate of the wave attenuation at a given frequency and mode. In addition, within certain resolution constraints, an estimate for lateral phase velocity variations across the array can be found.

The basic method is to center a Gaussian window about the maximum of the complex resolving function, followed by a Fourier transform over wavenumber. Asymmetries in the resolving function will result in deviations from zero phase in the spatial domain. The phase errors, evaluated at each station location, can be used as correction factors in the stack. Unfortunately, the presence of significant attenuation will cause these phase errors to be evaluated incorrectly. The analysis will show how analytic continuation of the stack in terms of the attenuation variable $\gamma_m(\omega)$ (Buland and Gilbert, 1978) can correct both phase errors and attenuation distortions, resulting in a stacked waveform close to the ideal Fourier kernel (3.2.11).

Assuming an isolated mode m, define the complex p-ω stack (3.2.3) in terms of wavenumber (at a constant frequency) as:

$$F(k) = A_m \sum_n e^{-\gamma_m(\omega)x_n} e^{ikx_n} e^{i\delta\phi_n} \qquad (3.2.24)$$

where $k = \omega(p - p_m)$. The wavenumber is defined such that the maximum of the stacked signal occurs at zero lag ($k = 0$). The term $\delta\phi_n$ signifies phase errors recorded at each station due to lateral changes in phase velocity across the array. For convenience, let the constant source phase associated with A_m be included with $\delta\phi_n$, making A_m a real constant. Window the stack with a Gaussian exponential function and Fourier transform the function over wavenumber:

$$g(x) = \int_{-k_c}^{k_c} e^{-\alpha k^2} \left(A_m \sum_n e^{-\gamma_m(\omega)x_n} e^{ikx_n} e^{i\delta\phi_n} \right) e^{-ikx}\, dk \quad .$$

It is assumed that the integral contribution is negligible outside the limits $|k| \geq k_c$. Rearranging gives

$$g(x) = A_m \sum_n e^{i\delta\phi_n} e^{-\gamma_m(\omega)x_n} \int_{-k_c}^{k_c} e^{-\alpha k^2} e^{ik(x_n - x)} dk \quad . \qquad (3.2.25)$$

The integral can be rewritten as

$$I = e^{-\frac{1}{4\alpha}(x_n - x)^2} \int_{-k_c}^{k_c} e^{-\alpha\left[k - i\frac{1}{2\alpha}(x_n - x)\right]^2} dk \quad .$$

If the contribution outside the limits is negligible, then (Herrmann, 1973):

$$I = \left(\frac{\pi}{\alpha}\right)^{\frac{1}{2}} e^{-\frac{1}{4\alpha}(x_n - x)^2} \quad .$$

Substitute into (3.2.25) for

$$g(x) = \left(\frac{\pi}{\alpha}\right)^{\frac{1}{2}} A_m \sum_n e^{i\delta\phi_n} e^{-\gamma_m(\omega)x_n} e^{-\frac{1}{4\alpha}(x_n - x)^2} \quad .$$

Algebraic manipulation results in

$$g(x) = \left(\frac{\pi}{\alpha}\right)^{\frac{1}{2}} A_m \; e^{(\alpha\gamma_m^2 - \gamma_m x)} \sum_n e^{i\delta\phi_n} e^{-\frac{1}{4\alpha}\left[x_n - (x - 2\alpha\gamma_m)\right]^2} \quad . \tag{3.2.26}$$

To interpert (3.2.26), first assume $\gamma_m = 0$. Since the terms outside the sum are real, the phase calculated at x will be a weighted average of the phases $(\delta\phi_n)$ recorded at adjacent station locations. When $\gamma_m \neq 0$, the same interpretation is valid, except that the phase function has been shifted to a position of decreasing x by an amount $-2\alpha\gamma_m$. This can be represented as

$$g(x) = |g(x)| e^{i\delta\hat{\phi}(x - 2\alpha\gamma_m)} \quad , \tag{3.2.27}$$

where $\delta\hat{\phi}$ is the smoothed estimate of phase at the position indicated in parentheses. Notice that as a function of attenuation, the entire process outlined above need only be calculated once. For different values of attenuation, the phase estimate is merely shifted along the x axis by the amount indicated.

This leads to an algorithm for determining the attenuation of the m'th mode at a given frequency. The original stack (3.2.24) can be analytically continued by multiplying each element in the sum by

$$e^{\gamma x_n} \quad e^{-i\delta\psi_n(\gamma)}$$

where

$$\delta\psi_n(\gamma) = \delta\hat{\phi}[x_n + 2\alpha\gamma] \quad .$$

In this equation γ is considered an independent variable. The phase estimate is found as follows: first calculate the smoothed estimate using the Fourier transform over k. Then, for successive values of γ, pick the phase at the position $x_n + 2\alpha\gamma$. Notice that when $\gamma = \gamma_m$ the phase will no longer be influenced by attenuation. Multiplying (3.2.24) by the above

gives:

$$F(k,\gamma) = A_m \sum_n e^{ikx_n} \; e^{(\gamma - \gamma_m)x_n} \; e^{i[\delta\phi_n - \delta\psi_n(\gamma)]} \quad . \qquad (3.2.28)$$

When $\gamma = \gamma_m$, distortions due to attenuation and phase error should be minimized, and the stack should approach the theoretical sum (3.2.11), assuming equally spaced stations. Therefore, the correct attenuation can be found by searching for the value of γ that minimizes the residual between the observed and theoretical stack.

Once the correct attenuation has been found, the estimated phases can be re-evaluated to determine the spatial distribution of wavenumber error. First, a phase unwrapping algorithm must be employed to remove 2π discontinuities from the phase estimate $\delta\hat{\phi}$. Methods of phase unwrapping will be discussed in detail in a later chapter. This is necessary to define the interstation wavenumber error:

$$\delta k_n = \frac{d\,\delta\hat{\phi}}{dx} \quad . \qquad (3.2.29)$$

This is a differential approximation to the assumption that phase error across the array is a cumulative function of the wavenumber error:

$$\delta\phi(x) = \int_0^x \delta k(x)dx \quad .$$

Equation (3.2.29) involves numerical differentiation, which can tend to be a noisy process, but this should not be a problem in this case since the phase estimates ($\delta\hat{\phi}$) have been smoothed by a Gaussian filter. The interstation wavenumber errors can either be used to correct the original wave slowness estimates (ωp_m) as a function of distance across the array, or they can provide an estimate of standard error in wave slowness:

$$\delta p_m = \frac{1}{\omega}\left(\frac{\sum \delta k_n^2}{N-1}\right)^{\frac{1}{2}} \tag{3.2.30}$$

Papoulis (1965, p. 151) shows how to translate (3.2.20) into an approximate phase velocity error:

$$\delta c_m = \frac{1}{p_m^2}\delta p_m = c_m^2 \delta p_m \quad . \tag{3.2.31}$$

It is interesting to note that the velocity error is proportional to the square of the measured phase velocity.

3.2.3 Discussion

In order to give dimensional perspective to the above analysis, a convenient unit of measure is the bandwidth of the major lobe of the ideal p-ω stack (3.2.13). In terms of wavenumber this will be:

$$B = \frac{4\pi}{L} \quad , \qquad L = N\Delta x \quad . \tag{3.2.32}$$

L will approximate the total width of the array ($[N-1]\Delta x$) when N is large. In the k domain, the width of the Gaussian window in (3.2.25) is defined as $2k_c$, enclosing the total area between the cutoff limits. A measure of the limit k_c is

$$e^{-\alpha k_c^2} = e^{-\pi} \approx .04 \quad . \tag{3.2.33}$$

Let N_b be the number of fundamental bandwidths (3.2.32) equal to the width of the Gaussian window:

$$N_b \frac{4\pi}{L} = 2k_c \quad . \tag{3.2.34}$$

Substituting (3.2.34) into (3.2.33) gives

$$\alpha \left(\frac{2\pi N_b}{L} \right)^2 = \pi \ ,$$

or:

$$\alpha = \frac{L^2}{4\pi N_b^2} \tag{3.2.35}$$

The resolution width of the Gaussian filter in the spatial domain (3.2.26) can be defined as the width of the approximate half-amplitude points:

$$e^{-\frac{1}{4\alpha}x^2} = e^{-\frac{\pi}{4}} \ . \tag{3.2.36}$$

Solving for $R = 2x$ in (3.2.36) and using (3.2.25) gives the filter resolution width:

$$R = \frac{L}{N_b} \ . \tag{3.2.37}$$

If the interstation wavenumber error (3.2.29) is used for spatial corrections to the phase velocity dispersion, (3.2.37) gives the finest resolution possible for the correction. As an example, if the array length is 50 kilometers long, and 2 bandwidths are used to construct the Gaussian window in the k domain, it is pointless to look for spatial detail finer than 25 kilometers in width. Notice that the periodicity condition (3.1.14) limits the maximum number of bandwidths to $N/2$, where N is the total number of stations. If multiple modes are present , this number may have to be considerably reduced to avoid modal interference.

Finally, if the Gaussian filter is only used for wavenumber errors, and attenuation measurements are not desired, the analytic continuation search for correct attenuation (3.2.28) may not be necessary. In terms of resolution, the distance error in spatial calculations of phase from (3.2.27)

is, (3.2.35), and (3.2.37):

$$x_{error} = 2\alpha\gamma_m = \frac{R^2\gamma_m}{2\pi}. \qquad (3.2.38)$$

If an upper bound for γ_m is estimated in (3.2.28), and if the calculated distance error is well within the resolution estimate R, the error produced by attenuation can be ignored since it cannot be resolved.

3.3 Summary

Theoretically, p-ω stacking is the most direct way of determining the dispersive characteristics of a wave field. It involves a direct reconstruction of the dispersed wave in the frequency-wavenumber domain which is uncontaminated by source phase effects. It was shown above that attenuation and lateral variations in wavenumber can also be found. However, p-ω stacking is subject to the same problems of aliasing as discrete Fourier transforms, since it constructs the wavefield as a sum of wavenumber harmonics (McMechan and Yedlin, 1981). Insufficient sampling in the spatial domain and unevenly spaced sampling locations can cause gross errors in the stack. In addition, large ambient noise levels can cause spurious peaks in the wavenumber-frequency domain, making determination of maximum wavenumber values difficult or impossible. Standard error analysis is possible (equation 3.2.30), but this requires estimating the center location of the wavenumber peak, which may be difficult in cases of low signal-to-noise ratios.

To overcome such difficulties, the approach of phase-matched filtering will be discussed, first in the context of a single-station method, and then generalized to multi-channel processing.

CHAPTER 4

PHASE-MATCHED FILTERING

4.1 Single channel phase-matched filtering

4.1.1 Theory

Phase-matched filtering (Herrin and Goforth, 1977) is a method of compressing, in the time domain, the energy of a particular mode of interest in a seismogram. This is done by removing the phase of a desired mode, resulting in a zero-phase signal with energy concentrated about zero-lag in the time domain. By time windowing this isolated mode, the effects of incoherent noise (random) and coherent noise (multi-pathing, higher modes, extraneous signals) can be removed from the spectrum. A necessary assumption for the matched filter is that the desired mode has a frequency bandwidth broad enough to result in a narrow time-domain signal.

Herrin and Goforth (1977) refer to the zero-phase time-domain signal as the "pseudo-autocorrelation function." Assuming that the seismogram is composed of propagating normal modes (3.1.4), the phase-matched filter can be expressed as

$$\psi_j(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \{e^{i\tilde{k}_j x}\} \sum_m A_m e^{i(\omega t - k_m x)} \, d\omega \ , \qquad (4.1.1)$$

where it is understood that A_m, k_m, and $\tilde{k}_j$ are functions of frequency. For the present, assume that both source and instrument have been removed from A_m, making it a real function. The term in brackets is the phase-matched filter, and $\tilde{k}_j$ is an estimate of wavenumber dispersion of

the j'th mode of interest. At this point, the filter need only approximate the j'th mode; it will be shown how this estimate can be iteratively improved. $\psi_j(t)$ is the time-domain pseudo-autocorrelation function. It can be rewritten as

$$\psi_j(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} A_j\, e^{i(\tilde{k}_j - k_j)x}\; e^{-i\omega t}\, d\omega \qquad (4.1.2)$$

$$+ \frac{1}{2\pi}\int_{-\infty}^{\infty} \sum_{m\neq j} A_m\; e^{i(\omega t - k_m x)} d\omega.$$

If $\tilde{k}_j \approx k_j$ the first integral will be approximately zero-phase and should be concentrated about zero-lag (provided it has sufficient bandwidth). Time windowing the first integral with a symmetric, zero-phase window $w(t)$ will remove the effects of other modes and noise, provided they are spatially separated from the mode of interest:

$$\psi_j(t)w(t) = w(t)\cdot\frac{1}{2\pi}\int_{-\infty}^{\infty} A_j\, e^{i\delta k\, x}\; e^{-i\omega t}\, d\omega, \qquad (4.1.3)$$

where

$$\delta k = \tilde{k}_j - k_j \quad . \qquad (4.1.4)$$

Taking the Fourier transform of the windowed pseudo-autocorrelation function will result in

$$A_j e^{i\delta k\, x} = \int_{-\infty}^{\infty} \psi_j(t)\; w(t) e^{-i\omega t}\, dt. \qquad (4.1.5)$$

Equation (4.1.5) is now the isolated amplitude spectrum of the desired mode. Notice that there may be a residual phase error ($\delta k\; x$), depending on the accuracy of the initial estimate of the wavenumber dispersion. A new estimate can be found using (4.1.4):

$$\tilde{k}_j^{new} = \tilde{k}_j - \delta k. \qquad (4.1.6)$$

The new value (4.1.6) can be substituted into (4.1.2) and the process repeated, resulting in a new estimate of the isolated spectrum (4.1.5). This can be continued until there is no further decrease in the residual phase (4.1.4).

In this manner a precise estimate of the wavenumber spectrum can be found, assuming that the original estimate ($\tilde{k}_j$) is reasonably close to the true mode. The amplitude spectrum (A_j) may be biased, however, due to the distorting effects of the particular window used. The effect of window bias on both phase and amplitude will be discussed in detail in section 4.1.3.

4.1.2 Multiple filter analysis

In order to find an initial estimate of wavenumber, Herrin and Goforth (1977) used the multiple filter technique (MFT). This method determines group velocities (3.1.3) in dispersed wavetrains, and only a single receiver is necessary for the analysis. Complete descriptions are given by Dziewonski *etal* (1969) and Herrmann (1973), and briefly reviewed here.

Given a seismogram composed of a sum of normal modes, let the complete spectrum be windowed at frequency ω_0 by a narrow Gaussian function and then Fourier transformed to the time domain to give a complex time signal $g(t)$:

$$g(t) = \int_{\omega_0-\omega_c}^{\omega_0+\omega_c} H(\omega-\omega_0) \sum_m A_m(\omega)\, e^{i(\omega t - k_m(\omega)x)}\, d\omega\,,$$

where

$$H(\omega) = \begin{cases} \exp(-\alpha\omega^2/\omega_0^2) & |\omega| \leq \omega_c \\ 0 & |\omega| > \omega_c \end{cases} .$$

Herrmann (1973) showed that the modulus of $g(t)$ is approximately

$$|g(t)| = \frac{\omega_0}{2\pi}\left(\frac{\pi}{\alpha}\right)^{\frac{1}{2}} \sum_m A_m(\omega) \exp\left[-\frac{\omega_0^2}{4\alpha}\left(t - \frac{x}{U_{0j}}\right)^2\right] , \quad (4.1.7)$$

where U_{0j} is the group velocity of the j'th mode at frequency ω_0 . Assuming sufficient modal separation, maxima of $|g(t)|$ will occur when

$$t_j = \frac{x}{U_{0j}}.$$

Since the distance from the source x is known, group velocities can be calculated as a function of frequency ω_0 .

Let the phase of the j'th mode be

$$\phi_j = k_j\, x. \quad (4.1.8)$$

Differentiating (4.1.8) gives

$$\frac{d\phi_j}{d\omega} = x\,\frac{dk_j}{d\omega} = \frac{x}{U_j} = t_j \ . \quad (4.1.9)$$

This demonstrates that differentiating the phase spectrum of an isolated mode gives the group delay (t_j) of that mode, which can be calculated directly from MFT (4.1.7). Reversing the process results in

$$\phi_j = \int_0^\omega t_j\, d\omega. \quad (4.1.10)$$

Equation (4.1.8) can then be used to calculate the initial estimate of wavenumber for use in the phase-matched filter. In practice, the integral (4.1.10) can have an additive constant without affecting the derivative

group delay. It is therefore quite possible for the initial wavenumber ($\tilde{k}_j$) to be in error by a large constant value. Fortunately, the next section will show that the windowed pseudo-autocorrelation function is not biased by constant phase errors. The residual phase in (4.1.5) can therefore be used to accurately compensate for this error. However, notice that the correction will only be valid to the nearest multiple of 2π. Equation (4.1.5) gives an infinity of possible wavenumbers, since

$$\exp(i\,\delta k\ x) = \exp(i\,\delta k\ x \pm i2n\pi)\,, \qquad n = 0,1,2,\ldots\ldots\ .$$

This is a fundamental ambiguity which is due to spatial sampling by a single station, and requires *a priori* knowledge of the location of the wavenumber at least at one frequency. This can be resolved with multichannel processing, which will be discussed in a later section.

4.1.3 Window bias

In the process of calculating the Fourier transform from (4.1.3) to (4.1.5), it is assumed that the window $w(t)$ does not distort the spectrum of the j'th mode. This is not true in practice, and the purpose of this section is to approximately calculate the bias in the frequency domain due to time domain windowing. The analysis is similar to Jenkins and Watts (1968, p. 247). Define the "pure" pseudo-autocorrelation function as the the single mode signal uncontaminated by noise:

$$\psi_j(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} A_j\, e^{i\,\delta k\ x}\ e^{-i\omega t}\, d\omega. \tag{4.1.11}$$

The bias in the frequency domain will be the transformed difference between the windowed pseudo-autocorrelation function and (4.1.11):

$$B(\omega) = \int_{-\infty}^{\infty} (w(t) - 1)\psi_j(t) e^{-i\omega t}\, dt. \tag{4.1.12}$$

At this point it is necessary to define the particular window desired. Harris (1978) reviews a broad class of spectral windows and their characteristics. One measure of window performance is the maximum sidelobe level. For a given window in the time domain, construct the Fourier transform and measure the amplitude of the maximum sidelobe. This gives an indication of the "smoothness" or convolutional rippling effect in the frequency domain. Windows with high sidelobes can cause distortional rippling in the frequency domain when there is a significant truncation of signal (or noise) in the time domain by the window. The rectangular window is the worst case, with a maximum sidelobe -13 dB below (almost 1/2) the main lobe. However, Jenkins and Watts (1968) show that windows that tend toward the rectangular have the least signal bias, under the condition that there is no significant truncation of the signal in the time domain.

Two windows are examined for bias: the Parzen (also called de la Valle - Poisson), and cosine windows. The cosine window is defined as:

$$w_c(t) = \begin{cases} \cos(\pi t/2M) & |t| \leq M \\ 0 & |t| > M \end{cases} \tag{4.1.13}$$

where M is the one-sided width of the cosine window. The maximum sidelobe level for the cosine window is -23 dB (7 percent) of the main lobe. The Parzen window is defined as:

$$w_p(t) = \begin{cases} 1 - 6(t/M)^2 + 6(|t|/M)^3 & |t| \leq M/2 \\ 2(1 - |t|/M)^3 & M/2 < |t| \leq M \\ 0 & |t| > M \end{cases} \tag{4.1.14}$$

where M is the one-sided width of the Parzen window. The maximum

sidelobe level for the Parzen window is -53 dB (.2 percent) of the main lobe. For a given width M, the Parzen window is a much smoother convolutional filter in the frequency domain due to low sidelobes.

To calculate the approximate bias, substitute the windows into (4.1.12) and keep terms only on the order of $1/M^2$ or more. For the cosine window,

$$B_c(\omega) = \frac{\pi^2}{8M^2} \int_{-\infty}^{\infty} -t^2 \, \psi_j(t) \, e^{-i\omega t} \, dt \quad + \quad O(1/M^4)$$

and the Parzen window,

$$B_p(\omega) = \frac{6}{M^2} \int_{-\infty}^{\infty} -t^2 \, \psi_j(t) \, e^{-i\omega t} \, dt \quad + \quad O(1/M^3) \, .$$

Notice that the infinite limits are kept for the integrals. This is with the assumption that M is wide enough to insure that there is insignificant truncation of the pseudo-autocorrelation function, so there is negligible signal outside the limits $+M$, $-M$. Making use of Fourier transform properties of differentiation gives (Papoulis, 1962, p. 16):

$$B_c(\omega) = \frac{\pi^2}{8M^2} \Psi''_j(\omega) \tag{4.1.15}$$

$$B_p(\omega) = \frac{6}{M^2} \Psi''_j(\omega) \tag{4.1.16}$$

where the double primes indicate the second derivative with respect to angular frequency of the pseudo-autocorrelation spectrum. Taking the ratio of (4.1.16) to (4.1.15) shows that the Parzen window has almost five times the bias of the cosine window, which appears to indicate that it may be a poor choice of a windowing function. However, further investigation of the spectral second derivative yields interesting results.

Define the spectrum from (4.1.11) as

$$\Psi_j(\omega) = A_j e^{i\delta k\, x} = \alpha e^{i\theta} . \tag{4.1.17}$$

The terms α and θ are chosen for notational convenience. If multiple-filter analysis is used to derive an initial estimate of phase (4.1.10), the first and second derivatives of ($\delta k \cdot x$) should be small since the first derivative of k_j is equal to the residual group delay. However, δk itself can be large due to the additive constant in integrating the group delay. Therefore, let θ be indeterminate, let θ' be small, and let θ'' also be small. Calculate the second derivative of the pseudo-autocorrelation spectrum for:

$$\Psi''_j(\omega) = [\alpha'' - \alpha(\theta')^2]e^{i\theta} + [\alpha\theta'' + 2\alpha'\theta']e^{i(\theta + \pi/2)} . \tag{4.1.18}$$

Using vector analysis to calculate the bias in amplitude and phase, and keeping only first order terms in θ' and θ'' results in the following expressions for the Parzen window. The amplitude bias is

$$B_{p\alpha} = \frac{6}{M^2}\,\alpha'' \tag{4.1.19}$$

and the phase bias is

$$B_{p\theta} = \frac{6(\alpha\theta'' + 2\alpha'\theta')}{M^2\alpha + 6\alpha''} . \tag{4.1.20}$$

Similar expressions for the cosine window are, for amplitude bias:

$$B_{c\alpha} = \frac{\pi^2}{8M^2}\,\alpha'' , \tag{4.1.21}$$

and for phase bias:

$$B_{c\alpha} = \frac{\pi^2(\alpha\theta'' + 2\alpha'\theta')}{8M^2\alpha + \pi^2\alpha''} . \tag{4.1.22}$$

Equations (4.1.20) and (4.1.22) show that the bias in phase is independent of constant phase values, and that if the first and second derivatives are small, the bias will also be small. Therefore, it is advisable to use smooth (low sidelobe) windows such as the Parzen for calculating residual phases in the matched filtering process. However, it should be clear that spectral amplitudes may be quite biased when there is significant curvature in the amplitude spectrum, for example, a narrow band spectrum or in the vicinity of sharply changing band edges. The fact that phase-matched filtering is an iterative process can remove residual bias in phase, but this will not help the bias in the amplitude spectrum, since it is independent of phase.

4.2 Multi-channel phased-matched filtering

4.2.1 Introduction

The analysis presented in the following section represents the heart of this dissertation. The concepts of p-ω stacking and single channel phase-matched filtering will be combined to form a new method of multi-channel processing, which will exploit salient features of the above techniques. In addition, statistical parameters from bivariate spectral analysis will be utilized to improve signal-to-noise measurements across the channels.

In section 3.2, it was shown how p-ω stacking, followed by frequency transformations and Gaussian filtering, could be used to determine phase velocity errors (3.2.30). However, this requires a determination of the mean phase velocity from an inspection of the maxima of a frequency - wave slowness plot, which can be inaccurate in the presence of large

ambient noise levels. It will be shown that multi-channel matched filtering can give both mean phase velocity and error estimates in the presence of extremely high incoherent noise levels, which can be quite useful when short array lengths or a small number of channels are used.

This is not to say that multi-channel phase-matched filtering is a replacement for p-ω stacking. The multi-channel method is excellent for identifying incoherent noise, but it cannot be used for mode identification. It suffers from the problem of single channel phase-matched filters, in that an initial estimate of the mode of interest must be supplied, in order for a pseudo-autocorrelation function to be constructed in the time domain. P-ω stacking can complement the multi-channel filter by delineating which modes are present in a signal, and by providing a starting estimate of the mode of interest to the filter.

4.2.2 General theory

Assume a linear array of N seismic stations, with unequal spacing possible between channels. Let the complex frequency response of each channel to a source function be

$$X_j(\omega)\,, \quad \text{where } j = 1,2,3..N\,.$$

The interstation Green's function between any two channels is defined as:

$$H_{ij}(\omega) = \frac{X_j(\omega)}{X_i(\omega)} = \left|\frac{X_j(\omega)}{X_i(\omega)}\right| e^{-i\,k(\omega)(x_j - x_i)} \qquad (4.2.1)$$

$H_{ij}(\omega)$ represents the response at station j to an impulse at station i, assuming a linear medium of propagation. This is for the noise-free case, where it is assumed that the signals $X(\omega)$ are single modes propagating

according to (3.1.1). For all combinations of the stations, the following Green's function matrix can be constructed:

$$\mathbf{G} = \begin{bmatrix} H_{11} & H_{12} & . & . & H_{1N} \\ H_{21} & H_{22} & & & . \\ . & & . & & . \\ . & & & . & . \\ H_{N1} & . & . & . & H_{NN} \end{bmatrix} . \qquad (4.2.2)$$

This matrix has several interesting features. The diagonal elements are zero-phase with amplitudes identically equal to one, representing Dirac impulses in the time domain. Each row represents the medium response to impulse sources at each channel location. On each row, the terms to the right of the diagonal are the Green's function medium response to the impulse on the diagonal. The terms to the left of the diagonal have only a mathematical meaning, in that they represent non-causal contributions to the diagonal in negative time. Notice that the true source has been totally removed by the Green's function operator (4.2.1). It simply divides out, being common to all signals $X_j(\omega)$. By the same reasoning, matched instrument responses are removed by the Green's function. Therefore, the elements of (4.2.2) represent true interstation transfer functions for each pair of channels.

In terms of the distance x_j as an independent variable, the phase response along the i'th row can be defined from (4.2.1) as

$$\phi_{ij}(\omega) = k(\omega)(x_j - x_i) \ . \qquad (4.2.3)$$

For a given frequency, (4.2.3) defines a straight line in terms of x_j, with the wavenumber $k(\omega)$ equal to the negative slope of the line. Notice that the slope remains the same for each row of (4.2.2); only the intercept

changes. Therefore, in the noise-free case, the phase velocity across the array can be solved for by first finding the wavenumber as a function of frequency for any of the rows of (4.2.2), and then using (3.1.2):

$$c(\omega) = \frac{\omega}{k(\omega)} .$$

In the presence of noise, the most general method of processing the signals $X_j(\omega)$ is multivariate spectral analysis (Jenkins and Watts, 1968, p. 458). The Green's function matrix (4.2.2) is the frequency domain equivalent of the impulse response matrix defined by Jenkins and Watts (1968, p. 469), and can be processed directly for gain and phase estimates using multivariate techniques. However, the number of calculations involved in multivariate processing can grow unacceptably large for even moderately sized seismic arrays, so an alternate method is preferable. Fortunately, the seismic phase response is controlled by a simple linear model (4.2.3) across the array, and by taking advantage of this assumption, it will be shown that the multivariate analysis can be reduced to a sequence of bivariate (two channel) processes followed by linear regression. In addition, by utilizing phase-matched filtering across the array, it is possible to isolate modes of interest from conflicting coherent signals, such as multipathing and higher modes.

4.2.3 Matched filtering and bivariate statistics

Assume that the signals $X_j(\omega)$ are contaminated by both random noise and interfering coherent signals. Assume also that an initial estimate $\tilde{c}(\omega)$ of the phase velocity of a mode of interest is given, either from average group velocities across the array, or from p-ω stacking. The interstation phase estimate of the mode can then be defined as

$$\tilde{\phi}_{ij}(\omega) = \frac{\omega}{\tilde{c}(\omega)}(x_j - x_i) \quad . \tag{4.2.4}$$

Notice that $\tilde{c}(\omega)$ is not spatially dependent, implying lateral homogeneity across the dimensions of the array. As will be shown, lateral variations will be incorporated as standard errors about the mean velocity.

Define the interstation cross-correlation function as

$$X_{ij}(\omega) = X_j{}^*(\omega)\, X_i(\omega) \, .$$

The star (*) superscript indicates taking the complex conjugate of X . The interstation pseudo-autocorrelation function can then be defined as in single channel processing (4.1.1):

$$\psi_{ij}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \{e^{i\tilde{\phi}_{ij}(\omega)}\}\, X_{ij}(\omega) e^{i\omega t}\, d\omega$$

where $\tilde{\phi}_{ij}(\omega)$ is defined by (4.2.4), and the term in brackets is the phase-matched filter. To isolate the mode of interest, window the pseudo-autocorrelation function, transform back to the frequency domain, and reset the phase:

$$\overline{X}_{ij}{}^P(\omega) = \{e^{-i\tilde{\phi}_{ij}(\omega)}\} \int_{-\infty}^{\infty} w(t)\psi_{ij}(t) e^{-i\omega t}\, dt \quad . \tag{4.2.5}$$

This defines the interstation phase-matched cross-correlation function. The bar indicates smoothing by the time window $w(t)$ and the superscript P indicates phase-matched filtering.

Define the interstation auto-correlation function as:

$$X_{ii}(\omega) = X_i{}^*(\omega)\, X_i(\omega) \, .$$

The smoothed interstation auto-correlation function is calculated by windowing X_{ii} in the time domain with $w(t)$, and then transforming back to

the frequency domain.

$$\psi_{ii}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_{ii}(\omega) e^{i\omega t} \, d\omega$$

$$\overline{X}_{ii}(\omega) = \int_{-\infty}^{\infty} w(t) \, \psi_{ii}(t) \, e^{-i\omega t} \, dt \; . \tag{4.2.6}$$

Notice that no matched filtering is necessary for the auto-correlation function since it is by definition a zero-phase signal.

Using (4.2.5) and (4.2.6), the phase-matched interstation Green's function is defined by

$$\overline{H}_{ij}{}^{P}(\omega) = \frac{\overline{X}_{ij}{}^{P}(\omega)}{\overline{X}_{ii}(\omega)} \tag{4.2.7}$$

Equation (4.2.7) is similar to the smoothed frequency response defined by Jenkins and Watts (1968, p. 432). It differs in that phase-matched filtering is incorporated into the above response. It should be noted that Jenkins and Watts (1968, p. 399) define a rudimentary form of a phase-matched filter, which they refer to as "phase alignment", calculated by a linear phase shift of the peak of the cross-correlation function in the time domain to zero lag. They use phase alignment to reduce the bias (4.1.20) in phase measurements of the windowed cross-correlation function. The present analysis extends this alignment to a complete phase-matched filter, in order to isolate surface wave normal modes.

Equation (4.2.7) can now be processed as in the noise-free case, using (4.2.1) through (4.2.3). In addition, statistics associated with bivariate processing can be used to weight the regression algorithm for each row of (4.2.2). The fundamental parameter for distinguishing between signal and

noise in bivariate analysis is the smoothed coherency spectrum, defined as

$$\overline{K}_{ij}^{\,2}(\omega) = \frac{|\overline{X}_{ij}^{\,P}(\omega)|^2}{\overline{X}_{ii}(\omega)\overline{X}_{jj}(\omega)} \quad . \tag{4.2.8}$$

The coherency spectrum varies between zero and one for any frequency ω, with maximum coherency (or noise-free signal) existing when $\overline{K}_{ij}^{\,2}(\omega)$ equals one.

Windowing with $w(t)$ in (4.2.5) and (4.2.6) plays an essential role in the ability of the coherency spectrum to distinguish between signal and noise. If the window is infinitely wide (or not used), implying no smoothing, it can be seen by inspection that (4.2.8) will identically equal one for all frequencies, which implies that time-domain windowing helps isolate signal from noise. Notice also that the phase-matched cross-correlation operator (4.2.5) selects only modes of interest to be windowed at zero-phase, while the auto-correlation function (4.2.6) sets the entire record to zero-phase. Therefore, unless the original record contains only the single mode of interest, the denominator of (4.2.8) will tend to have more energy than the numerator, causing the coherency to be less than one.

Utilizing the coherency spectrum, a 95% confidence interval can be calculated for the interstation phase spectrum (Jenkins and Watts, 1968, p. 435):

$$C_{ij}(\omega) = \sin^{-1}\left\{\left[\frac{2}{\nu-2}\, f_{2,\nu-2}(.95)\, \frac{1-\overline{K}_{ij}^{\,2}(\omega)}{\overline{K}_{ij}^{\,2}(\omega)}\right]^{\frac{1}{2}}\right\}, \tag{4.2.9}$$

Where ν is the number of degrees of freedom for the windowed cross-correlation function (4.2.5). For a Parzen window ν is defined as

$$\nu = 3.71\frac{T}{M} \quad , \tag{4.2.10}$$

where T is the total length of the record and M is the one-sided width of the Parzen window. Degrees of freedom for other window types are defined in Jenkins and Watts (1968, p. 252). The f term is the Fischer F_{ν_1,ν_2} distribution evaluated for 2 , ν–2 degrees of freedom at a 95 % probability point (Jenkins and Watts, 1968, p. 84).

To calculate standard deviations for phase, Berger *et al.* (1979, p. 278) used the following approximation with excellent results:

$$\sigma_{ij}(\omega) = \frac{C_{ij}(\omega)}{1.96} \quad . \tag{4.2.11}$$

The same method is used here, except when the coherency is quite low. As the coherency becomes small it is possible for the term in brackets in (4.2.9) to be greater than one, making it impossible to calculate the inverse sine function. A more robust, but less accurate estimate of the phase standard deviation is (Jenkins and Watts, 1968, p. 379):

$$\sigma_{ij}(\omega) \approx \left[\frac{\alpha M}{2T}\,\frac{1-\overline{K}_{ij}{}^2(\omega)}{\overline{K}_{ij}{}^2(\omega)}\right]^{\frac{1}{2}} \quad . \tag{4.2.12}$$

The constant α is determined by the type of window used. In this case it is calculated by insuring continuity between (4.2.11) and (4.2.12) when the term in brackets in (4.2.9) is equal to one. The term M/T has the same meaning as in (4.2.10).

4.2.4 Regression analysis

Given the successful modal isolation of a signal by the phase-matched interstation Green's function (4.2.7), it is now possible to treat

the phase across the array in terms of a linear regression problem, as in the noise-free case (4.2.3). Prior regression variances can be provided by the coherency estimates (4.2.11) and (4.2.12). Using (4.2.3), the regression parameters are set up as follows:

$$\phi_{ij} = k_i \; x_{ij} + b_i \tag{4.2.13}$$

where

ϕ_{ij} = the measured interstation phase from (4.2.5)

k_i = the wavenumber across the i'th row of (4.2.2)

x_{ij} = $x_j - x_i$, the interstation distance

b_i = constant phase term across array, due to possible initial phase offsets in ϕ_{ij}.

All terms except for x_{ij} in (4.2.13) should be considered as functions of frequency (ω). The specific details of measuring ϕ_{ij} at a given frequency will be discussed in the section below on phase unwrapping. Equation (4.2.13) defines, for each row of (4.2.2), a linear equation of ϕ_{ij} and x_{ij}, with k_i being the slope. It can be weighted by the variances (4.2.11, 4.2.12) as follows: let

$$w_{ij} = \frac{1}{\sigma_{ij}} \quad .$$

The weighted linear regression equation can be written as

$$w_{ij} \; \phi_{ij} = w_{ij} \; k_i x_{ij} + w_{ij} b_i \quad . \tag{4.2.14}$$

On the i'th row of (4.2.2), summing over all values of j gives the following regression sums (Lindeman *et al.*, 1980, p. 9):

$$S_{xx}^{(i)} = \sum_j w_{ij}^2 \; x_{ij}^2 - \frac{(\sum_j w_{ij}^2 \; x_{ij})^2}{\sum_j w_{ij}^2}$$

$$S_{\phi\phi}^{(i)} = \sum_j w_{ij}^2 \, \phi_{ij}^2 - \frac{(\sum_j w_{ij}^2 \, \phi_{ij})^2}{\sum_j w_{ij}^2}$$

$$S_{x\phi}^{(i)} = \sum_j w_{ij}^2 \, x_{ij} \phi_{ij} - \frac{\sum_j w_{ij}^2 \, x_{ij} \sum_j w_{ij}^2 \, \phi_{ij}}{\sum_j w_{ij}^2} \quad .$$

The least-squares regression estimate of the wavenumber along the i'th row is

$$k_i = \frac{S_{x\phi}^{(i)}}{S_{xx}^{(i)}} \quad , \tag{4.2.15}$$

and the wavenumber variance estimate from the least squares covariance matrix is

$$\sigma_i^2 = \frac{\sigma_{si}^2}{S_{xx}^{(i)}} \quad . \tag{4.2.16}$$

The "scaling variance" (σ_{si}^2) should approximately equal one if the interstation standard deviations (4.2.11, 4.2.12) are close to equaling the residuals between observed and computed phases in (4.2.13) (Wiggins, 1972, p. 258). However, the interstation standard deviations are highly dependent on the coherency spectrum (4.2.8), which in turn is a function of the window width used in smoothing the time domain cross-correlation operators. For instance, if a very wide window is used, the coherency spectrum will be close to one, causing the interstation standard deviations to be small. This will cause the covariance matrix term $1/S_{xx}^{(i)}$ to be small, resulting in low variance estimates for the wavenumber k_i. But, when dealing with noisy data, using wide windows can lead to large resi-

duals in the regression equation (4.2.13).

If narrow windows are used in calculating the coherency spectrum, the reverse situation can happen: the coherency spectrum approaches zero, resulting in a large covariance matrix. However, the narrow time domain windows lead to very smooth (and biased) phase estimates, and the result is low residuals in (4.2.13).

Considering the above, a suitable scaling for the covariance matrix is:

$$\sigma_{si}^2 = 1 + \frac{S_{\phi\phi}^{(i)} - \dfrac{S_{x\phi}^{(i)\,2}}{S_{xx}^{(i)}}}{N-2} \quad , \tag{4.2.17}$$

where N is the number of stations. The term to the right of the plus sign is the calculated standard variance due to the residuals in (4.2.14) (Lindeman *et al.* , 1980, p. 14). The above scaling variance compensates for wide windows by the dominance of the residual standard variance, while for narrow windows it will reduce to a minimum value of one, not zero.

Equations (4.2.15) and (4.2.16) represent wavenumber and associated errors (at a given frequency) along the i'th row of the Green's function matrix (4.2.2). In the noise-free case, or when random errors are equally distributed for all stations, the above estimates should be equal for all rows of (4.2.2). However, in the presence of lateral inhomogeneities or local site effects (such as dead receivers), the estimates will vary. Therefore, to calculate the overall wavenumber and variance, the following weighted mean values are constructed. From (4.2.16), let

$$w_i = \frac{1}{\sigma_i} \quad .$$

The weighted mean wavenumber across the rows is

$$\bar{k} = \frac{\sum_i w_i^2 \, k_i}{\sum_i w_i^2} \quad . \tag{4.2.18}$$

The corresponding total variance is

$$\sigma_{\bar{k}}^2 = \sigma_s^2 \, \bar{\sigma}^{\,2} \quad , \tag{4.2.19}$$

with the weighted mean variance equal to

$$\bar{\sigma}^{\,2} = \frac{\sum_i w_i^2 \, \sigma_i^2}{\sum_i w_i^2} = \frac{N}{\sum_i w_i^2} \quad . \tag{4.2.20}$$

The scaling variance (σ_s^2) in this case is defined as

$$\sigma_s^2 = \frac{\hat{\sigma}^2}{\bar{\sigma}^2} + \frac{S_{kk}}{N-1} \tag{4.2.21}$$

where

$$S_{kk} = \sum_i w_i^2 \, k_i^2 - \frac{(\sum_i w_i^2 \, k_i)^2}{\sum_i w_i^2}$$

and

$$\hat{\sigma}^{\,2} = \text{median}(\sigma_i^2) \quad .$$

The term on the right of the plus sign in (4.2.21) is the standard variance calculated from the weighted residuals between observed and computed wavenumbers. The term to the left rescales the wavenumber variance to the median value found along the rows of (4.2.2).

The choice of scaling the variance in terms of the median is not necessary, but it will give a more robust estimate when some of the stations exhibit very low coherence (such as dead channels). In cases where

the variance is more uniform across the rows of (4.2.2), the ratio of median to weighted mean variance in (4.2.21) can be set equal to one.

Finally, to determine mean phase velocities and standard errors, the following expressions from Papoulis (1965, p. 152) are useful. Let

$$y = f(x)$$

be a function of the random variable x, where η is the mean of x and σ_x^2 is the variance. Then the mean and variance of y can be approximated by

$$\overline{y} = f(\eta) + f''(\eta)\,\frac{\sigma_x^2}{2} \tag{4.2.22}$$

and

$$\sigma_{f(x)}^2 = [f'(\eta)]^2\,\sigma_x^2 \quad . \tag{4.2.23}$$

From (3.1.2), (4.2.18), (4.2.19) and (4.2.22), the mean phase velocity at a given frequency is:

$$\overline{c}(\omega) = \frac{\omega}{\overline{k}(\omega)} + \frac{\omega}{\overline{k}^{\,3}(\omega)}\,\frac{\sigma_{\overline{k}}^2}{2} \tag{4.2.24}$$

and using (4.2.23), the variance in phase velocity is:

$$\sigma_{\overline{c}}^2 = \left[\frac{\omega}{\overline{k}^{\,2}(\omega)}\right]^2 \sigma_{\overline{k}}^2 \quad . \tag{4.2.25}$$

Notice that the dependence on frequency is now being stated explicitly.

Equations (4.2.24) and (4.2.25) are the final result for the multi-channel phase-matched filter, and are in a suitable form for surface wave inversion, which will be discussed in the next chapter. At this point, it is also possible to set up an iterative algorithm for the phase velocities, as was done for the single-station method. The phase velocities (4.2.24) can

be substituted into (4.2.4) as refined initial estimates for the multi-channel matched filter, and the entire process repeated. This is useful in cases where the original estimate is quite inaccurate. As an example, when the fundamental-mode surface wave dominates the waveform and there is little conflicting signal energy, a first estimate for phase velocity can be found from the time domain lag of the peak in the interstation cross-correlation function $X_{ij}(\omega)$. This gives a constant phase velocity for the estimate, which can lead to large residual first and second derivatives in phase. The result will be a bias in phase, as shown by (4.1.20). By iterating the multi-channel matched filter several times, the bias can be significantly reduced. This will be demonstrated in chapter 6, where tests on real and synthetic data will be performed.

4.2.5 Phase unwrapping

A further problem requiring analysis in phase-matched filtering is that of phase unwrapping. By definition, the phase spectrum is equal to the inverse tangent of the imaginary divided by the real complex spectrum. However, inverse tangents are computed as modulo 2π functions, resulting in discontinuities at multiples of 2π. In order to calculate wavenumbers and phase velocities, the phase must be translated into a continuous, or "unwrapped" spectrum. Techniques for accomplishing this will be discussed below, with emphasis on a new method for multi-channel processing.

In the section on single-station phase-matched filtering, equation (4.1.5) gives the frequency-domain expression for the pseudo-autocorrelation function, with the residual phase error (4.1.4) being due to

approximations in the original wavenumber estimate. To find the phase error, the phase spectrum of the pseudo-autocorrelation spectrum must be unwrapped to prevent 2π discontinuities in the residual. Stoffa *et al.* (1974) present a useful method for unwrapping which depends on numerically differentiating and integrating the complex spectrum. Let

$$\phi(\omega) = \delta k \; x \; = \tan^{-1}(\frac{v}{u})$$

be the residual phase of the pseudo-autocorrelation spectrum, where u is the real and v is the imaginary part of the complex spectrum. Differentiating the inverse tangent function gives the group delay (4.1.9):

$$t_g(\omega) = \frac{d\phi}{d\omega} = \frac{u\frac{dv}{d\omega} - v\frac{du}{d\omega}}{u^2 + v^2}.$$

This expression requires the numerical differentiation of the real and imaginary parts of the spectrum. Numerically integrating t_g

$$\phi(\omega) = \int_0^{\omega} t_g(\omega) d\omega \qquad (4.2.26)$$

brings back the original phase spectrum, but now in an unwrapped form. This method is useful where the sampling interval is sufficiently dense to insure small numerical errors, and where iteration of the matched-filtering process will reduce any residual numerical errors. However, as was pointed out in the chapter on single station filtering, there is an ambiguity due to the integration constant for group delay. If the initial phase determined at the start of integration in (4.2.26) is incorrect, all subsequent phase values will be off by a constant. In practice, this error will be a multiple of 2π after the matched-filter process converges. Unfor-

tunately, this is a fundamental ambiguity in processing which cannot be resolved by the single station spectrum. Multiplying the spectrum by $\exp(\pm i2n\pi)$ will have no effect on the time-domain transform. The only way to solve this problem is to have *a priori* knowledge of the wavenumber at least at one frequency, or to resort to multi-channel processing, which separates the above periodicity into wider phase intervals.

A more subtle and difficult problem found in phase unwrapping is that of " π - discontinuities," distinguished from the above " 2π " discontinuities. This is caused by both the real and imaginary components of the complex spectrum changing sign simultaneously. It occurs only when the amplitude of the signal is close to zero, and can be due either to noise at this level, or "spectral nulls," caused by eigenfunction zero-crossings of earthquake source mechanisms (Aki and Richards, 1980, p. 318). When the real and imaginary components change sign, a π - discontinuity occurs in the phase spectrum. Such shifts can be easily identified by the group delay (4.1.9), which will be a Dirac delta function at the discontinuity, since the derivative of a step function is impulsive.

There is an ambiguity in unwrapping π - discontinuities, caused by the direction of the π phase shift. Either adding or subtracting π from the phase will result in the same modulo 2π location, as can be seen by inspection of a phasor plot of the complex spectrum (Figure 4.1a). There is no fundamentally preferred direction to the shift, since either addition or subtraction of π will result in the same time-domain transform. The above method of numerically differentiating the phase spectrum will automatically choose a direction for the phase shift based on values calculated by the numerical derivative. However, this is inconclusive since the

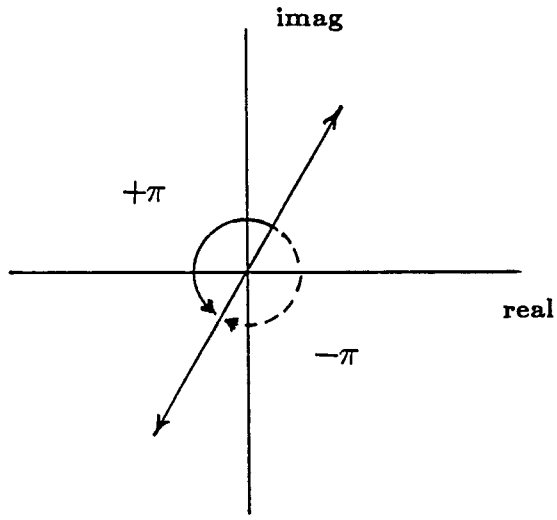

Figure 4.1a Phasor plot showing that adding $\pm\pi$ to the phase results in the same phasor location.

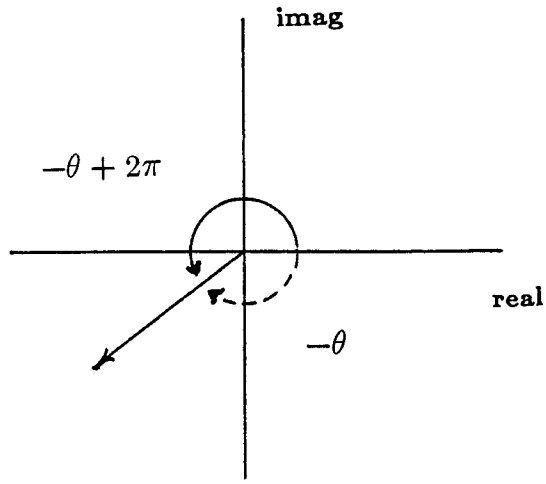

Figure 4.1b Phasor plot showing quadrant location where there is a $\pm 2\pi$ discontinuity.

direction can change with slight changes in noise levels. Again, for single-station processing, this problem can only be solved by *a priori* knowledge of phase velocity values at selected periods.

With multi-station processing, it is possible to completely solve the unwrapping problem, even in cases of high noise levels. As a foundation for this, another single-station method is introduced, which takes advantage of windowing in phased-matched filtering.

Using the Discrete Fourier Transform (DFT), a simple recursive formula for unwrapping the phase of the pseudo-autocorrelation function can be constructed. Let

$$\phi_l^p = \phi_{l-1}^p + \delta\phi_l^p \quad . \tag{4.2.27}$$

The subscripts indicate the discrete frequency in the DFT. The superscript "p" distinguishes (4.2.27) from the phase of the cross-correlation function (4.2.5), which has the initial phase estimate added back in. The phase increment is defined from

$$\delta\phi_l^p = \tan^{-1}\left(\frac{v_l}{u_l}\right) - \tan^{-1}\left(\frac{v_{l-1}}{u_{l-1}}\right) \tag{4.2.28}$$

where u and v are the real and imaginary part of the complex spectrum. To avoid modulo 2π discontinuities, the quadrant position of the l'th phase with respect to the $(l-1)$'th phase must be identified, as shown in Figure 4.1b.

Since phase-matched filtering is being used, the width of the time window used to isolate the mode can be used to calculate the maximum phase increment in (4.2.28). The relationship between frequency sampling and time sampling in the DFT is (Brigham, 1974)

$$\delta\omega = \frac{2\pi}{n\,\delta t} \tag{4.2.29}$$

where $\delta\omega$ is the frequency sampling interval, δt is the time sampling interval, and n is the number of samples. From (4.1.9) the group delay can be approximated by

$$\frac{\delta\phi_l^p}{\delta\omega} \approx t_l \quad . \tag{4.2.30}$$

Since there is no energy outside the time window of the pseudo-autocorrelation function, the maximum continuous group delay is

$$t_{\max} \leq m\,\delta t \tag{4.2.31}$$

where m is the number of samples in the one-sided width of the time window. Substituting (4.2.31) into (4.2.30) and solving for the phase increment gives

$$\delta\phi_{\max}^p \leq m\,\delta t\,\delta\omega = 2\pi\ m/n \quad , \tag{4.2.32}$$

where (4.2.29) is substituted for $\delta\omega$. For example, if m is equal to one-fourth the total array width n, the maximum phase increment will be less than or equal to $\pi/2$. As long as $\delta\phi^p$ satisfies the inequality in (4.2.32), the recursive formula (4.2.27) will be adequate for unwrapping single spectrum phases.

Phase increments which are greater than the maximum in (4.2.32) define π - discontinuities in the phase spectrum. At these points, there will be an ambiguity in the sign of $\delta\phi^p$, since either adding or subtracting π in the recursion formula (4.2.27) will give the same phasor location (see Figure 4.1a). As a result, the phase spectrum will be divided into continuous segments, with boundaries defined at frequency points where the

π discontinuities occur.

Multi-channel techniques can now be utilized to adjust the segments to the correct modulo 2π locations. Assuming an array of N stations, construct the interstation Green's function matrix (4.2.2). Let

$$\phi^p_{ijl} \ , \quad l = 1,2,...n \tag{4.2.33}$$

be the discrete phase spectrum of the pseudo-autocorrelation function corresponding to each Green's function H_{ij}, where l is the discrete frequency location and n is the number of points in the DFT. For each H_{ij}, unwrap the phase spectrum according to (4.2.27)

$$\phi^p_{ijl} = \phi^p_{ij\ l-1} + \delta\phi^p_{ijl} \ . \tag{4.2.34}$$

At locations where the phase increment in (4.2.34) is greater than the maximum allowed by (4.2.32), construct the following "jump table":

$$J_{ij}(I) \ , \quad I = 1,2,...N_{ij} \tag{4.2.35}$$

where N_{ij} is the total number of π - discontinuities found in the pseudo-autocorrelation function. The jump table defines frequency locations for the boundaries of continuous segments in each phase spectrum ϕ_{ij}.

To set up corrections to the segments, define the following terms:

ϕ_{ijl} = cross-correlation phase corresponding to (4.2.5),

σ_{ijl} = phase standard deviation calculated from (4.2.11) and (4.2.12),

$\hat{k}_l$ = multi-channel wavenumber estimate at frequency l,

$J1 = J_{ij}(I)$ = lower bound for the I'th segment,

$J2 = J_{ij}(I+1) - 1$ = upper bound for the I'th segment.

Notice that the superscript is now removed from the phase spectrum,

indicating that it is the complete phase spectrum for the interstation Green's function, not the residual spectrum of the pseudo-autocorrelation function. From (4.2.3) it is related to the wavenumber estimate by

$$\phi_{ijl} = \hat{k}_l \; x_{ij} \pm 2n^{(I)}\pi \quad . \tag{4.2.36}$$

The integer $n^{(I)}$ indicates that the phase can be incorrect by a multiple of 2π. The superscript (I) indicates that the multiple can change value at different frequency locations, due to π - discontinuities. Construct the weighted minimization function

$$M \;=\; \sum_{l=J1}^{J2} (\phi_{ijl} - \hat{k}_l x_{ij} + C^{(I)})^2 \; w_{ijl}^2$$

where $w_{ijl} = 1/\sigma_{ijl}$, and x_{ij} is the interstation distance. $C^{(I)}$ is an unknown constant over the frequency interval between $J1$ and $J2$, representing the difference between the unwrapped spectrum and the "true" phase estimate $\hat{k}_l x_{ij}$. The phase estimate does not have to be exact, but $\hat{k}_l$ should be within $\pm\pi$ of the true wavenumber. This will be discussed in detail below. Least-squares analysis on (4.2.36) in terms of the constant over the I'th segment yields

$$C^{(I)} \;=\; \sum_{l=J1}^{J2} w_{ijl}^2 \; (\phi_{ijl} - \hat{k}_l x_{ij}) \; / \; D \quad , \tag{4.2.37}$$

where

$$D \;=\; \sum_{l=J1}^{J2} w_{ijl}^2 \quad .$$

From (4.2.36), the constant should be a multiple of 2π, so define

$$2\, n^{(I)}\, \pi \;\approx\; C^{(I)}$$

where $n^{(I)}$ is the nearest 2π multiple to $C^{(I)}$. The phase for the I'th

segment can now be corrected with

$$\hat{\phi}_{ijl} = \phi_{ijl} + 2n^{(I)}\pi , \quad J1 \leq l \leq J2 . \tag{4.2.38}$$

The procedure is repeated for all segments in all the interstation Green's functions.

To calculate the wavenumber estimate $\hat{k}_l$, stacking methods must be used, such as defined in chapter 3. For example, if a $p-\omega$ stack is initially used to calculate the initial interstation phase estimate (4.2.4), it can also be used to define the above wavenumber estimate $\hat{k}_l$. However, when noise levels are extremely high, the side-lobes generated by the Fourier kernel (3.2.11) may significantly interfere with the true location of the wavenumber. In this case, a more robust stacking procedure can be used.

At a given frequency l, define the following double sum

$$F_l(k) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \cos(k\, x_{ij} - \phi_{ijl}) , \tag{4.2.39}$$

where

ϕ_{ijl} = cross-correlation phase spectrum from (4.2.36),

x_{ij} = interstation spacing,

k = wavenumber variable.

Referring to the Green's function matrix (4.2.2), the above sum is over all interstation combinations above the diagonal. This insures that all interstation phases will contribute to determining the wavenumber estimate. Substituting the wavenumber estimate (4.2.36) for the phase in (4.2.39) gives

$$F_l(k) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \cos[(k - \hat{k}_l)x_{ij}] . \tag{4.2.40}$$

The 2π multiple term now disappears since the cosine function is invariant to modulo 2π additions.

The stack defined by (4.2.39) has several features which make it useful for calculating wavenumber estimates under noisy conditions. First, it will reach a maximum when $k = \hat{k}_l$. This is clear from (4.2.40), since the individual cosine elements are maximized when $k = \hat{k}_l$. Second, the estimates for $\hat{k}_l$ will be more stable than those determined in a $p-\omega$ stack, since the unwrapped phase spectra used in (4.2.39) are determined from phase-matched interstation cross-correlation functions (4.2.5). The stability comes from the improvement in signal to noise by time windowing the pseudo-autocorrelation function in (4.2.5). Third, local maxima caused by sidelobes in (4.2.39) are smaller than the p-ω stack, and the width of the major lobe is narrower. This can be demonstrated for the special case of equally spaced stations. Under this condition, it can be shown (see Appendix II) that (4.2.40) reduces to a Fejer kernel (Papoulis, 1977, p. 72)

$$F_l(k) = \frac{\sin^2[\,(N)(k-\hat{k}_l)\delta x/2\,]}{2\sin^2[\,(k-\hat{k}_l)\delta x/2\,]} - \frac{N}{2}\ , \qquad (4.2.41)$$

where δx is the interstation spacing. The equivalent Fourier-series kernel (assuming negligible attenuation) for the $p-\omega$ stack is given by (3.2.11). Figure 4.2 is a plot of the two kernels, which clearly confirms the third assertion.

It should be emphasized that the purpose of the cosine stack (4.2.39) is to give robust wavenumber estimates after phase-matched filtering. It is most useful in cases where noise levels are so high that coherent energy is not discernible from $p-\omega$ plots. This will be demonstrated below in the section on multi-channel surface-wave applications.

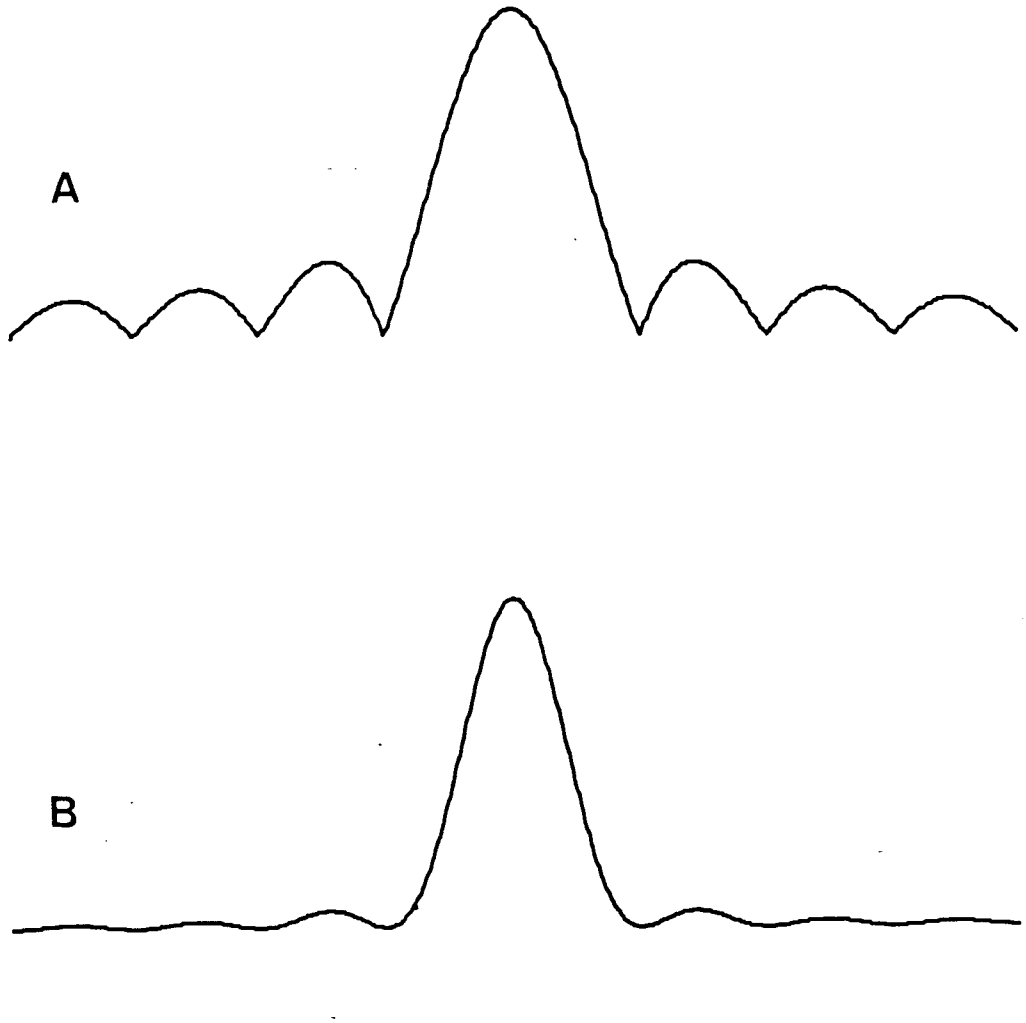

Figure 4.2 P-ω stacking kernels. (a) Fourier kernel. (b) Fejer kernel.

4.2.6 Summary

The above method for analyzing multi-channel phase velocity dispersion can be condensed into the following algorithm for reference. It is assumed for convenience that discrete frequencies are used, with subscripts given by "l", as in the section on phase unwrapping.

Step 1. Calculate the interstation cross-correlation function (4.2.5) and unwrap the corresponding pseudo-autocorrelation phase ϕ^p_{ijl}, as in (4.2.34).

Step 2. At frequency points where there are π - discontinuities, construct the jump table (4.2.35).

Step 3. Calculate the phase for the interstation cross-correlation function (4.2.5) using the unwrapped pseudo-autocorrelation phase and the original phase estimate used in the matched filter (4.2.4):

$$\phi_{ijl} = \phi^p_{ijl} - \tilde{\phi}_{ijl}$$

Step 4. Adjust the cross-correlation phase to the nearest correct 2π location (4.2.38) on each continuous segment defined by the jump table in step 2. This may involve constructing the cosine stack (4.2.39) for stable wavenumber estimates.

Step 5. Determine phase standard deviations (4.2.11) and (4.2.12) using interstation coherency measurements.

Step 6. Perform weighted regression analysis on the interstation phase estimates (4.2.14) , to find wavenumber estimates (4.2.15) and standard deviations (4.2.16) for the i'th row of the Green's function matrix (4.2.2).

Step 7. Find the weighted mean wavenumber across all rows (4.2.18) and the associated variance (4.2.19).

Step 8. Transform the mean wavenumber and variance into phase

velocity and velocity variance using (4.2.24) and (4.2.25).

This completes the technique for multi-channel phase velocity processing. In chapter 6, applications will be made to real and synthetic data sets, in order to test the method in a variety of noisy environments.

CHAPTER 5

LINEAR INVERSION THEORY

5.1 General theory

5.1.1 Introduction

Many geophysical problems can be modeled as a Fredholm integral of the first kind (Twomey, 1977, p. 5):

$$y(t) = \int_r A(t,r)\, x(r)\, dr \qquad (5.1.1)$$

where

t,r = coordinate systems for data space and model space, respectively,

$x(r)$ = model parameters desired, such as density, velocity, or Q,

$y(t)$ = observational measurements,

$A(t,r)$ = integral kernel relating model parameters to observations.

In the context of equation (5.1.1), the purpose of linear inversion is to find physically acceptable solutions to the model $x(r)$, which will in some sense satisfy the observational data $y(t)$.

Backus and Gilbert (1967, 1968) pointed out that although geophysical models can generally be regarded as continuous functions of r, "observational measurements" usually imply a discrete set of data points. In this case, (5.1.1) must be modified to

$$y_i = \int_r A_i(r)\, x(r)\, dr\ , \qquad i = 1,2,...m \qquad (5.1.2)$$

where m is the total number of observations. There are a variety of ways

of inverting (5.1.2) for x, given known data y_i. Three well known techniques are the Backus-Gilbert method, series expansion, and discrete modeling. These will be briefly outlined below.

5.1.2 Backus-Gilbert method

The Backus-Gilbert method (Backus and Gilbert, 1967, 1968) is in a sense the most general solution possible for (5.1.2), in that it requires no *a priori* parameterization of the model $x(r)$. The solution is continuous, and is constrained only by the maximum resolution obtainable by the particular Fredholm integral used.

The model is constructed at any continuous point r' as a linear combination of all the data points y_i. Specifically, both sides of equation (5.1.2) are summed as follows:

$$\sum_{i=1}^{m} h_i(r')\, y_i \;=\; \sum_{i=1}^{m} h_i(r') \int_r A_i(r)\, x(r)\, dr \;, \tag{5.1.3}$$

where $h_i(r')$ is a set of as yet unknown constants evaluated at point r'. Equation (5.1.3) can be written as

$$\sum_{i=1}^{m} h_i(r')\, y_i \;=\; \int_r R(r',r)\, x(r)\, dr \;, \tag{5.1.4}$$

where

$$R(r',r) \;=\; \sum_{i=1}^{m} h_i(r')\, A_i(r) \;. \tag{5.1.5}$$

The essence of the Backus-Gilbert method is to find a set of $h_i(r')$ that will allow equation (5.1.5) to approximate a Dirac delta function:

$$R(r',r) \;=\; \sum_{i=1}^{m} h_i(r')\, A_i(r) \;\approx\; \delta(r' - r) \;. \tag{5.1.6}$$

Substitute (5.1.6) into (5.1.4) for

$$\int_r R(r',r)\, x(r)\, dr \approx \int_r \delta(r' - r)\, x(r)\, dr = \hat{x}(r')\,, \quad (5.1.7)$$

which gives the model estimate as a linear combination of the data

$$\hat{x}(r') = \sum_{i=1}^{m} h_i(r')\, y_i\,. \quad (5.1.8)$$

Equation (5.1.5) is the "resolution kernel" for the model estimate (5.1.8). The estimate will be a weighted average of possible models about position r', as shown by equation (5.1.7).

Applying the Backus-Gilbert method requires constructing a least-squares minimization function in terms of the residuals between the resolution kernel and the Dirac delta function in (5.1.6). For each point r', a set of $h_i(r')$ is found which will minimize the sum of the squares of the residuals. Additional constraints to control the variance of the solution can be incorporated into the minimization function, leading to a "tradeoff" between resolution and variance. This procedure is discussed in detail by Backus and Gilbert (1967, 1968).

An advantage of the Backus-Gilbert method is that the form of the solution depends only on the original equation (5.1.2). No parameterization of the model is required, such as "block modeling" or discrete modeling. A disadvantage is that the inversion coefficients $h_i(r')$ must be reconstructed in a least-squares system of equations for every point r'. When the number of observations becomes large, the time to find a solution for many points in model space can become unacceptably long.

The other methods discussed below can be computationally more efficient than Backus-Gilbert, but at the expense of parameterizing the

form of the model.

5.1.3 Series expansion

For particular cases, it may be possible to parameterize the model in terms of a truncated series of linearly independent basis functions:

$$x(r) = \sum_{j=1}^{n} b_j \, f_j(r) \tag{5.1.9}$$

where

$f_j(r) =$ basis function,

$b_j =$ basis function coefficient,

$n =$ total number of coefficients.

For example, Woodhouse and Dziewonski (1984) took advantage of the sphericity of the earth and modeled the three-dimensional velocity distribution as a set of spherical harmonic basis functions.

Substituting (5.1.9) into (5.1.2) gives

$$y_i = \int_r A_i(r) \sum_{j=1}^{n} b_j \, f_j(r) \, dr$$

which can be rearranged as

$$y_i = \sum_{j=1}^{n} \{ \int_r A_i(r) f_j(r) dr \} \, b_j \, . \tag{5.1.10}$$

Considering the integral in brackets as elements of a matrix, (5.1.10) can be written in the simple form

$$y_i = \sum_{j=1}^{n} A_{ij} \, b_j \, . \tag{5.1.11}$$

Equation (5.1.11) can be inverted for b_j by standard least-squares

algorithms. Once the coefficients have been found, they can be substituted into equation (5.1.9) to form a continuous solution.

5.1.4 Discrete modeling

This method is applicable in cases where the model can be approximated by discrete, or piece-wise constant values over the model space. With no loss of generality, equation (5.1.2) can be written as a sum of integrals over subregions of the model space:

$$y_i = \sum_{j=1}^{n} \int_{r_j} A_i(r)\, x(r)\, dr \ . \tag{5.1.12}$$

The subscript r_j represents the integral boundaries on the j'th subregion. If $x(r)$ is constant within each subregion, it can be factored from the integrals in (5.1.12) for

$$y_i = \sum_{j=1}^{n} \left\{ \int_{r_j} A_i(r)\, dr \right\} x_j \ , \tag{5.1.13}$$

where x_j is the constant value of $x(r)$ in the subregion. This can be simplified to

$$y_i = \sum_{j=1}^{n} A_{ij}\, x_j \ , \tag{5.1.14}$$

where A_{ij} represents the integral in brackets.

Discrete modeling is frequently used for tomography problems (McMechan, 1983), where the velocity structure is subdivided into a large number of blocks, with a constant velocity assumed in each block. The forward problem (5.1.12) is the observed travel time as an integral function of wave slowness across the structure. The method is also used for

problems that assume a spherical or plane-layered earth structure. "Plane-layered structure" is defined as a medium which is laterally homogeneous, with a piece-wise constant depth variation. It is particularly applicable for plane-layered surface wave analysis, since the integral in (5.1.13) can be analytically evaluated. This will be discussed below, in the section on surface wave inversion.

All the above methods result in matrix equations similar to (5.1.14). Further analysis requires finding an inverse solution to (5.1.14) in terms of x_j. This can be complicated by instability in the system, occurring when small errors in y_i are magnified into large errors in x_j. The mathematical procedure of finding stable solutions to (5.1.14) will be discussed below in terms of orthogonal decompositions.

5.1.5 Least-squares analysis and singular value decomposition

Equation (5.1.14) assumes that the observed data will exactly equal the theoretical values, which is not true in practice. To compensate for errors, a residual term is added to (5.1.14):

$$y_i = \sum_{j=1}^{n} A_{ij}\, x_j \; + \; \epsilon_i \; . \tag{5.1.15}$$

It is convenient at this point to write (5.1.15) in vector-matrix form, so the entire set of observed data can be represented in one equation. Let

$$\mathbf{y} \; = \; \mathbf{A}\,\mathbf{x} \; + \; \boldsymbol{\epsilon} \; , \tag{5.1.16}$$

where

$\mathbf{y}$ = m x 1 vector of observations,

$\mathbf{x}$ = n x 1 vector of unknown parameters,

$\boldsymbol{\epsilon}$ = m x 1 vector of residuals,

$\mathbf{A}$ = m x n kernel matrix.

The convention used is that any bold-faced entry indicates a vector or matrix, while all other variables are scalar.

The principle of least-squares states that a valid set of model parameters ($\mathbf{x}$) is one which minimizes the sum of the squares of the residuals ($\boldsymbol{\epsilon}$). The minimization function can be represented mathematically as

$$M(\mathbf{x}) = \boldsymbol{\epsilon}^T\boldsymbol{\epsilon} = |\boldsymbol{\epsilon}|^2 = \sum_{i=1}^{m} \epsilon_i^2 \ . \qquad (5.1.17)$$

The terms on the right are the same, being three different ways of expressing the squared norm of the residuals. The superscript "T" indicates the transpose of $\boldsymbol{\epsilon}$, and the bars "$|\ |$" indicate the root mean square norm of $\boldsymbol{\epsilon}$.

Finding a minimum for M is simplified by the singular value decomposition (SVD), an orthogonal transformation of the $\mathbf{A}$ matrix. Lawson and Hanson (1974, p. 107) show that any arbitrary matrix can be transformed into

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T \qquad (5.1.18)$$

where

$\mathbf{U}$ = m x m orthogonal matrix,

$\mathbf{V}$ = n x n orthogonal matrix,

$\boldsymbol{\Lambda}$ = m x n upper left diagonal matrix.

"Orthogonal" means that the transpose of the matrix is also the inverse, and "upper left diagonal" means that $\boldsymbol{\Lambda}$ can be written as

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{5.1.19}$$

where

$\mathbf{\Lambda}_k$ = k x k diagonal matrix.

The terms along the diagonal are the singular values of the **A** matrix, and the rank of the **A** matrix is k, the number of non-zero singular values. From (5.1.19), the rank of the matrix obeys the inequality

$$k \leq min(n,m) . \tag{5.1.20}$$

Equations (5.1.18), (5.1.19), and (5.1.20) lead to a general classification for arbitrary matrices (Lawson and Hanson, 1974, p. 3). If

$n > m$ the matrix is overdetermined,

$n < m$ the matrix is underdetermined,

$n = m$ the matrix is even determined,

$k = min(n,m)$ the matrix has full rank,

$k < min(n,m)$ the matrix is rank deficient, or underconstrained.

The matrix may have full rank, but one or more of the singular values may be much less than the maximum. Given λ_i as any singular value of $\mathbf{\Lambda}$, if $\lambda_i << \lambda_{max}$, the matrix is poorly constrained. This condition is common in geophysical problems.

To transform the least-squares problem, substitute (5.1.18) into (5.1.16) for

$$\mathbf{y} = \mathbf{U\Lambda V}^T\mathbf{x} + \boldsymbol{\epsilon} .$$

Multiply both sides by $\mathbf{U}^T$, the inverse of **U** for

$$\mathbf{U}^T\mathbf{y} = \mathbf{\Lambda}\,\mathbf{V}^T\mathbf{x} + \mathbf{U}^T\boldsymbol{\epsilon}\,.$$

Make the following variable substitutions

$$\mathbf{g} = \mathbf{U}^T\mathbf{y}\,, \quad \mathbf{p} = \mathbf{V}^T\mathbf{x}\,, \quad \mathbf{e} = \mathbf{U}^T\boldsymbol{\epsilon} \tag{5.1.21}$$

for

$$\mathbf{g} = \mathbf{\Lambda p} + \mathbf{e}\,. \tag{5.1.22}$$

Equation (5.1.22) is equivalent to the original least-squares problem since the minimization function for (5.1.22) is the same as (5.1.17):

$$M(\mathbf{p}) = \mathbf{e}^T\mathbf{e} = \boldsymbol{\epsilon}^T\mathbf{U}\mathbf{U}^T\boldsymbol{\epsilon} = \boldsymbol{\epsilon}^T\boldsymbol{\epsilon}\,. \tag{5.1.23}$$

This demonstrates that the least-squares problem is invariant to multiplication by orthogonal matrices.

To find the least-squares solution, partition (5.1.22) according to the rank of **A**:

$$\begin{bmatrix}\mathbf{g}_k \\ \mathbf{g}_{m-k}\end{bmatrix} = \begin{bmatrix}\mathbf{\Lambda}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{p}_k \\ \mathbf{p}_{n-k}\end{bmatrix} + \begin{bmatrix}\mathbf{e}_k \\ \mathbf{e}_{m-k}\end{bmatrix}\,. \tag{5.1.24}$$

The subscripts indicate the dimensions of the vectors and matrix. From (5.1.23) and (5.1.24), the minimization function is

$$M(\mathbf{p}) = |\mathbf{e}|^2 = |\mathbf{e}_k|^2 + |\mathbf{e}_{m-k}|^2\,,$$

which from (5.1.24) is

$$M(\mathbf{p}) = |\mathbf{g}_k - \mathbf{\Lambda}_k\mathbf{p}_k|^2 + |\mathbf{g}_{m-k}|^2\,. \tag{5.1.25}$$

The minimum of (5.1.25) corresponds to the vector $\mathbf{p}_k$ which sets the first norm to zero, and this can be found immediately from

$$\mathbf{p}_k = \mathbf{\Lambda}_k^{-1}\,\mathbf{g}_k\,. \tag{5.1.26}$$

From (5.1.26), (5.1.24), and (5.1.21), the solution to the least-squares problem is

$$\mathbf{p} = \begin{bmatrix} \mathbf{p}_k \\ \mathbf{p}_{n-k} \end{bmatrix}, \qquad \mathbf{x} = \mathbf{V}\mathbf{p} . \tag{5.1.27}$$

Equation (5.1.25) is remarkable in that the least-squares solution can be found from algebraic considerations alone, without resorting to differential calculus. The solution is completely general; no restrictions are made on the dimensions or rank of **A**. Another feature of (5.1.25) is that the minimum residual can be determined directly from $\mathbf{g}_{m-k}$. Notice that for full rank underdetermined problems, $m = k$ and from (5.1.25), the minimum residual is zero.

Unless the problem is full rank and overdetermined ($n = k$), the solution will be non-unique. Only $\mathbf{p}_k$ is necessary to minimize the residual in (5.1.25), and inspection of (5.1.24) shows that $\mathbf{p}_{n-k}$ is arbitrary. Therefore, there are an infinite number of possible values for **p** in (5.1.27). This leads to the concept of the "minimum length" solution for underdetermined (or underconstrained) problems. If $\mathbf{p}_{n-k}$ is arbitrarily set to zero, a valid least-squares solution is, from (5.1.27)

$$\mathbf{x} = \mathbf{V}\begin{bmatrix} \mathbf{p}_k \\ \mathbf{0} \end{bmatrix}. \tag{5.1.28}$$

Substituting (5.1.26) into the minimum length solution (5.1.28), and transforming back to the original variables in (5.1.21) gives

$$\mathbf{x} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\,\mathbf{U}^T\mathbf{y} \tag{5.1.29}$$

where

$$\boldsymbol{\Lambda}^{-1} = \begin{bmatrix} \boldsymbol{\Lambda}_k^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

The dimensions of $\mathbf{\Lambda}^{-1}$ are n x m. This leads to the definition of the generalized inverse (Penrose, 1955):

$$\mathbf{H}_g = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}^T . \tag{5.1.30}$$

For full rank problems, there are two equivalent forms to the generalized inverse. If the problem is overdetermined, the inverse can be written

$$\mathbf{H}_o = (\mathbf{A}^T\mathbf{A})^{-1}\,\mathbf{A}^T \tag{5.1.31}$$

and for underdetermined problems, the inverse is

$$\mathbf{H}_u = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} . \tag{5.1.32}$$

These can be verified by substituting the singular value decomposition (5.1.18) into (5.1.31) and (5.1.32). Under the stated constraints (full rank, over or underdetermined), multiplying out the matrices shows that the two forms are exactly equal to the generalized inverse (5.1.30). For even determined systems, (5.1.31) and (5.1.32) are equivalent.

Another way of expressing the generalized inverse is in terms of a vector sum. Letting each column of $\mathbf{U}$ and $\mathbf{V}$ represent orthogonal vectors $\mathbf{u}_j$ and $\mathbf{v}_j$, matrix manipulation of (5.1.29) leads directly to the sum

$$\mathbf{x} = \sum_{j=1}^{k} \frac{\mathbf{u}_j^T\mathbf{y}}{\lambda_j}\,\mathbf{v}_j \tag{5.1.33}$$

This gives the solution vector as a sum of k orthogonal vectors $\mathbf{v}_j$. For poorly constrained problems, at least one of the singular values (λ_j) will be small, and this can lead to excessive magnification of the corresponding vector $\mathbf{v}_j$. This defines an unstable least squares problem, and the next section will discuss inverse solutions which control the instability.

5.1.6 Stochastic inversion

Generalized inversion gives solutions to rank deficient least squares problems, but it does not address poorly constrained problems, where singular values approach, but do not exactly equal zero. One method of dealing with small singular values is to set them equal to zero, reducing the rank of the problem. This is the "sharp cutoff" technique (Wiggins, 1972). In (5.1.33), only $\mathbf{v}_j$ vectors with relatively large singular values are included in the sum, in order to produce physically reasonable solutions. In practice, however, this method sometimes leads to unwanted "ripples" or oscillations in the solution, due to the abrupt truncation of the vector sum (5.1.33). The effect is similar to the problem of using ideal low-pass filters in Fourier analysis.

Another approach to instability is to constrain the norm of the solution vector. In (5.1.33), small singular values can produce solutions with large magnitudes. To control this, the Levenburg-Marquardt damped least-squares method (Levenburg, 1944; Marquardt, 1963) includes the norm of the solution vector as a part of the least-squares minimization function. This is implemented by appending a scaled identity matrix to the original least-squares problem:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \gamma \mathbf{I} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon}_\gamma \end{bmatrix}. \tag{5.1.34}$$

The scalar variable γ is undetermined at this point. If it is set equal to zero, (5.1.34) reduces to the original least-squares problem. The minimization function is

$$M(\mathbf{x}) = |\boldsymbol{\epsilon}|^2 + |\boldsymbol{\epsilon}_\gamma|^2$$

which from (5.1.34) is

$$M(\mathbf{x}) = |\mathbf{y} - \mathbf{A}\mathbf{x}|^2 + \gamma^2 |\mathbf{x}|^2 . \qquad (5.1.35)$$

As the value of γ^2 increases, more weight is put on minimizing the solution norm, and less on the least-squares residual. This insures stable solution vectors, but at the expense of larger least-squares residuals.

No matter what the rank and dimension of the original problem, equation (5.1.34) is overdetermined and full rank. This is obvious since the scaled identity matrix is even determined and full rank. It can be expressed as an independent least-squares problem

$$\hat{\mathbf{y}} = \hat{\mathbf{A}}\mathbf{x} + \hat{\boldsymbol{\epsilon}} ,$$

where the new variables are equivalent to the partitioned ones in (5.1.34). Since this problem is full rank and overdetermined, equation (5.1.31) is a valid inverse:

$$\mathbf{x} = (\hat{\mathbf{A}}^T \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}^T \hat{\mathbf{y}} .$$

Substituting back the original variables in (5.1.34) and multiplying out the partitioned matrices yields

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A} + \gamma^2 \mathbf{I})^{-1} \mathbf{A}^T \mathbf{y} .$$

This defines the Levenburg-Marquardt damped least-squares inverse

$$\mathbf{H}_d = (\mathbf{A}^T \mathbf{A} + \gamma^2 \mathbf{I})^{-1} \mathbf{A}^T . \qquad (5.1.36)$$

For non-zero γ, $\mathbf{H}_d$ is completely general in that no restrictions on rank or dimensions are necessary for the $\mathbf{A}$ matrix. If $\gamma = 0$, (5.1.36) reduces to (5.1.31), and the inverse exists only if $\mathbf{A}$ is full rank and overdetermined.

The orthogonal equivalent of (5.1.34) is found by substituting the singular value decomposition (5.1.18) into (5.1.34) and multiplying both sides by the partitioned orthogonal matrix

$$\begin{bmatrix} \mathbf{U}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{V}^T \end{bmatrix}.$$

Using the variable substitutions in (5.1.21), the orthogonally transformed system is written

$$\begin{bmatrix} \mathbf{g} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda} \\ \gamma\mathbf{I} \end{bmatrix} \mathbf{p} + \begin{bmatrix} \mathbf{e} \\ \mathbf{e}_\gamma \end{bmatrix}. \tag{5.1.37}$$

This is equivalent to (5.1.34) since the minimization function is invariant to multiplication by orthogonal matrices (see (5.1.23)). Considering (5.1.37) as an independent least-squares problem, it is full rank and over-determined, so (5.1.31) is a valid inverse. Following the same steps as the Levenburg-Marquardt inverse results in a solution

$$\mathbf{p} = (\mathbf{\Lambda}^T\mathbf{\Lambda} + \gamma^2\mathbf{I})^{-1}\,\mathbf{\Lambda}^T\,\mathbf{g}\,.$$

The inverse matrices are diagonal, so multiplying out the terms gives

$$\mathbf{p} = \tilde{\mathbf{\Lambda}}^{-1}\mathbf{g} = \begin{bmatrix} \tilde{\mathbf{\Lambda}}_k^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{g}\,. \tag{5.1.38}$$

where $\tilde{\mathbf{\Lambda}}$ has the same form as in the generalized inverse (5.1.29), but the individual singular values on the diagonal are modified as

$$\tilde{\lambda}_i = \lambda_i + \frac{\gamma^2}{\lambda_i}\,. \tag{5.1.39}$$

Substituting back the original variables in (5.1.21) gives

$$\mathbf{x} = \mathbf{V}\tilde{\mathbf{\Lambda}}^{-1}\mathbf{U}^T\mathbf{y}\,, \tag{5.1.40}$$

which leads to the definition of the stochastic inverse

$$\mathbf{H}_s = \mathbf{V}\tilde{\mathbf{\Lambda}}^{-1}\mathbf{U}^T . \tag{5.1.41}$$

It has exactly the same form as the generalized inverse, except that the singular values are modified as in (5.1.39). The effect on the solution can be seen by modifying the equivalent solution vector sum (5.1.33)

$$\mathbf{x} = \sum_{j=1}^{k} \frac{\mathbf{u}_j^T\mathbf{y}}{(\lambda_j+\gamma^2/\lambda_j)} \mathbf{v}_j . \tag{5.1.42}$$

This stabilizes poorly constrained problems by increasing the size of small singular values. The value of γ^2 can be varied by trial and error in (5.1.42) to produce reasonable physical models.

The drawback of (5.1.42) is increased minimum residuals for the original least-squares problem. From (5.1.38)

$$\mathbf{p}_k = \tilde{\mathbf{\Lambda}}_k^{-1}\,\mathbf{g}_k .$$

Putting this in the original least-squares minimization function (5.1.25) gives

$$M(\mathbf{p}) = |\mathbf{g}_k - \mathbf{\Lambda}_k\tilde{\mathbf{\Lambda}}_k^{-1}\mathbf{g}_k|^2 + |\mathbf{g}_{m-k}|^2 .$$

Writing this in summation form and rearranging terms results in

$$M(\mathbf{p}) = \sum_{i=1}^{k} \left[\frac{\gamma^2}{\lambda^2+\gamma^2}\right]^2 g_i^2 + \sum_{i=k+1}^{m} g_i^2 . \tag{5.1.43}$$

The first norm will be zero when γ is zero, leaving the original least-squares residual. As γ increases, the total residual also increases.

The term "stochastic" comes from an independent formulation of the inverse problem based on the statistics of the data and model covariance matrices (Jordan and Franklin, 1971; Aki and Richards, 1980, p.695). A

special case of this inverse is

$$\mathbf{H}_c = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \gamma^2\mathbf{I})^{-1} . \tag{5.1.44}$$

Substituting the singular value decomposition (5.1.18) into (5.1.44), and multiplying out the matrices shows that (5.1.44) is equal to the orthogonal inverse (5.1.41). As long as γ is not zero, the **A** matrix in (5.1.44) can be any rank and dimension.

Although the development is different for the above inverses, it should be emphasized that mathematically, (5.1.34), (5.1.41) and (5.1.44) are exactly equivalent for non-zero γ. Therefore, it is convenient to label all of them as "stochastic", unless there is a specific reason to differentiate between the methods.

5.1.7 Filtered inversion

The Levenberg-Marquardt method of appending to the original least squares problem can be generalized to the following form:

$$\begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \gamma\mathbf{F} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \epsilon \\ \epsilon_\gamma \end{bmatrix} . \tag{5.1.45}$$

where

$\mathbf{F}$ = n x n arbitrary matrix.

The matrix can be constructed to constrain individual elements of **x** in any desired manner. This differs from the stochastic inverse, which from (5.1.34) gives equal weight to each element of **x**. If **F** has an inverse, (5.1.45) can be transformed into a stochastic problem by the following operation:

$$\begin{bmatrix} \mathbf{A} \\ \gamma\mathbf{F} \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{A} \\ \gamma\mathbf{F} \end{bmatrix} (\mathbf{F}^{-1}\mathbf{F})\,\mathbf{x} = \begin{bmatrix} \mathbf{A}\mathbf{F}^{-1} \\ \gamma\mathbf{I} \end{bmatrix} \mathbf{F}\mathbf{x} .$$

Make the variable substitutions

$$\hat{\mathbf{x}} = \mathbf{F}\mathbf{x} , \qquad \hat{\mathbf{A}} = \mathbf{A}\mathbf{F}^{-1} \tag{5.1.46}$$

for

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{A}} \\ \gamma\mathbf{I} \end{bmatrix} \hat{\mathbf{x}} + \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon}_\gamma \end{bmatrix} . \tag{5.1.47}$$

This is an equivalent stochastic problem, operating on a "filtered" set of the unknowns, $\hat{\mathbf{x}}$. From the previous section, define the stochastic solution of the filtered model as

$$\hat{\mathbf{x}} = \hat{\mathbf{H}}\,\mathbf{y} .$$

Transform back to the original variables (5.1.46) for

$$\mathbf{x} = \mathbf{F}^{-1}\hat{\mathbf{H}}\,\mathbf{y} . \tag{5.1.48}$$

From (5.1.18) and (5.1.46) let

$$\mathbf{A}\mathbf{F}^{-1} = \hat{\mathbf{A}} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{V}}^T .$$

Substituting into (5.1.48) the stochastic inverses (5.1.36), (5.1.41), and (5.1.44) gives equivalent explicit forms for the filtered inverse

$$\mathbf{H}_{fd} = \mathbf{F}^{-1} (\hat{\mathbf{A}}^T\hat{\mathbf{A}} + \gamma^2\mathbf{I})^{-1} \hat{\mathbf{A}}^T \tag{5.1.49}$$

$$\mathbf{H}_{fc} = \mathbf{F}^{-1} \hat{\mathbf{A}}^T(\hat{\mathbf{A}}\hat{\mathbf{A}}^T + \gamma^2\mathbf{I})^{-1} \tag{5.1.50}$$

$$\mathbf{H}_{fs} = \mathbf{F}^{-1} \hat{\mathbf{V}}\, \tilde{\hat{\mathbf{\Lambda}}}^{-1}\, \hat{\mathbf{U}}^T \tag{5.1.51}$$

The filtered inverse is useful in problems where a direct minimization of the solution vector norm is inappropriate. Instead, it puts constraints on the norm of a linear transform of the solution vector ($\hat{\mathbf{x}}$). The next

section will discuss a specific example of this, which is useful in one-dimensional problems.

5.1.8 Differential inversion

In one-dimensional problems involving gradients, such as the velocity distribution in plane-layered media, a natural constraint is on the norm of the solution gradient, instead of the solution magnitude. This can be realized by a first-order difference filter operating on the solution vector.

Let

$$\mathbf{x}' = \mathbf{F}_d \, \mathbf{x} \, ,$$

where

$$\mathbf{F}_d = \begin{bmatrix} 1 & -1 & 0 & 0 & & & & 0 \\ 0 & 1 & -1 & 0 & . & . & . & 0 \\ 0 & 0 & 1 & -1 & & & & 0 \\ & . & & & . & & & 0 \\ & . & & & & . & & 0 \\ & . & & & & & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} . \qquad (5.1.52)$$

If z is the one-dimensional coordinate, then $\mathbf{x}'$ is the approximate differential of $\mathbf{x}$ at z. The inverse to $\mathbf{F}_d$ exists, being an integration operator defined by the upper triangular matrix

$$\mathbf{F}_d^{-1} = \begin{bmatrix} 1 & 1 & 1 & 1 & & & 1 \\ 0 & 1 & 1 & 1 & . & . & 1 \\ 0 & 0 & 1 & 1 & & & 1 \\ 0 & 0 & 0 & 1 & & & 1 \\ & . & & & . & & 1 \\ & . & & & & . & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} .$$

Equation (5.1.52) defines a linear transform of $\mathbf{x}$, so the filtered inverse of the previous section applies. From (5.1.46) and (5.1.47), the least-squares minimization function is

$$M(\mathbf{x}) = |\mathbf{y} - \mathbf{A}\mathbf{x}|^2 + \gamma^2 |\mathbf{x}'|^2 .$$

The constraint for stabilizing the least-squares problem is now in terms of the differential $\mathbf{x}'$. The solution vector can be found by the appropriate inverse in (5.1.49-5.1.51). Examples of differential inversion will be given in the section on surface wave inversion.

5.1.9 Data and model covariance

In order to completely describe observed data, it is necessary to know the error associated with each data point. This is accomplished by treating the observations as random variables, with known means and covariances. The model can also be treated as a set of random variables, with means and covariances determined as functions of the observations and the inverse operators.

Jenkins and Watts (1968, p. 72) define means and covariances in terms of expected values. Given a random variable x, the mean is

$$<x> = \int_{-\infty}^{\infty} x\, f(x)\, dx , \qquad (5.1.53))$$

where $f(x)$ is the probability density function of x. For two random variables x_1 and x_2, the covariance between them is defined as

$$<(x_1 - <x_1>)(x_2 - <x_2>)> =$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1 - <x_1>)(x_2 - <x_2>)\, f_{12}(x_1, x_2) dx_1 dx_2 , \qquad (5.1.54)$$

where $f_{12}(x_1, x_2)$ is the joint probability density function. The variance of a random variable is a special case of (5.1.54):

$$<(x - <x>)^2> = \int_{-\infty}^{\infty} (x - <x>)^2 f(x)\, dx . \qquad (5.1.55)$$

By definition, if the variables are independent, the joint probability density function is a multiple of the individual probability density functions:

$$f_{12}(x_1, x_2) = f_1(x_1) f_2(x_2) .$$

In this case , substituting into (5.1.54) and manipulating the integral gives

$$<(x_1 - <x_1>)(x_2 - <x_2>)> = <x_1 - <x_1>><x_2 - <x_2>> = 0 .$$

For a set of observations defined by the vector $\mathbf{y}$, the vector mean is defined as the mean of each element:

$$<\mathbf{y}> = (<y_1>, <y_2>, \ldots <y_m>)^T \qquad (5.1.56)$$

The transpose indicates that $\mathbf{y}$ is a column vector. The covariance is defined as the matrix

$$\mathbf{D} = < (\mathbf{y} - <\mathbf{y}>)(\mathbf{y} - <\mathbf{y}>)^T > . \qquad (5.1.57)$$

It is understood that the expected value operates on each element of the matrix, as in (5.1.56). An important feature of the covariance matrix is symmetry:

$$\mathbf{D} = \mathbf{D}^T . \qquad (5.1.58)$$

A feature of the expected value is linearity. Given two constant matrices $\mathbf{A}$ and $\mathbf{B}$,

$$< \mathbf{Ay} + \mathbf{By} > \; = \; \mathbf{A} <\mathbf{y}> \; + \; \mathbf{B} <\mathbf{y}> \; . \qquad (5.1.59)$$

This can be verified from (5.1.53) and (5.1.56).

To find the covariance matrix for the model, first assume that a least-squares inverse matrix has been found relating $\mathbf{x}$ to $\mathbf{y}$:

$$\mathbf{x} \; = \; \mathbf{H}\,\mathbf{y} \; . \qquad (5.1.60)$$

If $\mathbf{x}$ and $\mathbf{y}$ are considered to be random variables, the mean of (5.1.60) is

$$<\mathbf{x}> \; = \; <\mathbf{Hy}> \; = \; \mathbf{H}<\mathbf{y}> \; . \qquad (5.1.61)$$

From (5.1.57), define the model covariance as

$$\mathbf{C} \; = \; < (\mathbf{x} - <\mathbf{x}>)(\mathbf{x} - <\mathbf{x}>)^T > \; . \qquad (5.1.62)$$

Substitute (5.1.60) and (5.1.61) into (5.1.62)

$$\mathbf{C} \; = \; < (\mathbf{Hy} - \mathbf{H}<\mathbf{y}>)(\mathbf{Hy} - \mathbf{H}<\mathbf{y}>)^T > \; .$$

Factor out the inverse matrix

$$\mathbf{C} \; = \; < \mathbf{H}(\mathbf{y} - <\mathbf{y}>)(\mathbf{y} - <\mathbf{y}>)^T \mathbf{H}^T > \; .$$

Use the linearity of the expected value operator for

$$\mathbf{C} \; = \; \mathbf{H} <(\mathbf{y} - <\mathbf{y}>)(\mathbf{y} - <\mathbf{y}>)^T > \mathbf{H}^T \; ,$$

and substitute in (5.1.57) for the final result

$$\mathbf{C} \; = \; \mathbf{H}\,\mathbf{D}\,\mathbf{H}^T \; . \qquad (5.1.63)$$

As a special case, assume that the data are independent random variables, and that the variance for all the observations are equal. Under these conditions,

$$\mathbf{D} \; = \; \sigma^2 \mathbf{I} \; , \qquad (5.1.64)$$

where σ^2 is the data variance, and $\mathbf{I}$ is an identity matrix. Substituting

into (5.1.63) gives

$$\mathbf{C} = \sigma^2 \mathbf{H}\,\mathbf{H}^T \quad . \tag{5.1.65}$$

It must be emphasized at this point that (5.1.63) is an intermediate result, valid only for the special case (5.1.65). For the general data covariance matrix, the least-squares minimization function should be modified to give more weight to observations with small variances, and less weight to large variances. This results in "maximum likelihood" inversion (Menke, 1984, p. 79), which is discussed in the next section.

5.1.10 Maximum likelihood inversion

Maximum likelihood is a statistical method for estimating parameters of a given probability density function, based on samples of random variables. Let

$$P(\mathbf{y}) = f(\mathbf{y}, \theta)$$

be an assumed joint probability density function of $\mathbf{y}$, where θ represents known parameters of the function, such as the variance or mean. If $\mathbf{y}_0$ is a sample of the random variable $\mathbf{y}$, the likelihood function is defined as (Jenkins and Watts, 1968, p. 116):

$$L(\theta) = f(\mathbf{y}_0, \theta) \, .$$

The likelihood function differs from the probability density function in that the parameters (θ) are now assumed to be unknown, and the sampled values of the random variable are fixed. The maximum likelihood estimate is defined from

$$L(\hat{\theta}) = \max[\, f(\mathbf{y}_0, \theta) \,] \tag{5.1.66}$$

The estimate $\hat{\theta}$ is defined at the point which maximizes the probability of the sampled data $\mathbf{y}_0$.

To determine the mean of sampled observations, the maximum likelihood estimate is written

$$L(<\hat{\mathbf{y}}>) = \max[\, f(\mathbf{y}_0, <\mathbf{y}>)\,] \quad . \tag{5.1.67}$$

Assuming that the data is related to the model by (5.1.16)

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \boldsymbol{\epsilon}\,,$$

the mean value is

$$<\mathbf{y}> = \mathbf{A}<\mathbf{x}> + <\boldsymbol{\epsilon}> \quad .$$

If it is also assumed that

$$<\boldsymbol{\epsilon}> = 0 \tag{5.1.68}$$

then $\mathbf{A}<\mathbf{x}>$ is an unbiased estimate of the mean $<\mathbf{y}>$. Substituting into (5.1.67) gives the maximum likelihood estimate in terms of the model

$$L(<\hat{\mathbf{x}}>) = \max[\, f(\mathbf{y}_0, \mathbf{A}<\mathbf{x}>)\,] \quad . \tag{5.1.69}$$

Equation (5.1.69) is the basis for maximum likelihood inversion. While least-squares inversion finds a model by minimizing residuals, maximum likelihood determines the model by maximizing the probability of sampled data.

For many geophysical problems, a reasonable probability density function for observed data is the multivariate normal distribution (Menke, 1984, p. 30):

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{m/2}\,|\mathbf{D}\,|^{1/2}} \exp[\, -\frac{1}{2}(\mathbf{y} - <\mathbf{y}>)^T \mathbf{D}^{-1}(\mathbf{y} - <\mathbf{y}>)\,] \tag{5.1.70}$$

where m is the number of observations, and $\mathbf{D}$ is the data covariance matrix. Assuming that the model gives an unbiased estimate to the data (5.1.68), and that $\mathbf{y}_0$ is the sampled data, the maximum likelihood estimate of the model (5.1.69) is found from

$$L(<\hat{\mathbf{x}}>) = \max\left\{ \exp[-\frac{1}{2}(\mathbf{y}_0 - \mathbf{A}<\mathbf{x}>)^T \mathbf{D}^{-1}(\mathbf{y}_0 - \mathbf{A}<\mathbf{x}>)] \right\} . \quad (5.1.71)$$

Since the argument of the exponent is negative, (5.1.71) will have a maximum when the argument is minimized. This leads to the definition of the generalized least squares minimization function:

$$M(\mathbf{x}) = (\mathbf{y} - \mathbf{A}\mathbf{x})^T \mathbf{D}^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x}) = \boldsymbol{\epsilon}^T \mathbf{D}^{-1} \boldsymbol{\epsilon} , \quad (5.1.72)$$

where $\mathbf{y}$ now represents the sampled data, and $\mathbf{x}$ is the model mean.

Equation (5.1.72) can be transformed into a standard minimization function by modifying the original least-squares problem (5.1.16). Since $\mathbf{D}$ is symmetric (5.1.58), $\mathbf{D}^{-1}$ is symmetric, and a matrix $\mathbf{E}$ can be found such that

$$\mathbf{D}^{-1} = \mathbf{E}^T \mathbf{E} . \quad (5.1.73)$$

This can be verified by writing $\mathbf{D}$ in terms of a symmetric singular value decomposition (Jackson, 1972)

$$\mathbf{D} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T , \quad \mathbf{D}^{-1} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}^T . \quad (5.1.74)$$

Define $\mathbf{E}$ as

$$\mathbf{E} = \boldsymbol{\Lambda}^{-1/2} \mathbf{V}^T . \quad (5.1.75)$$

Then

$$\mathbf{E}^T\mathbf{E} = \mathbf{V}\boldsymbol{\Lambda}^{-1/2} \boldsymbol{\Lambda}^{-1/2}\mathbf{V}^T = \mathbf{D}^{-1} ,$$

and (5.1.73) is proved. To transform the original least-squares problem, multiply both sides of (5.1.16) by **E**:

$$\mathbf{Ey} = \mathbf{EAx} + \mathbf{E}\epsilon \,. \tag{5.1.76}$$

This defines a new least-squares problem

$$\bar{\mathbf{y}} = \bar{\mathbf{A}}\mathbf{x} + \bar{\epsilon} \,, \tag{5.1.77}$$

with a minimization function given by

$$M(\mathbf{x}) = \bar{\epsilon}^T\bar{\epsilon} = \epsilon^T\mathbf{E}^T\mathbf{E}\epsilon = \epsilon^T\mathbf{D}^{-1}\epsilon \,, \tag{5.1.78}$$

which is equivalent to (5.1.72).

The above shows that a least-squares problem can be transformed into a maximum likelihood problem by modifying the equations according to (5.1.76). Since this defines a new least-squares problem, the inverse matrices (5.1.34), (5.1.41), (5.1.44), or (5.1.49-5.1.51) can be used to find the model estimate:

$$\mathbf{x} = \bar{\mathbf{H}}\,\bar{\mathbf{y}} \,. \tag{5.1.79}$$

The bar above the inverse indicates that it is constructed for the transformed system (5.1.77). The model covariance matrix can be found from (5.1.63)

$$\bar{\mathbf{C}} = \bar{\mathbf{H}}\,\bar{\mathbf{D}}\,\bar{\mathbf{H}}^T \,, \tag{5.1.80}$$

where $\bar{\mathbf{D}}$ is the transformed data covariance matrix, defined from (5.1.57) as

$$\bar{\mathbf{D}} = < (\bar{\mathbf{y}} - <\bar{\mathbf{y}}>)(\bar{\mathbf{y}} - <\bar{\mathbf{y}}>)^T > \,.$$

Going back to the original coordinates gives

$$\bar{\mathbf{D}} = < (\mathbf{Ey} - \mathbf{E}<\mathbf{y}>)(\mathbf{Ey} - \mathbf{E}<\mathbf{y}>)^T > \,.$$

Factoring out $\mathbf{E}$ results in

$$\overline{\mathbf{D}} = \mathbf{E} <(\mathbf{y} - <\mathbf{y}>)(\mathbf{y} - <\mathbf{y}>)^T > \mathbf{E}^T \ ,$$

which from (5.1.57) is

$$\overline{\mathbf{D}} = \mathbf{E}\mathbf{D}\mathbf{E}^T \ . \tag{5.1.81}$$

From (5.1.74) and (5.1.75), the matrix in (5.1.81) can be written

$$\mathbf{E}\mathbf{D}\mathbf{E}^T = \mathbf{\Lambda}^{-1/2}\mathbf{V}^T \ \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \ \mathbf{V}\mathbf{\Lambda}^{-1/2} = \mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}\mathbf{\Lambda}^{-1/2} = \mathbf{I} \ ,$$

and (5.1.81) reduces to an identity matrix

$$\overline{\mathbf{D}} = \mathbf{I} \ . \tag{5.1.82}$$

Substituting (5.1.82) into (5.1.80) gives the final form for the maximum likelihood model covariance matrix:

$$\overline{\mathbf{C}} = \overline{\mathbf{H}}\,\overline{\mathbf{H}}^T \ . \tag{5.1.83}$$

It should be noted that for the special case (5.1.64) of a diagonal data covariance matrix with equal variances, (5.1.83) is equal to the least-squares model covariance matrix (5.1.65).

In terms of the filtered inverses (5.1.49-5.1.51), the maximum likelihood model covariance can be found by substituting into (5.1.83). Assume that the original problem has been transformed to a maximum likelihood problem by (5.1.76), defining the new least-squares problem

$$\overline{\mathbf{y}} = \overline{\mathbf{A}}\mathbf{x} + \overline{\boldsymbol{\epsilon}} \ .$$

Let

$$\overline{\mathbf{A}}\mathbf{F}^{-1} = \hat{\mathbf{A}} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{V}}^T \ .$$

The covariance matrices for (5.1.49-5.1.51) are

$$\overline{\mathbf{C}}_{fd} = \mathbf{F}^{-1} (\hat{\mathbf{A}}^T \hat{\mathbf{A}} + \gamma^2 \mathbf{I})^{-1} \hat{\mathbf{A}}^T \hat{\mathbf{A}} (\hat{\mathbf{A}}^T \hat{\mathbf{A}} + \gamma^2 \mathbf{I})^{-1} (\mathbf{F}^{-1})^T \quad (5.1.84)$$

$$\overline{\mathbf{C}}_{fc} = \mathbf{F}^{-1} \hat{\mathbf{A}}^T (\hat{\mathbf{A}} \hat{\mathbf{A}}^T + \gamma^2 \mathbf{I})^{-1} (\hat{\mathbf{A}} \hat{\mathbf{A}}^T + \gamma^2 \mathbf{I})^{-1} \hat{\mathbf{A}} (\mathbf{F}^{-1})^T \quad (5.1.85)$$

$$\overline{\mathbf{C}}_{fs} = \mathbf{F}^{-1} \hat{\mathbf{V}} \tilde{\hat{\mathbf{\Lambda}}}^{-2} \hat{\mathbf{V}}^T (\mathbf{F}^{-1})^T \quad (5.1.86)$$

where

$\tilde{\hat{\mathbf{\Lambda}}}^{-2}$ = nxn diagonal matrix with k non-zero elements defined by

$\tilde{\hat{\mathbf{\Lambda}}}_j^{-2} = \hat{\lambda}_j^2 / (\hat{\lambda}_j^2 + \gamma^2)^2$.

For non-zero γ, all these forms are equivalent. If $\mathbf{F}$ is an identity matrix, (5.1.84-5.1.86) reduce to equivalent stochastic covariance matrices.

5.1.11 Model resolution

The Backus-Gilbert method, outlined in equations (5.1.3-5.1.8), can also be applied to discrete matrix problems (Menke, 1984, p. 61; Twomey, 1977, p. 169). Assume the problem can be written as

$$\mathbf{y} = \mathbf{A}\mathbf{x} \ .$$

Multiply both sides by $\mathbf{H}$, an unknown inverse matrix

$$\mathbf{H}\mathbf{y} = \mathbf{H}\mathbf{A}\mathbf{x} \ .$$

This can be written as

$$\mathbf{H}\mathbf{y} = \mathbf{R}\mathbf{x} \ , \quad (5.1.87)$$

where

$$\mathbf{R} = \mathbf{H}\mathbf{A} \quad (5.1.88)$$

is the resolving kernel of the system. If the elements of the $\mathbf{H}$ matrix can

be found such that

$$\mathbf{R}\,\mathbf{x} \approx \mathbf{I}\,\mathbf{x} = \hat{\mathbf{x}}\,, \tag{5.1.89}$$

then substituting into (5.1.87) gives the solution

$$\hat{\mathbf{x}} = \mathbf{H}\,\mathbf{y}\ .$$

From (5.1.89), the resolving kernel constructs the estimate $\hat{\mathbf{x}}$ as an average over the solution space $\mathbf{x}$.

The Backus-Gilbert method differs from least-squares and maximum likelihood inversion in that the inverse is found independently of the data. The elements of $\mathbf{H}$ are constructed by minimizing the difference, subject to variance constraints, between the resolving kernel and an identity matrix, instead of between observed and theoretical data. However, maximum likelihood and least-squares inverses can be used to define the resolving kernel (5.1.88). The meaning is the same as in (5.1.89), except that the concept of resolution is not directly addressed in constructing the inverse matrices.

Assume that the original problem has been transformed to a maximum likelihood problem by (5.1.76), defining a new least-squares problem

$$\bar{\mathbf{y}} = \overline{\mathbf{A}}\,\mathbf{x} + \bar{\epsilon}\ .$$

From (5.1.88), the resolving kernel of the transformed system is

$$\overline{\mathbf{R}} = \overline{\mathbf{H}}\,\overline{\mathbf{A}}\ . \tag{5.1.90}$$

For filtered inverses, let

$$\overline{\mathbf{A}}\,\mathbf{F}^{-1} = \hat{\mathbf{A}} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{V}}^{T}\ .$$

Substituting (5.1.49-5.1.51) into (5.1.90) gives the resolving kernels:

$$\overline{\mathbf{R}}_{fd} = \mathbf{F}^{-1} (\hat{\mathbf{A}}^T \hat{\mathbf{A}} + \gamma^2 \mathbf{I})^{-1} \hat{\mathbf{A}}^T \hat{\mathbf{A}} \mathbf{F} \tag{5.1.91}$$

$$\overline{\mathbf{R}}_{fc} = \mathbf{F}^{-1} \hat{\mathbf{A}}^T (\hat{\mathbf{A}} \hat{\mathbf{A}}^T + \gamma^2 \mathbf{I})^{-1} \hat{\mathbf{A}} \mathbf{F} \tag{5.1.92}$$

$$\overline{\mathbf{R}}_{fs} = \mathbf{F}^{-1} \hat{\mathbf{V}} \hat{\Sigma} \hat{\mathbf{V}}^T \mathbf{F} \tag{5.1.93}$$

where

$\hat{\Sigma}$ = nxn diagonal matrix with k non-zero elements defined by

$\hat{\Sigma}_j = \hat{\lambda}_j^2/(\hat{\lambda}_j^2+\gamma^2)$.

As in the case of the covariance matrices, (5.1.91-5.1.93) are equivalent for non-zero γ.

5.2 Surface-wave inversion

5.2.1 Integral forms

In chapter 2, variational methods were used to find first-order phase velocity perturbations to Love and Rayleigh waves. These were in terms of perturbations to density (ρ), shear velocity (β), and in the case of Rayleigh waves, compressional velocity (α). The integral form for Love waves is given by (2.2.6), and for Rayleigh waves by (2.2.12). Given phase velocity as a measured quantity, the integral equations can be directly set up for inversion, as shown in section 5.1. However, the unknowns α, β, and ρ are not independent of each other, as the integrals imply. Compressional and shear velocity are related by Poisson's ratio r_p, according to (Bullen and Bolt, 1985):

$$\frac{\beta}{\alpha} = \left(\frac{1 - 2r_p}{2 - 2r_p} \right)^{\frac{1}{2}} .$$

In addition, compressional velocities can be empirically related to density in the upper crust by the Nafe-Drake relation (Talwani *et al.*, 1959), and at depth by Birch's law (Birch, 1964). Der and Landisman (1972) pointed out that attempting to invert for all the unknowns simultaneously significantly degrades their resolution. However, they also showed that changes to phase velocities are dominated by shear velocity perturbations, with relatively small contributions coming from density and compressional velocity perturbations.

Since the integrals are only accurate to first order, several inversion iterations may be required to converge to a solution. An alternate scheme to simultaneously inverting all parameters is to use the dominance of shear velocity perturbations. Holding compressional velocity and density fixed, the integral equations can be inverted for shear velocity alone. The solution can be used to update compressional velocity using Poisson's ratio, and then density using the Nafe-Drake relation and Birch's law. The integrals can be recalculated in terms of the updated quantities, and the procedure repeated until convergence is achieved.

The Love wave equation (2.2.6) in terms of shear velocity perturbations alone is

$$\delta c(\omega) = \int_0^\infty A_L(\omega,z)\, \delta\beta(z)\, dz \quad , \tag{5.2.1}$$

where

$$A_L(\omega,z) = \frac{\mu c}{\beta k W}\left[k^2 l_1 + \left(\frac{dl_1}{dz}\right)^2\right] , \tag{5.2.2}$$

and

$$W = k\int_0^\infty \mu l_1^2 \, dz \quad . \tag{5.2.3}$$

It is understood that all terms except μ in (5.2.2) and (5.2.3) are frequency dependent. The Rayleigh wave equation (2.2.12) can be written as

$$\delta c(\omega) = \int_0^\infty A_R(\omega,z)\, \delta\beta(z)\, dz \quad , \tag{5.2.4}$$

where

$$A_R(\omega,z) = \frac{\mu c}{\beta k W}\left[\left(kr_2 + \frac{dr_1}{dz}\right)^2 + 4kr_1\frac{dr_2}{dz}\right] , \tag{5.2.5}$$

and

$$W = k\int_0^\infty [(\lambda + 2\mu)r_1^2 + \mu r_2^2]dz + \int_0^\infty (\mu r_2\frac{dr_1}{dz} - \lambda r_1\frac{dr_2}{dz})dz \quad . \tag{5.2.6}$$

Again, all terms in (5.2.5) and (5.2.6) are frequency dependent, except λ and μ.

5.2.2 Discrete modeling

Equations (5.2.1) and (5.2.4) are in the form of a Fredholm integral of the first kind, which is suitable for linear inversion. If measured phase velocities are sampled at discrete points, the equations can be expressed as (5.1.2):

$$\delta c_i = \int_0^\infty A_i(z)\, \delta\beta(z)\, dz \quad , \qquad i = 1,2,....m \tag{5.2.7}$$

where m is the number of frequency observations and the kernel $A_i(z)$ corresponds to either A_L or A_R. If the velocity structure is assumed to be plane layered, (5.2.7) can be partitioned as in (5.1.13):

$$\delta c_i = \sum_{j=1}^{n} A_{ij}\ \delta\beta_j \quad , \qquad j = 1,2,\ldots n \tag{5.2.8}$$

where n is the number of shear velocity layers and

$$A_{ij} = \int_{z_j}^{z_{j+1}} A_i(z)\ dz \quad . \tag{5.2.9}$$

To set up (5.2.8) as an iterative problem, let β_j^o $(j=1,n)$ be an initial guess to the shear velocity structure. Let c_i^o be the corresponding theoretical phase velocity, and let c_i^m be the measured phase velocity at the i'th frequency. Then, if

$$\delta c_i = c_i^m - c_i^o \quad ,$$

the problem is to find the perturbations $\delta\beta_j$ that minimizes the residuals in

$$\delta c_i = \sum_{j=1}^{n} A_{ij}\ \delta\beta_j + \epsilon_i \quad . \tag{5.2.10}$$

The residual term is added since the first-order sum may not exactly equal δc_i, the phase velocity perturbation. This is now a standard least-squares problem, and can be put in vector-matrix form as (5.1.16):

$$\delta\mathbf{c} = \mathbf{A}\ \delta\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad , \tag{5.2.11}$$

where the dimensions are the same as for (5.1.16). Once $\delta\boldsymbol{\beta}$ is found, the original estimate can be updated as

$$\boldsymbol{\beta}^u = \boldsymbol{\beta}^o + \delta\boldsymbol{\beta} \quad .$$

The updated shear velocities can now be used to recalculate the matrix $\mathbf{A}$, and the process repeated.

5.2.3 Partial derivatives and group velocities

Another way of setting up the perturbation problem (5.2.8) is to consider phase velocity as a function of discrete shear velocities:

$$c_i^m = c_i(\beta) + \epsilon_i \quad ,$$

where the subscript represents the i'th frequency and the residual compensates for errors between the observed and theoretical phase velocity. This can be expanded in a Taylor series about about an initial model β^o as

$$c_i^m = c_i(\beta^o) + \sum_{j=1}^{n} \frac{\partial c_i}{\partial \beta_j} \delta\beta_j + \epsilon_i \quad . \tag{5.2.12}$$

Equating (5.2.12) with (5.2.10) shows that the partial derivatives are formally equivalent to the integral (5.2.9):

$$\frac{\partial c_i}{\partial \beta_j} \equiv \int_{z_j}^{z_{j+1}} A_i(z)\, dz \quad . \tag{5.2.13}$$

A possible approach to the perturbation problem would be to simply calculate the partial derivatives in (5.2.12) numerically. However, using a method developed by D. G. Harkrider, Wang (1981, p. 80) showed that integrals such as (5.2.13) can be solved analytically. This results in enormous improvements in efficiency and stability over numerical methods.

Aki and Richards (1980, p. 291) calculated theoretical group velocities using variational methods. Group velocity measurements can be found using methods such as the multiple-filter technique (4.1.7). Thus, a similar expression to (5.2.12) can be found by expanding group velocity in a Taylor series about a starting model β^o:

$$U_i^m = U_i(\beta^o) + \sum_{j=1}^{n} \frac{\partial U_i}{\partial \beta_j} \delta\beta_j + \epsilon_i \quad . \tag{5.2.14}$$

Unfortunately, it is not possible to analytically define group velocity partial derivatives. However, using central differences, Rodi *et al.* (1975) gave a stable approximation in terms of phase velocity partial derivatives:

$$\frac{\partial U_i}{\partial \beta_j} = P_i \frac{\partial c_i}{\partial \beta_j} + Q_i \left(\frac{\partial c_{i+1}}{\partial \beta_j} - \frac{\partial c_{i-1}}{\partial \beta_j} \right) \tag{5.2.15}$$

where

$$P_i = \frac{U_i}{c_i} \left(2 - \frac{U_i}{c_i} \right)$$

and

$$Q_i = \frac{\omega_i U_i^2}{2h \; c_i^2} \quad . \tag{5.2.16}$$

In (5.2.16), h is the discrete frequency sampling interval, and ω_i is the i'th frequency. Equation (5.2.14) can now be put into matrix form and inversion theory can be used to find shear velocity structure.

If group velocities give additional independent information to that found from phase velocities, equation (5.2.14) can be combined with (5.2.10) for a simultaneous inversion. Der and Landisman (1972) advocated this, stating that group velocity inversion improves structural resolution, under the condition that group velocity variances are similar in size to phase velocity variances.

To put this in perspective, construct a measured phase velocity data set over many equally spaced frequency intervals (h), such as found from phase matched filtering. Define the following matrix:

$$\mathbf{G} = \begin{bmatrix} 0 & 0 & 0 & 0 & & & & 0 \\ -Q_2 & P_2 & Q_2 & 0 & . & . & . & 0 \\ 0 & -Q_3 & P_3 & Q_3 & & & & 0 \\ & . & & & . & & & 0 \\ & . & & & & . & & 0 \\ & . & & & & -Q_{m-1} & P_{m-1} & Q_{m-1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} . \quad (5.2.17)$$

Transform the phase velocity equation (5.2.11) by

$$\mathbf{G}\ \delta\mathbf{c} = \mathbf{G}\ \mathbf{A}\ \delta\beta + \mathbf{G}\ \epsilon\ . \quad (5.2.18)$$

This is equivalent to (5.2.14), with

$$(\mathbf{G}\ \mathbf{A})_{ij} = \frac{\partial U_i}{\partial \beta_j}$$

defined by (5.2.15), and

$$(\mathbf{G}\ \delta\mathbf{c})_i = U_i^m - U_i(\beta^o)\ . \quad (5.2.19)$$

This transformation makes it possible to interpret group velocity inversion in terms of phase velocities. The least-squares minimization function for (5.2.18) is

$$M(\delta\beta) = \epsilon^T\ \mathbf{G}^T\mathbf{G}\ \epsilon\ . \quad (5.2.20)$$

Since $\mathbf{G}$ is not an orthogonal matrix, the minimization function (5.2.20) will not be the same as for phase velocities, which means that a different least-squares solution is possible for group velocities. This implies that for a simultaneous inversion, group velocities can contribute independent information, under the condition that the statistics for both sets of data are similar.

To determine the joint statistics, rewrite (5.2.19) as

$$\delta U_i = P_i \delta c_i + Q_i \ (\delta c_{i+1} - \delta c_{i-1}) \ . \qquad (5.2.21)$$

Let δU and δc be random variables with statistics described by the measured values. Define the variances as

$$\sigma_i^u = VAR\,(U_i^m) \ , \qquad \sigma_i^c = VAR\,(c_i^m) \ .$$

Since the group velocity is a linear combination of the phase velocities in (5.2.21), it can be shown that the group variance is (Jenkins and Watts, 1968, p. 73)

$$\sigma_i^u = P_i^2 \ \sigma_i^c \ + \ Q_i^2(\sigma_{i-1}^c + \sigma_{i+1}^c) \ - \ 2Q_i^2 \ COV(c_{i-1}^m \, , \ c_{i+1}^m)$$

$$+ P_i Q_i \ COV(c_i^m, \ c_{i+1}^m) \ - \ P_i Q_i \ COV(c_i^m, \ c_{i-1}^m) \ . \qquad (5.2.22)$$

where "COV" represents the covariance between the enclosed variables. Now, for a small value of h in (5.2.16), the phase velocities at adjacent frequency points will be quite close. This implies that for locally stationary statistics, the covariances in (5.2.22) will be almost identical, and the same for the variances. Under this condition, (5.2.22) reduces to

$$\sigma_i^u \approx P_i^2 \sigma_i^c \ + \ 2Q_i^2 \ [\sigma_i^c - COV(c_{i-1}^m \, , \ c_{i+1}^m)] \ . \qquad (5.2.23)$$

The correlation coefficient of adjacent phase velocities is defined as (Jenkins and Watts, 1968, p. 149)

$$\rho_i(2h) = \frac{COV(c_{i-1}^m \, , c_{i+1}^m)}{\sigma_i^c} \qquad (5.2.24)$$

The value "$2h$" indicates the spacing between adjacent phase velocities. If (5.2.24) is equal to one, the phase velocities are completely correlated, implying that any change in the first velocity will correspond to an equal change in the second. This is not unreasonable for smoothly varying phase velocities, when the value of h is small. If (5.2.24) is equal to zero,

the phase velocities are totally uncorrelated, implying that the phase velocity signal is equivalent to "white" noise. Substituting (5.2.24) into (5.2.23) gives

$$\sigma_i^u \approx P_i^2 \sigma_i^c + 2Q_i^2 \sigma_i^c [1 - \rho_i(2h)] \quad . \tag{5.2.25}$$

For highly correlated phase velocities, (5.2.24) will reduce to

$$\sigma_i^u \approx P_i^2 \sigma_i^c \ ,$$

and both group and phase velocities will be on the same order of magnitude. However, due to the presence of the h term in the denominator of Q_i, even low levels of noise in the phase velocity will result in unacceptably high variances in group velocity. This is intuitively obvious from the numerical approximation (5.2.18), which amplifies uncorrelated noise through the process of numerical differentiation.

The above analysis has the following implications. If coherent group velocity measurements are available with reasonable variances, the minimization function (5.2.19) indicates that group velocities will contribute independent information to the inversion problem. However, equation (5.2.25) implies that under noisy conditions, such as found with short array lengths in multi-channel processing, the use of phase velocities alone is preferable.

The next section will test the concepts presented in the last two chapters on real and synthetic data sets. Particular attention will be paid to subjecting multi-channel data sets to very high incoherent noise levels.

CHAPTER 6

SURFACE-WAVE APPLICATIONS

6.1 Differential inversion

Surface wave inversion is a poorly conditioned inversion problem when many thin layers are used to model the velocity medium. Typically, for a twenty layer medium, less than ten of the eigenvalues will be effectively non-zero, with the rest being five orders of magnitude smaller (or less) than the maximum. This requires strong constraints on the inversion, resulting in decreased resolution in the model.

Section 5.1.8 discussed the use of a first-order difference filter $\mathbf{F}_d$ to constrain the gradient of the solution vector $\mathbf{x}$ in linear inversion. In terms of a plane-layered medium, this is equivalent to constraining the difference between adjacent layers in the medium. This is an excellent constraint for surface-wave applications, since it allows simple, unbiased starting models for inversion, and the ability to include sharp layer discontinuities where *a priori* information exists on their location.

As defined by (5.1.52), $\mathbf{F}_d$ is not a true differential operator, since the bottom row is not identically zero, but has a one in the last column of the row. This is required to keep the matrix from becoming singular, but it has the effect in surface wave modeling of putting a direct constraint on the bottom layer (half-space), rather than the difference between the bottom layer and the one above. For highly constrained problems, this can result in the velocity for the half-space remaining close to the initial model.

To insure that the initial model does not bias the inversion, and to allow sharp discontinuities between selected layers, the following modification to the differential filter $\mathbf{F}_d$ is made. Let

$$\mathbf{F}_{dw} = \mathbf{W}\,\mathbf{F}_d \tag{6.1.1}$$

where $\mathbf{W}$ is a diagonal weighting matrix of the form

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & 0 & 0 & & & 0 \\ 0 & \epsilon & 0 & 0 & . & . & 0 \\ 0 & 0 & 1 & 0 & & & 0 \\ 0 & 0 & 0 & 1 & & & 0 \\ & . & & & . & & 0 \\ & . & & & & . & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \epsilon \end{bmatrix} .$$

The value of ϵ is a small, but non-zero number placed on rows corresponding to selected discontinuities, where little constraint in the inversion is desired. The reason for making ϵ non-zero is to keep $\mathbf{F}_{dw}$ from becoming singular. $\mathbf{F}_{dw}$ can now be used in the context of a filtered inversion, with the inversion matrices being given by (5.1.49-5.1.51).

It has been suggested that differential inversion results in "smoother" models than stochastic inversion, since the constraint is on layer differences, instead of the layers themselves (Claerbout, 1976, p. 120). This is not strictly correct, since "smoothness" is directly a function of the damping factor γ, not on the type of constraint used. For instance, if for a given γ, the differential inversion results in a smoother model than the stochastic inversion, the value of γ can be made smaller to bring out more detail in the differential inversion. In addition, the use of a weighting scheme such as (6.1.1) can result in sharp detail at selected discontinuities, although this must be considered as *a priori* information, not as a

consequence of the inversion.

6.2 Single-station group velocity inversion

To test differential inversion on surface waves, a set of group velocities were averaged, corresponding to a teleseismic path crossing Saudi Arabia, from an earthquake source in the Gulf of Aden to WWSSN station Tabriz, in northwestern Iran (group velocity data courtesy of Hafidh A. A. Ghalib). The distance across the path is 2662 km.

Figure 6.1 is a plot of the observed and inverted group velocity using differential inversion. In this run, no discontinuity weighting was applied. However, the half-space was allowed to freely move by setting $\epsilon = 0.2$. Figure 6.2 gives the shear velocity results and corresponding resolving kernels. The starting model was a sedimentary layer over a 5.2 km/sec shear velocity half space. Even with the low resolution expected on a fundamental-mode Rayleigh wave inversion, the Conrad discontinuity at 13 km, and the Moho at approximately 37 km, are discernible in the form of gradients. It should be noted that if the half-space was not weighted at 100 km, the approximately constant velocity between 50 and 100 km would appear to be an increasing gradient with depth, due to the half-space being constrained near the starting model of 5.2 km/sec.

Using the apparent gradients as a guide, another inversion was performed, with weights ($\epsilon = 0.1$) applied at 13 and 40 km. Figure 6.3 gives the group velocity results, showing no discernible change from Figure 6.1. This emphasizes the non-uniqueness of the surface-wave solution, and the fact that there is not enough information in the data to resolve sharp discontinuities. Figure 6.4 gives the shear velocity model with resolving

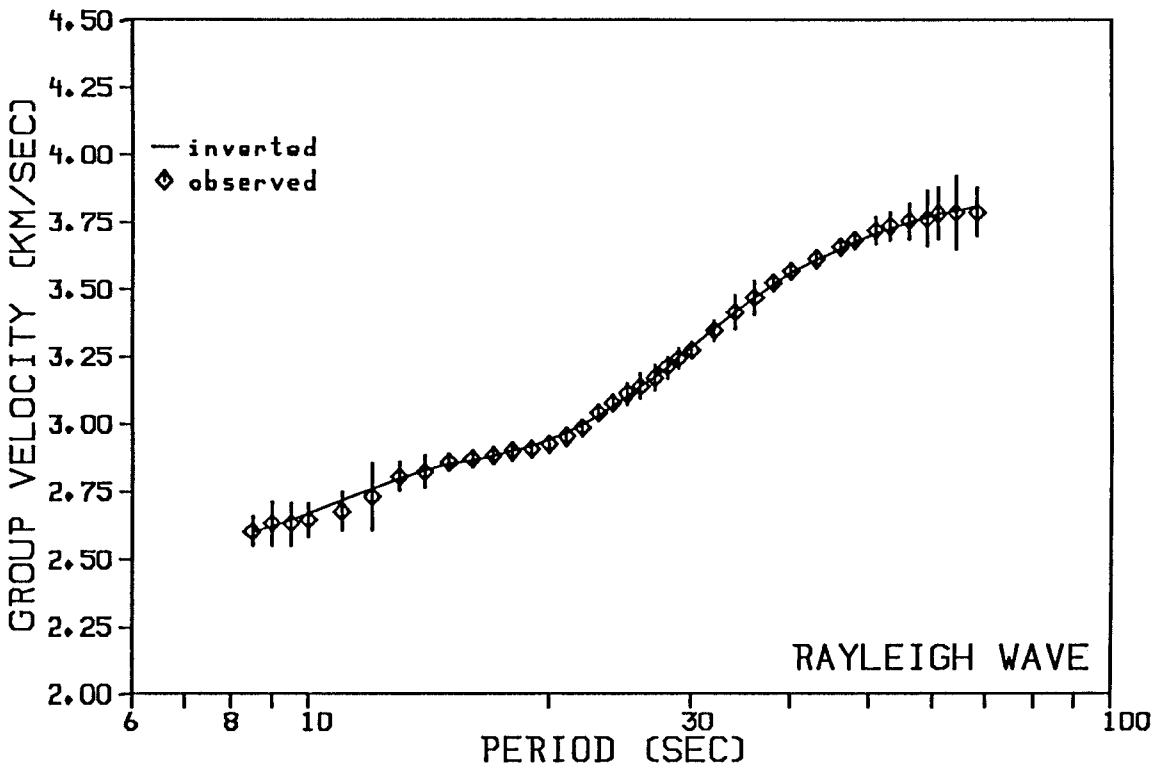

Figure 6.1 Comparison of observed and inverted group velocities for the Saudi Arabian plate inversion - no discontinuity weighting.

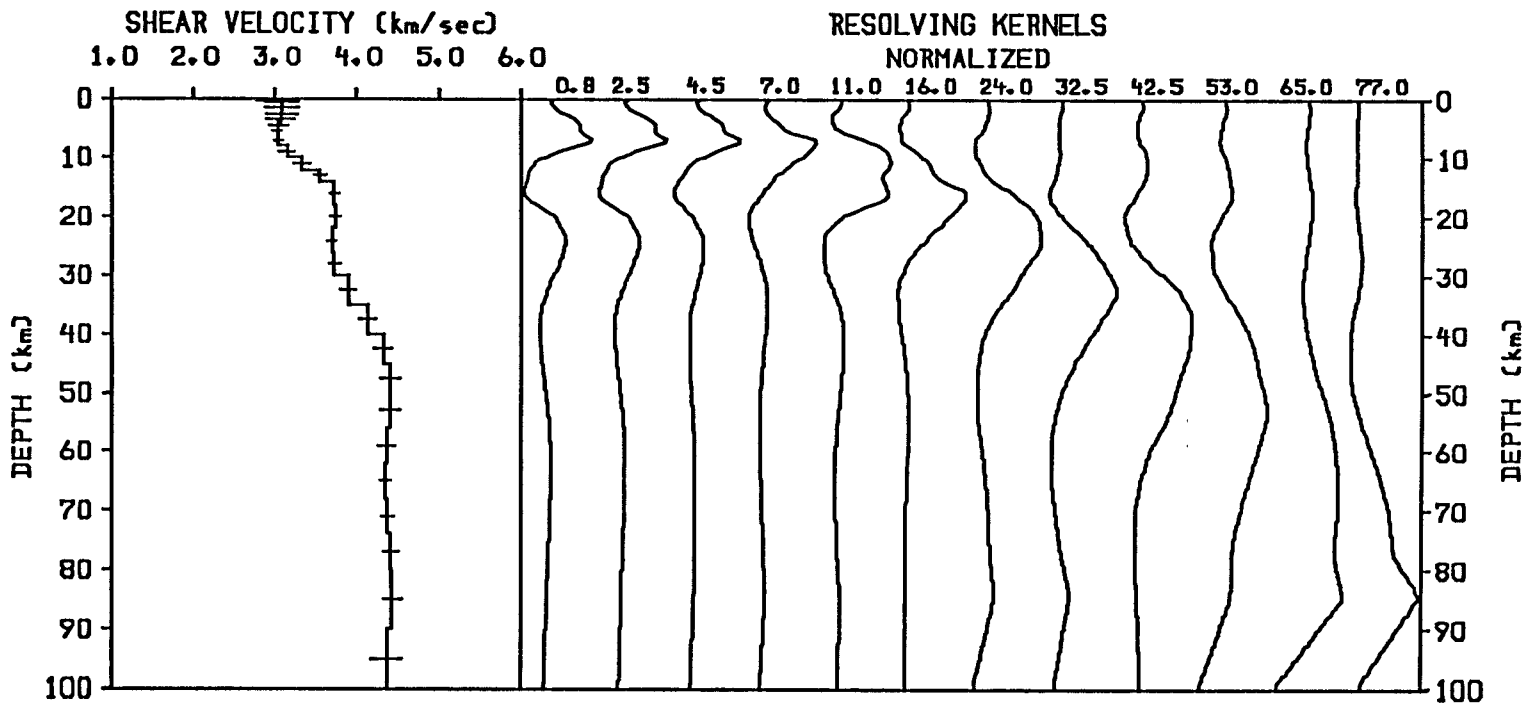

Figure 6.2 Shear velocities and resolving kernels for the Saudi Arabian plate inversion. Discontinuities were not emphasized on this inversion.

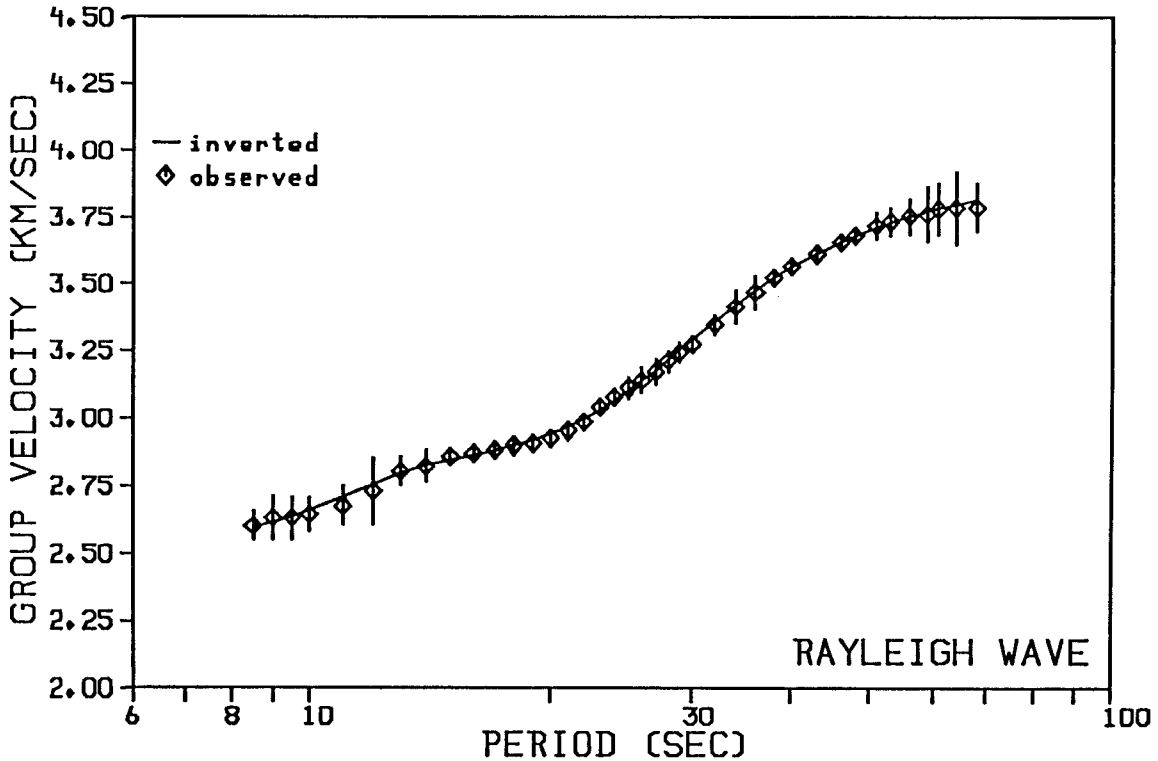

Figure 6.3 Comparision of observed and inverted group velocities for the Saudi Arabian plate inversion with discontinuity weighting.

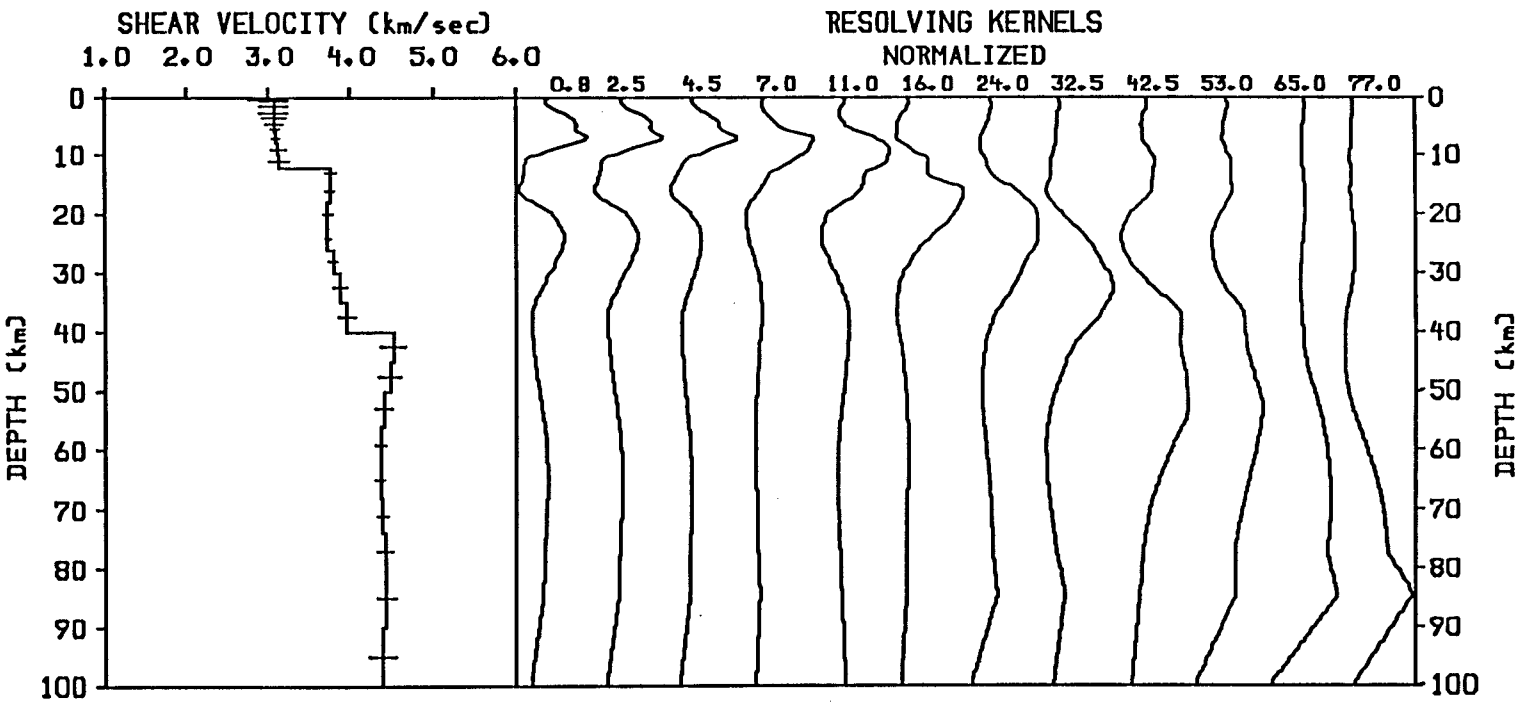

Figure 6.4 Shear velocities and resolving kernels for the Saudi Arabian plate inversion with emphasis on discontinuities. Notice the sharp velocity changes at 13 and 40 km depths.

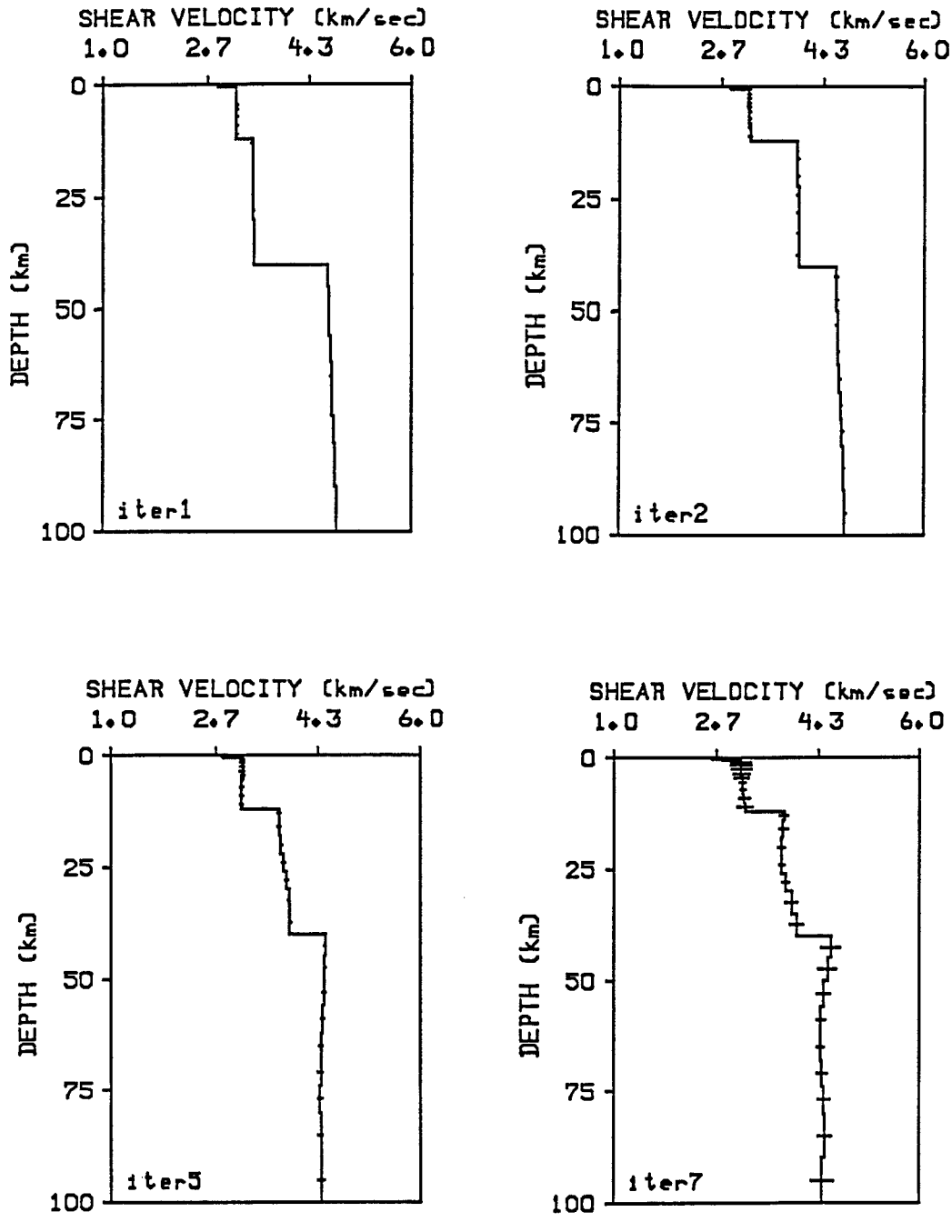

Figure 6.5 Intermediate steps in the Saudi Arabian plate inversion with discontinuity weighting. The iteration number is shown in the lower left corner of each plot.

kernels, with distinct breaks now obvious at 13 and 40 km. The gradients forced by the smooth solution in Figure 6.2 have disappeared.

It should be noted that initial model for the second run was the same as the first; no discontinuities were present, except for a sedimentary layer at the surface. Figure 6.5 shows the intermediate steps in the inversion. The character of the gross velocity structure is established by the second iteration, and subsequent iterations only bring out minor improvements in the detail of the solution.

6.3 Multi-channel data processing

In 1978, the U.S. Geological Survey conducted a deep seismic refraction survey transversing the Saudi Arabian shield (Healy, *et al.*, 1982). Twenty-second analog recordings of the explosion events were made on standard 2 hz geophones, which were subsequently digitized.

Some of the record sections exhibit pronounced fundamental-mode Rayleigh waves in the frequency range between 1 and 10 hz, to distances beyond 50 km from the sources. One of these, corresponding to shot point 3 northeast (Healy, *et al.*, 1982), was selected for testing the multi-channel phase-matched filter developed in chapter 4. Figure 6.6 is a plot of the raw data of the record section, clearly showing the pronounced fundamental-mode Rayleigh waves. All data traces were normalized to the same amplitude. To accentuate the Rayleigh waves, the data were passed through a 9 hz, low-pass, zero-phase, seven-pole Butterworth filter. The result is shown in Figure 6.7.

Using the multi-channel phase-matched filter, phase velocities were picked between 0.17 and 0.81 seconds, as shown on Figure 6.8. For an

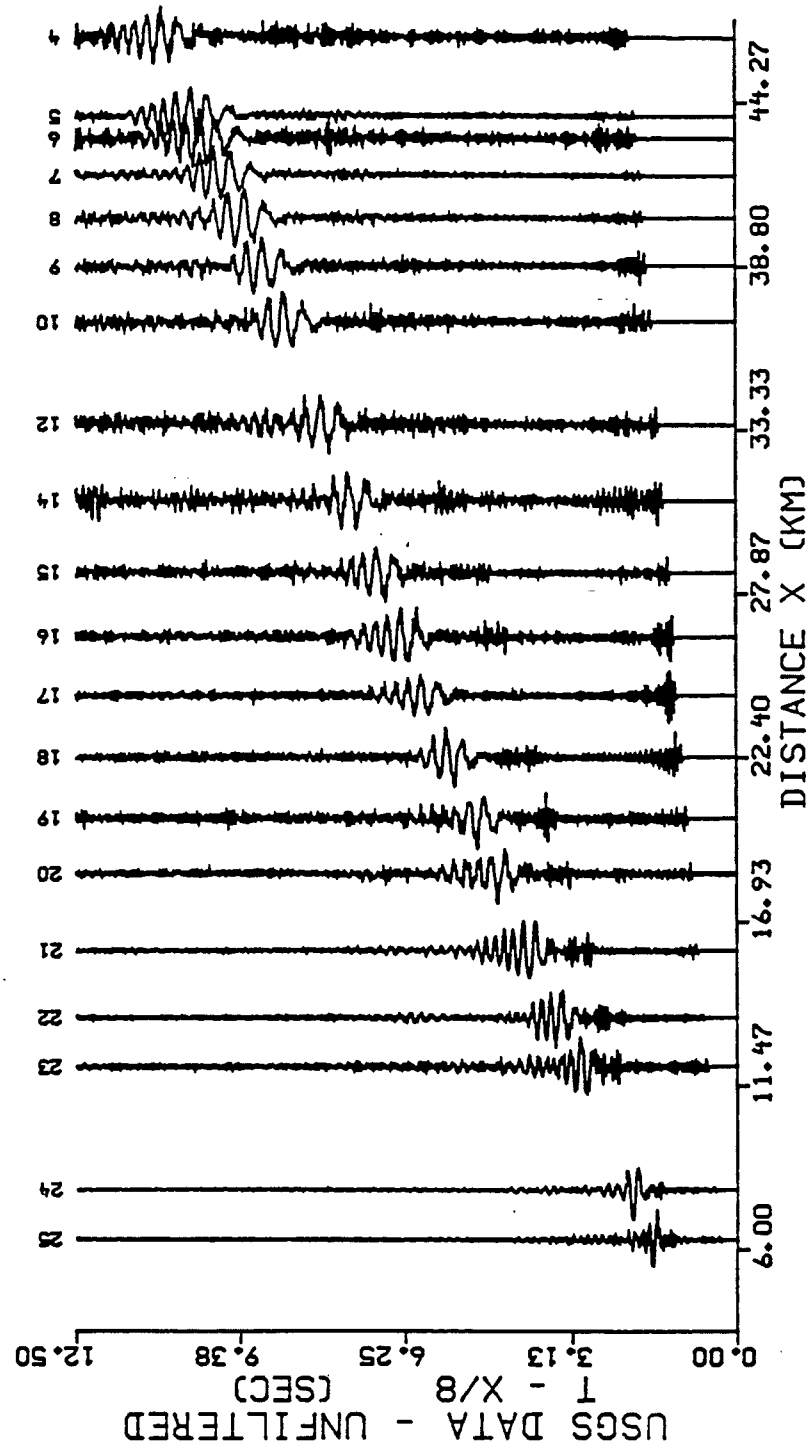

Figure 6.6 Multi-channel data set: raw data.

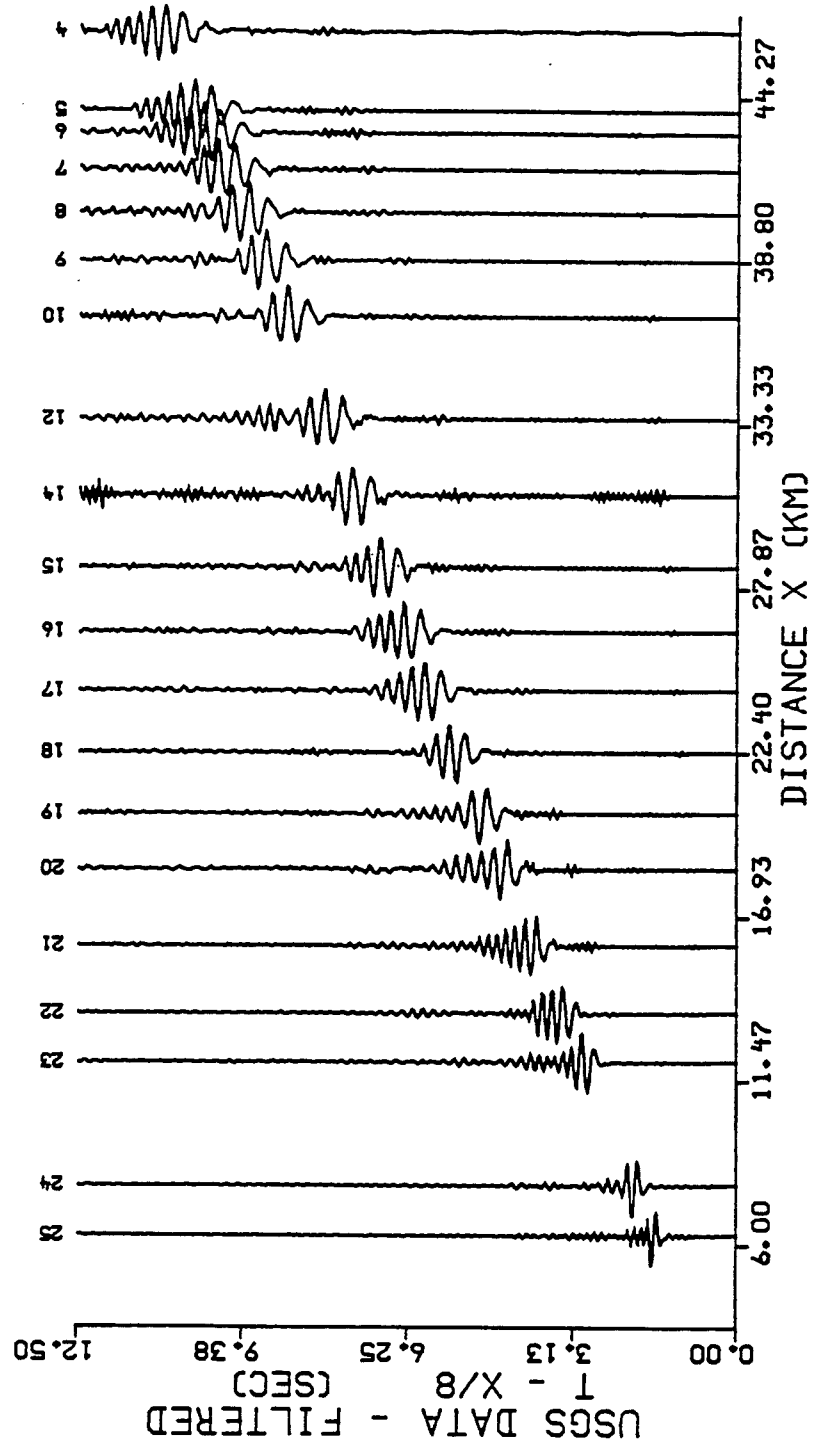

Figure 6.7 Multi-channel data set filtered with a 9 hz, seven pole, zero phase, low-pass Butterworth filter.

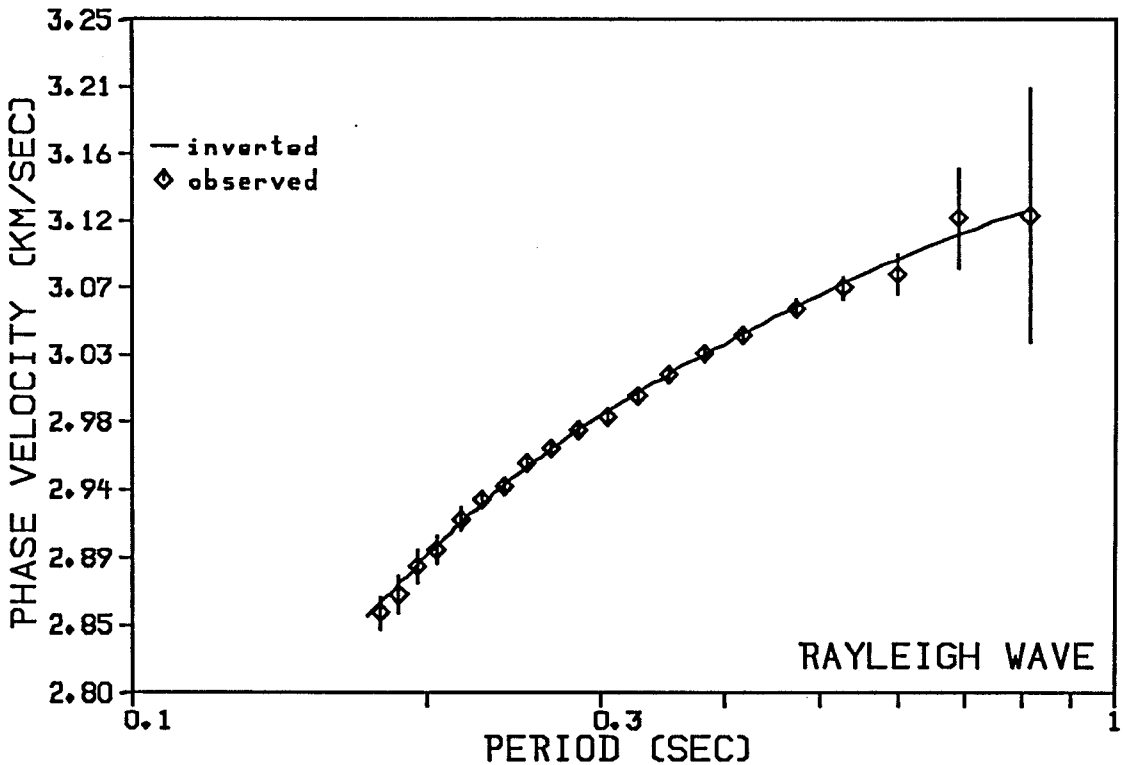

Figure 6.8 Comparision of observed and inverted phase velocities for the multi-channel data set. "Inverted" refers to the phase velocities calculated from the velocity inversion model.

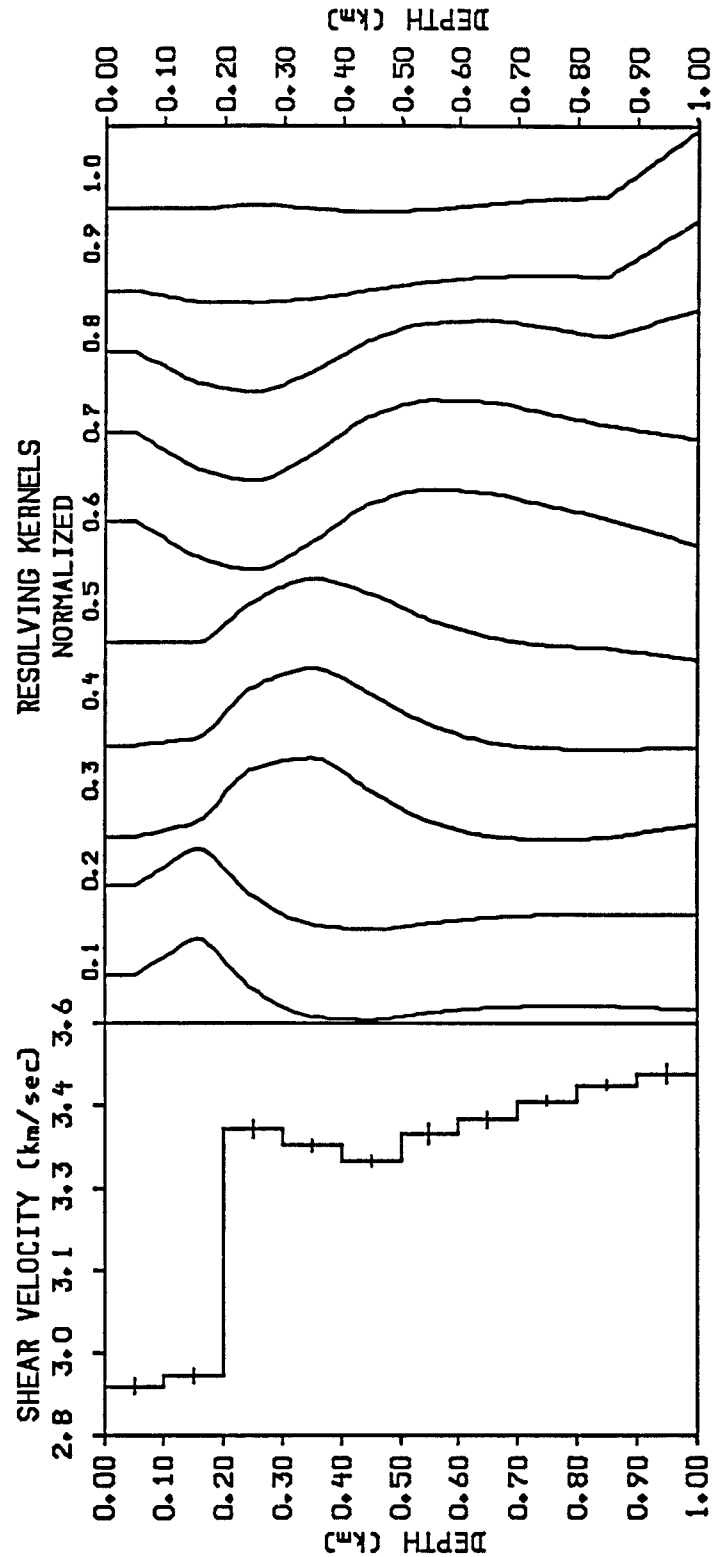

Figure 6.9 Shear velocities and resolving kernels from the phase velocity inversion of the multi-channel data. The discontinuity at 0.2 km was emphasized.

initial guess of the phase-matched filter, the peaks of the cross-correlations between waveforms were used to align the correlations at zero-lag for time-domain windowing (Jenkins and Watts, 1968, p. 399). Subsequent iterations used the phase velocities found by the multi-channel process for an initial guess. The theoretical phase velocities found from inverting a ten-layer shear velocity model (starting from a half-space at 3.5 km/sec), are also shown on Figure 6.8. Figure 6.9 gives the final shear velocity model and corresponding resolving kernels. The inversion was run twice; first to identify the discontinuity at 0.2 km depth, and then with a discontinuity weighting of $\epsilon = 0.1$ at 0.2 km.

Since the shear velocity model is quite simple, it was reparameterized as two layers over a half-space, and the inversion repeated. No constraints were necessary ($\gamma = 0$). A record section of fundamental-mode synthetic seismograms was then generated, using a program developed by Dr. Robert B. Herrmann (program ADDMOD; personal communication, R. B. Herrmann, 1985). Values of Q_β for the three layer model were found by trial and error using ADDMOD. The synthetic seismograms were significantly changed only by Q_β in the top layer; it was found to have a best fit value of 30. The explosion source was modeled as a Dirac impulse in displacement, corresponding to a source expansion followed by immediate collapse. It was found that attempting to model the source as a step function in displacement resulted in synthetic seismograms with long period energy far in excess of the actual data; no values of Q_β would compensate for this.

The synthetic seismogram record section is shown in Figure 6.10. It is so close to the actual record section (Figure 6.8), that an overlay was

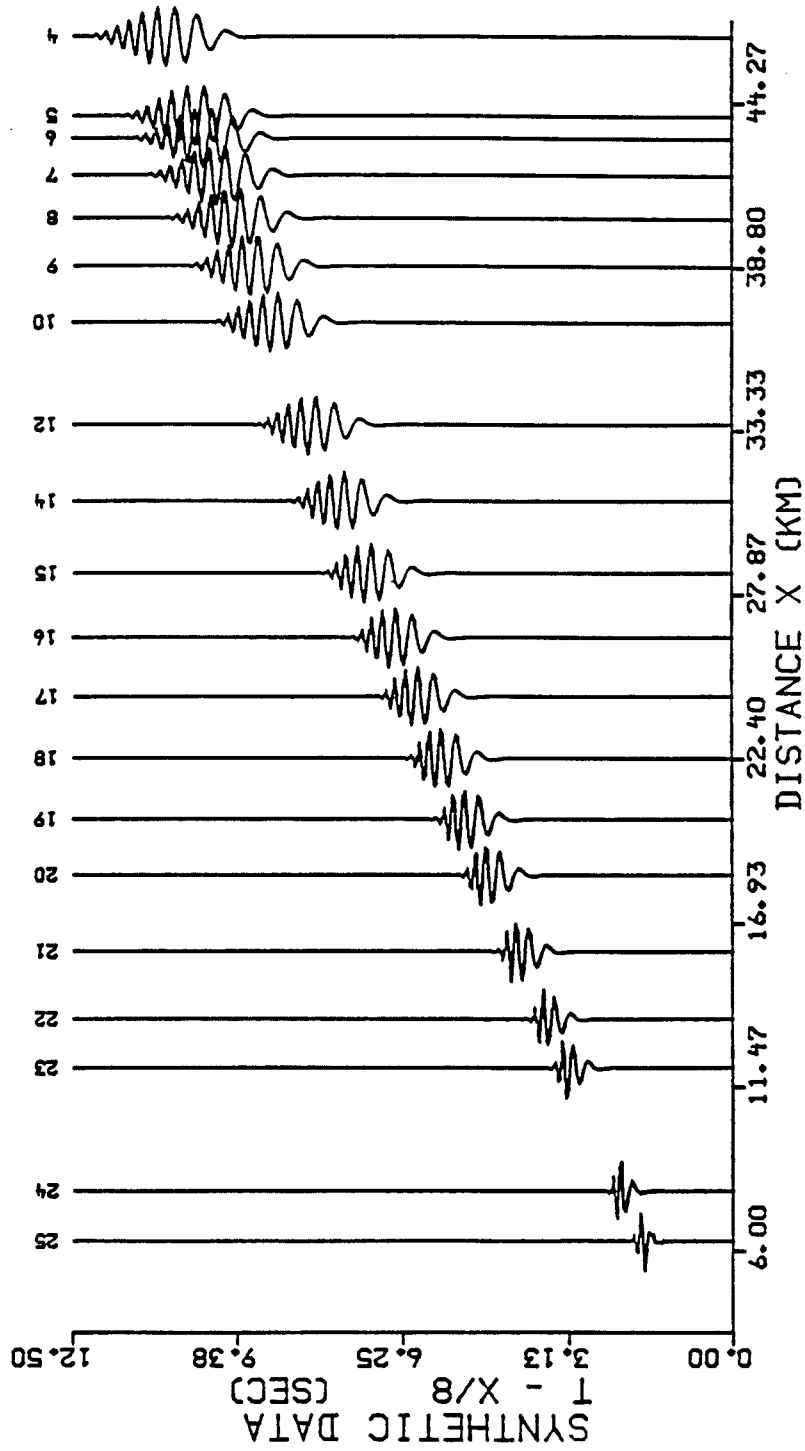

Figure 6.10 Fundamental mode synthetic seismograms produced from phase velocity inversion of the multi-channel data.

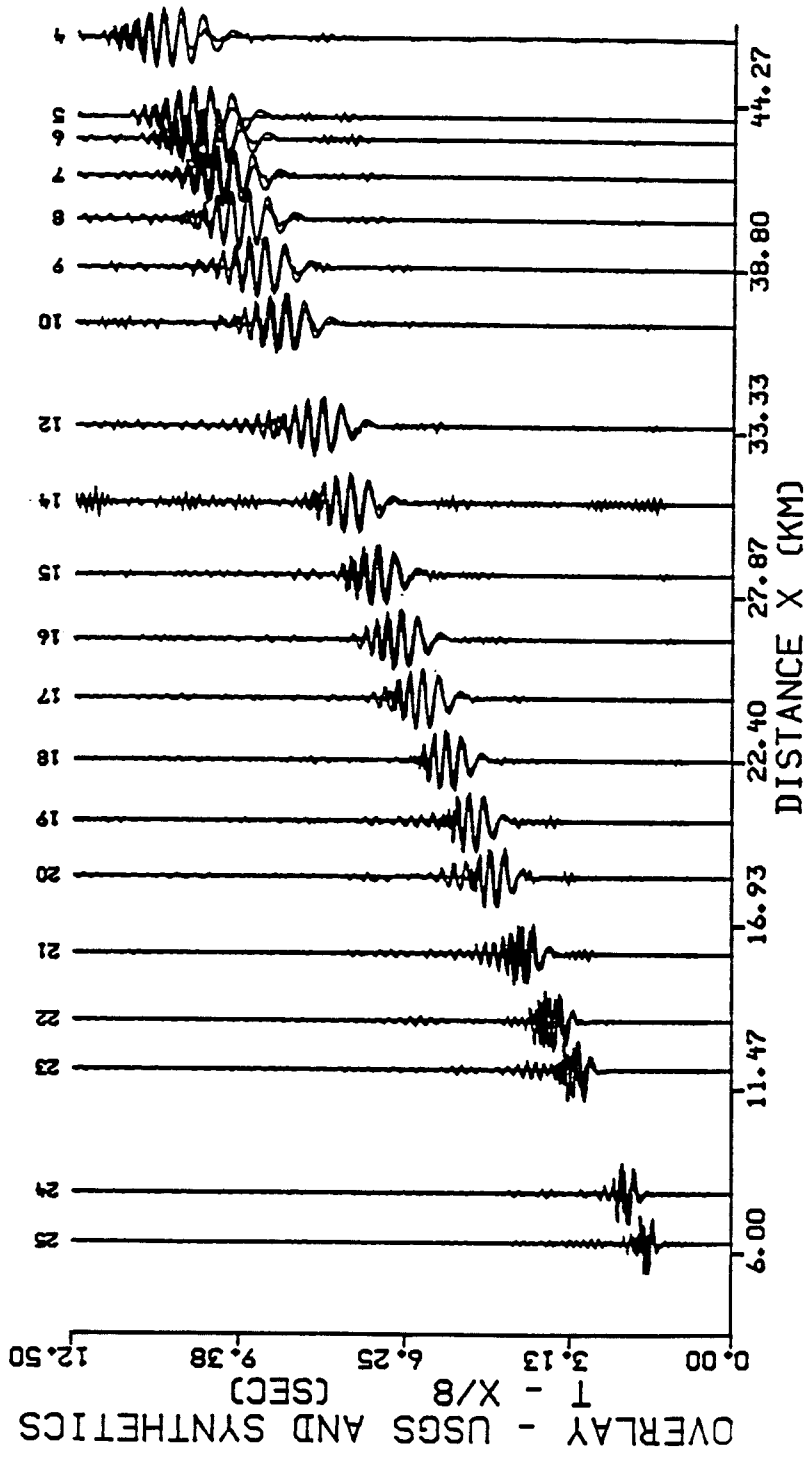

Figure 6.11 Overlay of the fundamental mode synthetic seismograms and the filtered multi-channel data set.

made of the two sections in Figure 6.11. This clearly demonstrates the effectiveness of the multi-channel filter.

6.4 Multi-channel noise study

To more rigorously test the multi-channel phase-matched filter, a noise study was made on the synthetic seismograms developed in the previous section. This was done for two reasons: to test the limit of the multi-channel filter's ability to distinguish signal from noise, and to insure that phase velocity standard deviations calculated by the filter would map over into correct standard deviations found for surface wave inversion models.

Uncorrelated (white) Gaussian noise was added to each trace in the record section. The noise level was calculated in terms of the peak amplitude of the traces. For instance, a 10% noise level would mean that one standard deviation of the Gaussian noise corresponds to 10% of the peak amplitude of the traces. Four increasing levels of noise were used: 10%, 20%, 50%, and 100%. The multi-channel filter was applied to each of these levels for phase velocities and corresponding standard deviations, and then surface wave inversion was performed, to see how much the calculated models would differ from the theoretical one used to generate the original synthetics.

Figures 6.12 through 6.15 show the synthetics with the noise levels added. Notice that at the 100% level, the signal cannot be distinguished from the background noise. Figures 6.16 through 6.19 show the multi-channel filter results. The symbols with the error bars indicate the phase velocities found by the multi-channel filter. The solid line is the theoreti-

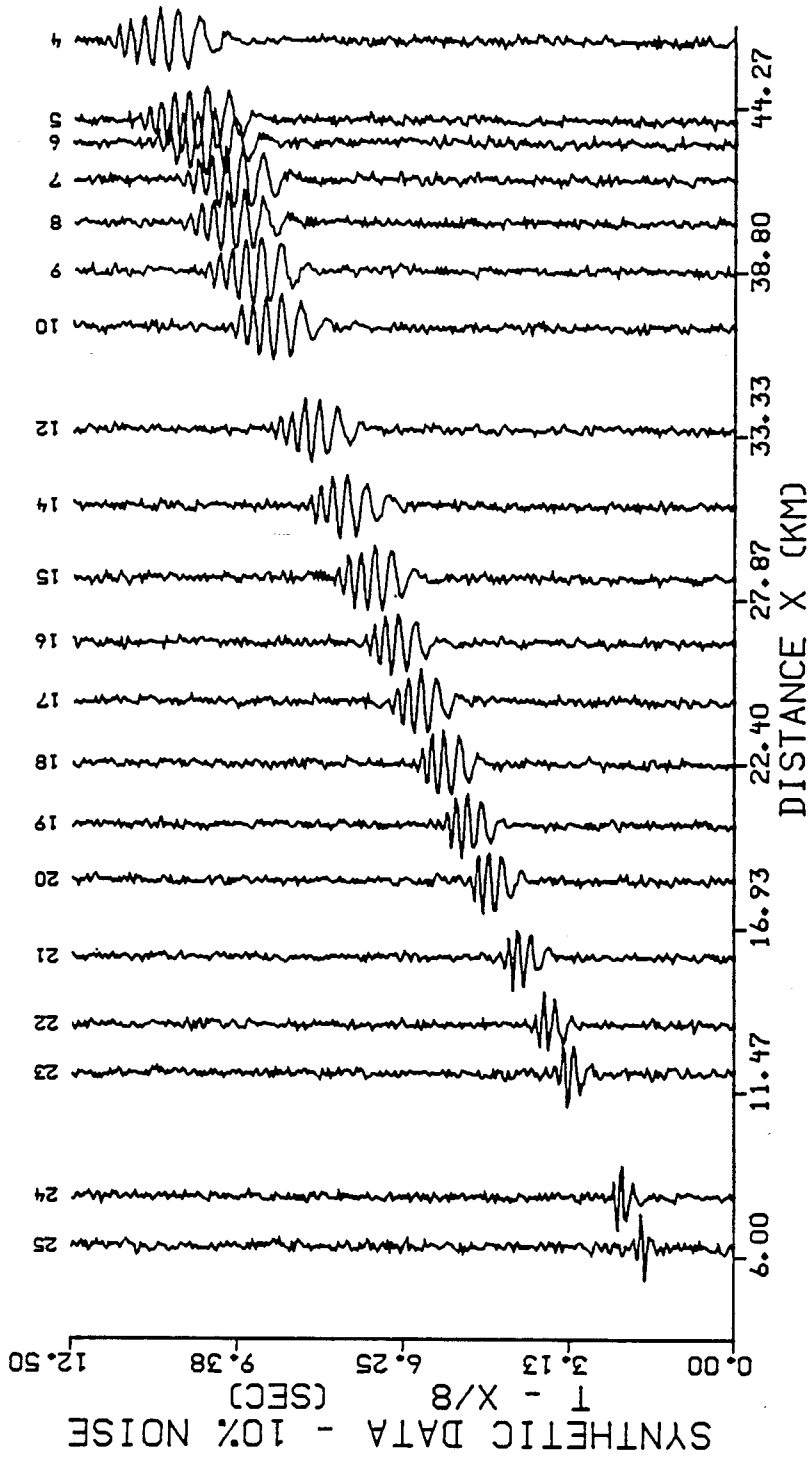

Figure 6.12 Fundamental mode synthetic seismograms with Gaussian noise equaling 10% of the peak amplitude.

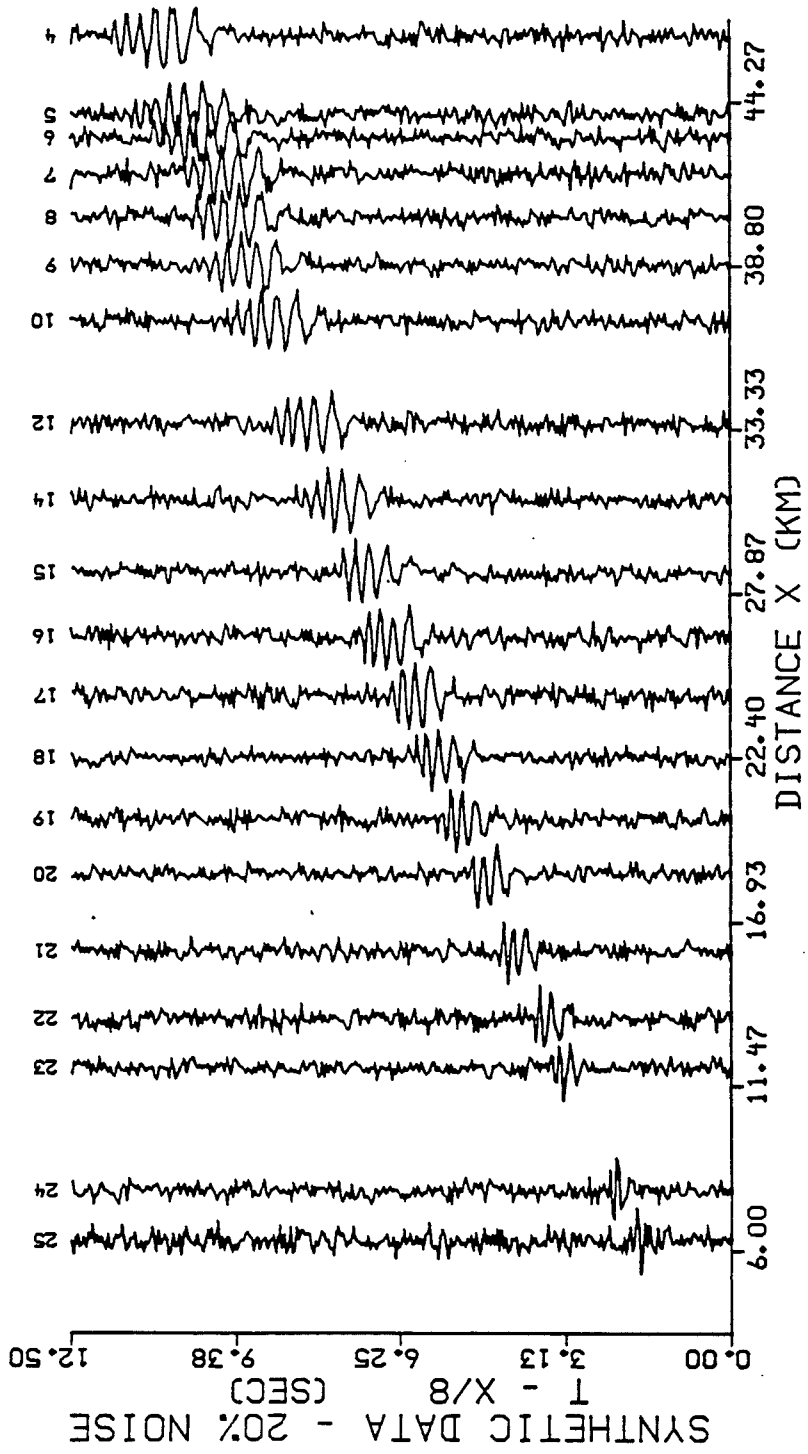

Figure 6.13 Fundamental mode synthetic seismograms with Gaussian noise equaling 20% of the peak amplitude.

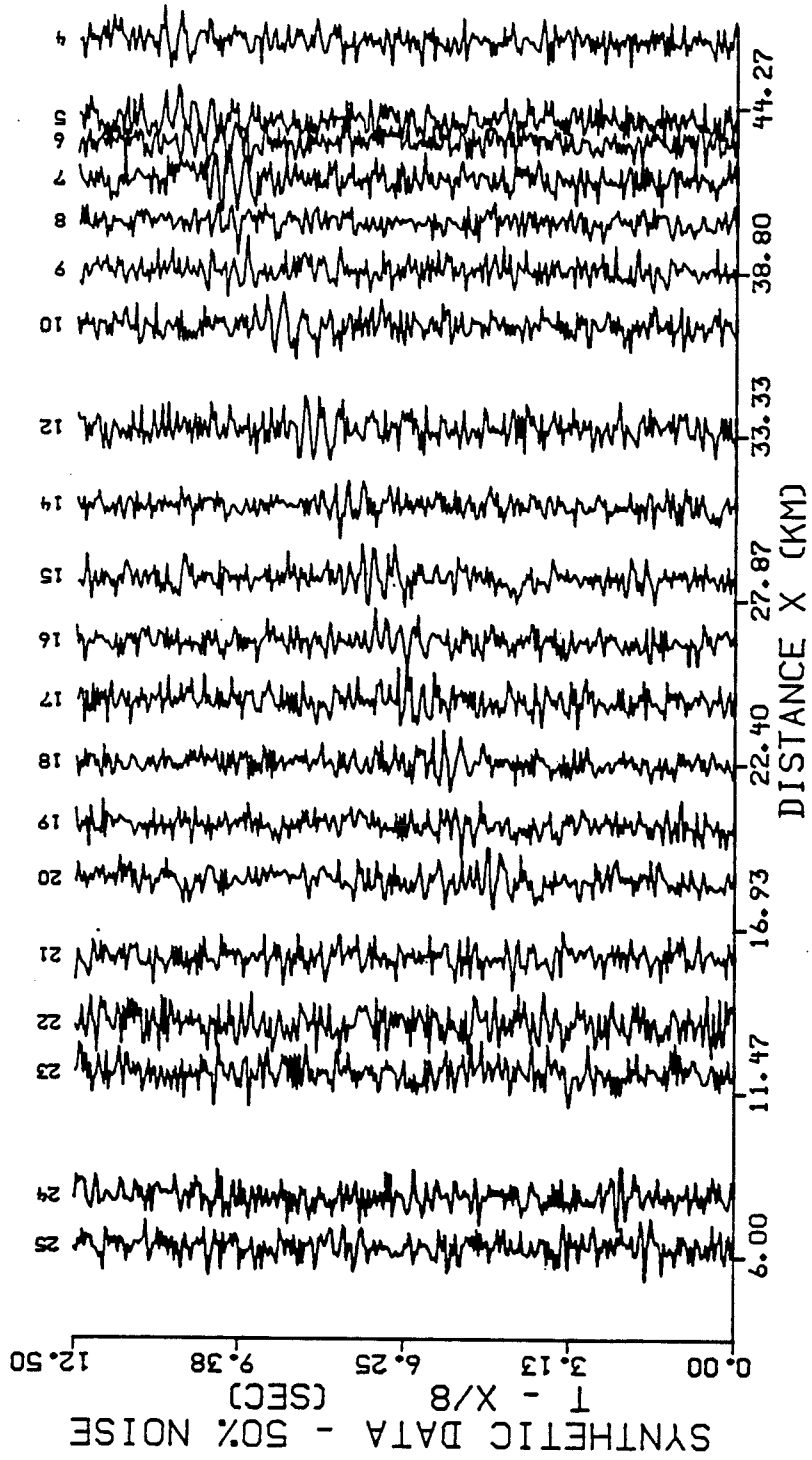

Figure 6.14 Fundamental mode synthetic seismograms with Gaussian noise equaling 50% of the peak amplitude.

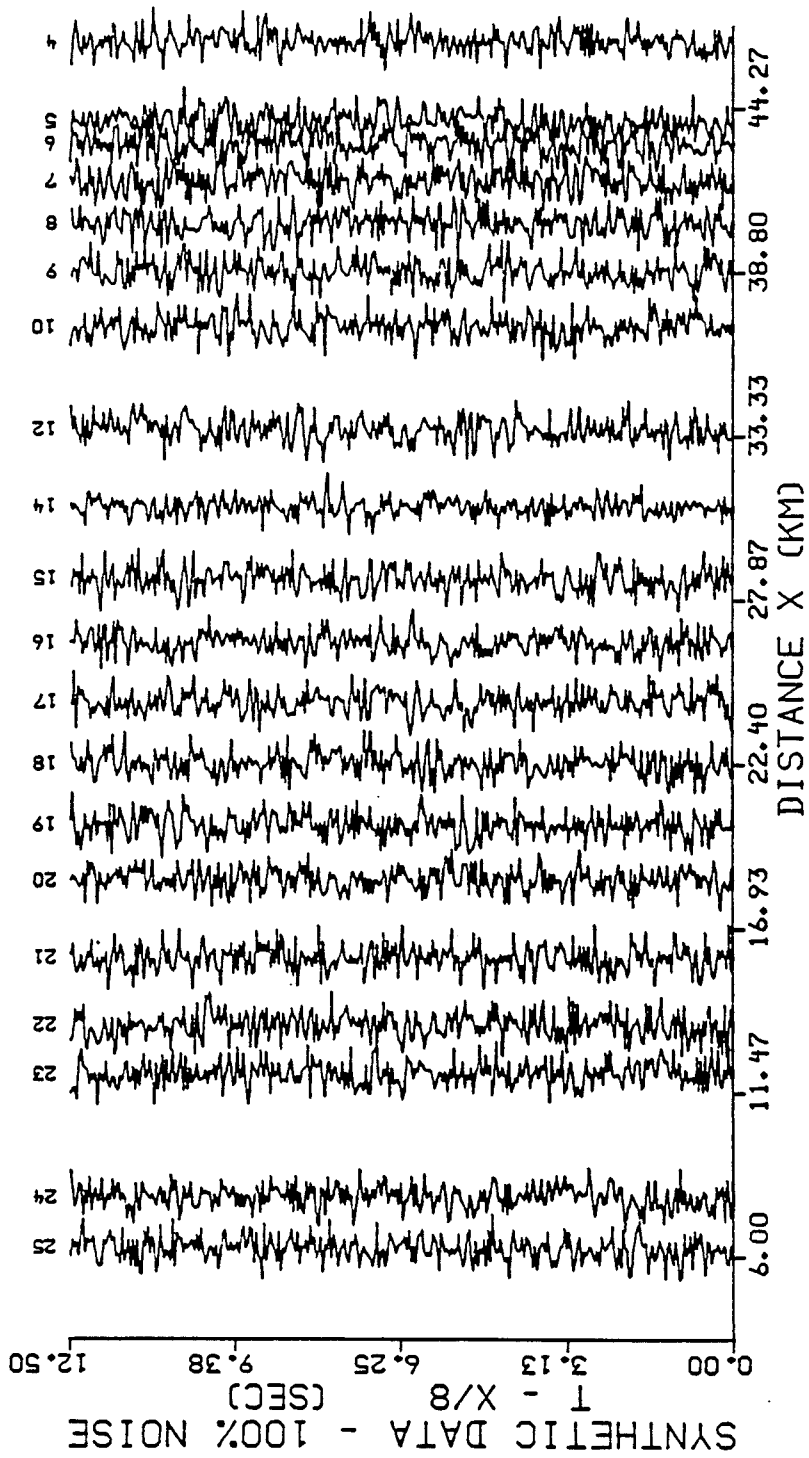

Figure 6.15 Fundamental mode synthetic seismograms with Gaussian noise equaling 100% of the peak amplitude.

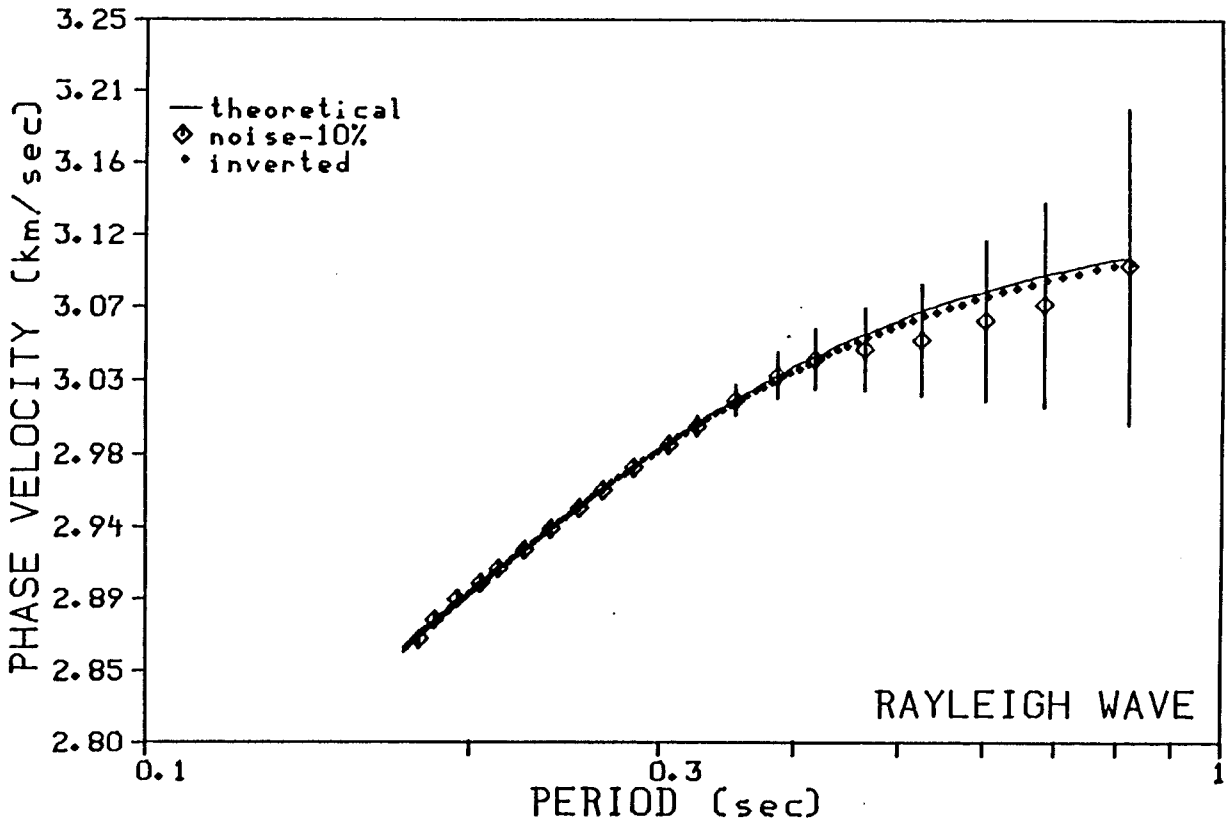

Figure 6.16 Phase velocity dispersion curves for theoretical, 10% noise level, and inverted data set. Error bars correspond to one standard deviation of the noisy data set.

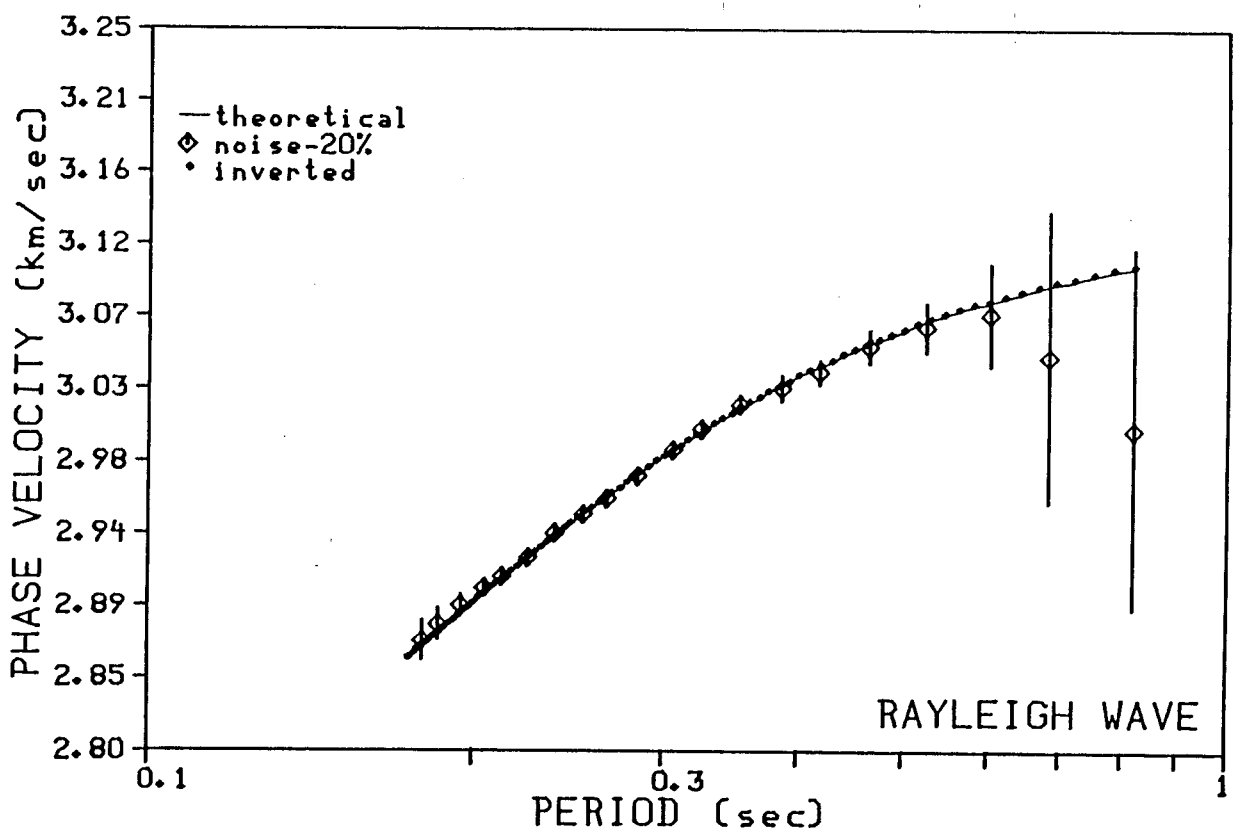

Figure 6.17 Phase velocity dispersion curves for theoretical, 20% noise level, and inverted data set. Error bars correspond to one standard deviation of the noisy data set.

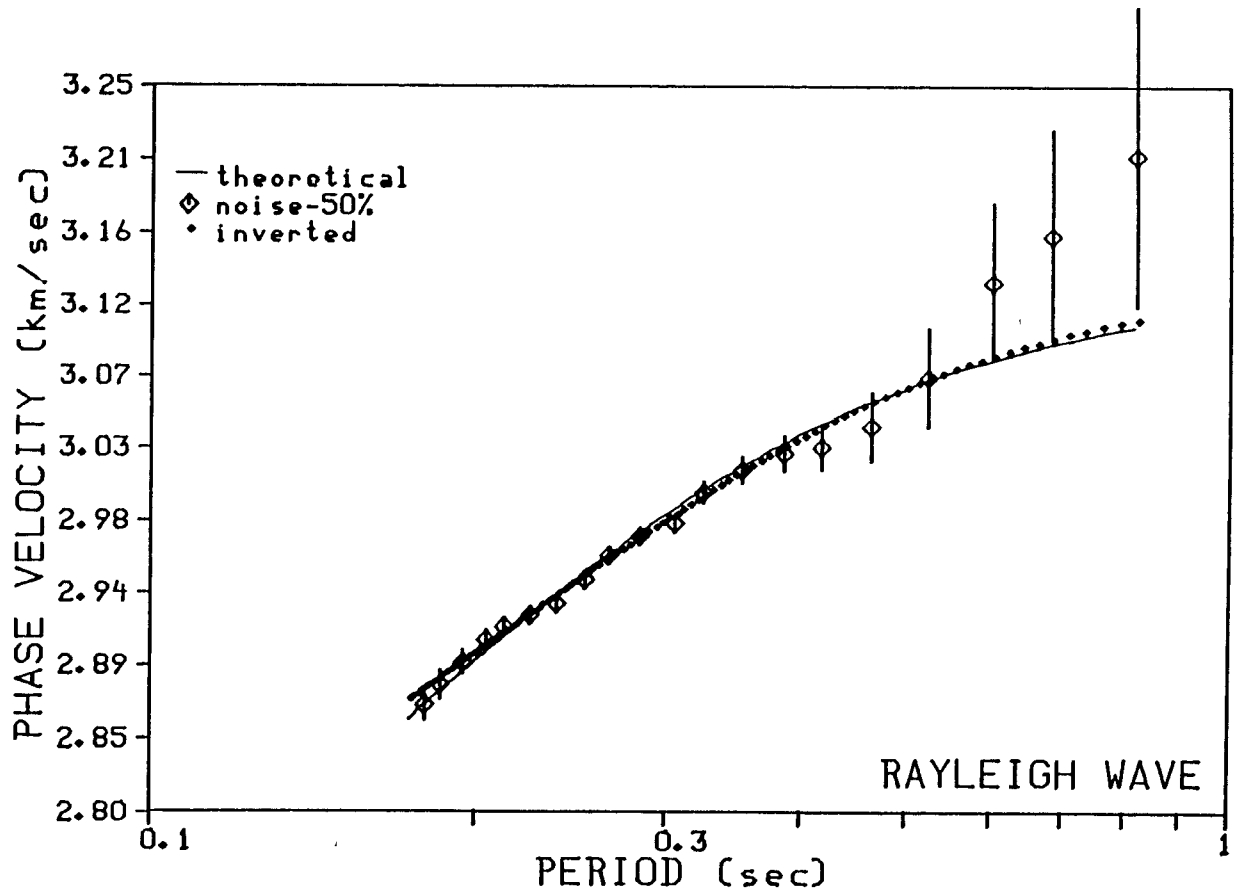

Figure 6.18 Phase velocity dispersion curves for theoretical, 50% noise level, and inverted data set. Error bars correspond to one standard deviation of the noisy data set.

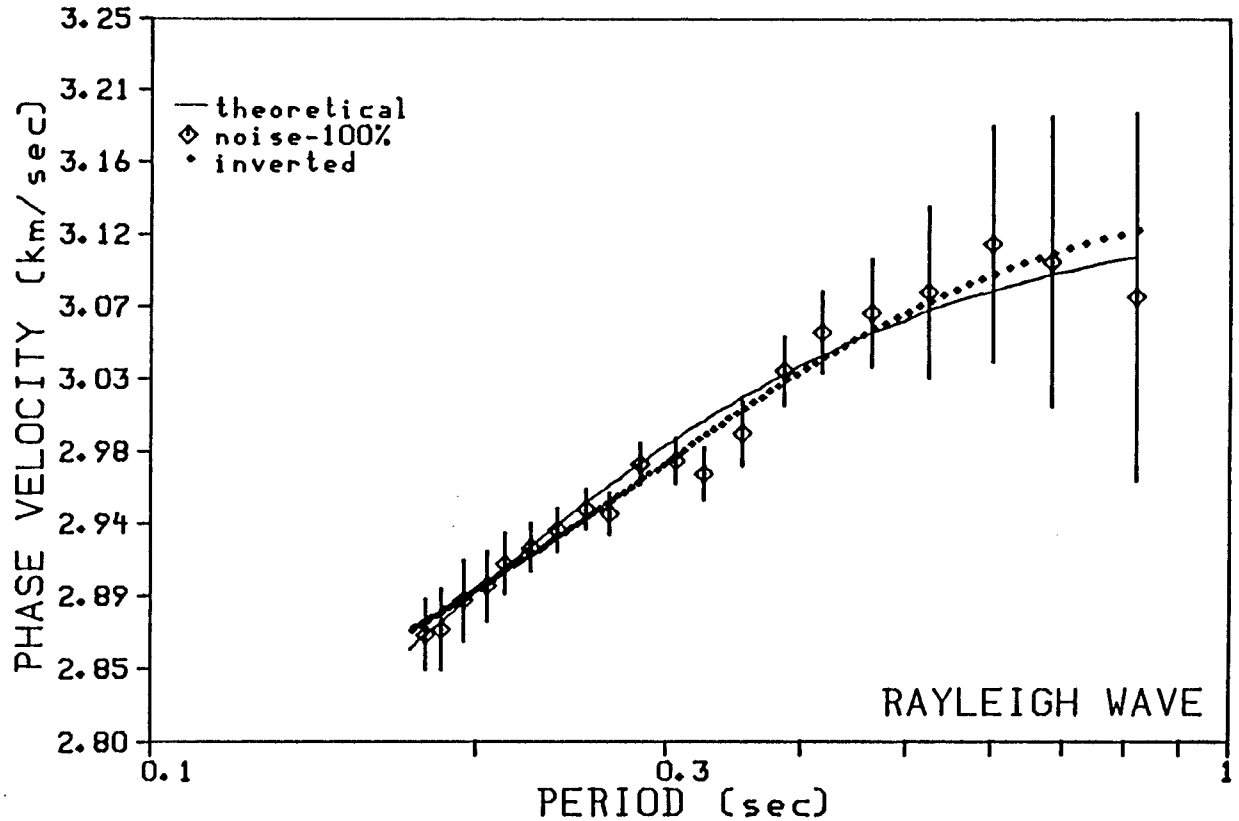

Figure 6.19 Phase velocity dispersion curves for theoretical, 100% noise level, and inverted data set. Error bars correspond to one standard deviation of the noisy data set.

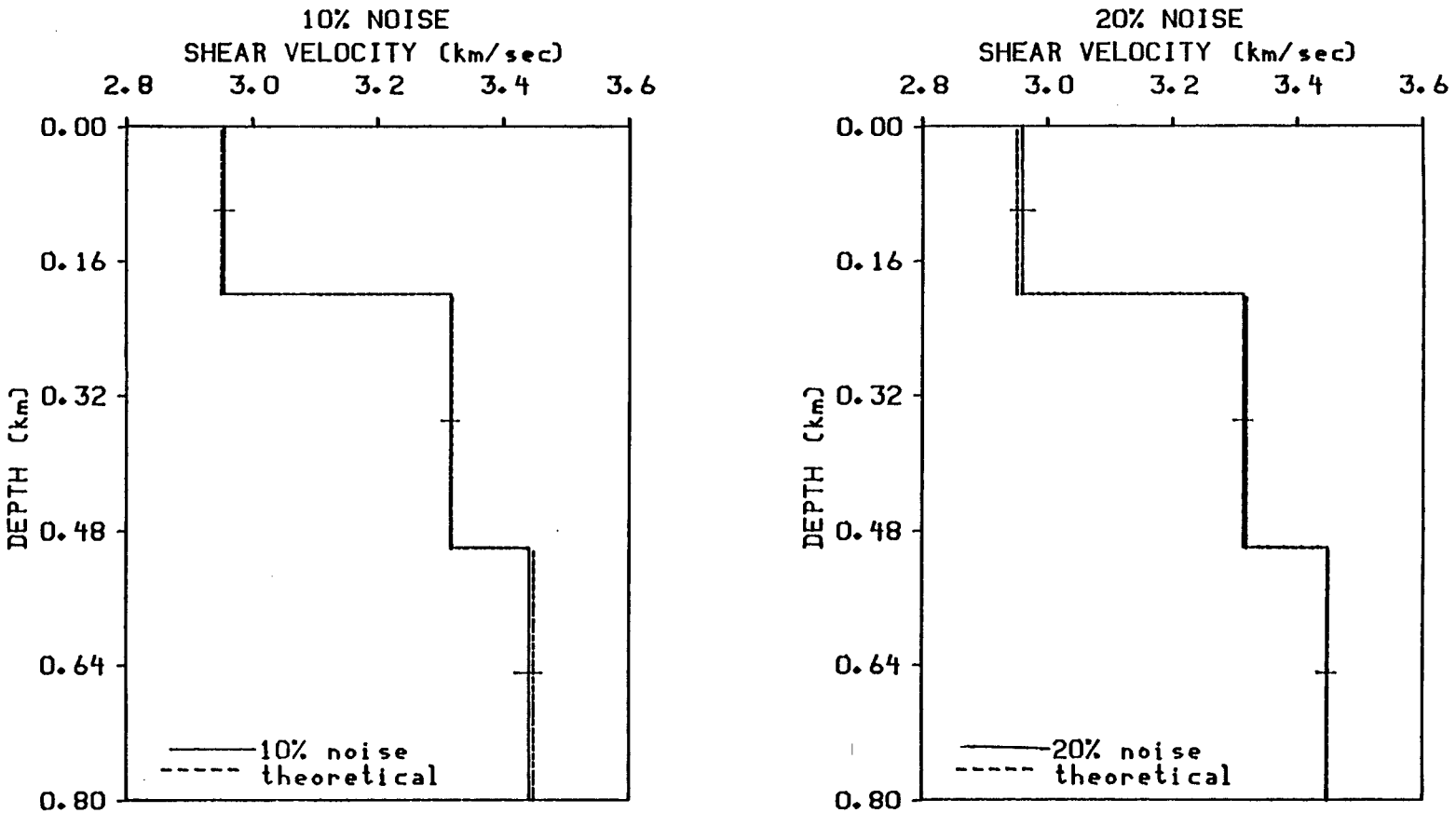

Figure 6.20 Theoretical and inverted models for 10 and 20% noise levels. Error bars correspond to one standard deviation of the inverted data.

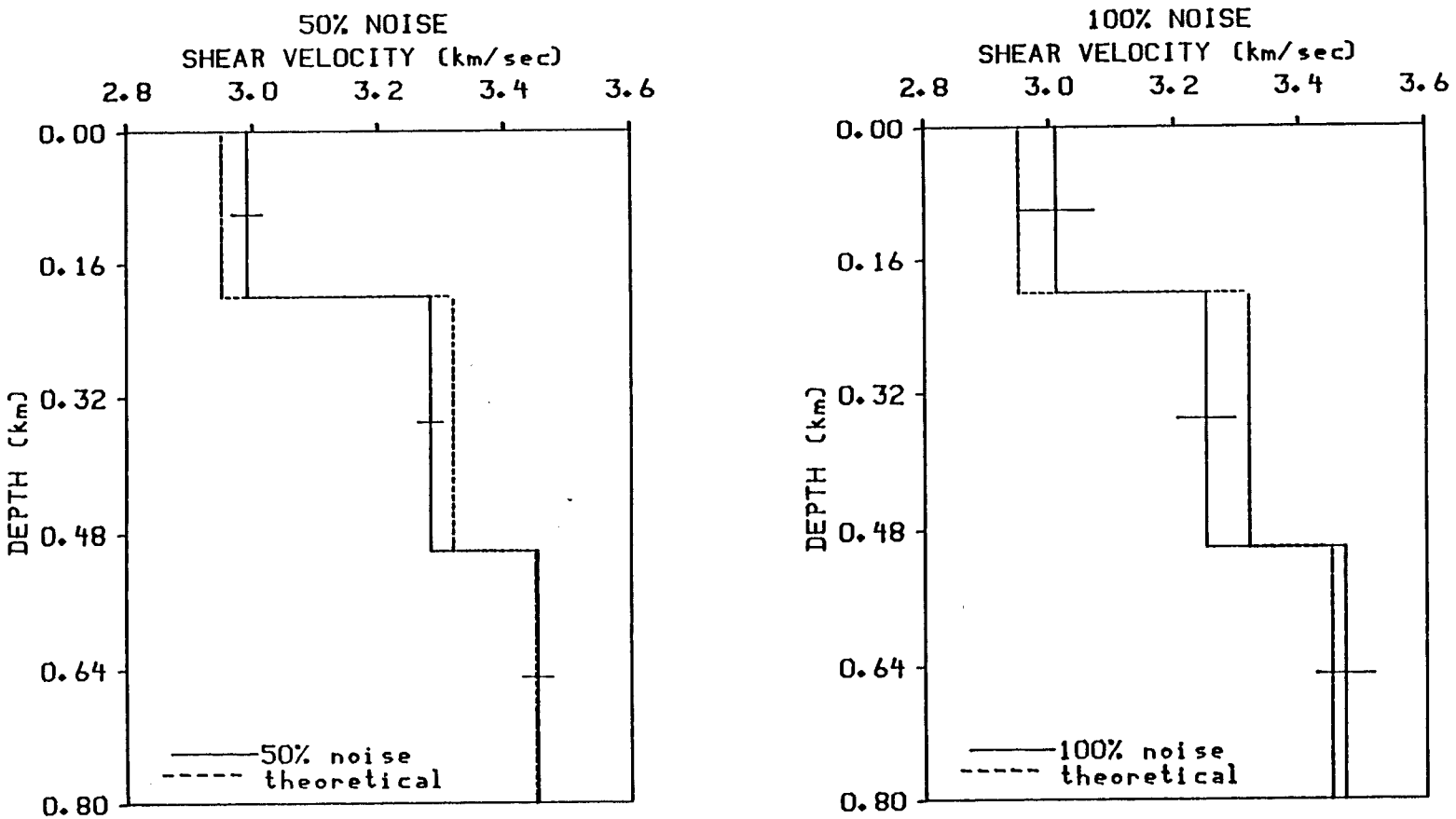

Figure 6.21 Theoretical and inverted models for 50 and 100% noise levels. Error bars correspond to one standard deviation of the inverted data.

cal phase velocity for the three-layer model used to generate the synthetics. The small, closely spaced symbols indicate the phase velocities found by inverting the filter results.

On all the traces, phase velocities found by the multi-channel filter correlate well with the theoretical values up to 0.6 seconds. Beyond that the signal deteriorates, but the filter compensates correctly with large standard deviations. The most unexpected result is at the 100% noise level. Even though the traces are indistinguishable from the noise, the multi-channel filter picks phase velocities which closely follow the theoretical trend. This is surprising, since the initial phase velocity estimates for the phase-matched filter are simply constant velocities picked from the lags of the cross-correlation functions.

It should be noted that the cosine stack (4.2.39), corresponding to the Fejer kernel (4.2.42), was used for the above analysis. A simpler stack corresponding to the Fourier kernel (3.2.11) was also tested, but the results failed to converge to theoretical phase velocities at high noise levels.

Figures 6.20 and 6.21 are the inversion results for the noise study. Since the model was parameterized as two layers over a half space, no damping was necessary in the inversions ($\gamma = 0$). The important result here is that actual deviations of the inverted models correspond closely to predicted standard deviations of the model.

As a final test, ten of the traces were completely replaced with Gaussian noise, as shown in Figure 6.22. This was done to simulate "dead" channels, or severe site effects. The locations of the channels were selected randomly from the data set. Figure 6.23 shows the results of the multi-

channel phase-matched filter. The close correlation of the predicted phase velocities to the theoretical values, and the low standard deviations, indicate that the filter effectively ignored the noisy channels. This indicates robust filter performance under severe field conditions.

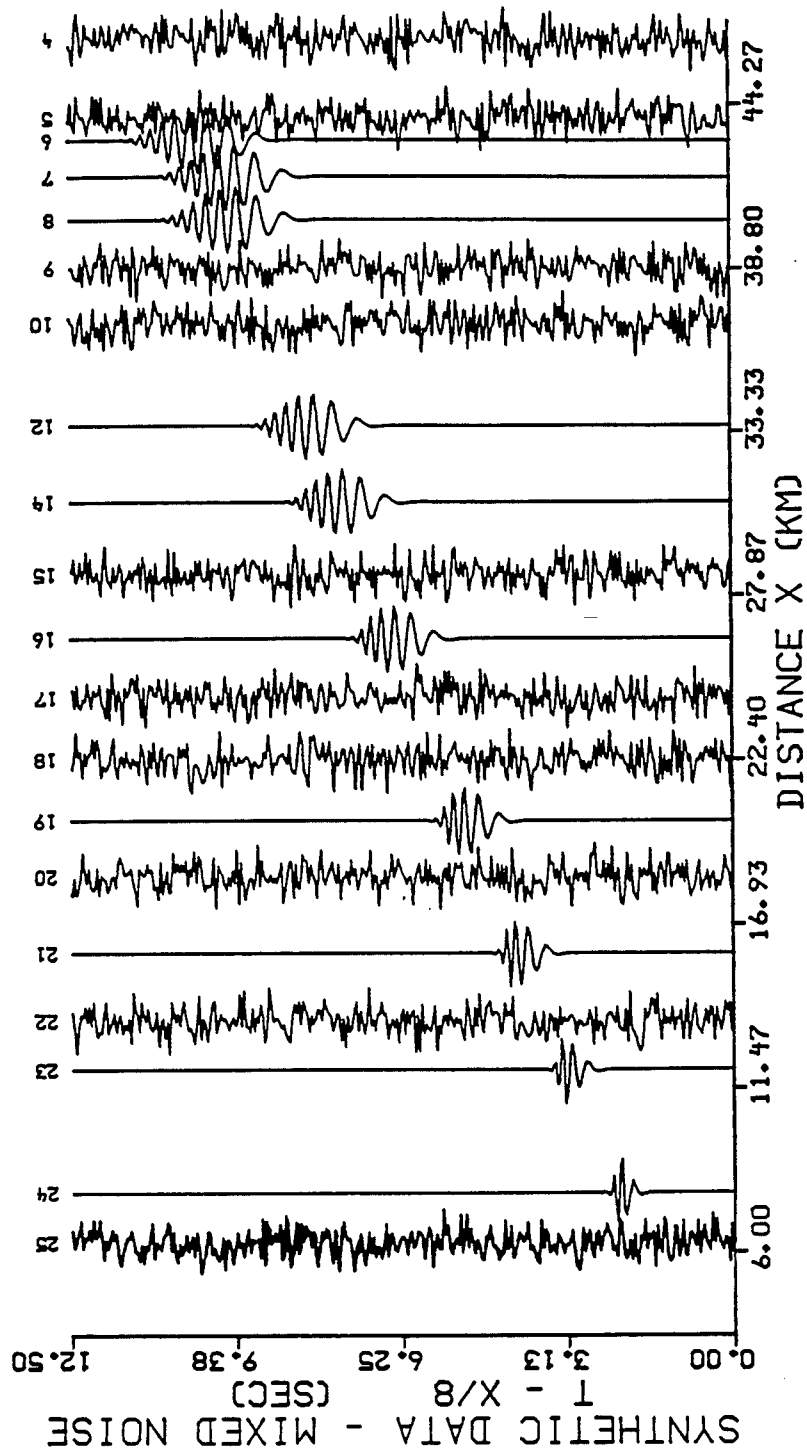

Figure 6.22 Fundamental mode synthetic seismograms with Gaussian noise substituted for randomly selected channels.

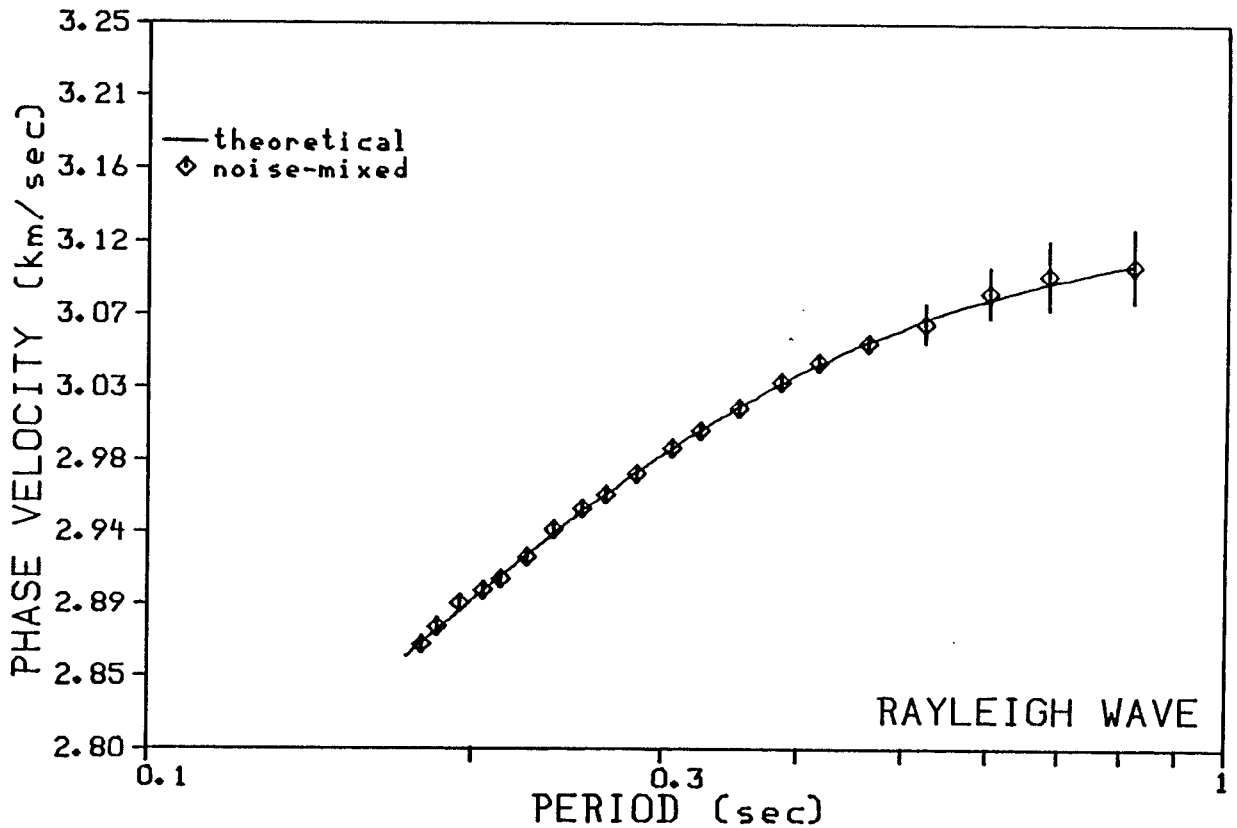

Figure 6.23 Theoretical phase velocities compared to the results of the mixed noise data set. Error bars correspond to one standard deviation of the noisy data set.

CHAPTER 7

SUMMARY AND CONCLUSIONS

The central purpose of this study was to develop a multi-channel process for measuring phase velocities under adverse conditions, and then to invert these data for earth structure where little *a priori* information exists. The analysis was successful, even under circumstances where signals were barely discernible in noise.

The work was initially conceived as a method of stripping off the shallow crustal structure (weathered layer) in exploration reflection seismology. Typically, many geophones are spread out at even intervals from a source, and the resultant record sections often display predominant fundamental and higher-mode Rayleigh waves (ground roll). If the dispersive effects of the ground roll can be recovered, they can be inverted for shallow structure.

For use in reflection seismology, an "average" shallow structure across the dimensions of the array is insufficient. The method must be able to account for lateral inhomogeneities. In chapter 3, this was done theoretically by windowing wavenumber kernels with Gaussian windows and transforming to the spatial domain. However, this method is computationally intensive, and requires much interaction from the analyst to properly construct the wavenumber windows on the p-ω stack. Another alternative is to find the homogeneous structure across overlapping segments of the array, and then to combine the results for a laterally changing structure. This is conceptually simpler than the method presented in

chapter 3, and is easier to automate, but it has the major drawback of making the dispersive analysis highly sensitive to ambient noise levels. This is intuitively obvious, since as the array segments decrease, the amount of observable dispersion across the dimensions of the array segments also decrease. In addition, fewer instruments are available for measuring dispersion in each segment.

To compensate for this, the multi-channel phase-velocity filter was developed, which uses coherency as a method of distinguishing signal from noise, followed by regression analysis on measured phases at each station. Much of the effort in design was concentrated on statistical weighting, phase unwrapping, and wavenumber stacking methods which are robust at high noise levels. The results, which are demonstrated in chapter 6, indicate that this is a viable method for use in exploration seismology.

Another potential use of the multi-channel filter is in acoustical well logging, where Stonely waves propagate along the cylindrical boundaries of boreholes. The variational equations for relating Stonely waves to structural parameters were not developed here, but follow a similar analysis to that of chapter 2.

Two areas of further research in multi-channel filtering are group velocity analysis and Q determination. At the end of chapter 5, equation (5.2.25) indicates that unless phase velocity correlation is almost one, group velocity variances will be unacceptably high, making them unsuitable for inversion. However, under quiet conditions the group velocity could significantly improve the resolution in surface wave inversion, so an actual measure of the group velocity variance could be useful. This may be accomplished by taking advantage of the dense spectral spacing in

discrete Fourier transforms, used in the multi-channel phase-matched filter. If the statistics across the spectrum are assumed to be *locally* stationary, a simple average about a given frequency could be used to estimate the correlation coefficient defined by (5.2.24):

$$\rho_i(2h) \approx \frac{\sum_{i=l-r}^{l+r} (c_{i-1} - \mu_{i-1})(c_{i+1} - \mu_{i+1})}{\sum_{i=l-r}^{l+r} (c_i - \mu_i)^2} ,$$

where l is the discrete frequency, $2r$ is the range of frequencies in which stationarity is assumed, c_i is the measured phase velocity, and μ_i is the local phase velocity mean value. The latter can be found from the theoretical phase velocities determined in surface wave inversion. The correlation estimate is substituted into (5.2.25) to determine group velocity variances, and a combined phase/group velocity inversion can then be accomplished, using (5.2.19) as a numerical estimate of group velocity perturbations.

It may be possible to develop a method similar to the multi-channel phase-matched filter for attenuation measurements, or Q values. Jenkins and Watts (1968, p. 435) define an estimate for 95% confidence intervals in bivariate spectral amplitudes, using the coherency spectrum, similar to the phase spectrum. By using logarithms of the amplitudes, a set of regression equations similar to that in chapter 4 can be constructed for determining attenuation values, which can then be inverted for intrinsic Q structure. However, as pointed out in chapter 3, more care must be taken in constructing the phase-matched filter for amplitude measurements, since the amplitude spectrum is more susceptible to bias than the phase

spectrum (see pp. 42-43).

APPENDIX I

SELF-ADJOINT LOVE AND RAYLEIGH WAVE EQUATIONS

A-I-1 General theory

Let **D** define a linear differential operator, which has a vector solution **y** to the equation

$$\mathbf{D}\,\mathbf{y} = 0 \ . \tag{A-I-1}$$

D is defined as self-adjoint if the following condition holds (Butkov, 1968, p. 341):

$$\int_z \mathbf{f}^{\mathrm{T}}\,(\mathbf{Dg})\,dz = \int_z \mathbf{g}^{\mathrm{T}}\,(\mathbf{Df})\,dz \ . \tag{A-I-2}$$

The vector functions **f** and **g** do not have to be solutions to (A-I-1), but they are required to satisy appropriate boundary conditions on (A-I-1). Another way of expressing the self-adjoint condition is to define the operator

$$\Psi(\mathbf{f},\mathbf{g}) = \int_z \mathbf{f}^{\mathrm{T}}(\mathbf{Dg})\,dz \ . \tag{A-I-3}$$

From (A-I-2), **D** is self-adjoint if

$$\Psi(\mathbf{f},\mathbf{g}) = \Psi(\mathbf{g},\mathbf{f}) \ .$$

A-I-2 Love waves

From chapter 2, the Love wave equation is

$$\frac{d}{dz}\left(\mu\frac{dl_1}{dz}\right) = (k^2\mu - \omega^2\rho)l_1 \ .$$

Define the operator (A-I-3) as

$$\Psi(f,g) = \int_0^\infty \left[f \frac{d}{dz}\left(\mu \frac{dg}{dz}\right) - (k^2\mu - \omega^2\rho)\, f\, g \right] dz \ . \qquad \text{(A-I-4)}$$

The displacement and stress boundary conditions for Love waves are

$$f = 0 \ , \qquad \frac{dg}{dz} = 0$$

at $z = \infty$ and $z = 0$, respectively. Integrate the left side of A-I-4 under the integral by parts and apply the boundary conditions for

$$\Psi(f,g) = -\int_0^\infty \left[\mu \frac{df}{dz}\frac{dg}{dz} + (k^2\mu - \omega^2\rho)\, f\, g \right] dz \ . \qquad \text{(A-I-5)}$$

From (A-I-5) it is obvious by inspection that

$$\Psi(f,g) = \Psi(g,f) \ .$$

A-I-3 Rayleigh waves

From chapter 2, the equation of motion for Rayleigh waves is

$$\frac{d}{dz}\left(\mathbf{A}\frac{d\mathbf{r}}{dz} + k\mathbf{B}\mathbf{r}\right) = k\mathbf{B}^{\mathbf{T}}\frac{d\mathbf{r}}{dz} + k^2\mathbf{C}\mathbf{r} - \omega^2\rho\mathbf{r} \ ,$$

where

$$\mathbf{A} = \begin{bmatrix} \mu & 0 \\ 0 & (\lambda+2\mu) \end{bmatrix} , \ \mathbf{B} = \begin{bmatrix} 0 & \mu \\ -\lambda & 0 \end{bmatrix} , \ \mathbf{C} = \begin{bmatrix} (\lambda+2\mu) & 0 \\ 0 & \mu \end{bmatrix} , \ \mathbf{r} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} .$$

Define the operator (A-I-3) as

$\Psi(\mathbf{f},\mathbf{g}) =$

$$\int_0^\infty \left[\mathbf{f}^{\mathbf{T}}\frac{d}{dz}\left(\mathbf{A}\frac{d\mathbf{g}}{dz} + k\mathbf{B}\mathbf{g}\right) - k\ \mathbf{f}^{\mathbf{T}}\mathbf{B}^{\mathbf{T}}\frac{d\mathbf{g}}{dz} - k^2\,\mathbf{f}^{\mathbf{T}}\mathbf{C}\mathbf{g} + \omega^2\rho\,\mathbf{f}^{\mathbf{T}}\mathbf{g} \right] dz \ . \qquad \text{(A-I-6)}$$

The displacement and stress boundary conditions for Rayleigh waves are

$$\mathbf{f} = 0 \ , \qquad \mathbf{A}\frac{d\mathbf{g}}{dz} + k\mathbf{B}\mathbf{g} = 0$$

at $z = \infty$ and $z = 0$, respectively. The stress term is equivalent to the radial and vertical stress eigenfunctions, and can be identified by substitution into the displacement-stress matrix (2.2.1). Integrate the left side of (A-I-6) under the integral by parts and apply the boundary conditions for

$\Psi(\mathbf{f},\mathbf{g})$ =

$$-\int_0^\infty \left[\frac{d\mathbf{f}^{\mathrm{T}}}{dz}\mathbf{A}\frac{d\mathbf{g}}{dz} + k\frac{d\mathbf{f}^{\mathrm{T}}}{dz}\mathbf{B}\mathbf{g} + k\mathbf{f}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\frac{d\mathbf{g}}{dz} + k^2\mathbf{f}^{\mathrm{T}}\mathbf{C}\mathbf{g} - \omega^2\rho\mathbf{f}^{\mathrm{T}}\mathbf{g} \right] dz \ . \qquad \text{(A-I-7)}$$

Multiplying out the matrix terms will demonstrate that (A-I-7) is a scalar integral. Therefore, since the transpose of a scalar is equal to itself,

$$\Psi^T(\mathbf{f},\mathbf{g}) = \Psi(\mathbf{f},\mathbf{g}) \ . \qquad \text{(A-I-8)}$$

The transpose of (A-I-7) is

$\Psi^T(\mathbf{f},\mathbf{g})$ =

$$-\int_0^\infty \left[\frac{d\mathbf{g}^{\mathrm{T}}}{dz}\mathbf{A}\frac{d\mathbf{f}}{dz} + k\frac{d\mathbf{g}^{\mathrm{T}}}{dz}\mathbf{B}\mathbf{f} + k\mathbf{g}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\frac{d\mathbf{f}}{dz} + k^2\mathbf{g}^{\mathrm{T}}\mathbf{C}\mathbf{f} - \omega^2\rho\mathbf{g}^{\mathrm{T}}\mathbf{f} \right] dz \ . \qquad \text{(A-I-9)}$$

The two terms with $\mathbf{B}$ and $\mathbf{B}^{\mathrm{T}}$ have switched position for clarity. Comparing (A-I-9) with (A-I-7) shows that

$$\Psi^T(\mathbf{f},\mathbf{g}) = \Psi(\mathbf{g},\mathbf{f}) \ . \qquad \text{(A-I-10)}$$

Equating (A-I-10) with (A-I-8) gives the final result:

$$\Psi(\mathbf{f},\mathbf{g}) = \Psi(\mathbf{g},\mathbf{f}) \ ,$$

which proves that the Rayleigh wave equation is self-adjoint.

APPENDIX II

FEJER KERNEL CALCULATIONS

The cosine stack defined in chapter 5 for multi-channel processing is

$$F_l(k) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \cos[(k - \hat{k}_l)x_{ij}] \ ,$$

where l refers to the frequency, and x_{ij} is the interstation spacing between channels i and l, and N is the total number of stations. For equally spaced stations, this can be written as

$$F_l(\mathrm{k}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \cos[\ \mathrm{k}\ (j-i)\delta x] \qquad \text{(A-II-1)}$$

where $\mathrm{k} = k - \hat{k}_l$, and δx is the interstation spacing. This section will show that the double sum is equal to a Fejer kernel (Papoulis, 1977, p. 72), in the specific form

$$F_l(\mathrm{k}) = \frac{\sin^2[\ N\mathrm{k}\ \delta x/2\ ]}{2\ \sin^2[\ \mathrm{k}\ \delta x/2\ ]} - \frac{N}{2} \ . \qquad \text{(A-II-2)}$$

By expanding (A-II-1) term by term, it can be shown that the double sum in (A-II-1) is equivalent to

$$F_l(k) = \sum_{r=1}^{N} \sum_{s=1}^{N-r} \cos(\mathrm{k}\ r\ \delta x) \ . \qquad \text{(A-II-3)}$$

Since the sum over s does not appear inside the cosine term, the double sum can be condensed to the single sum

$$F_l(\mathrm{k}) = \sum_{r=1}^{N-1} (N - r) \cos(\mathrm{k}\ r\ \delta x) \ . \qquad \text{(A-II-4)}$$

Factoring N from the sum gives

$$F_l(\mathrm{k}) = N \sum_{r=1}^{N-1} (1 - \frac{r}{N}) \cos(\mathrm{k}\ r\ \delta x) \ . \qquad \text{(A-II-5)}$$

Papoulis defines the Fejer kernel as

$$\sum_{r=-2N}^{r=2N} \left(1 - \frac{|r|}{2N+1}\right) e^{ik\,r\,\delta x} = \frac{\sin^2[\,(N+1/2)\,k\,\delta x\,]}{(2N+1)\sin^2[\,k\,\delta x/2\,]} \quad . \text{(A-II-6)}$$

By changing the limits of the sum from $2N$ to $N-1$, equation (A-II-6) will have the form

$$\sum_{r=-(N-1)}^{r=(N-1)} \left(1 - \frac{|r|}{N}\right) e^{ik\,r\,\delta x} = \frac{\sin^2[\,N\,k\,\delta x/2\,]}{N\sin^2[\,k\,\delta x/2\,]} \quad . \text{(A-II-7)}$$

The complex exponential in (A-II-7) can be expressed as

$$e^{ik\,r\,\delta x} = \cos(k\,r\,\delta x) + i\,\sin(k\,r\,\delta x) \quad ,$$

and by using the even and odd properties of the cosine and sine functions, (A-II-7) can be written as

$$1 + 2\sum_{r=1}^{N-1} \left(1 - \frac{r}{N}\right) \cos(k\,r\,\delta x) = \frac{\sin^2[\,N\,k\,\delta x/2\,]}{N\sin^2[\,k\,\delta x/2\,]} \quad . \text{(A-II-8)}$$

Rearranging terms gives

$$N\sum_{r=1}^{N-1} \left(1 - \frac{r}{N}\right) \cos(k\,r\,\delta x) = \frac{\sin^2[\,N\,k\,\delta x/2\,]}{2\sin^2[\,k\,\delta x/2\,]} - \frac{N}{2} \quad . \text{(A-II-9)}$$

The sum in (A-II-9) is equal to the left side of (A-II-5), so that by equating the two,

$$F_l(k) = \frac{\sin^2[\,N\,k\,\delta x/2\,]}{2\sin^2[\,k\,\delta x/2\,]} - \frac{N}{2} \quad . \text{(A-II-10)}$$

and (A-II-2) is proved.

BIBLIOGRAPHY

Aki, K., and P. Richards (1980). *Quantitative Seismology: Theory and Methods,* W. H. Freeman and Co.

Arfken, G. (1973). *Mathematical Methods for Physicists,* Academic Press, New York.

Bache, T. C., W. L. Rodi, and D. G. Harkrider (1978a). Crustal structures inferred from Rayleigh wave signatures of NTS explosions, *Bull. Seism. Soc. Am. 68,* 1399-1413.

Backus, G., and F. Gilbert (1967). Numerical applications of a formalism for geophysical inverse problems, *Geophys. J. R. astr. Soc. 13,* 247-276.

Backus, G., and F. Gilbert (1968). The resolving power of gross earth data, *Geophys. J. 16,* 169-205.

Berger, J., D. C. Agnew, R. L. Parker, and W. E. Farrell (1979). Seismic system calibration: parts I and II, *Bull. Seism. Soc. Am. 69,* 271-288.

Birch, F. (1964). Density and composition of mantle and core, *J. Geophys. Res. 69,* 4377-4388.

Bloch, S., and A. L. Hales (1968). New techniques for the determination of surface wave phase velocities, *Bull. Seism. Soc. Am. 58,* 1021-1034.

Brigham, E. O. (1974). *The Fast Fourier Transform,* Prentice-Hall, Inc., New Jersey.

Buland, R., and F. Gilbert (1978). Improved resolution of complex eigenfrequencies in analytically continued seismic spectra, *Geophys. J. R Astr. Soc. 52,* 457-470.

Bullen, K. E., and B. A. Bolt (1985). *An Introduction to the Theory of Seimology,* Cambridge University Press, Cambridge.

Butkov, E. (1968). *Mathematical Physics,* Addison-Wesley Publishing Company, Inc., Massachusetts.

Churchill, R. V., J. W. Brown, and R. F. Verhey (1974). *Complex Variables and Applications,* McGraw Hill Book Company, New York.

Claerbout, Jon F. (1976). *Fundamentals of Geophysical Data Processing with Applications to Petroleum Prospecting,* McGraw-Hill Book Company, New York.

Der, Z. A., and M. Landisman (1972). Theory for errors, resolution, and separation of unknown variables in inverse problems, with application to the mantle and the crust in Southern Africa and Scandinavia, *Geophys. J. R. astr. Soc. 27,* 137-178.

Dziewonski, A., S. Bloch, and M. Landisman (1969). A technique for the analysis of transient seismic signals, *Bull. Seism. Soc. Am. 59,* 427-444.

Dziewonski, A., and A. Hales (1972). Numerical analysis of dispersed seismic waves, *in* B. A. Bolt (editor), *Methods of computational physics, vol. 11,* Academic Press, New York.

Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform, *Proceedings of the IEEE 66,* 51-83.

Haskell, N. S. (1953). The dispersion of surface waves on multilayered media, *Bull. Seis. Soc. Am. 43,* 17-34.

Healy, J. H., W. D. Mooney, H. R. Blank, M. E. Gettings, W. M. Kohler, R. J. Lamson, and L. Leone (1982). Saudi Arabian seismic deep-refraction profile, final project report: *U.S. Geol. Surv. Saudi Arabian Mission open-file rep.,* 2-37.

Herrin, E., and T. Goforth (1977). Phase-matched filters: application to the study of Rayleigh waves, *Bull. Seis. Soc. Am. 67,* 1259-1275.

Herrmann, R. B. (1973). Some aspects of band-pass filtering of surface waves, *Bull. Seism. Soc. Am. 63,* 663-671.

Jackson, D. D. (1972). Interpretation of inaccurate, insufficient and inconsistent data, *Geophys. J. 28,* 97-109.

Jeffreys, H. (1961). Small corrections in theory of surface waves, *Geophys. J. R. Astr. Soc. 6,* 115-117.

Jenkins, Gwilym M., and D. G. Watts (1968). *Spectral Analysis and Its Applications,* Holden Day, San Francisco.

Jordan, T. H., and J. Franklin (1971). Optimal solutions to a linear inverse problem in geophysics, *Proceedings of the National Academy of Sciences 70,* 291-293.

Keilis-Borok, V. I., M. G. Neigauz, and G. V. Shkadinskaya (1965). Application of theory of eigenfunctions to the calculation of surface wave velocities, *Revs. Geophys. 3,* 105-109.

Kennett, B. L. N. (1983). *Seismic Wave Propagation in Stratified Media,* Cambridge University Press, Cambridge.

Lanczos, C. (1961). *Linear Differential Operators,* D. Van Nostrand Co., London.

Landisman, M., Dziewonski, A., and Y. Sato (1969). Recent improvements in the analysis of surface wave observations, *Geophys. J. R. Astr. Soc. 17,* 369-403.

Lawson, C. L., and R. J. Hanson (1974). *Solving Least Squares Problems,* Prentice-Hall, Inc., New Jersey.

Levenberg, L. (1944). A method for the solution of certain non-linear problems in least squares, *Quart. Appl. Math. 2,* 164-168.

Lindeman, R. H., P. F. Merenda, and R. Z. Gold (1980). *Introduction to Bivariate and Multivariate Analysis,* Scott, Foresman and Company, Ill.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Indust. Appl. Math. 11, No. 2,* 431-441.

McMechan, G. A. (1983). Seismic tomography in boreholes, *Geophys. J. R. Astr. Soc. 74,* 601-612

McMechan, G. A., and M. J. Yedlin (1981). Analysis of dispersive waves by wave field transformation, *Geophysics 46,* 869-874.

Menke, W. (1984). *Geophysical Data Analysis: Discrete Inverse Theory,* Academic Press, Florida.

Papoulis, A. (1962). *The Fourier Integral and its Applications,* McGraw-Hill Book Company, New York.

Papoulis, A. (1965). *Probability, Random Variables, and Stochastic Processes,* McGraw-Hill Book Company, New York.

Papoulis, A. (1977). *Signal Analysis,* McGraw-Hill Book Company, New York.

Penrose, R. (1955). A generalized inverse for matrices, *Proc. Cambridge Phil Soc. 51,* 406-413.

Rodi, W. L., P. Glover, T. M. C. Li, and S. S. Alexander (1975). A fast, accurate method for computing group velocity partial derivatives for Rayleigh and Love modes, *Bull. Seism. Soc. Am. 65,* 1105-1114.

Russell, D. R. (1980). Constrained inversion techniques applied to surface wave analysis, M. S. thesis, University of Texas at El Paso.

Sato, Y. (1955). Analysis of dispersed surface waves I, *Bull. Earthq. Res. Inst. Tokyo Univ. 34,* 33-47.

Stoffa, P. L., P. Buhl, and G. M. Bryan (1974). The application of homomorphic deconvolution to shallow-water marine seismology - part I: models, *Geophysics 39,* 401-416.

Takeuchi, H., and M. Saito (1972). Seismic surface waves, *in* B. A. Bolt (editor), *Methods of computational physics, vol. 11,* Academic Press, New York.

Talwani, M., G. H. Sutton, and J. L. Worzel (1959). Crustal section across the Puerto Rico Trench, *J. Geophys. Res. 64,* 1545-1555.

Thomson, W. T. (1950). Transmission of elastic waves through a stratified solid medium, *J. Applied Phys. 21,* 89-93.

Twomey, S. (1977). *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements,* Elsevier Scientific Publishing Co., Amsterdam.

Wang, C. Y. (1981). Wave theory for seismogram synthesis, Ph.D. Thesis, St. Louis University.

Wiggins, R. (1972). The general linear inverse problem: implications of surface waves and free oscillations on earth structure, *Rev. Space Phys. 10,* 251-284.

Woodhouse, J. H., and A. M. Dziewonski (1984). Mapping the upper mantle: three dimensional modeling of the earth structure by inversion of seismic waveforms, *J. Geophys. Res., 89, B7,* 5953-5986.

Biography of the Author

David Ray Russell was born on September 20, 1950, in Branson, Missouri, U.S.A. His father was a foreign service officer for the U.S. Department of State, and was required to work overseas. As a result, the author grew up in Africa, living in Libya, Jordan, Egypt, Nigeria, and Ghana, from 1951 to 1968.

Between 1969 and 1972, the author served on active duty with the U.S. Army, working as a radar technician for air defense missiles. From 1973 to 1975, he was a computer programmer for the U.S. State Department, in Washington, D.C.

The author attended the University of Texas at El Paso from 1975 to 1980 as a physics major, receiving a Bachelor's degree in 1978, and a Master's degree in 1980. From 1980 to the present, the author has attended St. Louis University, studying in the Department of Earth and Atmospheric Sciences. He is married, and has a daughter and son.