INFORMATION TO USERS

Pavlis, Gary Lee

PROGRESSIVE INVERSION

University of Washington   PH.D.   1982

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy.
Problems encountered with this document have been identified here with a check mark __√__ .

1.   Glossy photographs or pages _____

2.   Colored illustrations, paper or print _____

3.   Photographs with dark background _____

4.   Illustrations are poor copy _____

5.   Pages with black marks, not original copy _____

6.   Print shows through as there is text on both sides of page _____

7.   Indistinct, broken or small print on several pages __✓__

8.   Print exceeds margin requirements _____

9.   Tightly bound copy with print lost in spine _____

10.   Computer printout pages with indistinct print _____

11.   Page(s) _____ lacking when material received, and not available from school or author.

12.   Page(s) _____ seem to be missing in numbering only as text follows.

13.   Two pages numbered _____. Text follows.

14.   Curling and wrinkled pages _____

15.   Other_____

Progressive Inversion

by

Gary Lee Pavlis
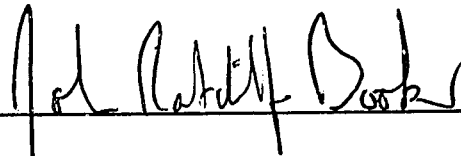
A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

1982

Approved by _____

Chairperson of the Supervisory Committee

Program Authorized

to offer Degree _____ Geophysics Program_____

Date _____ March 15, 1982_____

Doctoral Dissertation

Signature _____

Date _____ March 15, 1982 _____

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

NOTATION

CONVENTIONS

(1) Uppercase, boldface letters are used to symbolize matrices.

(2) Lowercase, boldface roman letters are used to symbolize vectors.

(3) Elements of a matrix or vector are symbolized by italic letters with subscripts.

(4) Partitions of a matrix or vector are written as boldface letters with subscripts.

(5) Uppercase, boldface letters with the ↔ symbol above them (e.g. $\overset{\leftrightarrow}{\mathbf{G}}$) are used to symbolize collections of functions as in equation (4.10).

(6) Elements of entities like $\overset{\leftrightarrow}{\mathbf{G}}$ are functions. A single member of the set is referenced using backet symbols surrounding the same root symbol with subscripts (e.g. $<G_i|$).

(7) Subsets of collections of functions like $\overset{\leftrightarrow}{\mathbf{G}}$ are denoted by the same symbol set in boldface but with a subscript (e.g. $\overset{\leftrightarrow}{\mathbf{G}}_i$).

(8) The δ symbol before anything symbolizes an infinitesimal perturbation to that entity.

(9) The Δ symbol before anything symbolizes a finite perturbation to that entity.

(10) The $\frown$ symbol over anything (e.g. $\hat{x}$) denotes an estimate of that quantity derived from the data.

(11) An underlined symbol (e.g. $<\underline{N_i}|$ or $\underline{\ddot{N}}$) always denotes a quantity derived from "quelled" (see text) kernel functions.

SYMBOLS

The following symbols are used unambiguously. Symbols used only in passing in the text are not included here. Most symbol definitions have an appended number which refers to the equation number where that symbol is defined in the text.

General

R^n Vector space of n-vectors with real coefficients

$R^{m \times n}$ The set of all real $m \times n$ matrices

$R(A)$ Range (column space) of matrix A (Definition 2.1)

$N(A)$ Null space of matrix A (Definition 2.2)

$\|x\|$ Norm of x where x is an unspecified mathematical object (vector, function, etc.). Subscripts on the final $\|$ symbol denote different types of norms. $\|x\|_2$ or $\|x\|$ denotes the L2 norm (2.11). $\|x\|_w$ denotes the weighted Euclidean norm (Appendix A).

$<A|B>$ Inner product of function $<A|$ with function $|B>$. In this dissertation this symbol generally implies the definition

$$<A|B> = \int_0^1 A(r)B(r)dr$$

Integers

M Total number of arrivals in data set   (3.23)

M_b Number of data blocks (section 4, Chapter 4)

M_n Number of "annulled data" $(M_n = M - N)$ (4.16)

m_e Number of earthquakes in data set   (3.23)

m_{ot} Number of explosion sources with known spatial location but for which the origin time is unknown or for which a shot point correction is to be calculated.(section 2, Chapter 5)

N Number of discrete parameters in data set (generally $N = 4m_e + n_s + m_{ot}$)   (3.28)

n_s Number of seismic stations in the network   (3.21)

Scalars

B Bound on L2 norm of derivative of the slowness perturbation (5.48)

L Deepest depth that velocity is to be estimated at   (4.3)

l Resolution length   (5.27)

τ Nondimensional vertical position variable   (4.5)

u Slowness

v Velocity

κ Roughness length   (5.51)

σ^2 Variance of a random variable

Vectors and Matrices

$\mathbf{A}^\dagger$ Moore-Penrose inverse of $\mathbf{A}$ (2.33)

$\mathbf{A}^M$ Levenberg-Marquardt (stochastic) inverse of $\mathbf{A}$ (2.38)

$\mathbf{A}^+$ Pseudoinverse of $\mathbf{A}$ (2.37)

$\mathbf{C}$ Covariance of discrete parameter estimates (3.70)

$\mathbf{D}$ Covariance of data (3.69)

$\mathbf{H}$ Generalized inverse (type not specified)   (3.16)

$\mathbf{h}'$ Coordinates of earthquake hypocenter (a four vector) (3.1)

$\mathbf{I}$ Identity matrix

$\mathbf{n}$ Annulled data vector (4.17)

Q	Orthogonal matrix of QR decomposition (theorem 2.1)
r	Residual vector (2.12)
$\mathbf{r}^s$	Residual vector including station corrections (3.20)
S	Station correction partial derivative matrix (3.21)
s	Vector of station corrections (3.21)
U	Orthogonal matrix of left singular vectors from a singular value decomposition of a matrix (theorem 2.2)
$\mathbf{U}_R$	Columns of **U** spanning range of matrix (2.21)
$\mathbf{U}_N$	Columns of **U** spanning null space of dual matrix (2.21)
V	Orthogonal matrix of right singular vectors from a singular value decomposition of a matrix (theorem 2.2)
0	A zero matrix (i.e. a matrix whose coefficients are all zero)
$\boldsymbol{\Lambda}$	Diagonal matrix of singular values from a singular value decomposition of a matrix (theorem 2.2)

Functions

$\langle A\,(r,r_0)|$	Backus-Gilbert averaging function for estimate of slowness perturbation at depth r_0 (5.23)
$\mathbf{\hat{G}}$	Collection of Frechet kernels for original data (4.10)
$\langle H\,(r-r_0)|$	Heaviside step function with step at r_0

$\hat{\textbf{N}}$ Collection of data kernels for annulled data (4.18)

$\underline{\textbf{N}}$ Inner product matrix of annulled data kernels (4.26)

$\hat{\textbf{X}}$ Collection of prediction kernels for discrete parameter esti-

mates (5.42)

$|\delta u>$ Slowness perturbation function (5.13)

$<\delta(r-r_0)|$ Delta function centered at r_0

$<\varepsilon|$ Prediction error function (5.13)

$<\underline{\Psi}_i|$ Orthonormal functions from spectral decomposition (winnow-

ing) (4.29)

ACKNOWLEDGEMENTS

The full list of the names of individuals I wish that I could thank is too long to be practical. Several individuals help has, however, been so outstanding that they must be recognized. At the top of this list is my wonderful wife, Mary Lynn. Words cannot express the debt I owe her. She supported me through my years in graduate school not only financially but also emotionally in ways that only she and I will ever know.

Next on the list is my advisor, John Booker. John and I worked well together on this project. The reason, I think, is that he had the good sense to leave me alone to let me do things my own way (that is the only way most of us will accept anyway), but I could always depend on him to answer any questions I had or to give me guidance when I needed it.

At this point the list of individuals and organization I need to thank widens enormously. The research reported here was financed by the National Science Foundation through grant EAR-7919605. Special thanks are also due to Alan Walter and Craig Weaver for providing me with access to the data from the Coso network. During the many hours of work which culminated in this dissertation I profited greatly from

conversations with so many individuals it is impossible to name them all. Of these, however, I would especially like to thank John Knapp. He gave me a lot of support as we struggled side by side to finish our own PhD dissertations.

I would like to recognize the contributions of Tom Jordan and Steve Roecker to this work. They were both generous enough to send me preprints of their papers describing their related work without my asking. Many of their ideas helped shape my own in ways that deserve more than a simple citation. Their openness to share new ideas is the kind of attitude that helps science move forward.

Finally, I want to thank the people of Shoreline United Methodist Church who strengthened me directly through social interaction and indirectly through their prayers. For through them I gained strength and knowledge from the One greatest source of all.

CHAPTER 1
INTRODUCTION

One of the most fundamental measurements in seismology is the arrival time of body waves recorded by an array of seismic stations. These data have been used to solve two classic problems in seismology. The first is investigate the internal structure of the earth by using these data to infer how seismic velocity varies within the earth's interior (a remote sensing problem) [Jordan, 1979]. The second is to determine the location (hypocenter) of the focus of an earthquake [Lee and Lahr, 1972]. The literature on methods of handling these two separate problems is vast (see e.g. Lee and Stewart [1981] or Bullen [1965,pp. 345-346]). Surprisingly little attention has been focused, however, on analyzing both problems at the same time. That is, arrival times from earthquakes generally contain information about both the hypocenter of the earthquake and the seismic velocity structure of the region. Peters and Crosson [Peters and Crosson ,1972; Peters, 1973; and Crosson, 1976a] were the first to investigate this problem quantitatively. The equations one must solve in this problem are nonlinear. They chose to linearize these equations with the velocity model specified as a finite number of parameters. Peters [1973] investigated models parameterized as constant velocity layers and a linear velocity gradient model. Crosson [1976a] extended Peters' work with layered models into a practical algorithm. This allowed them to solve these equations by a "simultaneous" nonlinear least squares method. Their work was important as a pioneering effort to solve this problem but it suffers from two major flaws:

(1) Least squares procedures normally require that the system of equations to be solve must be overconstrained [Jackson,1972]. The net result in this case is that application of Peters and Crosson's method usually requires that the number of parameters used to specify the velocity model (number of layers) be small. This can be criticized on the grounds that it is difficult to ascertain the effect of this rather arbitrary *a priori* form that is imposed on the velocity model. This problem can make interpreting the results of this technique difficult [Knapp, 1976, 1982].

(2) The linear equations that have to be solved grow in size proportional to the number of earthquakes in the data set. This leads to large computer memory requirements with the large data sets typical of local earthquake networks.

The work described in this dissertation eliminates problem (1) and substantially reduces problem (2). The key to these improvements is a special matrix operator I have termed the annulling transformation[1] because of its connection to the null space of the matrix from which it is derived (see appendix B). Application of this transformation to a set of earthquake arrival time data yields two significant results that are closely connected to problems (1) and (2) above.

(1) As I noted above the arrival time recorded at a given seismic station from an earthquake depends on both the hypocenter of the earthquake and the velocity structure. The most significant feature of the annulling transformation is that it yields a set of

---

[1] The idea of applying this transformation to aid in solving this problem appears to have been conceived almost simultaneously by three separate groups [Pavlis and Booker, 1981; Spencer and Gubbins, 1981; and Rodi et. al., 1980]. The form of the equations used by Spencer and Gubbins [1981] is different from that of Pavlis and Booker [1981] but they are essentially identical [Roecker, 1982].

independent averages of the original data, which I call the annulled data, that are locally independent of the hypocenter of the earthquake (When station corrections are used their contribution can also be annihilated by the same process.). The result is a set of equations that depend only upon the velocity model which can be inverted directly for the velocity structure. This is significant because it allows me to break free of the need to parameterize the velocity model. Instead, the velocity can be allowed to be an arbitrary function (In this work velocity is allowed to vary only in the vertical direction.) that can be estimated by the methods described by Backus and Gilbert [1967, 1969] and Johnson and Gilbert [1972]. Although it is more difficult to calculate travel times with models that vary continuously with depth, their use has been found (in the case of Hawaii at least) to frequently yield superior earthquake locations and focal mechanism solutions [Klein, 1981].

(2) The calculations required to implement the annulling transformation can be arranged to exploit the special structures in the matrix equations that one must ultimately solve. This leads to a considerable reduction in computer memory requirements in the practical implementation of this procedure. This, combined with a winnowing procedure [Gilbert, 1971] in the velocity model construction, results in a considerable degree of data compression (a 90% reduction was typical with data studied here). The result is that the procedure described here is capable of handling quite large data sets; a claim formerly made only for parameterized least squares methods.

Following the lead of Roecker [1982] I have dubbed the procedure I describe here with the title PRIMEL (Progressive Inversion and Multiple Event Location). I have adopted the term "progressive" in describing this procedure to contrast it with Peters and Crosson's "simultaneous" procedure. The algorithm described here is

"progressive" because the use of the annulling transformation permits one to adjust the three fundamentally different parts of the model (hypocenters, station corrections, and the velocity model) in three separate steps;

(1) Hypocenters are estimated by conventional (least squares) single event location methods.

(2) Station corrections are estimated from a set of annulled data calculated from the residuals of step (1) using a new procedure called progressive multiple event location (PMEL).

(3) The velocity model is constructed from a second set of annulled data derived from the residuals of step (2).

Thurber [1981] has shown that if the velocity model is parameterized, step (3) is solved by least squares, and no iterative improvement is used in steps (1) and (2); then this procedure reduces to the same result as Peters and Crosson's method. The actual algorithm used here is different, however, because steps (1) and (2) are generally performed iteratively. This is the fundamental reason for adopting the term progressive, since the hypocenters and station corrections are estimated in separate, iterative steps.

Finally, considering the length of this document it seems appropriate to provide a road map for the reader. First, this dissertation contains a considerable number of equations. Consequently, I have included a summary of notation to help the reader keep the many symbols straight. Within the main text, chapter 2 is a brief review of a number of relevant terms and concepts from linear algebra and the theory of generalized inverses. This material was provided primarily for those readers unfamiliar with the ideas of the range and null space of a matrix that are fundamental to understanding the annulling transformation and the nature of generalized inverses. A major companion to chapter 2 is appendix B, which develops a general theory for "mixed discrete-continuous inverse problems" based on the use of the annulling transformation.

Chapter 3 then reviews current methods of single and multiple event location. This review is provided primarily to place in its proper persective a new multiple event location scheme that is also described in chapter 3. This new procedure, which I term Progressive Multiple Event Location (PMEL), is based fundamentally on the annulling transformation. It is significant in its own right, however, because it has two major advantages over its only real competition; the method of joint hypocenter determination (JHD) introduced by Douglas [1968] and Dewey [1972].

(1) PMEL operates within a small, fixed storage area that is small enough to be implementable even on a micro computer. This is in contrast to JHD which suffers the same explosive storage growth problem as Peters and Crosson's simultaneous inversion procedure.

(2) PMEL does not require the auxiliary constraints needed for numerical stability by JHD.

For these reason I believe PMEL makes other multiple event location schemes obsolete.

The principle usage of PMEL in this study is as a building block in a larger scale solution construction scheme I have dubbed PRIMEL. PRIMEL (described in chapter 4) finds a single solution (hypocenters, station corrections, and a velocity model) that fits the data. That solution is, however, fundamentally nonunique [Backus and Gilbert, 1968, 1970]. Consequently, chapter 5 takes up the more fundamental question of appraising the uniqueness of this solution. I show how the uniqueness of the velocity model can be assessed by an application of the trade-off analysis of Backus and Gilbert [1970] and Johnson and Gilbert [1972]. In addition, a new technique is described for assessing the errors in the discrete parameters (hypocenters and station corrections) that takes into account the effect of the fundamental ambiguity introduced by the the nonuniqueness of the velocity model. Finally, chapters 6 and 7

demonstrate the validity of the methods developed in the preceding
chapters by demonstrating their use in the analysis of synthetic
data (chapter 6) and real data from two different seismic networks
(chapter 7).

CHAPTER 2
LEAST SQUARES
AND
ORTHOGONAL TRANSFORMATIONS

1. INTRODUCTION

One of the most widely used methods of analysis in all branches of science and engineering is the method of least squares. In particular least squares plays a central role in the work presented here because of its intimate connection with all currently used algorithms for locating earthquake hypocenters (see chapter 3). Because of that, it is useful to review some properties of least squares analysis that will be of use in subsequent chapters. The approach I have taken appeals primarily to the geometry of least squares because it is hoped that this will provide the reader with a better intuitive feeling for the important projective properties of least squares that will be exploited extensively in subsequent chapters. The approach taken here is relatively standard. Entire books are written on the subject of least squares analysis and the material of this chapter is only a summary of material relevant to this dissertation. For a more extensive treatment of the material presented here the reader is referred to the books of Stewart [1973], Ben-Israel and Greville [1974], and Lawson and Hanson [1974] that the author has found particularly useful.

2. THE GEOMETRY OF OVERDETERMINED SYSTEMS OF LINEAR EQUATIONS

2.1. Range and Null Space

We are interested in the general problem of solving a set of linear equations of the form

$$Ax=b \tag{2.1}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{b} \in \mathbb{R}^m$ [1]. For the purpose of this work we are always concerned with the case when $m > n$. In that case (2.1) is said to be **overdetermined**. To understand the nature of over-determined problems it is illustrative to consider a simple concrete example.

Suppose we are given the following set of equations

$$x_1 - 3x_2 = -3$$

$$4x_1 + 2x_2 = -4 \tag{2.2}$$

$$3x_1 + 2x_2 = 6$$

which we can write in matrix form as

$$\begin{bmatrix} 1 & -3 \\ 4 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -3 \\ -4 \\ 6 \end{bmatrix} \tag{2.3}$$

However, it is illuminating to write (2.3) in yet another way

$$\begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix} x_1 + \begin{bmatrix} -3 \\ 2 \\ 2 \end{bmatrix} x_2 = \begin{bmatrix} -3 \\ -4 \\ 6 \end{bmatrix} \tag{2.4}$$

Equation (2.4) is educational because it shows the fundamental nature of **any** solution to the set of equation of (2.2). That is, every solution defines some **linear combination** of the two vectors

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix} \quad \text{and} \quad \mathbf{a}_2 = \begin{bmatrix} -3 \\ 2 \\ 2 \end{bmatrix} \tag{2.5}$$

A particular solution is defined by how we chose the weights, x_1 and

---

[1] I will use a common notation convention in which $\mathbb{R}^n$ denotes the vector space of all possible n vectors with real components. Similarly I will use $\mathbb{R}^{m \times n}$ to denote the vector space of all $m \times n$ matrices with real coefficients.

x_2, of the linear combination. It is useful to consider the set of all possible solutions given formally by[2]

$$R(\mathbf{A})=\{\mathbf{y}: \mathbf{y}=\mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 \text{ for all } x_1 \in R \text{ and } x_2 \in R\} \qquad (2.6)$$

This has an important geometric interpretation that is sketched in figure 2.1. Each of the vectors, $\mathbf{a}_1$ and $\mathbf{a}_2$ defined by (2.5) have three components. If we erect a coordinate system with three orthogonal axes, then $\mathbf{a}_1$ and $\mathbf{a}_2$ can be drawn as vectors in the usual way. This has been done in figure 2.1. The set of all solutions, $R(\mathbf{A})$, defined by (2.6) is the unique plane defined by the two vectors $\mathbf{a}_1$ and $\mathbf{a}_2$. That plane is a fundamental property of the matrix operator of our example called its **range** or **column space** [Lawson and Hanson, 1974, p. 235]. In the language of linear algebra $R(\mathbf{A})$ is termed a **subspace** of R^3 of dimension 2. The two vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ define a **basis** for $R(\mathbf{A})$ and are said to **span** $R(\mathbf{A})$, which means $R(\mathbf{A})$ is completely defined by linear combinations of $\mathbf{a}_1$ and $\mathbf{a}_2$ [Stewart, 1973, pp. 12-18]. Every matrix $\mathbf{A} \in R^{m \times n}$ has associated with it four related subspaces. Two are properties of the problem defined by equation (2.1). They are defined as follows [Ben-Israel and Greville, 1974, p. 14]

Definition 2.1

$R(\mathbf{A})=\{\mathbf{y} \in R^m : \mathbf{y}=\mathbf{Ax} \text{ for some } \mathbf{x} \in R^n\}$, is called the **range** of $\mathbf{A}$.

Definition 2.2

$N(\mathbf{A})=\{\mathbf{x} \in R^n : \mathbf{Ax}=0\}$ is called the **null space** of $\mathbf{A}$.

Two others subspaces are properties of the dual problem, $\mathbf{A}^T \mathbf{y}=\mathbf{x}$. These are define as

---

[2] Equation (2.6) is written using some standard shorthand from linear algebra. The symbol ":" stands for "such that" and the brackets "{ }" are a shorthand for "the set of all"

Definition 2.3

$R(\mathbf{A}^T) = \{x \in R^n : x = \mathbf{A}^T y \text{ for some } y \in R^m\}$, is called the **range** of $\mathbf{A}^T$.

Definition 2.4

$N(\mathbf{A}^T) = \{y \in R^m : \mathbf{A}^T y = 0\}$ is called the **null space** of $\mathbf{A}^T$.

The range and null space of a matrix and its transpose are inseparablely linked. To see how, it is illustrative to consider our example again. It turns out that, for that example, $N(\mathbf{A}^T)$ is completely defined by a vector that is perpendicular to $R(\mathbf{A})$. In other words, $N(\mathbf{A}^T)$ is the part of three space that is perpendicular to the plane $R(\mathbf{A})$. This concept is easily generalized for any arbitrary matrix $\mathbf{A} \in R^{m \times n}$ and yields two fundamental relations in the theory of generalized inverses [Ben-Israel and Greville, 1974, p. 64]

$$N(\mathbf{A}^T) = R(\mathbf{A})^\perp \qquad (2.7)$$

and for the dual problem

$$N(\mathbf{A}) = R(\mathbf{A}^T)^\perp \qquad (2.8)$$

the $\perp$ symbol in (2.7) implies three properties

(1) $R(\mathbf{A}) \cap N(\mathbf{A}^T) = \{0\}$

(2) $R(\mathbf{A}) \cup N(\mathbf{A}^T) = R^m$

(3) For all $y_r, y_n \in R^m$ with $y_r \in R(\mathbf{A})$ and $y_n \in N(\mathbf{A}^T)$, $y_r^T y_n = 0$

The corresponding relationships for the dual problem are similar and can be had simply by interchanging $\mathbf{A}$ and $\mathbf{A}^T$ and replacing R^m by R^n everywhere. Properties (1) and (2) define $R(\mathbf{A})$ and $N(\mathbf{A}^T)$ as **complimentary subspaces** of R^m [Stewart, 1973, p. 17]. They are termed complimentary because they share only a single common member (the zero vector), and yet together they generate all of R^m. Property (3) is a stronger requirement. It says that any vector in $R(\mathbf{A})$ is perpendicular to any other vector in $N(\mathbf{A}^T)$. Because of (3), $R(\mathbf{A})$ and $N(\mathbf{A}^T)$ are called **orthogonal compliments** [Stewart, 1973, p. 215]. The significance of this is that any vector $y \in R^m$ can always

be decomposed into two perpendicular vectors

$$\mathbf{y} = \mathbf{y}_r + \mathbf{y}_n \qquad (2.9)$$

where $\mathbf{y}_r \in R(\mathbf{A})$ and $\mathbf{y}_n \in N(\mathbf{A}^T)$. The importance of (2.9) cannot be overstressed as it is the foundation of the theory presented in this dissertation. We will return to it again on several occasions.

2.2. Least Squares

In the previous section the concept of the range and null space of a matrix and its transpose was introduced. The range and null space of a matrix are unique properties of the matrix itself and are independent of any particular solution of some set of linear equations the matrix might be associated with. In this section, however, the discussion will be limited to the study of the special solution method that is commonly called least squares.

To begin this discussion, consider again the simple example introduced in the last section. Because the narrative there was focused on the general properties of all possible solutions, I conveniently ignored the other key part of equation (2.4). That is, the right hand side vector

$$\mathbf{b} = \begin{bmatrix} -3 \\ -4 \\ 6 \end{bmatrix}$$

If $\mathbf{b}$ is plotted with $\mathbf{a}_1$ and $\mathbf{a}_2$, the resulting geometry is that shown in figure 2.2. The main point to observe from figure 2.2 is that the right hand side vector, $\mathbf{b}$, does not lie in the plane $R(\mathbf{A})$. Yet, it was shown in the previous section that **all** solutions must lie in the plane $R(\mathbf{A})$. Thus, no solution can possible reproduce the right hand side vector, $\mathbf{b}$, exactly. The equations of the example are said to be **inconsistent**. What is the least squares solution then? A **least squares** solution of the general $m \times n$ system of equations (2.1) is defined as
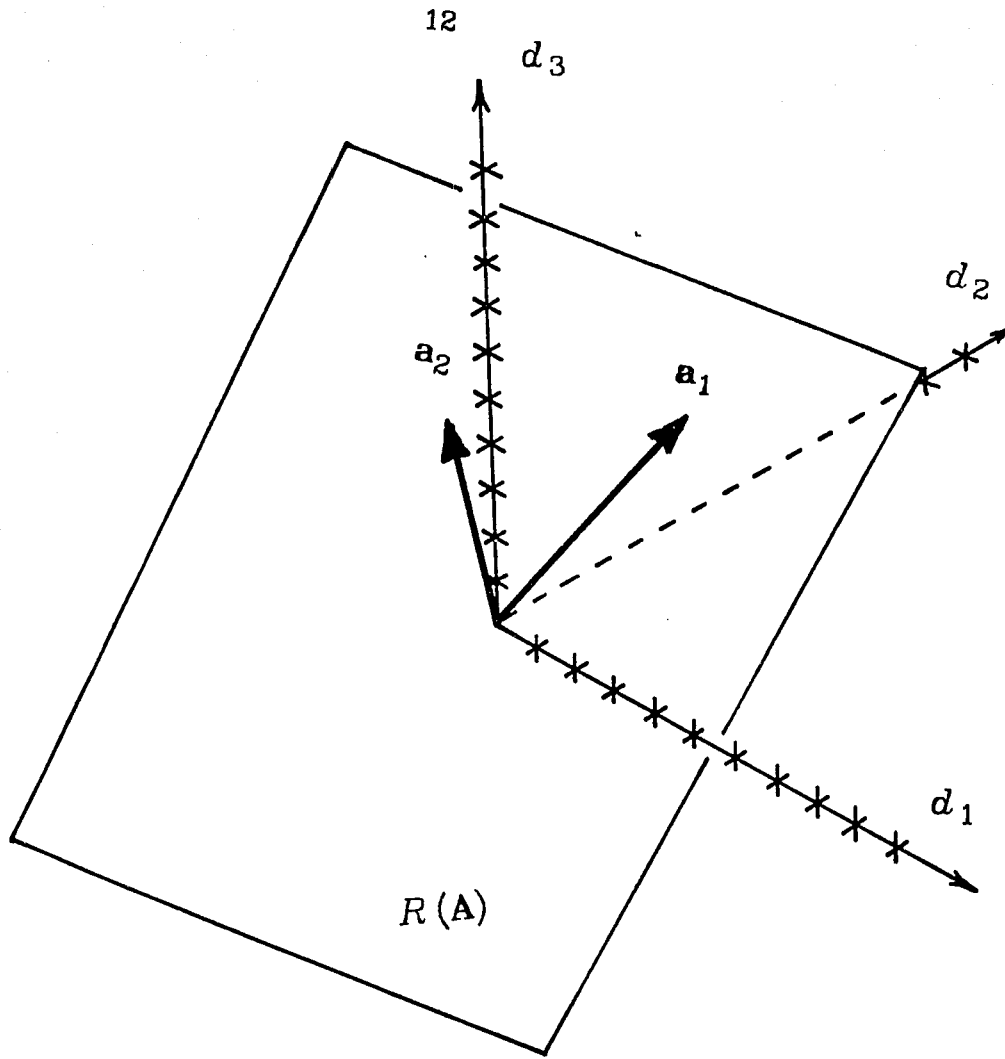
Figure 2.1. Geometry of matrix operator for overdetermined example problem. The plane $R(\mathbf{A})$ is formed by the two vectors $\mathbf{a}_1$ and $\mathbf{a}_2$. This plane lies behind the positive d_1 and d_3 axes and in front of the positive d_2 axis in this isometric projection.
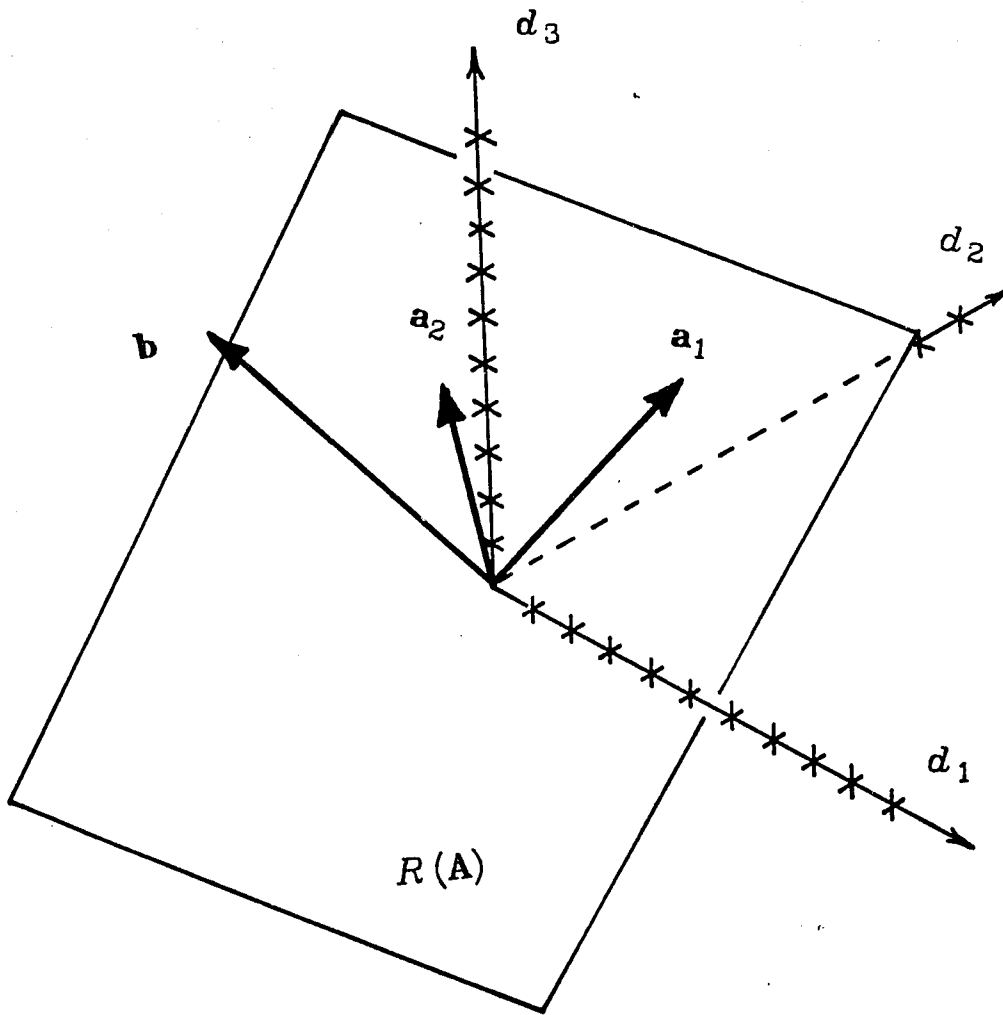
Figure 2.2. Geometry of set of overdetermined linear equations of example problem. The data vector, **b**, lies above the plane $R(\mathbf{A})$ in this view.

$$\mathbf{x}_{l.s.} = \min_{\mathbf{x} \in R^n} \left[\|\mathbf{b} - \mathbf{Ax}\|_2^2 \right] \tag{2.10}$$

where the symbol $\| \ \|_2$ denotes the L2 norm defined as

$$\|\mathbf{y}\|_2 = \left[\sum_{i=1}^{m} y_i^2 \right]^{1/2} \quad \text{and} \quad \|\mathbf{y}\|_2^2 = \left[\|\mathbf{y}\|_2 \right]^2 = \mathbf{y}^T \mathbf{y} \tag{2.11}$$

In our example, we already saw that $\mathbf{Ax}$ was restricted to lie in the plane $R(\mathbf{A})$ of figure 2.2. Since $\mathbf{b}$ does not lie in $R(\mathbf{A})$, criterion (2.10) says that the least squares solution is the unique one for which the solution vector, $\mathbf{Ax}$, is "closest" to the right hand side vector, $\mathbf{b}$. For least squares, "closest" means close in the sense of the L2 norm. For our example, this has the simple geometric interpretation sketched in figure 2.3. The least squares solution is the **orthogonal projection** of the right hand side vector onto the plane $R(\mathbf{A})$. That is, if you were to reorient your viewing position of figure 2.3 so that you were looking perpendicular to the plane $R(\mathbf{A})$, then the vector $\mathbf{Ax}_{l.s.}$ would be the image of $\mathbf{b}$ projected onto the plane $R(\mathbf{A})$. This is a significant observation because it means that a least squares solution accomplishes the decomposition of equation (2.9). That is, the residual vector

$$\mathbf{b} - \mathbf{Ax}_{l.s.} = \mathbf{r} \in N(\mathbf{A}^T) \tag{2.12}$$

and

$$(\mathbf{Ax}_{l.s.})^T \mathbf{r} = 0 \tag{2.13}$$

In other words, the residual vector, $\mathbf{r}$, lies totally in $N(\mathbf{A}^T)$ and has no projections in $R(\mathbf{A})$. $\mathbf{r}$ is the same as the vector $\mathbf{y}_n$ of (2.9). One way of viewing a set of equations like (2.2) is that we have three **data values** (right hand sides) that we are trying to fit with only two **parameters.** Since we know that normally two data points are sufficient to determine two parameters, the data are redundant. The preceding analysis of our sample problem shows that the least squares solution of (2.2) extracts only two independent pieces of
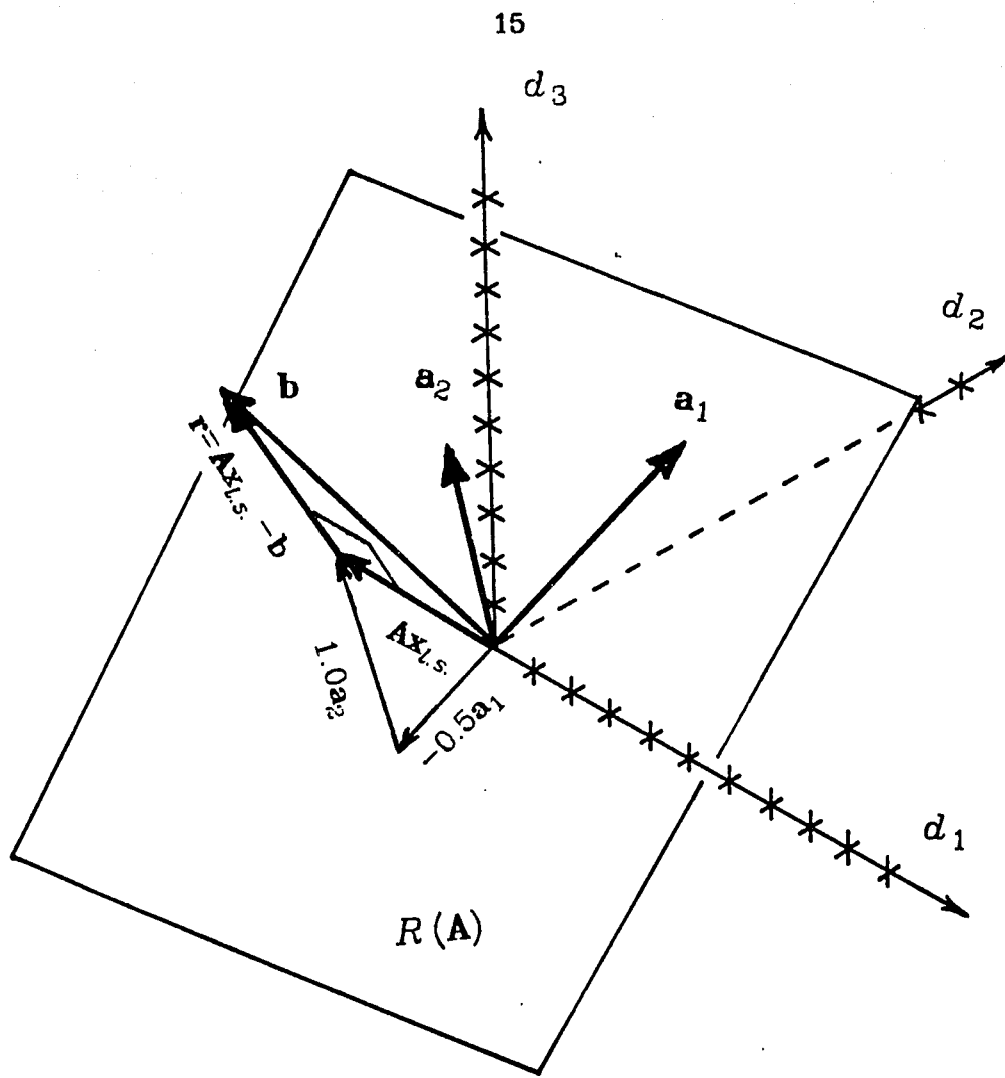
Figure 2.3. Geometry of least squares solution of example problem. The least squares solution, $\mathbf{Ax}_{l.s.}$, is the orthogonal projection of the data vector,$\mathbf{b}$, onto $R(\mathbf{A})$. The residual vector,$\mathbf{r}$, lies orthogonal to $R(\mathbf{A})$.

information from the data vector. The information that is left over, the residual, contains the one remaining independent piece of information. This idea generalizes readily to higher order systems. A least squares solution of an arbitrary $m \times n$ system of equations with $m > n$, extracts at most n pieces of information from the data vector. There always remain in the dregs of the data vector (the residuals) at least $m - n$ pieces of information that can be extracted.

3. ORTHOGONAL TRANSFORMATIONS

If we look at the geometry of our sample problem again in figure 2.3, one might observe that a great deal of the complications that occur there are due to a bad orientation of the coordinate system. That is, if we could reorient the coordinate system so two of the coordinate axes lay in the plane R (**A**) and one axis was perpendicular to R(**A**) (i.e. in $N(\mathbf{A}^T)$), then we would find the problem easier to solve. In this section I will review the principles of a method that does precisely that. This method is called the QR factorization. For a more detailed treatise the reader is referred to chapter 5 of Stewart [1973].

The QR factorization is based on the use of orthogonal transformations. Orthogonal transformations are synonymous with orthogonal matrices [Stewart, 1973, p.212] that are defined as follows

Definition 2.5

A matrix $\mathbf{U} \in R^{m \times m}$ is **orthogonal** if and only if $\mathbf{U}^T \mathbf{U} = \mathbf{U} \mathbf{U}^T = \mathbf{I}^{m \times m}$

where $\mathbf{I}^{m \times m}$ denotes an m by m identity matrix. Orthogonal transformations are fundamental to L2 norm problems because of the following two properties:

$$(\mathbf{U}\mathbf{x})^T \mathbf{U}\mathbf{y} = \mathbf{x}^T \mathbf{y} \quad \text{and} \quad \|\mathbf{U}\mathbf{x}\|_2^2 = \|\mathbf{x}\|_2^2 \qquad (2.14)$$

for all $\mathbf{x}, \mathbf{y} \in R^m$. The proof of both these relations follows immediately from definition 2.5. What this says is that an orthogonal transformation does not alter the size (when measured by the L2

norm) of a vector or the angular relation (the dot product) of two vectors. That basically implies that an orthogonal transformation does not distort Euclidean space.

Any orthogonal matrix can be written as the product of a finite number of two elementary operations; **plane rotations (Givens rotation matrix)**

$$\mathbf{P}_{ij} = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos\theta & \cdots & \sin\theta & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & -\sin\theta & \cdots & \cos\theta & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \qquad (2.15)$$

and the **elementary reflectors (Householder transformations)**

$$\mathbf{H} = \mathbf{I} - 2\frac{\mathbf{u}\mathbf{u}^T}{\|\mathbf{u}\|_2^2} \qquad (2.16)$$

Plane rotations perform a rigid rotation of the coordinate system and elementary reflectors reflect the coordinate system through a plane of symmetry whose normal is $\frac{\mathbf{u}}{\|\mathbf{u}\|_2}$.

The elementary orthogonal matrices (2.15) and (2.16) are important as building blocks to accomplish the QR decomposition. The QR decomposition is defined by the following theorem that also guarantees its existence [Lawson and Hanson, 1974, p. 11]

Theorem 2.1 QR factorization

For every $\mathbf{A} \in \mathbb{R}^{m \times n}$ there is an $m \times m$ orthogonal matrix $\mathbf{Q}$ such that $\mathbf{Q}^T \mathbf{A} = \mathbf{R}$ is zero below the main diagonal.

The matrix $\mathbf{Q}$ can be calculated using either plane rotations or elementary reflectors. (For the application presented here

elementary reflectors are preferred because they automatically produce an orthogonal basis for $N(\mathbf{A}^T)$ whose usefulness will become apparent in the two succeeding chapters.) The details of algorithms to accomplish the QR decomposition are well known and will not be inflicted upon the reader[3]. The essence of the method is that given an $m \times n$ matrix $\mathbf{A}$ a set of $r = \min(m,n)$ elementary reflectors can be calculated such that the matrix $\mathbf{Q}$ of the QR factorization can be written as the product

$$\mathbf{Q} = \mathbf{H}_r \mathbf{H}_{r-1} \cdots \mathbf{H}_2 \mathbf{H}_1 \qquad (2.17)$$

To make this more concrete let us return again to the example problem (2.2). After doing the arithmetic of (2.17) we get for this example

$$\mathbf{Q} = \begin{bmatrix} -.196 & .974 & .112 \\ -.785 & -.087 & -.613 \\ -.588 & -.208 & .781 \end{bmatrix} \qquad (2.18)$$

$\mathbf{Q}$ reduces the (2.3) to the upper triangular system

$$\begin{bmatrix} -5.099 & -2.156 \\ 0 & -3.514 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -.196 \\ -3.82 \\ 6.81 \end{bmatrix}$$

which yields the solution quoted earlier in figure 2.3.

$x_1 = -0.499$

$x_2 = 1.09$

by simple back substitution.

The major purpose for this discussion is to show what $\mathbf{Q}$ does geometrically. I have stated that $\mathbf{Q}$ simply reorients the coordinate

---

[3] Businger and Golub [1965] were the first to publish a computer algorithm using elementary reflectors in least squares problems. For a more transparent treatise see Stewart, 1973, pp. 230-245.

system. That reorientation is sketched in figure 2.4 where I show the original coordinate axes labeled as d_1, d_2, and d_3 and their images after the transformation by **Q** that are labeled as d'_1, d'_2, and d'_3. The directions of the axes d'_1, d'_2, and d'_3 are defined by the unit vectors that are the columns of the matrix **Q** in (2.18). Note that the d'_1 and d'_2 axes now lie in the plane $R(\mathbf{A})$ and d'_3 is perpendicular to $R(\mathbf{A})$. This is a convenient reorientation because now the solution vector will lie in the plane of d'_1 and d'_2 and the least squares residual vector will lie parallel to d'_3. This says that in the transformed coordinate system the decomposition (2.9) is automatic. For instance, the two projections for our sample problem are

$$\mathbf{y}_r = \begin{bmatrix} 0.196 \\ -3.82 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{y}_n = \begin{bmatrix} 0 \\ 0 \\ 6.81 \end{bmatrix}$$

The major point is that when viewed in this transformed coordinate system the residual vector is seen to have only one nonzero component. This concept generalizes to higher order systems. For an $m \times n$ system of equations with $m > n$ the residual vector will always have a total of $m - n$ nonzero components after transformation by the matrix **Q** of the QR decomposition.

4. NON FULL RANK PROBLEMS

4.1. Singular value decomposition

The example problem I have been discussing has an important simplification; the matrix **A** is full rank. The geometric interpretation of the rank of a matrix is that

$$\text{rank}(\mathbf{A}) = \text{dimension}(R(\mathbf{A}))$$

In the case of the example discussed above $R(\mathbf{A})$ was a plane so rank(**A**) was two. That problem was full rank because that is the
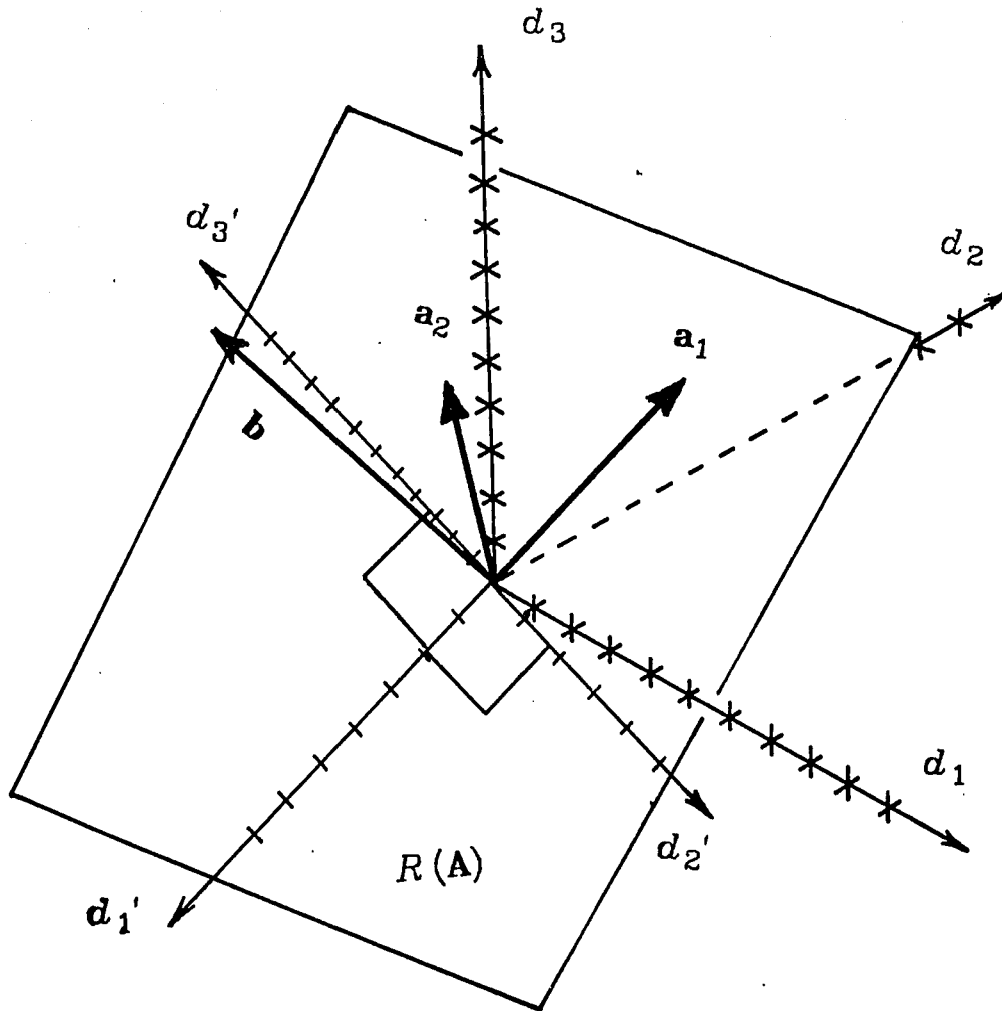
Figure 2.4. Coordinate reorientation produced by the orthogonal matrix **Q** resulting from a QR decomposition of the matrix **A** of example problem. d'_1 lies in $R(\mathbf{A})$ and is parallel to $\mathbf{a}_1$. d'_2 lies in $R(\mathbf{A})$ but is perpendicular to $\mathbf{a}_1$. d'_3 is orthogonal to the plane $R(\mathbf{A})$.

maximum possible dimension of $R(\mathbf{A})$ for any matrix with only two columns. A simple version of a non full rank problem is a modified version of (2.2)

$$\begin{bmatrix} 1 & -2 \\ 4 & -8 \\ 3 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -3 \\ -4 \\ 6 \end{bmatrix} \tag{2.19}$$

For this example,

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix} \text{ and } \mathbf{a}_2 = \begin{bmatrix} -2 \\ -8 \\ -6 \end{bmatrix}$$

The geometry of this problem is shown in figure 2.5. The crux of the situation here is that $\mathbf{a}_2$ and $\mathbf{a}_1$ are now simple multiples of one another ($-2\mathbf{a}_1 = \mathbf{a}_2$). $\mathbf{a}_2$ and $\mathbf{a}_1$ are linearly dependent and the subspace $R(\mathbf{A})$ has degenerated to a single line, a subspace of dimension one. Consequently its orthogonal complement, $N(\mathbf{A}^T)$, is a subspace of dimension two defined by any plane normal to the line $R(\mathbf{A})$. In addition, the dual subspaces, $N(\mathbf{A})$ and $R(\mathbf{A}^T)$, given by definitions 2.2 and 2.3 now occupy an important role (I intentionally ignored them for the initial example because $N(\mathbf{A})$ for that example had only one member (the zero vector) and had little relevance to the problem.). Definition 2.3 states the $R(\mathbf{A}^T)$ is defined as all possible linear combinations of the rows of $\mathbf{A}$. For the present example, the rows of $\mathbf{A}$ define vectors with two components that can be plotted on a plane. Doing this yields the geometry shown in figure 2.6. The three row vectors are all collinear so $R(\mathbf{A}^T)$ is a subspace of dimension one and its orthogonal compliment is also a subspace of dimension one.
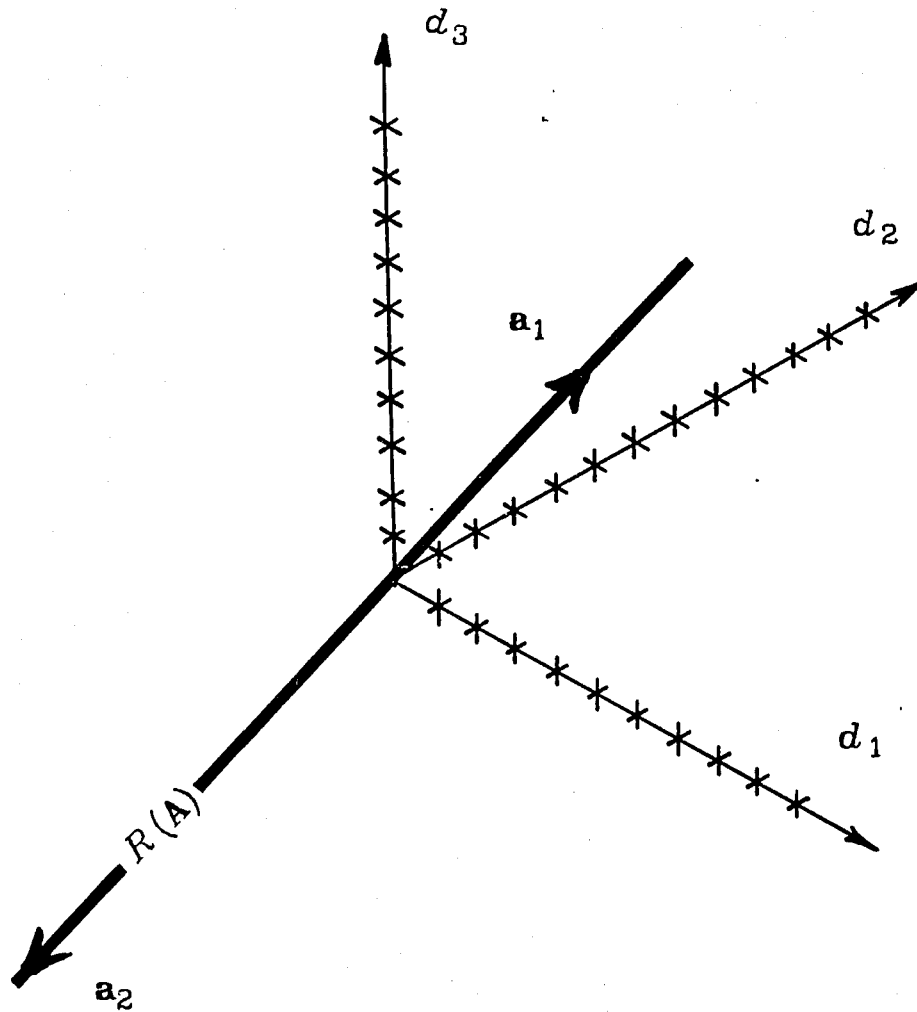
Figure 2.5. Geometry of non full rank example problem. Here $A \in \mathbb{R}^{3 \times 2}$ has a rank of one and $R(A)$ degenerates to a line.
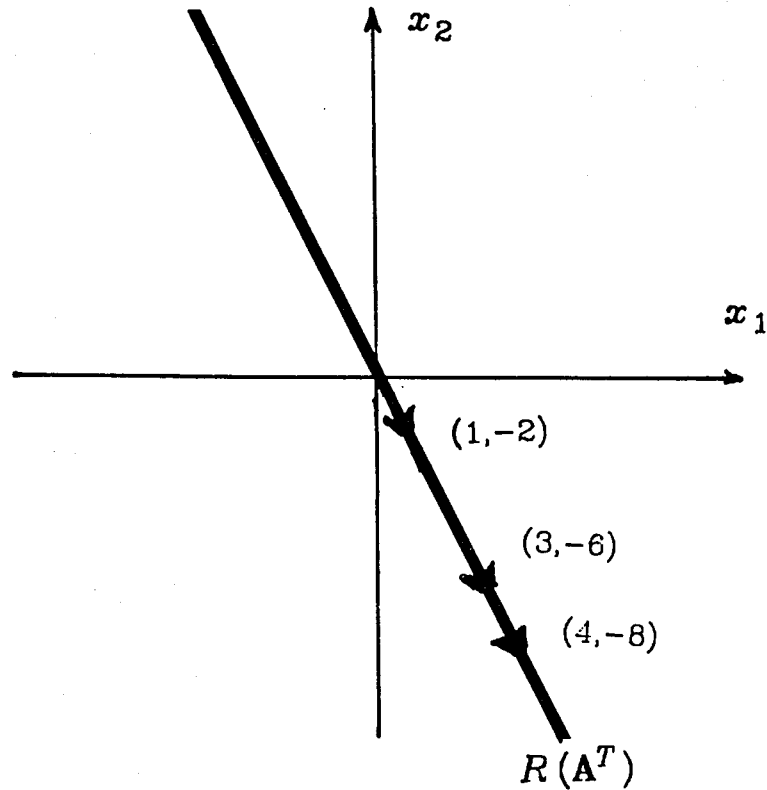
Figure 2.6. Geometry of dual problem for non full rank example. (1,-2), (3,-6), and (4,-8) are the row vectors of the matrix **A** in this example. They are all colinear in this case because $rank(\mathbf{A}^T)=rank(\mathbf{A})=1$. Hence $R(\mathbf{A}^T)$ is a line.

The geometry of the data space (figure 2.5) and the geometry of the parameter space (figure 2.6) of this example will help provide a geometric interpretation to a second important orthogonal decomposition called the singular value decomposition. The singular value decomposition[4] is defined by the following theorem that also guarantees its existence [Lawson and Hanson, 1974, p.18-20].

Theorem 2.2 (Singular Value Decomposition): Given $A \in R^{m \times n}$ with rank $(A) = r$, then there is an $m \times m$ orthogonal matrix **U**, an $n \times n$ orthogonal matrix **V**, and an $m \times n$ diagonal matrix Λ such that

$$U^T A V = \Lambda \quad \text{and} \quad A = U \Lambda V^T \tag{2.20}$$

Here the diagonal entries of Λ can be arranged to be nonincreasing; all these entries are nonnegative; and exactly r of them are strictly positive.

The diagonal entries of Λ are called the **singular values** of **A** (For a proof and a more complete discussion of this theorem see Lawson and Hanson [1974, pp. 18-20].).

To clarify theorem 2.2 it is instructive to write (2.20) in an expanded, partitioned form

$$A = \begin{bmatrix} U_R & U_{N_1} & U_{N_2} \end{bmatrix} \begin{bmatrix} \Lambda_R & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_R^T \\ V_{N_1}^T \end{bmatrix} \tag{2.21}$$

where $U_R \in R^{m \times r}$, $U_{N_1} \in R^{m \times (n-r)}$, $U_{N_2} \in R^{m \times (m-n)}$, $V_R \in R^{n \times r}$,

---

[4] The singular value decomposition is closely connected with an eigenvalue-eigenvector decomposition of the two nonnegative definite matrices $A^T A$ and AA^T. This interpretation is also very instructive and can be found in a well written form in chapter 3 of Lanczos [1961]. The rise in popularity of the direct decomposition of equation (2.20) was a direct consequence of the publication of a fast, stable algorithm by Businger and Golub [1969] that utilized elementary reflectors and the QR algorithm of Francis [1961, 1962].

$\mathbf{V}_{N_1} \in \mathbb{R}^{n \times (n-r)}$, and $\Lambda_R \in \mathbb{R}^{r \times r}$ is of the form

$$\Lambda_R = \begin{bmatrix} \Lambda_1 & 0 & . & . & . & 0 \\ 0 & \Lambda_2 & . & . & . & 0 \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ 0 & 0 & . & . & . & \Lambda_r \end{bmatrix} \tag{2.22}$$

The partitioning of equation (2.21) is not arbitrary It is related as follows to the range and null space of **A** and its dual:

(1) The columns of $\mathbf{U}_R$ form an orthonormal basis for $R(\mathbf{A})$.

(2) The columns of $\mathbf{U}_{N_1}$ and $\mathbf{U}_{N_2}$ together form an orthonormal basis for $N(\mathbf{A}^T)$ [5].

(3) The columns of $\mathbf{V}_R$ (rows of $\mathbf{V}_R^T$) form an orthonormal basis for $R(\mathbf{A}^T)$.

(4) The columns of $\mathbf{V}_{N_1}$ (rows of $\mathbf{V}_{N_1}^T$) form an orthonormal basis for $N(\mathbf{A})$.

One view of the singular value decomposition is that it is a QR decomposition taken one step farther. That is, it was noted in the previous section that the QR decomposition can be thought of as a convenient reorientation of the coordinates of the data space. The singular value decomposition goes one step farther by also reorienting the coordinates of the parameter space to yield the simple form (2.21).

To clarify this I again appeal to the example problem (2.19). After some simple arithmetic, one can find the singular value decomposition of the matrix **A** in (2.19) is

---

[5] As a practical matter $\mathbf{U}_{N_1}$ $\mathbf{U}_{N_2}$ are indistinguishable. The reason I have chosen to write them as two separate entities is because $\mathbf{U}_{N_2}$ always exists but $\mathbf{U}_{N_1}$ does not. $\mathbf{U}_{N_1}$ only exists when **A** is not full rank $(r < n)$. This notation emphasizes this fact.

$$
\begin{bmatrix} 1 & -2 \\ 4 & -8 \\ 3 & -6 \end{bmatrix} = \begin{bmatrix} -.196 & -.785 & -.588 \\ -.785 & .486 & -.386 \\ -.588 & -.386 & .711 \end{bmatrix} \begin{bmatrix} 11.4 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -.447 & .894 \\ .894 & .447 \end{bmatrix} \quad (2.23)
$$

As with the QR factorization each of the orthogonal matrices of of the singular value decomposition represent a coordinate reorientation. **U** represents a reorientation of the data space similar to the matrix **Q** of the QR decomposition, and the matrix **V** represents a similar reorientation of the parameter space. The reorientation defined by **U**, for this example, is sketched in figure 2.7. The original coordinate axes are shown labeled as d_1, d_2, and d_3 and their images after the transformation **U** are labeled d'_1, d'_2, and d'_3 where the directions of d'_1, d'_2, and d'_3 are defined by the unit vectors that are the columns of **U**. Similarly, figure 2.8 shows the corresponding reorientation of the parameter space by $\mathbf{V}^T$. The original coordinate axes labeled x_1 and x_2 are transformed by $\mathbf{V}^T$ to x'_1 and x'_2. However, here the directions of x'_1 and x'_2 are defined by the rows of $\mathbf{V}^T$ (columns of **V**). The affect of these reorientations shown in figure 2.7 and 2.8 is the same in both cases. d_1 and x_1 are reoriented such that images, d'_1 and x'_1, lie in $R(\mathbf{A})$ and $R(\mathbf{A}^T)$ respectively. d'_2 and d'_3, on the other hand, are reoriented to lie perpendicular to $R(\mathbf{A})$ and their images are a natural set of coordinates (an orthonormal basis) for $N(\mathbf{A}^T)$. similarly, x_2 is reoriented perpendicular to $R(\mathbf{A}^T)$ and is a natural coordinate for $N(\mathbf{A})$. Thus, **U** and **V** represent a reorientation of the data and parameter spaces into principle coordinate systems. This leads directly to a simple set of relationships linking the data and parameter spaces that I will now discuss.
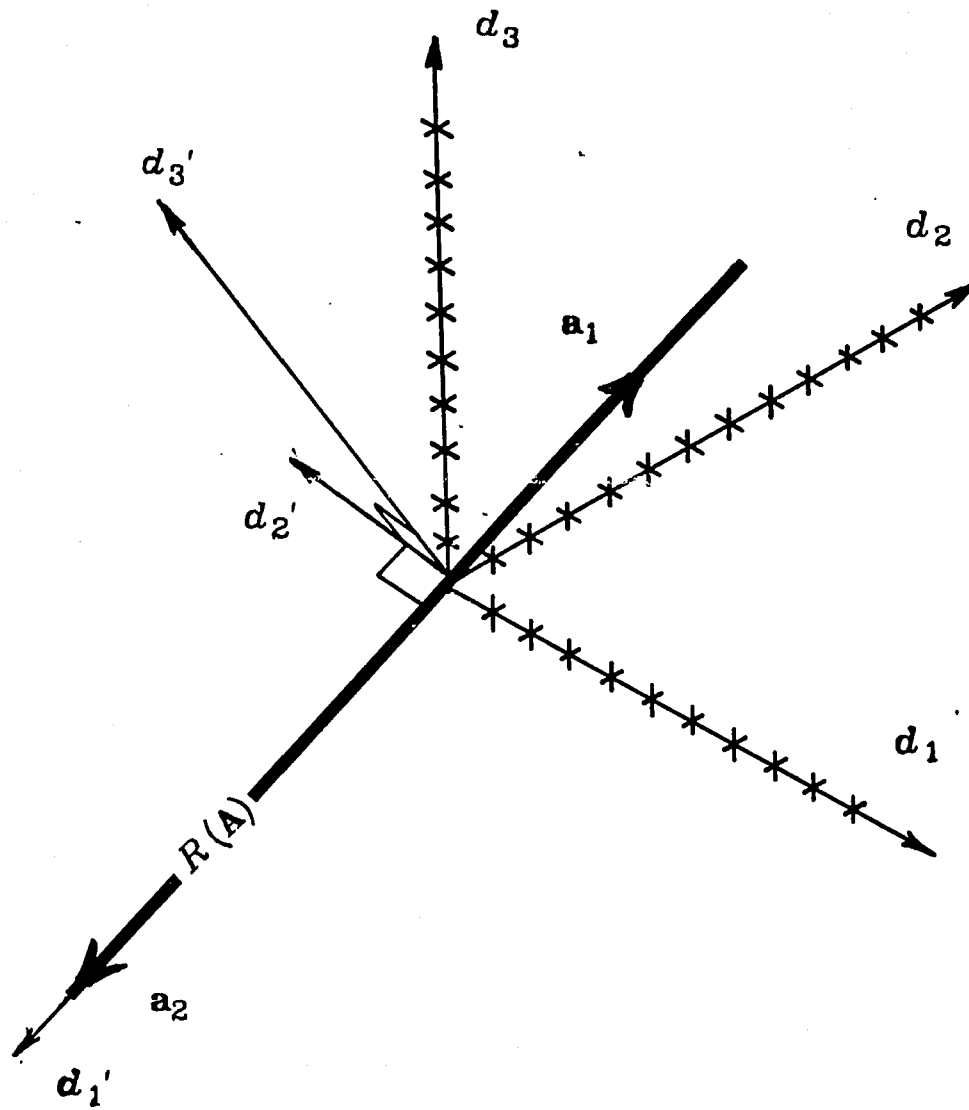
Figure 2.7. Reorientation of axes produced by the matrix **U** resulting from a singular value decomposition of the matrix **A** of example non full rank problem. d'_1 lies along the line $R(\mathbf{A})$. d'_2 and d'_3 are orthogonal to $R(\mathbf{A})$.
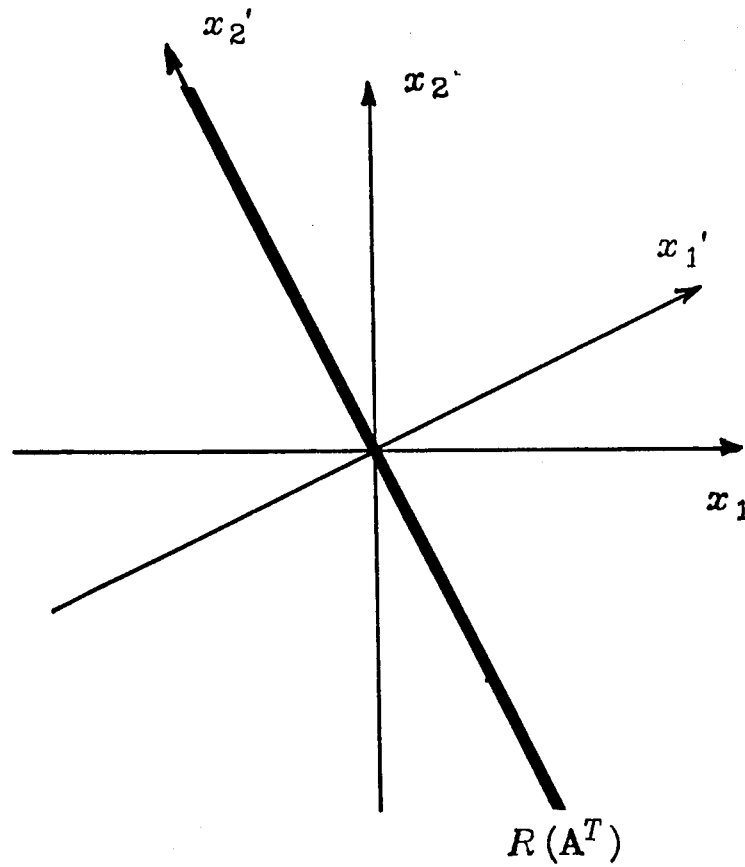
Figure 2.8. Coordinate reorientation produced by the orthogonal matrix **V** resulting from a singular value decomposition of the matrix **A** of example non full rank problem. x'_1 lies parallel to the line $R(\mathbf{A}^T)$ and x'_2 lies orthogonal to $R(\mathbf{A}^T)$

4.2. Least squares generalized inverses

4.2.1. The Moore-Penrose inverse

The singular value decomposition was introduced in the previous section because of its intimate connection with the theory of generalized inverses and especially the unique generalized inverse called the Moore-Penrose inverse[6]. Following Lanczos [1961] and Jackson [1972], I will now demonstrate that the Moore-Penrose inverse follows in a natural way from the singular value decomposition.

From equation (2.20) we see (2.1) can be rewritten as

$$\mathbf{U}\Lambda\mathbf{V}^T\mathbf{x}=\mathbf{b} \qquad (2.24)$$

Since $\mathbf{U}$ is orthogonal, this can also be written as

$$\Lambda\mathbf{V}^T\mathbf{x}=\mathbf{U}^T\mathbf{b} \qquad (2.25)$$

It is convenient to define the following change of variables:

$$\mathbf{b'}=\mathbf{U}^T\mathbf{b} \qquad (2.26)$$

and

$$\mathbf{x'}=\mathbf{V}^T\mathbf{x} \qquad (2.27)$$

In the previous section we saw from a simple example that $\mathbf{U}$ and $\mathbf{V}$ can be thought of as transformations that perform a rigid reorientation of the coordinate system of the data space (R^m) and the parameter space (R^n) respectively. The change of variable defined by (2.26) and (2.27) are nothing but an application of these

---
[6] The Moore-Penrose inverse has an almost absurd number of aliases. It is also referred to as the general reciprocal [Moore, 1920], generalized inverse [Penrose, 1955], intrinsic inverse [Lanczos, 1961], pseudoinverse [Lawson and Hanson, 1974], and probably others.

transformations to obtain the coordinates of **b** and **x** in the reoriented coordinate systems. The remarkable property of the reorientations defined by **U** and **V** is the form the equations (2.25) take after the change of variables (2.26) and (2.27).

$$\Lambda_1 x'_1 = b'_1$$

$$\Lambda_2 x'_2 = b'_2$$

$$\vdots$$

$$\Lambda_r x'_r = b'_r \tag{2.28}$$

$$0 x'_{r+1} = b'_{r+1}$$

$$\vdots$$

$$0 x'_n = b'_n$$

Thus, in the reoriented coordinate systems the parameters and the data are simple multiples of one another. This is why I referred to them as "principle coordinates" earlier. Because of this simple form, inverting these equations is relatively easy. There is, however, a complication we have to consider. The last $n-r$ relationships in (2.28) cannot be made into equalities for any choice of x'_i $(i = r+1, r+2, \ldots, n)$. Thus, in a sense, the choice of the numbers x'_i in these relations is arbitrary. The Moore-Penrose inverse simply sets them to zero. With this definition we can write the reciprocal relations to (2.28) as

$$x'_1 = \frac{b'_1}{\Lambda_1}$$

$$x'_2 = \frac{b'_2}{\Lambda_2}$$

$$\vdots$$

$$x'_r = \frac{b'_r}{\Lambda_r} \tag{2.29}$$

$$x'_{r+1} = 0$$

$$\vdots$$

$$x'_n = 0$$

It is convenient to rewrite (2.28) in matrix form as

$$\mathbf{x}' = \Lambda^{-1}\mathbf{b}' \tag{2.30}$$

where $\Lambda^{-1} \in R^{n \times m}$ is of the form

$$\Lambda^{-1} = \begin{bmatrix} \frac{1}{\Lambda_1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\Lambda_2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\Lambda_r} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{2.31}$$

After substituting back into (2.30) for $\mathbf{x}'$ and $\mathbf{b}'$ from equations

(2.26) and (2.27), one obtains

$$\mathbf{V}^T\mathbf{x}=\Lambda^{-1}\mathbf{U}^T\mathbf{b}$$

which is equivalent to

$$\mathbf{x}=\mathbf{V}\Lambda^{-1}\mathbf{U}^T\mathbf{b} \qquad (2.32)$$

since **V** is orthogonal. The matrix $\mathbf{V}\Lambda^{-1}\mathbf{U}^T$ is the Moore-Penrose inverse. For shorthand we write it as

$$\mathbf{A}^\dagger=\mathbf{V}\Lambda^{-1}\mathbf{U}^T \qquad (2.33)$$

so equation (2.32) is equivalent to

$$\mathbf{x}=\mathbf{A}^\dagger\mathbf{b} \qquad (2.34)$$

The Moore-Penrose inverse is fundamental in the theory of generalized inverses. It is fundamental because of its uniqueness, since every matrix $\mathbf{A}\in\mathbb{R}^{m\times n}$ has one and only one $\mathbf{A}^\dagger$ [Penrose, 1955]. Its uniqueness arises out of a dual minimization criterion that is an alternate way of defining $\mathbf{A}^\dagger$. That is,

(1) The solution (2.34) is a least squares solution (defined by equation (2.10)) of equation (2.1) [Ben-Israel and Greville, 1974, p. 104].

(2) When $rank\,(\mathbf{A})<n$, there are infinitely many least squares solution of (2.1). Of all possible least squares solution of (2.1), the solution (2.34) is the unique one that also minimizes $\|\mathbf{x}\|_2^2$ [Penrose, 1955].

The second is achieved automatically when **A** is full rank since the least squares solution is then itself unique. Both minimizations have a geometric interpretation. To see this, I appeal again to the specific problem of equation (2.19), whose singular value decomposition was given as equation (2.23). Applying equation (2.33), the Moore-Penrose inverse of the matrix (2.19) is

$$\mathbf{A^t} = \begin{bmatrix} -.447 & .894 \\ .894 & .447 \end{bmatrix} \begin{bmatrix} \dfrac{1}{11.4} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -.196 & -.785 & -.588 \\ -.785 & .486 & -.386 \\ -.588 & -.386 & .711 \end{bmatrix} \qquad (2.35)$$

$$= \begin{bmatrix} .0077 & .0308 & .0231 \\ -.0153 & -.0616 & -.0462 \end{bmatrix}$$

which yields a solution by (2.34) of

$$x_1 = -0.0077 \qquad (2.36)$$

$$x_2 = 0.015$$

A geometric interpretation of this solution is shown in figures 2.9 and 2.10. The important observations to make are the following:

(1) The solution (2.36) is the orthogonal projection of the data vector $\mathbf{b}$ onto $R(\mathbf{A})$.

(2) The solution vector lies totally in $R(\mathbf{A}^T)$.

Observation (1) is a statement that the solution is a least squares solution. There are infinitely many least squares solutions to this particular problem. This is true because there are infinitely many combinations of the columns of this matrix that can produce the same vector in the data space since the columns vectors are collinear. (2) makes the solution unique because the solution (2.36) is the only one with no component in $N(\mathbf{A})$.

4.2.2. Practical least squares generalized inverses

The Moore-Penrose inverse introduced in the previous section occupies a key position in the theory of generalized inverses. It occupies this position because of its many remarkable theoretical properties (see Ben-Israel and Greville, 1974). However, as a practical method of solving systems of overdetermined equations, it is often both difficult to determine and frequently highly undesirable.
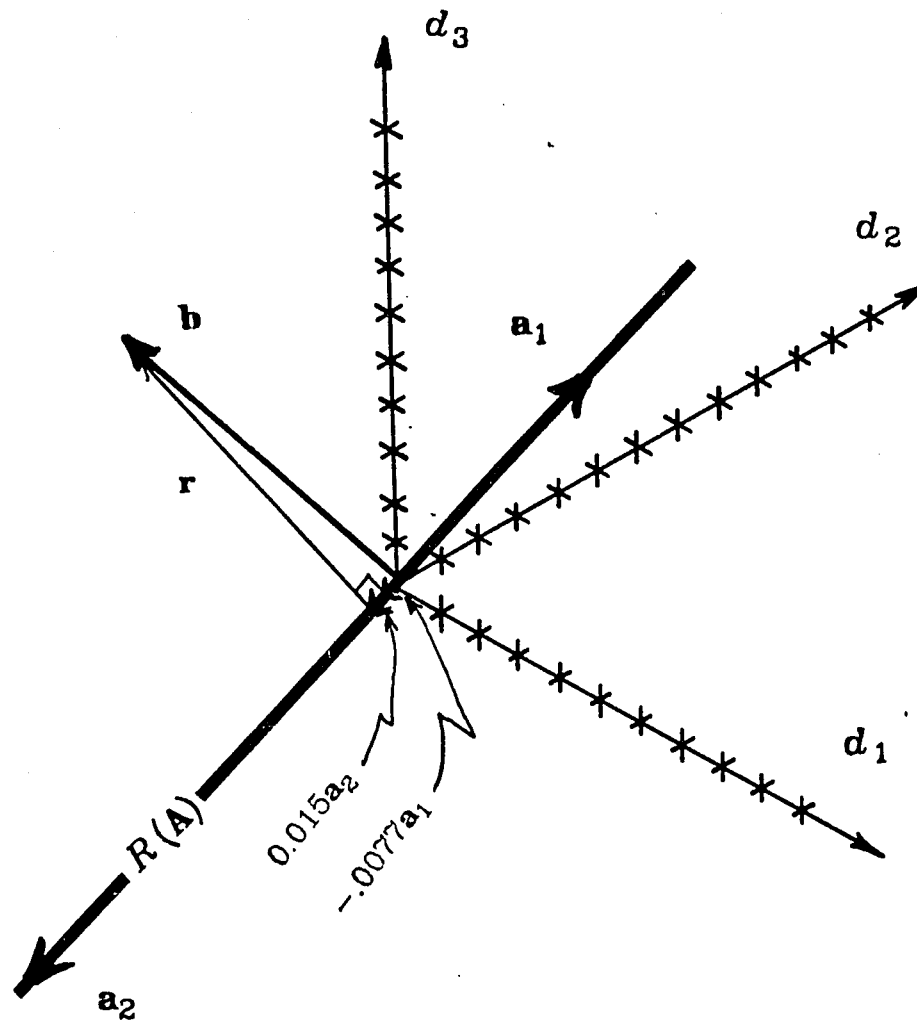
Figure 2.9. Geometry of Moore-Penrose inverse solution for non full rank example problem. Solution is again the orthogonal projection of the data vector, **b**, onto R (**A**).
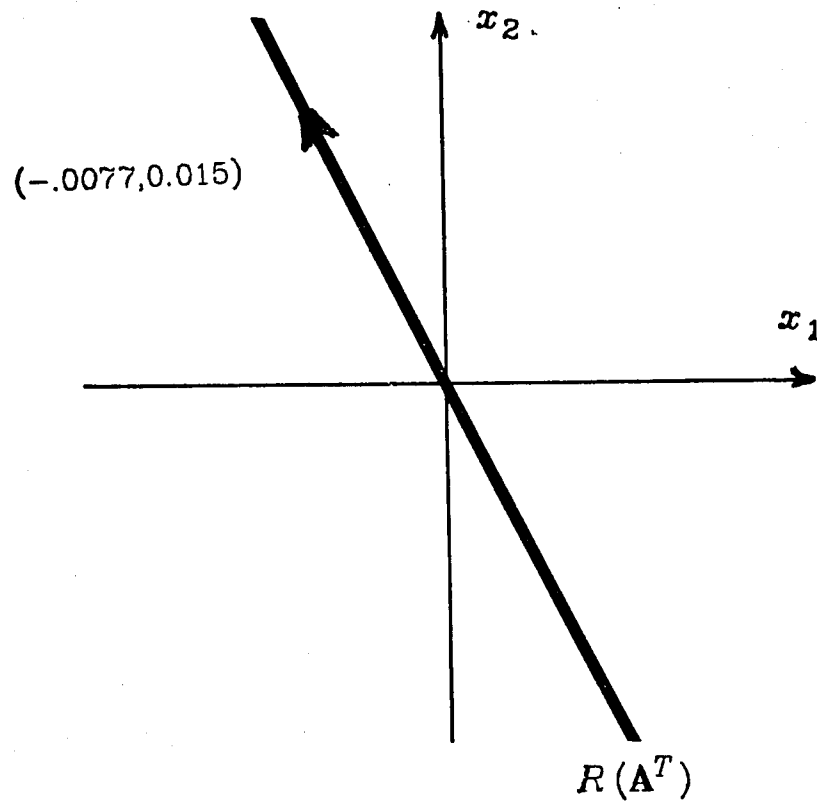
Figure 2.10. Geometry of Moore-Penrose inverse solution for non full rank example problem in parameter space. This solution is the unique least squares solution that minimizes $\|\mathbf{x}\|^2 = \sqrt{x_1^2 + x_2^2}$.

The reason for this is the practical problem of how to handle small singular values. That is, because matrices are always manipulated on a computing machine that has a finite precision, the singular values one is actually able to calculate are subject to rounding errors. Because of that, truly zero singular values almost never emerge from a singular value calculation. It is possible to determine when singular values are negligible to machine precision (see Lawson and Hanson, 1974, p. 95) permitting a best approximation to A^t. Unfortunately, however, that solution is not always desirable because of the error magnification produced by small, but significant, singular values. This problem is apparent from equation (2.29). That is, given some fixed precision of the data values, b'_i, on the right hand side, the errors in the estimates, x'_i, increases inversely with Λ_i. Hence small Λ_i can produce unacceptably large errors in the estimates x'_i.

.The problem of how to handle small, but significant, singular values is a difficult one that has not truly been solved [Dongarra et. al., 1979]. In the work presented here I have utilized what are probably the two most commonly used methods for analyzing practical least squares problems. These are the **pseudoinverse** and the **Levenberg-Marquardt** (also called **stochastic**) inverse. The pseudoinverse is almost identical to the Moore-Penrose inverse. The only significant difference is the manner by which the rank, r, of the system is established. In the previous section I defined A^t under the tacit assumption that r was a well defined number. As I noted above, however, this is not always the case in practice. The pseudoinverse is defined in exactly the same way as A^t in equation (2.33) but with the rank set as a number k that I will call the **pseudorank**. The pseudorank is defined by Lawson and Hanson as follows:

"We define the pseudorank k of a matrix $\mathbf{A}$ to be the rank of the matrix $\widetilde{\mathbf{A}}$ that replaces $\mathbf{A}$ as a result of a specific computational algorithm. Note that pseudorank is not a unique property of the matrix $\mathbf{A}$ but also depends on other factors such as the details of the computational algorithm, the value of tolerance parameters used in the computation, and the effects of machine round-off errors." [Lawson and Hanson, 1974, pp. 77-78]

When $k=r$ the pseudoinverse is identical to the Moore-Penrose inverse. Otherwise they are not the same, with the difference dependent upon how different k and r are. Thus, a good working definition of the pseudoinverse, $\mathbf{A}^+$, is

$$\mathbf{A}^+=(\widetilde{\mathbf{A}})^\dagger=\mathbf{V}\Lambda^+\mathbf{U}^T \tag{2.37}$$

where $\widetilde{\mathbf{A}}$ is the matrix defined in Lawson and Hanson's definition given above and Λ^+ is the same as Λ^{-1} defined by (2.31) with $r=k$.

The Levenberg-Marquardt inverse is usually defined as [Crosson, 1976a]

$$\mathbf{A}^m=(\mathbf{A}^T\mathbf{A}+\theta^2\mathbf{I})\mathbf{A}^T \tag{2.38}$$

where θ is an adjustable constant usually called a damping parameter. $\mathbf{A}^m$ has a simple relation to the singular value decomposition of $\mathbf{A}$. This leads to the alternate form of $\mathbf{A}^m$

$$\mathbf{A}^m=\mathbf{V}\Lambda^m\mathbf{U}^T \tag{2.39}$$

[Lawson and Hanson, 1974, pp. 188-194] where $\mathbf{V}$ and $\mathbf{U}$ are the orthogonal matrices of (2.20) that result from a singular value decomposition of the matrix $\mathbf{A}$ and

$$\mathbf{\Lambda}^m = \begin{bmatrix} \dfrac{\Lambda_1}{\Lambda_1^2 + \theta^2} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \dfrac{\Lambda_2}{\Lambda_2^2 + \theta^2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \dfrac{\Lambda_r}{\Lambda_r^2 + \theta^2} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \qquad (2.40)$$

$\mathbf{A}^m$ handles zero (or very small) singular values the same way $\mathbf{A}^+$ does implicitly, since as $\Lambda \to 0$, $\dfrac{\Lambda}{\Lambda^2 + \theta^2} \to 0$ for any finite θ.

Although their forms are completely different $\mathbf{A}^+$ and $\mathbf{A}^m$ have a fundamental property in common. They both find a solution that lies in the subspace $R(\mathbf{A})$ of $\mathbb{R}^m$ as any proper solution should. This fact is clear from the definitions of $\mathbf{A}^+$ and $\mathbf{A}^m$ given by equations (2.37) and (2.39). These equations show that both $\mathbf{A}^+$ and $\mathbf{A}^m$ produce a solution derived from linear combinations of the columns of the matrix $\mathbf{U}_R$ (defined in (2.21)) that form an orthonormal basis for $R(\mathbf{A})$. This is an important observation because that is a necessary condition for any generalized inverse to be used as part of the separation technique that is the basis of most of the work presented in this dissertation (see Appendix B).

5. SUMMARY

Given the amount of material reviewed in this chapter, it seems appropriate to emphasize the most important points the reader should glean from it. That is, a least squares solution of a set of m linear equations with n unknowns $(m > n)$ is the solution that is the orthogonal projection of the data vector, $\mathbf{b}$, onto the range of the matrix $\mathbf{A}$. This solution effectively splits the data vector into two

orthogonal components; the solution vector, $\mathbf{AA^{\dagger}b}$, and the residual vector, $\mathbf{r}$ defined by equation (2.12). Both vectors, in general, have m nonzero components. However, the coordinate reorientation defined by the matrix $\mathbf{Q}$ of a QR decomposition (theorem 2.1) or the matrix $\mathbf{U}$ of the singular value decomposition (theorem 2.2) show that $\mathbf{AA^{\dagger}b}$ actually has at most n nonzero components and $\mathbf{r}$ has at least $m-n$ nonzero components. This result is not surprising since overdetermined equations clearly contain redundant information. Only n pieces of information are extracted in a least squares solution. The $m-n$ pieces of information that are left over are the residuals.

CHAPTER 3
HYPOCENTER LOCATION

1. INTRODUCTION

Location of earthquake hypocenters is one of the most basic data analysis problems in earthquake seismology. Current location algorithms can be grouped into two major categories that I refer to here as single and multiple event location procedures. As the name suggests, single event location methods operate on data from one earthquake at a time oblivious to the data from other earthquakes. Multiple event locations, on the other hand, consider data from many events simultaneously. The first major section of this chapter is a review of the basic principles of single event location. I have presented this material primarily to place multiple event location procedures in their proper perspective. That is, all multiple event locations methods that are currently in use are only a simple extension of single event location procedures. I consider in some detail two particular multiple event location These are the master event method [Everden, 1969] and the method of joint hypocenter determination (JHD) [Douglas, 1967]. This helps set the stage for the introduction of a new multiple event location that I have utilized in this study. This new location scheme, which I call progressive multiple event location, is a descendent of a similar procedure recently proposed by Jordan and Sverdrup [1981] and the recent work of Pavlis and Booker [1980]. The most important feature of this new procedure is its remarkably efficient usage of computer storage. Unlike JHD the storage requirements of this procedure do not grow with the number of events in the data set but instead the algorithm operates within a small, fixed storage area. In the actual work that I am presenting here I have only used this algorithm as an integral part of the larger scale algorithm that I will discuss in the two succeeding chapters. Nonetheless, I believe it may well find a much wider application because of its efficient usage of computer storage.

2. SINGLE EVENT LOCATION

2.1. Geiger's Method

The basic algorithm of all conventional single event earthquake location procedures is the iterative least squares method originally proposed by Geiger [1910]. The algorithm he proposed was not widely applied at the time because the calculations involved were too tedious to perform by hand. Consequently, the idea was almost forgotten but was quickly rediscovered and extended by a number of workers (Bolt[1960],Flinn[1960], Nordquist [1962], and Herrin et. al. [1962]) in the early 1960's after the introduction of high speed digital computers that could be programmed to do the necessary calculations. I will briefly review the fundamentals of Geiger's procedure here because it is essential that the reader understand the principles of the algorithm. This is so because this nonlinear least squares algorithm, with various modern extensions (for example, Buland [1976] and Klein [1978]), is the most widely used method of earthquake location. Furthermore, Geiger's algorithm occupies a key position in the bigger scale algorithm I have used here. Those readers intimately familiar with the principles of Geiger's procedure may safely skip this section. Those wishing a more extensive treatise are referred to Lee and Stewart [1981].

The input data for earthquake location are the arrival times of different phases recorded by each seismograph in the earthquake network. In principle, any arbitrary phase could be used. However, with regional or local networks, to which this study is restricted, the data is always restricted to the arrival times of the direct compressional wave and the arrival time of the direct shear wave.

Geiger's procedure is based on two primary assumptions:

(1) A complete set of travel time curves are available.

(2) The travel time curves are known exactly and the only source of error is statistical error from measuring arrival times.

Assumption (1) can be restated in mathematical symbolism as follows: we assume that given a point source at a position $\mathbf{h}'$, we can calculate the arrival time of the disturbance at any position (h_1, h_2, h_3) in the medium as[1]

$$t = h_4' + T(h_1, h_2, h_3, h_1', h_2', h_3') \tag{3.1}$$

h_4' is the "origin time", which is the time the source began radiating elastic wave energy. The function T is the "travel time table" referred to in assumption (1) above.

The goal of hypocenter location is to estimate the four numbers h_1', h_2', h_3', and h_4' from a set of m observed arrival times $t_i, i = 1, 2, \cdots, m$. The crux of the problem is that although the origin time, h_4', is linearly related to the observed arrival times, the

---

[1] With small scale earthquake networks the coordinate system generally used is cartesian with

h_1 = positive northward

h_2 = positive eastward

h_3 = depth

h_4 = time

In teleseismic locations spherical coordinates are more conventional so that in that case we would have

h_1 = latitude or colatitude

h_2 = longitude

h_3 = depth

h_4 = time

I will use the h_i symbolism here for greater generality. The choice of the letter h was made to emphasize that the vector $\mathbf{h}'$ (the "hypocenter") is a four vector.

function T is, in general, quite complicated and the observations, t_i, are nonlinear functions of the spatial coordinates of the source. Earthquake location is thus a problem in nonlinear optimization [Lee and Stewart,1981]. Geiger's method solves the optimization problem by a first order expansion of equation (3.1). That is, we expand (3.1) in a Taylor series about some initial estimate of the hypocenter for each of the observed arrivals. This yields m equations of the form

$$t_i = h_4' T(h_{1_i}, h_{2_i}, h_{3_i}, h_1', h_2', h_3') \qquad (3.2)$$

$$+ \left. \frac{\partial T}{\partial h_1'} \right|_{\mathbf{h}_i, \mathbf{h}'} \delta h_1' + \left. \frac{\partial T}{\partial h_2'} \right|_{\mathbf{h}_i, \mathbf{h}'} \delta h_2' + \left. \frac{\partial T}{\partial h_3'} \right|_{\mathbf{h}_i, \mathbf{h}'} \delta h_3' + \left. \frac{\partial T}{\partial h_4'} \right|_{\mathbf{h}_i, \mathbf{h}'} \delta h_4'$$

$$+\Omega(2) \qquad i = 1, 2, \cdots, m$$

where $\Omega(2)$ is the collection of all higher order terms. All the partial derivatives are evaluated at some current estimate of the true, unknown hypocenter vector, $\mathbf{h}'$ and $T(h_{1_i}, h_{2_i}, h_{3_i}, h_1', h_2', h_3')$ is the calculated travel time based on a hypocenter at (h'_1, h'_2, h'_3) to the recording station at position $(h_{i_1}, h_{i_2}, h_{i_3})$. I define the residual in the usual way as

$$r_i = t_i - T(h_{1_i}, h_{2_i}, h_{3_i}, h_1', h_2', h_3') - h_4' \qquad (3.3)$$

That is, r_i is the observed arrival time, t_i , minus the calculated arrival time based on the current estimate of the hypocenter. Then, by rearranging and dropping higher order terms, equation (3.2) becomes

$$r_i = \left. \frac{\partial T}{\partial h_1'} \right|_{\mathbf{h}_i, \mathbf{h}'} \delta h_1' + \left. \frac{\partial T}{\partial h_2'} \right|_{\mathbf{h}_i, \mathbf{h}'} \delta h_2' + \left. \frac{\partial T}{\partial h_3'} \right|_{\mathbf{h}_i, \mathbf{h}'} \delta h_3' + \delta h_4' \quad (3.4)$$

$$i = 1, 2, \cdots, m$$

since $\dfrac{\partial T}{\partial h_4'} = 1$. These m equations can be written in matrix form as

$$r = A\delta h \qquad (3.5)$$

where I have dropped the prime on δh for convenience and the coefficients of the matrix A are defined by $A_{ij} = \left.\dfrac{\partial T}{\partial h'_j}\right|_{h_i, h'}$ Geiger's procedure assumes $m \geq 4$ and solves (3.5) by least squares. That is, it finds the estimate of δh, which I write as $\widehat{\delta h}$, that minimizes the L2 norm (see equation (2.11)). of the residual vector, r_i. That is

$$\widehat{\delta h} = \min_{\delta h \in R^4}\left[\|r - A\delta h\|_2^2\right] \qquad (3.6)$$

where R^4 denotes the vector space of all possible 4 vectors. The solution of the minimization problem (3.6) yields the classic normal equations

$$\widehat{\delta h} = (A^T A)^{-1} A^T r \qquad (3.7)$$

A proof that (3.7) does indeed solve the minimization problem (3.6) can be had in almost any text on linear algebra or basic data analysis (see for example, Stewart [1973],p. 220). Note that the residuals defined by (3.3) are nonlinear functions of the hypocenter coordinates. The perturbation, $\widehat{\delta h}$, estimated by (3.7) is, however, based on a linearization (3.3) that neglects the higher order terms. Consequently, the estimate obtained by (3.7) of the hypocenter, $h' + \widehat{\delta h}$, will usually be inadequate and the hypocenter estimate will need to be updated iteratively until it is "good enough". Geiger's procedure uses the following iterative algorithm[2]:

---

[2] This outline of Geiger's algorithm is written in an informal language similar to that described by Stewart [1973,pp. 82-93]. This algorithm and a number of others are described here using this informal language. The intent is to show the skeleton of the algorithm unobscured by minor details required in a formal language such as FORTRAN, but without being so terse as to be incomprehensible.

ALGORITHM GEIGER

1) $\mathbf{h}'_j = \mathbf{h}'_{initial}$

2) Loop from 3) to 6) until convergence

    3) For $i = 1, 2, \cdots, m$

        1i) $r_i = t_i - T(h_{1_i}, h_{2_i}, h_{3_i}, h_1', h_2', h_3') - h'_4$

        2i) $A_{ij} = \dfrac{\partial T}{\partial h'_j}$ $(j = 1, 2, 3, 4)$

    4) $\delta \widehat{\mathbf{h}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{r}$

    5) $\mathbf{h}' = \mathbf{h}' + \delta \widehat{\mathbf{h}}$

6) End loop

Note that the matrix of partial derivatives, **A**, and the residual vector, **r** , defined by equation (3.5) are updated in this algorithm with each iteration. $\delta \mathbf{h}$ is always estimated by linear steps based on the current best estimate of the hypocenter. This is the classic Gauss-Newton method of nonlinear optimization [Lee and Stewart, 1981].

2.2. Modern Improvements

ALGORITHM GEIGER described in the previous section has three major flaws:

(1) It does not recognize that all the data used to estimate the hypocenter may not be of equal quality.

(2) $\delta \widehat{\mathbf{h}}$ calculated by equation (3.7) will frequently fail because the matrix $\mathbf{A}^T \mathbf{A}$ is singular and cannot be inverted by elementary methods. [Lee and Stewart, 1981]

(3) The assumption of Geiger's procedure that the travel time curves are perfect is never totally justified.

The first problem is solved readily by giving each of the m equations of (3.4) a different weight. That is, each of the m equations is multiplied by some constant, w_i , to yield m new equations

$$(r_w)_i \equiv w_i r_i \tag{3.8}$$

$$= w_i \left[\left. \frac{\partial T}{\partial h_1'} \right|_{\mathbf{h_i},\mathbf{h'}} \delta h_1' + \left. \frac{\partial T}{\partial h_2'} \right|_{\mathbf{h_i},\mathbf{h'}} \delta h_2' + \left. \frac{\partial T}{\partial h_3'} \right|_{\mathbf{h_i},\mathbf{h'}} \delta h_3' + \delta h_4' \right]$$

This process is conveniently illustrated by defining a diagonal matrix $\mathbf{W}$ as

$$W_{ij} = w_i \delta_{ij} \tag{3.9}$$

where δ_{ij} is the Kronecker delta. With definition (3.9) the m equations of (3.8) are easily seen to be equivalent to the following matrix equation

$$\mathbf{Wr} = \mathbf{WA} \delta \mathbf{h} \tag{3.10}$$

or

$$\mathbf{r}_w = \mathbf{A}_w \delta \mathbf{h} \tag{3.11}$$

where $\mathbf{r}_w = \mathbf{Wr}$ and $\mathbf{A}_w = \mathbf{WA}$. It is shown in appendix A that weighting of this form does not change the nature of this problem fundamentally. The matrix $\mathbf{W}$ can be considered as a change of variables. This change of variables changes the norm used to measure the size of the residual vector from the simple L2 norm to the ellipsoidal or weighted Euclidean norm defined as [Ben-Israel and Greville, 1974,p. 128]

$$\|\mathbf{x}\|_w = \left[(\mathbf{Wx})^T (\mathbf{Wx}) \right]^{1/2} \tag{3.12}$$

$$= \left[\mathbf{x}^T (\mathbf{W}^T \mathbf{W}) \mathbf{x} \right]^{1/2}$$

$$= \left[\sum_{i=1}^{n} x_i^2 w_i^2 \right]^{1/2}$$

since for this case $\mathbf{W}$ is always diagonal. `It is shown in appendix A that when equation (3.11) is solved as a least squares problem it is equivalent to the following solution

$$\delta \widehat{\mathbf{h}}_w = \min_{\delta \mathbf{h} \in \mathbf{R}^4} \left[\| \mathbf{r} - \mathbf{A} \delta \mathbf{h} \|_w^2 \right] \tag{3.13}$$

Hence, a least squares solution of (3.12) solves the different minimization problem defined by (3.13). For the remainder of this work I will drop the w subscripts on $\mathbf{r}$ and $\mathbf{A}$ for convenience. Weighting will henceforth be assumed to be inherently present. The simple least squares solution (no weighting) is just the special case of $\mathbf{W}=\mathbf{I}$.

A major variation between competing earthquake location algorithms is how that algorithm calculates the weights, w_i. A great variety of method exist for calculating these weights in some prescribed optimum sense (See Anderson [1978] for a good review of different weighting schemes.). However, all weighting currently used can be summarized as the product of three terms that I write as

$$w_i = QP(r_i)D(d_i) \tag{3.14}$$

These three terms carry the following conventional names:

Q quality weighting [Lee and Lahr, 1972] [Klein, 1978]

$P(r_i)$ penalty function [Anderson, 1978] or residual weighting [Klein, 1978].

$D(d_i)$ distance weighting [Klein, 1978]

Q is a weight assigned, as the name suggest, according to how clear the arrival of the particular phase being observed is. Q is often assigned purely subjectively. A superior method of setting Q is to base it on an estimate of the errors in measuring the arrival time. (This scheme is used, for example, in the hypocenter location

program LQUAKE used for routine earthquake locations in western Washington [Rogers and Crosson, 1976].) In this case Q is defined as

$$Q = \frac{1}{\sigma_i} \qquad (3.15)$$

where σ_i is the estimate of the expected standard error in picking the ith arrival[3].

$P(r_i)$ is a weighting function commonly used for suppressing the effect of outliers in the data. A variety of $P(r_i)$ functions are possible [Anderson, 1978]. They are all functions of the size of the residuals designed to strongly downweight arrivals that show large residuals.

$D(d_i)$ is a function of the estimated distance, d_i, from the source to the ith seismometer. $D(d_i)$ is always a decreasing function of distance so that the effect of $D(d_i)$ is to downweight the observations of more distant stations. This is used to suppress errors caused by two practical problems:

(1) Errors in picking arrival times increase with distance from the source because the amplitudes of the arrivals decrease with distance because of geometric spreading and attenuation.

(2) In practice the travel time curves are not perfect as Geiger's procedure assumes. This error tends to increase with distance.

The second major problem with Geiger's procedure is that estimation of δx by equation (2.8) can often fail because the matrix $\mathbf{A}^T\mathbf{A}$ is numerically singular. Table 3.1 lists four numerical methods that

---

[3] Errors in picking P arrivals are statistically independent and approximately normally distributed [Buland, 1976]. If estimates of the expected standard error, σ_i , are the same as the true standard error, and if the travel time tables are perfect, then applying Geiger's method using weights based on equation (3.15) alone will yield a maximum likelihood estimate of the true hypocenter [Jordan and Sverdrup, 1981].

have been applied extensively to stabilize location algorithms when the system of equations to be solved is ill conditioned.

Table 3.1.

Hypocenter Inversion Algorithms	
Author(s)	Numerical Algorithm
Lee and Lahr [1972]	Step-wise multiple regression (variation of Gaussian elimination with pivoting) on normal equations
Bolt [1970] and Klein [1978]	pseudoinverse calculated by singular value decomposition
Rogers and Crosson [1976] and Crosson [1976a]	Levenberg-Marquardt damped least squares
Buland [1976]	QR decomposition with Curry-Altman step length damping

The details of these various algorithms can be found by consulting the references listed in Table (3.1). For the present, it is more important to realize what they all have in common. They are all practical least squares solution algorithms, where by "practical" I mean they will not fail dramatically when **A** is ill conditioned. All hypocenter location algorithms currently in use form an estimate of the hypocenter perturbation as

$$\delta \hat{\mathbf{h}} = \mathbf{H} \mathbf{A} \delta \mathbf{h} = \mathbf{H} \mathbf{r} \qquad (3.16)$$

where **H** is a $4 \times m$ matrix that is a generalized inverse [Lee and

Stewart, 1981] of **A** . All of the procedures listed in Table 3.1, however, are a particular type of generalized inverse. A compact way of showing this relationship is through the singular value decomposition introduced in section 4.1 of chapter 2. Theorem 2.2 guarantees that we can write **A** as

$$\mathbf{A}=\mathbf{U}\Lambda\mathbf{V}^T \qquad (3.17)$$

The generalized inverse for all the procedures listed in Table 3.1 can be written as

$$\mathbf{H}=\mathbf{V}\Lambda^H\mathbf{U}^T \qquad (3.18)$$

where Λ^H is a matrix of the form

$$\Lambda^H=\left[\Lambda_R^H 0\right] \qquad (3.19)$$

where $\Lambda^H \in \mathbb{R}^{4\times4}$ is different for each algorithm. Of special interest for the purpose of this work is the Levenberg-Marquardt inverse. (The Levenberg-Marquardt inverse is important to this work because I have used the hypocenter location program LQUAKE written by Crosson. LQUAKE uses the Levenberg-Marquardt inverse.) The Levenberg-Marquardt inverse can be written in the form of equation (3.18) with $\Lambda_R^H=\Lambda^m$ where Λ^m is as defined in equation (2.40). Solutions of the form (3.16) are not always, strictly speaking, least squares solutions but are practical approximations to the least squares solution that do not fail when **A** is ill conditioned. The crucial observation for the purpose of this work is that all generalized inverses currently used for hypocenter location can be written in the form (3.18). In chapter 2, I pointed out that the matrix **U** in (3.18) is just a rigid reorientation of the coordinates of the space the residual vector lies in. However, it reorients the space such that the first four components of the transformed residual vector, $\mathbf{U}^T\mathbf{r}$, lie in the range of **A** , $R(\mathbf{A})$, and the other $m-4$ components lie perpendicular to it in $N(\mathbf{A}^T)$. From the form of (3.18) it is clear that

only the first four components of the vector $\mathbf{U}^T \mathbf{r}$ are used to estimate $\delta \mathbf{h}$. The other $m - 4$ components lie in directions orthogonal to $R(\mathbf{A})$, and are not used in estimating $\delta \mathbf{h}$.

The third major problem with Geiger's original procedure is the assumption that travel times can be calculated with negligible error. Travel times can only be calculated precisely if a good P and S wave velocity model is known. There is strong evidence that the seismic velocity structure of the earth varies significantly in all three spatial dimensions [Jordan, 1979]. For routine earthquake locations, however, the assumption is nearly always made (Engdahl and Lee [1976], Thurber and Ellsworth [1980], and Thurber [1981] are some exceptions.) that the seismic velocity varies only in the vertical direction. This is a practical simplification that is required because of the computational problems involved in calculating travel times in a laterally inhomogeneous medium. Seismologists have long recognized this problem, and have traditionally attempted to remedy the problem by the simple, yet practical approach of station corrections. Stations corrections are a set of fixed constants associated with the location of each seismometer in the network that are added to the calculated travel time for all arrivals recorded at that station. This leads to a redefinition of the residual as[4]

---

[4] The matrix $\mathbf{S}$ in equation (3.20) is used primarily to provide a connection between this section and the discussion on multiple event location that follows. In any real single event location algorithm it is never explicitly formed since its only purpose is indexing. That is, the only role of $\mathbf{S}$ in (3.20) is to associate the kth station correction with the ith arrival. In single event location this process is always accomplished by a linear search method. In multiple event location, however, we shall see that $\mathbf{S}$ must be formed and manipulated explicitly.

$$r_i^s = t_i - h_4' - T(h_{1_i}, h_{2_i}, h_{3_i}, h_1', h_2', h_3') - \sum_{k=1}^{n_s} S_{ki} s_k \qquad (3.20)$$

$$= t_i - t_i' - \sum_{k=1}^{n_s} S_{ki} s_k$$

where

n_s = total number of stations in the network

s_k = station correction assigned the kth seismometer

$$S_{ik} = \begin{cases} 1 & \text{when the } i^{th} \text{ arrival is observed at the } k^{th} \text{ seismometer} \\ 0 & \text{otherwise} \end{cases} \qquad (3.21)$$

$$t_i' = T(h_{1_i}, h_{2_i}, h_{3_i}, h_1', h_2', h_3') + h_4'$$

= calculated arrival time table with hypocenter at **h′**

With this change the m equations of (3.5) can be rewritten as

$$\mathbf{r^s = t - t' - Ss = A\delta h} \qquad (3.22)$$

Station corrections are a practical approximation to the effects of lateral inhomogeneity. If $T(h_{1_i}, h_{2_i}, h_{3_i}, h_1', h_2', h_3')$ is calculated using a one dimensional velocity model, then station corrections are the simplest possible way to account for the effects of a laterally varying velocity structure. They will give a good estimate of the actual travel time only if the velocity anomaly they reflect is localized near the station they are associated with [Crosson, 1976b]. In that case, they are similar to static corrections used in reflection seismology (see Dobrin [1976, pp. 211-222]).

The modern improvements to ALGORITHM GEIGER can be summarized in the following modified algorithm sketch I call ALGORITHM MODHYLOC.

ALGORITHM MODHYLOC

1) $\mathbf{h}'=\mathbf{h}'_{initial}$

2) Loop from 3) to 7) until convergence

    3) For $i=1,2,\cdots,m$

        1i) $r_i=t_i-T(h_{1_i},h_{2_i},h_{3_i},h_1',h_2',h_3')-h_4'$

        2i) Scan station table for station correction, s_i

        3i) $r_i=r_i-s_i$

        4i) $w_i=QP(r_i)D(d_i)$

        5i) $r_i=r_iw_i$

        6i) $A_{ij}=w_i\dfrac{\partial T}{\partial h'_j}$ $(j=1,2,3,4)$

    4) Calculate generalized inverse, $\mathbf{H}$

    5) $\delta\widehat{\mathbf{h}}=\mathbf{Hr}$

    6) $\mathbf{h}'=\mathbf{h}'+\delta\widehat{\mathbf{h}}$

7) End Loop

This algorithm differs significantly from ALGORITHM GEIGER in its details. I reemphasize, however, that the two algorithms are fundamentally related. The generalized inverse, $\mathbf{H}$, in ALGORITHM MODHYLOC is always calculated based on some **least squares** criterion.

3. MULTIPLE EVENT LOCATION

3.1. Previous Work

3.1.1. Introduction

In the previous section I reviewed the concept of station corrections and described how they are used in routine earthquake locations. A question naturally arises, however. How does one estimate what they are? Several methods have been used to accomplish this task. In this section, however, I will review the two principle methods that are currently employed for this purpose; the **master event location method** [Everden, 1969] and the method of **joint hypocenter determination** (JHD) developed by Douglas [1967] and Dewey [1972]. Other methods have been successfully employed for estimating station corrections, but I restrict the discussion here to the master event and JHD techniques for two reasons:

(1) They are probably the two most commonly used methods.

(2) Both are intimately connected to the progressive multiple event location method that I introduce in the section that follow this one.

3.1.2. Master Event Location

The master event location technique [Everden, 1969] is, without doubt, the simplest method for estimating station corrections. In this approach, the residuals from some "master event" (usually and explosion or a well recorded earthquake) are equivalenced to the station corrections. These station corrections are then used to relocate earthquakes by conventional single event methods. The master event method has proved successful in teleseismic locations (e.g. Everden [1969]) and in local network studies of clusters of events in mainshock-aftershock sequences (e.g. Yelin and Crosson [1982]). The methods major advantage is its simplicity. It suffers,

however, from two major problems.

(1) Normally, one requires that the "master event" be recorded at every station that will be used. This is not always possible.

(2) The station corrections derived from the "master event" have a relatively large uncertainty since they are based on only a single data point for each station.

Finally, it should be pointed out that the master event technique has the distinction of being being both a single event location method and a multiple event location method at the same time. That is, it is a multiple event location method because its major goal is to provide precise relative locations of a set of earthquakes. It is also basically a single event location procedure, however, since the basic algorithm it uses is the same as conventional single event location schemes. The point is that all modern single event location procedures are, in a sense, "master event" methods since the use of station corrections is nearly universal. The difference is that the station corrections are not always based on data from only a single event.

3.1.3. Joint Hypocenter Determination

Probably, the best method currently available for estimating station corrections is the method of joint hypocenter determination (JHD) developed by Douglas [1967] and Dewey [1972] and applied by many authors, primarily to teleseismic locations. JHD is a simple extension of the single event location methods described in the previous two sections. However, we will see it shares all the problems involved in single event location described in section 1.2 above plus some others.

If we wish to estimate the n_s station corrections from arrival time data, it is immediately obvious from equation (3.19) that first arrivals from a single event are insufficient. n_s is the total number of stations in the seismic network, which is the maximum number

of equations possible in (3.19) from first arrival data alone. Consequently, we are immediately led to combine the data from more than one event. Suppose, we have available a set of m_e events. If we define m_i as the number of arrivals recorded for the ith earthquake, then the total number of arrivals for the entire data set is

$$M = \sum_{i=1}^{m_e} m_i \qquad (3.23)$$

The multiple event location problem is to use this data to estimate the hypocenters of the m_e events ($4m_e$ parameters) and the n_s parameters, s_i, that are the station corrections. An argument identical to that in section 1.1 that led to equation (3.3) leads here to

$$(r_i^s)_j = (t_i)_j - T(h_{1_i}, h_{2_i}, h_{3_i}, h_1'_j, h_2'_j, h_3'_j) - \sum_{k=1}^{n_s} S_{ik} s_k \qquad (3.24)$$

$$\approx \frac{\partial T}{\partial h'_{1_j}} \delta h'_{1_j} + \frac{\partial T}{\partial h'_{2_j}} \delta h'_{2_j} + \frac{\partial T}{\partial h'_{3_j}} \delta h'_{3_j} + \delta h'_{4_j} + \sum_{k=1}^{n_s} S_{lk} \delta s_k$$

$$i = 1,2,\ldots,m_i$$

$$j = 1,2,\ldots,m_e$$

$$l = 1,2,\ldots,n_s$$

where I define the following

$(h_{l_1}, h_{l_2}, h_{l_3}, h_{l_4})$ = spatial coordinates of lth station

$(h_{l_1}, h_{l_2}, h_{l_3}, h_{l_4}) = \mathbf{h}'_j$ = hypocenter of jth earthquake

δs_k = perturbation to kth station correction.

S_{ik} is defined by equation (3.18). The large number of indices in equation (3.24) obscure the special form of these equations. It is convenient to group the data from individual events. Doing this yields a set of m_e matrix equations

$$r_j^s = t_j - t'_j - S_j s = A_j \delta h_j + S_j \delta s \qquad (3.25)$$

where t'_j is as defined in (3.18) and A_j is the matrix of partial derivatives as in equation (3.5). Equation (3.25) is, in fact, the same as (3.19) but with an additional term involving the perturbation to the station corrections, $S_j \delta s$, and an added indexing label, j. When the m_e equations of (3.25) are combined the result is

$$
\begin{bmatrix} r_1^s \\ r_2^s \\ r_3^s \\ \cdot \\ \cdot \\ \cdot \\ r_{m_e}{}^s \end{bmatrix}
=
\begin{bmatrix}
A_1 & 0 & 0 & \ldots & 0 & S_1 \\
0 & A_2 & 0 & \ldots & 0 & S_2 \\
0 & 0 & A_3 & \ldots & 0 & S_3 \\
\cdot & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & & \cdot & \cdot \\
\cdot & \cdot & \cdot & & \cdot & \cdot \\
0 & 0 & 0 & \ldots & A_{m_e} & S_{m_e}
\end{bmatrix}
\begin{bmatrix} \delta h_1 \\ \delta h_2 \\ \cdot \\ \cdot \\ \cdot \\ \delta h_{m_e} \\ \delta s \end{bmatrix}
\qquad (3.26)
$$

For a convenient shorthand notation, I write the full set of equations of (3.26) as

$$r^s = Bx \qquad (3.27)$$

where $r^s \in R^M$, $B \in R^{M \times N}$, and $x \in R^N$. with N defined as

$$N = 4m_e + n_s \qquad (3.28)$$

The method of joint hypocenter determination solves (3.26) iteratively, exactly as in Geiger's method. That is, perturbations to the hypocenters of the m_e events and the perturbations to the station corrections are calculated by a least squares solution of (3.26). As in single event location, since (3.24) is based on a linearization

that neglects higher order terms, a sequence of least squares solutions of equations of the form of (3.26) are used in an iterative scheme to converge to a final solution. The basic scheme of JHD is identical to Geiger's method; the Gauss-Newton method. However, with JHD this means that the $4m_e$ hypocenter parameters and the n_s station corrections are all adjusted simultaneously.

In section 1.2 above I discussed three problems with Geiger's method for single event earthquake location and how modern location algorithms avoid them. Of these, weighting is usually used in JHD [Dewey, 1972] in the same way, and for the same purpose as in single event location. Inadequate knowledge of the travel time curves is ubiquitous and the very reason for doing JHD is to help reduce that problem. It is the singularity of the matrix equation (3.27), however, that is the most significant problem in implementing JHD. As it stands equation (3.27) is, in fact, **always** singular. Physically, this indicates a fundamental ambiguity in this problem. The average value of all the station corrections can be set to any arbitrary value, but the resulting effect can be completely compensated for by a corresponding change in the origin time of all the events in the data set. Douglas [1967] recognized this problem in his original paper proposing the use of the JHD method. He proposed a solution using one of the following constraints

(1) Fix the station correction for one reference station.

(2) Solve (3.27) subject to the constraint $\sum_{i=1}^{n_s} s_i = 0$.

This helps, but it is not always sufficient. In particular, when all the events are localized in a small region, the system of equations may be numerically singular [Douglas, 1967]. Physically, this happens because in this case the information contained in the data is highly redundant. Consider the limiting case where every event in the data set occurred at the same spatial position. These data are then equivalent to a single event recorded at every station but of varying

quality. Thus we have the equivalent of only n_s observations, but we want to estimate $n_s + 4m_e$ parameters. The JHD method avoids this problem by specifying one of the events in the data set (say event 1 for convenience) as the **master event** whose location is held fixed. Thus, in JHD, (3.26) is solved by least squares but subject to a second constraint that

$$\delta x_1 = 0$$

which is normally sufficient to permit a solution.

The principle advantage of JHD, as we have seen, is its simplicity since it is simply an extension of the conventional single event location techniques. However, it suffers from two major drawbacks.

(1) The system of equations (3.26) that has to be solved grows rapidly in size as more data is added.

(2) The requirement of a fixed master event is rather arbitrary [Jordan and Sverdrup, 1981].

I will now describe a new method of multiple event location that eliminates both of these problems.

3.2. Progressive Multiple Event Location

3.2.1. Introduction

The multiple event location scheme I will describe in this section is closely related to the simplest imaginable method of estimating station corrections. That is, we could estimate the station corrections as follows:

(1) Estimate the hypocenters of the m_e events by a conventional single event location algorithm using some initial estimate of the station corrections, s_i.

(2) Estimate a perturbation, δs_i, to each station correction as the average of all weighted residuals recorded at that station.

Optionally, one could then use the new station corrections, $s_i + \delta s_i$, and repeat (1) and (2) until the correction becomes small. This method is pleasing because of its simplicity, but it has a severe flaw. The hypocenters of all m_e events are not all equally well constrained. Because of that, a large weighted residual for one event may not be as significant as that from another event because the residuals can be biased by a bad location estimate. The simple scheme outlined above ignores this and gives each weighted residual equal weight. The new method of multiple event location described here is like the simple method outlined above in the sense that it is what Jordan [1980] has termed hierarchic, and Roecker [1982] calls progressive. Both terms are descriptive because unlike JHD, the unknown parameters are not all adjusted simultaneously by Gauss-Newton steps, but instead the hypocenters are adjusted independently from the station corrections. A hierarchy is implied in which greater emphasis is placed on estimating the hypocenters and only the dregs of the residuals are used to estimate the station corrections. The major difference between the procedure advocated here and the simpler algorithm discussed above is that instead of using the raw residuals, this procedure uses only those components of the residual vector that lie in the subspace $N(\mathbf{A}^T)$, since they are unbiased by possible mislocation errors. The procedure I describe here is a direct descendent of a method of multiple event location recently proposed by Jordan and Sverdrup [1981] and a new method of time term analysis for refraction data developed by Rohay [1982]. The procedure I propose here is very similar to that described by Jordan and Sverdrup [1981] with one major exception. Their study was focused on the analysis of the relative locations of small clusters of events recorded at teleseismic distances. This permitted them to solve the problem in a single

linear step as the nonlinearity of the partial derivatives does not affect the solution significantly in their case [Jordan and Sverdrup, 1981]. I have adopted this procedure with a primary interest in the analysis of data from local seismic networks. My experience with analysis of local network data indicates that nonlinear effects are significant and an iterative solution is generally required. I would argue that a "progressive" approach is a sensible way to proceed with this problem because the observed arrival times are linear functions of the station corrections but are nonlinear functions of the hypocenter coordinates. Hence it seems sensible to place greater emphasis on constraining the hypocenters, since only the spatial coordinates of the hypocenters are nonlinear functions of the data.

The approach I have followed here is to lead the reader through the manipulations involved in implementing this algorithm. At the same time I will demonstrate the key properties of this algorithm that make it work. After that I will sketch an overview of the whole algorithm and summarize its properties and relevance.

3.2.2. Single event residuals and the multiple event location problem

The method I am advocating here is a process of alternately locating individual earthquakes by conventional single event methods, followed by estimation of perturbations to station corrections using the resulting residuals. Consequently, it is necessary to see what happens to the overall system of equations of the multiple event location problem, (3.26), when each earthquake is located individually. To see this, it is useful to rewrite the set of equations (3.26) as

$$\mathbf{r}^s = \mathbf{A}\delta\mathbf{h} + \mathbf{S}\delta\mathbf{s} \qquad (3.29)$$

where $\mathbf{r}^s$ is as defined in (3.26)[5]

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{m_e} \end{bmatrix} \qquad (3.30)$$

$$\delta\mathbf{h} = \begin{bmatrix} \delta\mathbf{h}_1 \\ \delta\mathbf{h}_2 \\ \cdot \\ \cdot \\ \cdot \\ \delta\mathbf{h}_{m_e} \end{bmatrix} \qquad (3.31)$$

and

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{S}_{m_e} \end{bmatrix} \qquad (3.32)$$

The matrix $\mathbf{A}$ defined in equation (3.30) has a very special structure called **block diagonal** [Golub and Plemmons, 1981]. (This term is potentially a bit confusing since $\mathbf{A}$ is not even close to any diagonal form. The name arises from the form the matrix $\mathbf{A}^T\mathbf{A}$ of the normal equations of $\mathbf{A}$.) The block diagonal form is significant because it shows that in the multiple event location problem, individual events are coupled only through the station corrections. One view

---

[5] I am using a notation convention here in which partitions of a larger matrix are denoted by a boldface letter that is the same as the larger matrix of which it is a part but labeled with a subscript. The special structure of the matrices considered here allow me to use only one subscript instead of the two that are usually required to identify individual pieces of a partitioned matrix. This subscript can thus be considered as a simple label that identifies the particular event that entity is associated with.

of single event location procedures is that they solve (3.29) ignoring the last term involving the station corrections, $S\delta s$. When that is done, the locations of individual earthquakes become completely uncoupled and locating each event in isolation is justified. That approach, however, is obviously somewhat blind as it is tantamount to the assumption that the station corrections are known exactly. Nonetheless, it is instructive to study what happens to the set of equations (3.30) when each earthquake is located separately and the station correction term, $S\delta s$, is included. We saw above that single event location of each of the m_e events involves solving the set of equations

$$\mathbf{r}_j^s = \mathbf{A}_j \delta \mathbf{h}_j \qquad j = 1,2,3, \ldots , m_e \qquad (3.33)$$

using a least squares solution of the general form (3.16)[6]. This process is repeated until the solution converges. In practice, "convergence" usually implies that $\|\delta \mathbf{h}_j\|$ is small. Hence, when all m_e events are located individually, we would find that for each event

$$\delta \mathbf{h}_j = \mathbf{H}_j \mathbf{r}_j^s \approx 0 \qquad (3.34)$$

where the matrix $\mathbf{H}_j \in \mathbb{R}^{4 \times m_j}$ is a generalized inverse of the form (3.18). Because of the block diagonal form of $\mathbf{A}$, the final solution (the hypocenters of all m_e events) could be considered as a matrix multiplication of (3.29) by a matrix $\mathbf{H} \in \mathbb{R}^{4m_e \times M}$, which I define as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{H}_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & \mathbf{H}_{m_e} \end{bmatrix} \qquad (3.35)$$

---

[6] In the actual implementation of this algorithm, I have utilized the Levenberg-Marquardt inverse, $\mathbf{A}^m$, defined by equation (2.39). As a practical matter the pseudoinverse (defined by equation (2.37)) may actually be preferable for the purpose of this algorithm because its range and null space are better defined.

where the individual nonzero partitions in (3.35) are the matrices H_j of equation (3.34). This matrix multiplication yields

$$Hr^s = HA\delta h \qquad (3.36)$$

or after rearranging

$$\delta\hat{h} = Hr^s \qquad (3.37)$$

where I define

$$\delta\hat{h} = HA\delta h = R\delta h \qquad (3.38)$$

R is a matrix usually called the resolution matrix (e.g. Wiggins[1972] or Crosson [1976a]). For least squares problems such as this one, R is some approximate identity matrix. R reduces to an identity matrix only when A is full rank and $H=A^\dagger$, the Moore-Penrose inverse defined in equation (2.33)[7]. No matter what R is, it can be seen from (3.34) that when each of the m_e events is located initially by standard single event methods, we will find

$$\delta\hat{h} = Hr^s \approx 0 \qquad (3.39)$$

H is a matrix with $4m_e$ rows and M columns. Consequently, (3.39) declares that the net result of locating the m_e events individually is to cause those $4m_e$ linear combinations of the residuals defined by (3.39) to vanish. What this says as a practical matter is that only these $4m_e$ parcels of information contained in the original data are utilized to locate the events[8]. I will now show how the information

---

[7] In practice H is always calculated from a matrix A using finite precision arithmetic. Thus even when A is full rank and $H=A^\dagger$, $R=A^\dagger A$ is not strictly an identity matrix because of roundoff error [Lawson and Hanson, 1974, pp. 90-99]. The amount R differs from an identity in this case is dependent upon how ill conditioned A is.

[8] $4m_e$ is actually the maximum number of independent data that are utilized in locating the m_e earthquakes. The number will be less when the rows of H are linearly dependent.

that remains in the residuals can be extracted in an unbiased way to estimate the station corrections.

To understand the interplay of the process of single event location and the resulting residuals for all m_e events, it is useful to rewrite (3.35) in terms of an orthogonal decomposition similar to (3.18). To do this we first note that from the definition (3.18) that each nonzero block of $\mathbf{H}, \mathbf{H}_i \in R^{4 \times m_i}$, can be written as

$$\mathbf{H}_i = \mathbf{V}_i \Lambda_i^H \mathbf{U}_i^T \qquad i = 1, 2, \ldots, m_e \tag{3.40}$$

where $\mathbf{V}_i \in R^{4 \times 4}$, $\Lambda_i^H \in R^{4 \times m_i}$, and $\mathbf{U}_i^T \in R^{m_i \times m_i}$ are as defined in (3.18). The only difference is that I have added the subscript i as a label. Because of the form of the Λ_i^H matrices it is useful to rewrite (3.40) in the partitioned form

$$\mathbf{H}_i = [\mathbf{V}]_i [\Lambda_R^H 0] \begin{bmatrix} \mathbf{U}_R^T \\ \mathbf{U}_N^T \end{bmatrix}_i \tag{3.41}$$

where $\Lambda_R^H \in R^{4 \times 4}$ is as in equation (3.19). The row partitioning of $\mathbf{U}_i^T$ is defined congruently so $\mathbf{U}_R^T \in R^{4 \times m_i}$ and $\mathbf{U}_N^T \in R^{(m_i - 4) \times m_i}$. This partitioning is consistent with that introduced in chapter 2 in equation (2.21). This form, however, combines the partitions labeled with the subscripts R and N_1 to account for the form of Λ_R^H that I have used here. Because of that I have dropped the subscript on N and so I write $\mathbf{U}_N^T = \mathbf{U}_{N_2}^T$ for convenience. With these definitions it is easy to see that $\mathbf{H}$ can be written in the alternate form

$$\mathbf{H} = \mathbf{V} \Lambda^H \mathbf{U}^T = \mathbf{V} [\Lambda_R^H 0] \begin{bmatrix} \mathbf{U}_R^T \\ \mathbf{U}_N^T \end{bmatrix} \tag{3.42}$$

where $\mathbf{V} \in R^{4m_e \times 4m_e}$ is of the form

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{V}_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & \mathbf{V}_{m_e} \end{bmatrix} \tag{3.43}$$

with the nonzero blocks along the diagonal of $\mathbf{V}$ being the matrices, $\mathbf{V}_i \in R^{4\times4}$ in equation (3.41). Similarly, $\Lambda^H \in R^{4m_s \times M}$ and $\mathbf{U}^T \in R^{M \times M}$ are defined as

$$\Lambda^H = [\Lambda_R^H \mid 0] = \left[\begin{array}{c|c} \begin{matrix} (\Lambda_R^H)_1 & 0 & \cdots & 0 \\ 0 & (\Lambda_R^H)_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & (\Lambda_R^H)_{m_s} \end{matrix} & \mathbf{0} \end{array} \right] \qquad (3.44)$$

and

$$\mathbf{U}^T = \left[\begin{matrix} \mathbf{U}_R^T \\ \hline \mathbf{U}_N^T \end{matrix} \right] = \left[\begin{array}{cccc} (\mathbf{U}_R^T)_1 & 0 & \cdots & 0 \\ 0 & (\mathbf{U}_R^T)_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & (\mathbf{U}_R^T)_{m_s} \\ \hline (\mathbf{U}_N^T)_1 & 0 & \cdots & 0 \\ 0 & (\mathbf{U}_N^T)_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & (\mathbf{U}_N^T)_{m_s} \end{array} \right] \qquad (3.45)$$

The reader may well think the only thing that has been accomplished at this point is converting the simpler matrix in equation (3.35) into one that is at least three times more complicated. However, do not despair for it should become apparent directly that quite the opposite is true. (3.42) actually leads to a simpler set of relationships between the hypocenter parameters and the station correction parameters. To see this, it is instructive to multiply the equations in (3.29) by the decomposed form of $\mathbf{H}$ defined in (3.42). The first step of that process is to multiply (3.29) by the orthogonal matrix $\mathbf{U}^T$. This leads to

$$\begin{bmatrix} (\mathbf{r}_R^s)_1 \\ (\mathbf{r}_R^s)_2 \\ \cdot \\ \cdot \\ (\mathbf{r}_R^s)_{m_e} \\ \hline (\mathbf{r}_N^s)_1 \\ (\mathbf{r}_N^s)_2 \\ \cdot \\ \cdot \\ (\mathbf{r}_N^s)_{m_e} \end{bmatrix} = \left[\begin{array}{cccc} (\mathbf{A}_R)_1 & 0 & \cdots & 0 \\ 0 & (\mathbf{A}_R)_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & (\mathbf{A}_R)_{m_e} \\ \hline 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & 0 \end{array}\right] \begin{bmatrix} \delta\mathbf{h}_1 \\ \delta\mathbf{h}_2 \\ \cdot \\ \cdot \\ \delta\mathbf{h}_{m_e} \end{bmatrix} + \left[\begin{array}{c} (\mathbf{S}_R)_1 \\ (\mathbf{S}_R)_2 \\ \cdot \\ (\mathbf{S}_R)_{m_e} \\ \hline (\mathbf{S}_N)_1 \\ (\mathbf{S}_N)_2 \\ \cdot \\ (\mathbf{S}_N)_{m_e} \end{array}\right] \begin{bmatrix} \delta\mathbf{s}_1 \\ \delta\mathbf{s}_2 \\ \cdot \\ \cdot \\ \delta\mathbf{s}_{m_e} \end{bmatrix}$$

$$(3.46)$$

where

$$(\mathbf{r}_R^s)_i = (\mathbf{U}_R^T)_i \, \mathbf{r}_i^s \in \mathbf{R}^4 \qquad\qquad (3.47)$$

$$(\mathbf{A}_R)_i = (\mathbf{U}_R^T)_i \, \mathbf{A}_i \in \mathbf{R}^{4\times4} \qquad\qquad (3.48)$$

$$(\mathbf{S}_R)_i = (\mathbf{U}_R^T)_i \, \mathbf{S}_i \in \mathbf{R}^{4\times n_s} \qquad\qquad (3.49)$$

$$i = 1, 2, \ldots, m_e$$

and

$$(\mathbf{r}_N^s)_i = (\mathbf{U}_N^T)_i \, \mathbf{r}_i^s \in \mathbf{R}^{m_i - 4} \qquad\qquad (3.50)$$

$$(\mathbf{S}_N)_i = (\mathbf{U}_N^T)_i \, \mathbf{S}_i \in \mathbf{R}^{(m_i - 4)\times n_s} \qquad\qquad (3.51)$$

$$i = 1, 2, \ldots, m_e$$

Equation (3.46) contains two distinctly different types of equations marked by the horizontal dividing line in (3.46). The solution process used to estimate the hypocenter parameters handles these two types of equations totally differently. To see how, consider the product $\mathbf{V\Lambda}^H$ of (3.42). We observe that it has the partitioned form

$$\mathbf{V\Lambda}^H = [\mathbf{V\Lambda}^H \mid \mathbf{0}] \tag{3.52}$$

$$= \begin{bmatrix} (\mathbf{V\Lambda}_R^H)_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (\mathbf{V\Lambda}_R^H)_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & (\mathbf{V\Lambda}_R^H)_{m_e} \end{bmatrix} \mathbf{0}$$

where each $(\mathbf{V\Lambda}_R^H)_i \in R^{4\times4}$. Thus, we see the product $\mathbf{V\Lambda}_R^H$ is a $4m_e \times 4m_e$ matrix. When we multiply (3.46) by $\mathbf{V\Lambda}^H$ we will produce the same result as when we multiplied (3.29) by $\mathbf{H}$.

$$\begin{bmatrix} \mathbf{H}_1 \mathbf{r}^s \\ \mathbf{H}_2 \mathbf{r}^s \\ \vdots \\ \mathbf{H}_{m_e} \mathbf{r}^s \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R}_{m_e} \end{bmatrix} \begin{bmatrix} \delta\mathbf{h}_1 \\ \delta\mathbf{h}_2 \\ \vdots \\ \delta\mathbf{h}_{m_e} \end{bmatrix} + \begin{bmatrix} \mathbf{H}_1\mathbf{S}_1 \\ \mathbf{H}_2\mathbf{S}_2 \\ \vdots \\ \mathbf{H}_{m_e}\mathbf{S}_{m_e} \end{bmatrix}[\delta\mathbf{s}] \tag{3.53}$$

where the partitions $\mathbf{R}_i \in R^{4\times4}$ are defined as

$$\mathbf{R}_i = \mathbf{H}_i \mathbf{A}_i$$

Thus, the matrix $\mathbf{R}$ that is the first term on the right hand side of (3.53) is the same as $\mathbf{R}$ in equation (3.38). The significant point to observe from (3.53) is that the hypocenters are estimated only from the first $4m_e$ residuals of (3.46). The other $M - 4m_e$ residuals are not used to estimate the hypocenters because they are forced to be zero in (3.53) and hence are ignored in estimating the hypocenters[9].

---

[9] This result should come as no surprise to those who have read chapter 2 of this dissertation. It was noted there that orthogonal matrices represent a rigid reorientation of the vector space in which the data vector, $\mathbf{r}^s \in R^M$, lies. From the way I defined it, the matrix $\mathbf{U}$ in (3.45) is the matrix one would obtain from a singular value decomposition of the matrix $\mathbf{A}$ in equation (3.30). In the dis-

This suggests we can estimate the station corrections by solving

$$\mathbf{r}_N^s = \mathbf{S}_N \, \delta \mathbf{s} \tag{3.54}$$

where $\mathbf{r}_N^s \in \mathbf{R}^{M-4m_e}$ and $\mathbf{S}_N \in \mathbf{R}^{(M-4m_e) \times n_e}$ are defined as

$$\mathbf{r}_N^s = \begin{bmatrix} (\mathbf{r}_N^s)_1 \\ (\mathbf{r}_N^s)_2 \\ \cdot \\ \cdot \\ \cdot \\ (\mathbf{r}_N^s)_{m_e} \end{bmatrix} \tag{3.55}$$

and

$$\mathbf{S}_N = \begin{bmatrix} (\mathbf{S}_N)_1 \\ (\mathbf{S}_N)_2 \\ \cdot \\ \cdot \\ \cdot \\ (\mathbf{S}_N)_{m_e} \end{bmatrix} \tag{3.56}$$

The actual method used to solve (3.54) is somewhat a matter of choice. In the implementation I have made of this procedure I have used the pseudoinverse solution (see Chapter 2) of (3.54) that I write as

$$\delta \hat{\mathbf{s}} = \mathbf{S}_N^+ \mathbf{S}_N \, \delta \mathbf{s} = \mathbf{S}_N^+ \mathbf{r}_N^s \tag{3.57}$$

which as we have seen is closely related to the Moore-Penrose

---

cussion in chapter 2 it was demonstrated in a heuristic fashion that $\mathbf{U}$ produces a special reorientation of $\mathbf{R}^M$. The first $4m_e$ rows of $\mathbf{U}^T$ (those denoted as $\mathbf{U}_R^T$ in equation (3.45)) form an orthonormal basis for the range of $\mathbf{A}$. The remaining $M-4m_e$ rows of $\mathbf{U}^T$ (denoted $\mathbf{U}_N^T$ in (3.45)) are an orthonormal basis for the rest of $\mathbf{R}^M$, $N(\mathbf{A}^T)$. Thus the process of multiplying (3.29) by $\mathbf{U}^T$ produces a special reorientation of $\mathbf{R}^M$. The first $4m_e$ components of the vector $\mathbf{U}^T \mathbf{r}^s$ that I have denoted by the symbol $\mathbf{r}_R^s$ are the projections of $\mathbf{r}^s$ onto $R(\mathbf{A})$. Similarly, the remaining $M-4m_e$ components of $\mathbf{U}^T \mathbf{r}^s$ that I have denoted by the symbol $\mathbf{r}_N^s$ are the projections of $\mathbf{r}^s$ onto the $N(\mathbf{A}^T)$.

inverse[10]. It should be noted that a simple solution of (3.54) by the normal equations is not feasible because S_N is always singular. This singularity is the manifestation here of a problem I noted earlier with JHD. The average value of the station corrections is indeterminate because it trades off arbitrarily with the origin times of the m_e events[11]. I would argue that the pseudoinverse is a sensible choice because it finds the unique solution of (3.54) that minimizes $\|\delta s\|_2$. Thus, it makes the station corrections as small as possible in the mean square sense. This is desirable because normally we want the station corrections to be the smallest size possible that is consistent with the data. If one prefers the constraint

$$\sum_{i=1}^{n_s} s_i = 0 \qquad (3.58)$$

it can always be imposed *a posteriori* since the solution is independent of the mean station correction. As a practical matter I have found that (3.58) is nearly always satisfied approximately anyway provided the initial guess for the station corrections satisfied (3.58).

---

[10] Practical experience with this procedure to date has shown that generally the pseudoinverse and the Moore-Penrose inverse are identical for S_N. This occurs because the tendency is to get $n_s - 1$ singular values of the matrix S_N that are roughly the same order of magnitude and a single very small singular value of the size of the floating point precision of the computer being used. Any reasonable tolerance parameters would, in this case, set the pseudorank to $n_s - 1$ which I noted above is the maximum rank of S_N. There is some reason to suspect that this may not be universally true, however, since $rank\,(S_N) = n_s - 1$ is probably indicative of well balanced data in which each station records about the same total number of arrivals.

[11] This problem could also be solved by including refraction data from an explosion source with known origin time or fixing the location of one "master event" as in JHD. I feel the use of the pseudoinverse is preferable to this because it does not require a master event but will, nonetheless, use events with known hypocenters as a type of "master event" implicitly when they are available.

This completes the basic description of the progressive multiple event location procedure. The description has, to this point, involved a considerable amount of manipulation involving partitioned matrices with complicated structures. Consequently, it seems appropriate to emphasize the key points the reader should glean from these manipulations. This is most readily accomplished by summarizing the four key steps in implementing this procedure.

(1) Locate each earthquake individually be conventional single event location methods.

(2) Calculate the matrix $\mathbf{U}$ defined in (3.45).

(3) Use $\mathbf{U}$ to calculate the matrix $\mathbf{S}_N$ in equation (3.56) and the vector $\mathbf{r}_N^s$ in equation (3.55).

(4) Calculate perturbations to the station corrections by (3.57).

These steps are repeated until the procedure converges. Overall, this procedure is similar to the simple method of estimating station corrections using raw residuals described in section 2.2.1 above. That is, both estimate the hypocenters and station corrections in separate calculations. The fundamental difference is that this procedure does not use the raw residuals to estimate the station corrections, but instead uses only the $M-4m_e$ numbers $\mathbf{r}_N^s$ in equation (3.55) that are ignored in estimating the hypocenters.

3.2.3. A practical algorithm

The algorithm I described in the previous section is not very practical as it stands, because the matrix equation (3.46) can attain an overwhelming size when the amount of data is large. This is the same storage problem that I noted earlier with a practical implementation of the JHD method. That is, the equations grow in size by m_i rows and 4 columns for each earthquake added to the data set. This explosive growth in the storage requirements of multiple event location schemes has limited their usefulness to small data sets.

One of the major advantages of the progressive multiple event location method is that the special structure of the equations can be exploited to produce an algorithm that operates in a small, fixed storage area of size of the order $6n_s + 2n_s^2$ where n_s is the number of stations in the network. This has great practical importance because it means this algorithm could be used for routine earthquake locations even with a micro-computer. This section is primarily devoted to showing how this storage reduction can be accomplished.

The first observation to make is that the matrix $\mathbf{A}$ in (3.30) is mostly zeros. The block diagonal form of $\mathbf{A}$ leads to an equally simple form for the matrix $\mathbf{U}$ in (3.45) used to form the decomposition in (3.46) which is the fundamental basis for the progressive method. The key observation to make is that there is no reason to explicitly form the matrix $\mathbf{A}$ or the matrix $\mathbf{U}$. Instead, it is natural to work with each event in the data set individually. The process has two parts. The first is the process of single event location. This involves the now familiar process of solving (3.33) repeatedly, using (3.16), until convergence is achieved. This operation requires storage of only of $5n_s$ numbers ($\mathbf{A}_i$ is of size $4n_s$ and $\mathbf{r}^s$ is of size n_s.). The second major step is to calculate the matrix $\mathbf{U}$ of (3.45) and to multiply (3.29) by it. Fortunately, this step is also amenable to efficient usage of computer storage. The block diagonal form of $\mathbf{A}$ leads to the simple form (3.45) of $\mathbf{U}$. The form of $\mathbf{U}$ is such that it, like $\mathbf{A}$, need ever be formed explicitly. The reason is that we are not particularly interested in what each of the submatrices $\mathbf{U}_i$ actually are, but all we really need are the products defined in equation (3.47) to (3.51). These products are all various pieces of the product of $\mathbf{U}_i^T$ with the quantities $\mathbf{r}_i^s, \mathbf{A}_i$, and $\mathbf{S}_i$. All of these quantities are associated only with the ith event. Thus the products in (3.47) to (3.51) can be calculated individually, event by event, with no need to explicitly form any of the large matrices in (3.46). Furthermore, an even

greater storage reduction is possible because $\mathbf{U}_i^T$ need never be explicitly calculated either. Because the only use we make of $\mathbf{U}_i^T$ is to form the products (3.47) to (3.51), it is possible to arrange the calculations such that $\mathbf{U}_i^T$ itself is never actually formed, but only the products (3.47) to (3.51) are calculated. An example of this is an algorithm for calculating the singular value decomposition of a matrix written by Lawson and Hanson [1974,pp.260-262,295-300] that they call SVDRS[12]. SVDRS applied to a matrix $\mathbf{A}_i$ can be used to return the following[13]

$$\mathbf{r}_i^s \leftarrow \begin{bmatrix} (\mathbf{r}_R^s)_i \\ (\mathbf{r}_N^s)_i \end{bmatrix} = \mathbf{U}_i^T \mathbf{r}_i^s$$

$$\mathbf{S}_i \leftarrow \begin{bmatrix} (\mathbf{S}_R)_i \\ (\mathbf{S}_N)_i \end{bmatrix} = \mathbf{U}_i^T \mathbf{S}_i$$

and

$$\mathbf{A}_i \leftarrow [\mathbf{V}]$$

Details of how this can be accomplished can be found in the Lawson and Hanson reference. This algorithm makes efficient usage of storage because no space needs to be allocated to hold the matrix $\mathbf{U}_i^T$. Thus, this step of the algorithm can be accomplished with only $5n_s + n_s^2$ storage locations (Actually additional space is required by

---

[12] I should point out that this same storage saving could be accomplished using a simpler QR decomposition (Lawson and Hanson [1974,pp.254-256,290-291] also give an algorithm for doing this.) rather than a singular value decomposition. The difference in computation effort for this problem is, however, negligible because $\mathbf{A}_i$ has only four singular values. Timing data to verify this can be found in Dongarra et. al. [1979].

[13] The $\leftarrow$ symbol denotes what Stewart [1974, pp. 88-89] calls a dynamic equivalence operator. Its meaning is that the quantity on the right of the $\leftarrow$ symbol is associated with the location of the quantity on the left. It can be loosely interpreted as an overwriting operation.

SVDRS to hold the singular values and as a work space. These are, however, negligible for this problem.).

The only remaining step is calculation of the perturbations to the station corrections by equation (3.57). Directly forming the equation (3.54) is not a very attractive prospect, since S_N is a $(M - 4m_e) \times n_s$ matrix and can attain a very large size if M is very large. Fortunately, this problem too can be avoided. I know of two ways to accomplish this;

(1) S_N^+ can be calculated from normal equations which can be accumulated in blocks. This approach calculates the matrix $S_N^T S_N$

as $\quad S_N^T S_N = \sum\limits_{i=1}^{m_e} (S_N^T)_i (S_N)_i \quad$ and the vector $\quad S_N^T r_N^s \quad$ as

$S_N^T r_N^s = \sum\limits_{i=1}^{m_e} (S_N^T)_i (r_N^s)_i$ and calculates δs as $\delta \hat{s} = (S_N^T S_N)^+ S_N^T r_N^s$. (See

Ben-Israel and Greville [1974,pp.110-111] for a theoretical justification.)

(2) QR decomposition of S_N by sequential accumulation [Lawson and Hanson, 1974, chapter 27] with S_N^+ derived from a singular value analysis of the matrix R_s from QR decomposition of S_N.

Either algorithm is acceptable but I would argue that (2) is preferable primarily because of the superior numerical stability of modern algorithms for accomplishing a QR decomposition. This is particularly important in light of the fact that S_N is always singular and hence $S_N^T S_N$ cannot be inverted by elementary methods anyway. A sketch of an algorithm to accomplish a QR decomposition of a matrix by sequential accumulation (SEQHT) can be found in Lawson and Hanson [1974, pp. 210]. I will not burden the reader with a detailed description of this algorithm here, but I refer the reader to their text which contains a fairly intelligible description of this procedure. The main point here is what SEQHT is good for. It permits us to avoid forming the matrix S_N explicitly, but instead we only need to provide an extra space to hold the $n_s \times n_s$ upper triangular

matrix $\mathbf{R}_s$ of the QR decomposition of $\mathbf{S}_N$. By utilizing this algorithm $\mathbf{R}_s$ can be accumulated event by event using a sequence of elementary reflectors calculated from the submatrices $(\mathbf{S}_N)_i$. The storage required to accomplish this step is $2n_s^2$. n_s^2 locations are required to hold $\mathbf{S}$ and n_s^2 locations are required to hold $\mathbf{R}_s$ [14]. The final result, after all m_e events are processed, is the equivalent of calculating a QR decomposition from $\mathbf{S}_N$ as

$$\mathbf{S}_N = \mathbf{Q}_s \mathbf{R}_s \tag{3.59}$$

with $\mathbf{Q}_s \in \mathbb{R}^{M \times n_s}$ and $\mathbf{R}_s \in \mathbb{R}^{n_s \times n_s}$. $\mathbf{Q}_s$, like the matrix $\mathbf{U}_i^T$ discussed above, is not explicitly calculated but instead is used to reduce (3.54) to the form

$$\mathbf{R}_s \delta s = \mathbf{Q}_s^T \mathbf{r}_N^s = \mathbf{r}_Q \tag{3.60}$$

Once the reduction to (3.60) is accomplished, δs can be calculated by

$$\delta \hat{s} = \mathbf{R}_s^+ \mathbf{r}_Q$$

where the pseudoinverse of $\mathbf{R}_s$, $\mathbf{R}_s^+$, is readily obtained from a singular value decomposition of $\mathbf{R}_s$.

The preceding discussion is summarized in the following brief sketch of a procedure I call ALGORITHM PMEL:

---

[14] More complicated coding could be used to reduce the storage requirements for $\mathbf{R}_s$ to $\dfrac{n_s^2}{2}$ since only the upper triangle is used.

ALGORITHM PMEL

1) Initialize s and $\mathbf{h}_i$, $i = 1, 2, \ldots, m_e$

2) Loop from 3) to 6) until convergence

   3) For $i = 1, 2, \ldots, m_e$

      1i) Execute ALGORITHM MODHYLOC (section 1.2 above) using s. Return: $\mathbf{h}_i, \mathbf{A}_i$, and $\mathbf{r}_i^s$

      2i) Execute ALGORITHM SVDRS [Lawson and Hanson, 1974, pp. 260-267, 295-300] SVDRS returns:

$$\mathbf{r}_i^s \leftarrow \mathbf{U}_i^T \mathbf{r}_i^s$$

$$\mathbf{A}_i \leftarrow \mathbf{V}$$

$$\mathbf{S}_i \leftarrow \mathbf{U}_i^T \mathbf{S}_i$$

      3i) Execute ALGORITHM SEQHT on $(\mathbf{S}_N)_i$ [Lawson and Hanson, 1974, p. 210]. Accumulate matrix $\mathbf{R}_s$ and vector $\mathbf{r}_Q$ of QR decomposition of $\mathbf{S}_N$.

   4) Calculate $\mathbf{R}_s^+$ by singular value decomposition of $\mathbf{R}_s$.

   5) $\delta\hat{\mathbf{s}} = \mathbf{R}_s^+ \mathbf{r}_Q$

6) End loop

---

the major storage areas required by this algorithm are illustrated in figure 3.1. Figure 3.1 is useful primarily as an introduction to figure 3.2. Figure 3.2 is intended to clarify this rather complex algorithm by showing what areas of storage are active at each major step of the algorithm. We see there that in the earthquake location phase, step 1i), only $\mathbf{A}_i$ and $\mathbf{r}_i^s$ are used. Blockwise multiplication by the $\mathbf{U}_i^T$ matrices, step 2i), uses only the spaces labeled $\mathbf{A}_i, \mathbf{r}_i^s$, and $\mathbf{S}_i$. Sequential accumulation of $\mathbf{R}_s$, step 3i), uses the upper triangle of $\mathbf{R}_s$ and the bottom $m_i - 4$ rows of $\mathbf{S}_i$ [15]. Finally, the singular value decomposition of $\mathbf{R}_s$ and calculation of $\delta\hat{\mathbf{s}}$, steps 4) and 5), use the array $\mathbf{R}_s$, the vector $\mathbf{r}_Q$, and small work spaces (These are omitted

---

[15] Step 3i) is somewhat oversimplified in figure 3.2. Actually, the first few events have to be handled specially (see Lawson and Hanson [1974, pp. 208-212]).
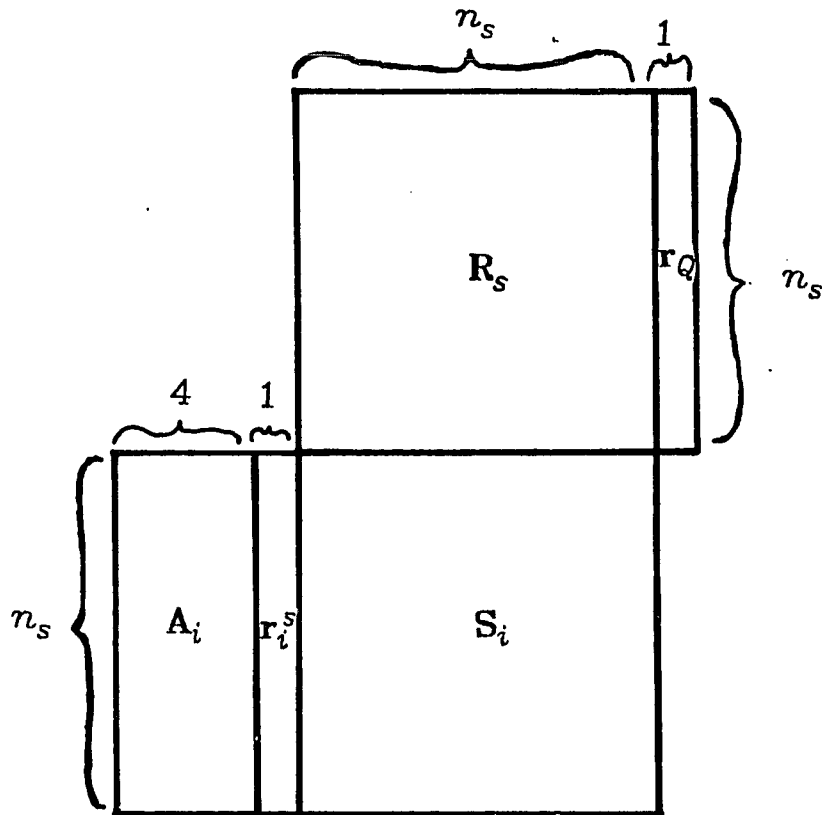
Figure 3.1. Major storage areas required by progressive multiple event location algorithm. This figure is a schematic in which each element of an a given array can be viewed as a small square at the proper position for a normal matrix tableau. For example, S_i is an $n_s \times n_s$ array. The element $(S_{11})_i$ can be visualized as occupying the upper left hand corner of S_i.

Step 1i)

Step 2i)

Step 3i)

Step 4) and 5)
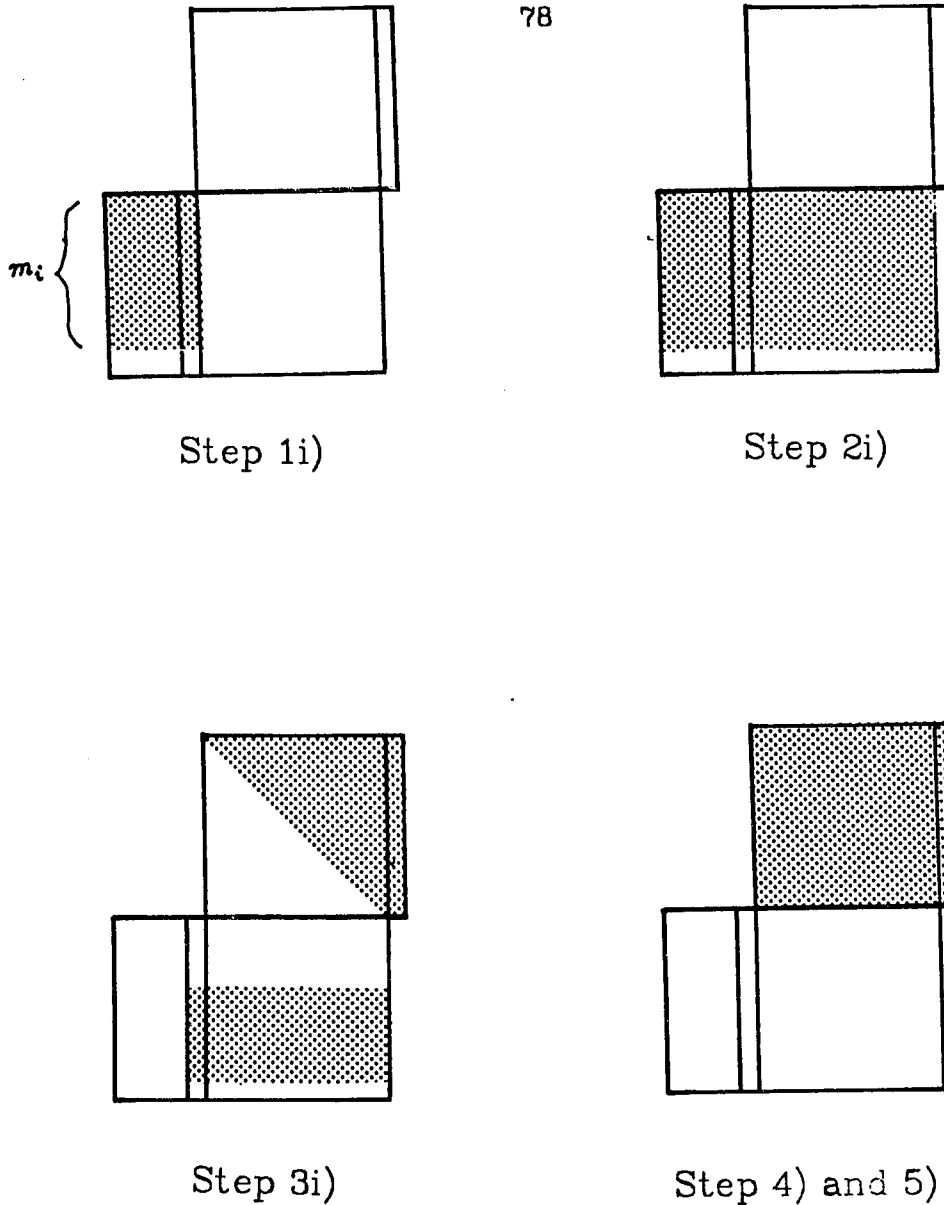
Figure 3.2. Storage areas referenced at each major step of PMEL. Major storage areas are as pictured in figure 3.1. Stippled areas indicate areas that are used during each indicated step of ALGORITHM PMEL. m_i, the number of arrival times recorded for the ith event, is generally less than n_s. Because of this, only m_i rows of the arrays $\mathbf{A}_i, \mathbf{S}_i$, and $\mathbf{r}_i^S$ are used for the ith event. m_i varies with each event so the level to which these arrays are filled is variable. This is emphasized here by showing the bottom section of these arrays unstippled.

in figure 3.2. Either $\mathbf{A}_i$ or $\mathbf{S}_i$ could be used for this purpose.).

The key observation to make from figure 3.2 is that this algorithm operates in a fixed storage area. It does not require added storage for each new earthquake added to the data set like JHD. Consequently, it is capable of digesting virtually unlimited quantities of data while operating within memory requirements that are readily obtainable even on a mini-computer. I have actually utilized this procedure only as a part of a larger scale procedure that also inverts for P-wave velocity structure (see chapter 4). However, because of its remarkably small computer memory requirements, I believe it may find use in three other ways.

(1) Station corrections calculated by PMEL represent an average time correction for ray paths joining all sources to a given receiver. They thus represent some average path anomaly. If data from a seismograph network were sorted by the epicenters of the sources, then significant changes of station corrections calculated for data from different regions could be used to infer the existence of lateral velocity variations. This has been attempted previously by other workers (see for example Knapp and Smith [1979] or Spieth [1981]) using other multiple event location methods. The advantage of PMEL is that it is capable of digesting large data sets and does not require any *ad hoc* constraints.

(2) The procedure has promise as a method of data compression in inversion for three dimensional velocity structure. When this procedure is applied to small clusters of earthquakes it becomes meaningful to calculate what Jordan and Sverdrup [1981] call the "hypocentroid" of the cluster. The station corrections one calculates can then be thought of as the average path anomaly from the "hypocentroid" to each station. One view of ALGORITHM PMEL, in this case, is that it is a general, optimum method of stacking data from a set of closely

spaced events of variable quality to estimate the average path anomaly for the entire cluster. Event clusters are common features of micro-earthquake data. This algorithm provides a means of combining the data from hundreds of earthquakes to look like a single, high quality event.

(3) SEQHT could be modified to operate in an updating fashion. Each earthquake recorded by the network could be used to update R_s in real time. R_s would then represent the accumulated knowledge of the station corrections from all past events. Similarly, data later found to be erroneous could be selectively deleted by the reverse process. For a description of how this can be accomplished see Lawson and Hanson [1974, pp. 225-231].

I have now completed a description of a practical algorithm for doing progressive multiple event locations and described some of this algorithms possible uses. Its principle advantage is that it operates in a small, fixed working area in computer memory. Since a universal principle is that "there is no free lunch" the skeptic may well ask what price we have paid to accomplish this. The skeptic is right. This was not a gift. I will now demonstrate through the use of the resolution matrix [Wiggins, 1972] that the price we have paid is the hypocenter estimates are coupled to the estimates of the station corrections. However, I will show that this is a small price to pay because it is only an admission of a fact that is usually ignored.

3.2.4. Error analysis

Although ALGORITHM PMEL is actually a complex procedure involving iterations on two separate levels (Single event location is an iterative procedure and the entire algorithm is a larger scale loop.), the final estimate can still be viewed as a linear perturbation from a reference solution using a generalized inverse solution of

equation (3.26). Generalized inverses for problems characterized by a finite number of parameters have associated with them three important matrix operators. These are the matrices usually referred to as the resolution matrix, covariance matrix, and the information density matrix (see for example, Wiggins [1972] or Crosson [1976a]). These matrices are important as tools for assessing the reliability of parameter estimates produced by generalized inverse solutions. I will derive here the form for the resolution matrix and covariance matrix for ALGORITHM PMEL and show that both have a special structure. I will not derive the form for the information density matrix because it has no special structure. It could be calculated, if desired, directly from its definition (see e.g. Crosson [1976a]). However, doing so would normally be of dubious value since it is a matrix of size $M \times M$ and would thus defeat the major purpose of this algorithm which is storage reduction.

The resolution and covariance matrices are both calculated from the generalized inverse matrix used to estimate the parameters. Thus, the first order of business is to determine the form of the generalized inverse that is used by ALGORITHM PMEL. This means we seek the form of the matrix $\mathbf{H}_B \in R^{N \times M}$ that we use to estimate the parameter vector via a relation of the form

$$\hat{\mathbf{x}} = \mathbf{H}_B \, \mathbf{B} \mathbf{x} = \mathbf{H}_B \, \mathbf{r}^s \tag{3.61}$$

where $\mathbf{x} \in R^N$ and $\mathbf{B} \in R^{M \times N}$ are as defined in (3.27). $\hat{\mathbf{x}}$ is the actual estimate of the true, unknown parameter vector $\mathbf{x}$. From equations (3.46), (3.53), and (3.57) we see that $\mathbf{H}_B$ for the progressive multiple event location solution is of the form

$$\mathbf{H}_B = \begin{bmatrix} \mathbf{H} \\ \mathbf{S}_N^+ \mathbf{U}_N^T \end{bmatrix} \tag{3.62}$$

where $\mathbf{H} \in R^{4m_s \times M}$ is defined in (3.35), $\mathbf{S}_N^+$ is defined in (3.57), and $\mathbf{U}_N^T$ is defined in (3.45).

Having now specified the form of the generalized inverse, H_B, we turn now to calculation of the resolution matrix. The resolution matrix for x is defined as [Crosson, 1976a]

$$R_B = H_B B \qquad (3.63)$$

Hence (3.61) could be also written as

$$\hat{x} = R_B x = H_B r^s \qquad (3.64)$$

(3.64) demonstrates the usefulness of the resolution matrix. R_B measures how well determined the solution is, because whatever the true solution x is, the estimate $\hat{x}$ is always related to it by (3.64). The usual measure used to ascertain how well determined $\hat{x}$ is, is to compare R_B to an identity matrix. This is sensible since if R_B were an identity matrix it would indicate x is known unambiguously. However, R_B is an identity matrix only in the special case when B is full rank and H_B is the least squares (Moore-Penrose) inverse, $H_B = B^\dagger = (B^T B)^{-1} B^T$. I have already noted several times that B is never full rank so the least squares solution is not unique and R_B can never be reduced to an identity matrix for this problem. This is indicative of a fundamental ambiguity that cannot be removed without auxiliary information[16]. An advantage, in my opinion, of the progressive procedure is that it expresses this ambiguity directly in a manner that makes sense intuitively. To see this, note that if we substitute for B from equation (3.26) and H_B from equation (3.62) into the definition of R_B in (3.63) we obtain

---

[16] A JHD solution will generally result in a solution for which R_B is a close approximation to an identity matrix. This is accomplished in JHD through the use of a set of auxiliary constraints that can be criticized as being somewhat *ad hoc* [Jordan and Sverdrup, 1981].

$$R_B = \begin{bmatrix} \begin{array}{cccc|c} R_1 & 0 & \cdots & 0 & H_1 S_1 \\ 0 & R_2 & \cdots & 0 & H_2 S_2 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & \cdots & R_{m_s} & H_{m_s} S_{m_s} \\ \hline \multicolumn{4}{c|}{0} & S_N^+ S_N \end{array} \end{bmatrix} \tag{3.65}$$

where the matrices $R_i \in R^{4 \times 4}$ are as in equation (3.53). The fact that the lower left partition is a zero matrix follows directly from the definition of U_N^T. The form of the lower right partition follows directly from the definition of S_N since

$$S_N = U_N^T S$$

Obviously, (3.65) is usually not even remotely close to an identity matrix. Nonetheless, (3.65) does indeed express the correct ambiguity that is inherent in progressive multiple event location. To see this, recall that the hypocenters are estimated in this procedure by single event location using station corrections derived from the previous iteration. This involves calculating $\delta \hat{h}_i$ for each event as

$$\delta \hat{h}_i = H_i A_i \delta h_i$$

$$= R_i \delta h_i \tag{3.66}$$

$$= H_i r_i^s$$

(3.66) is solved repeatedly until convergence when $\delta \hat{h}_i \approx 0$. However, if we substitute for the definition of r_i^s in (3.25) this implies that

$$R_i \delta h_i + H_i S_i \delta s \approx 0 \tag{3.67}$$

This is precisely the ambiguity declared by (3.65). It declares that changes in the hypocenter of the ith event, $\delta\widehat{\mathbf{h}}_i$, can be completely compensated for by a change in the station corrections, δs. (3.65) simply states this ambiguity of the single event location process that is an integral part of this procedure.

The lower partition of (3.65) has a quite different interpretation. To clarify this substitute (3.65) into (3.64) to obtain

$$\begin{bmatrix} \delta\widehat{\mathbf{h}} \\ \delta\widehat{\mathbf{s}} \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{HS} \\ \mathbf{0} & \mathbf{S}_N^+\mathbf{S}_N \end{bmatrix} \begin{bmatrix} \delta\mathbf{h} \\ \delta\mathbf{s} \end{bmatrix} \qquad (3.68)$$

where I have simplified (3.65) by combining all the upper partitions into those I have written as $\mathbf{R}$ and $\mathbf{HS}$. (3.68) demonstrates the remarkable fact that the estimate of the station corrections is independent of the hypocenters[17]. This is in marked contrast to the hypocenter estimates that we have seen are intimately connected to the station corrections. This is a significant observation. The difficult parameters to estimate in this problem are the hypocenters because they are nonlinear functions of the data. The station corrections, by contrast, are simple linear functions of the data. Thus it is intuitively clear that it is useful to obtain an estimate of s that does not depend on the hypocenters since they are more prone to error.

Having disposed of the resolution matrix, I turn now to a brief description of the covariance matrix for this procedure. The covariance matrix is important because it can be used to assess how second moment statistics of errors in the data propagate into the second moment statistics of errors in the estimates of the parameters. Obviously, the first requirement then is that we have to know something about the second moment statistics of the data errors.

---

[17] Realistically, $\delta\widehat{\mathbf{s}}$ is only locally independent of the hypocenters since $\mathbf{A}_i$ is based on a local linear approximation.

That is we assume we know the covariance of the data defined as

$$D_{ij} = \mathrm{E}[r_i^s r_j^s] \qquad i = 1, 2, \ldots, M$$

$$j = 1, 2, \ldots, M$$

where $\mathrm{E}[\,]$ denotes the expectation operator (see for example, Hoel [1971]). The M^2 numbers D_{ij} define a covariance matrix $\mathbf{D} \in \mathbb{R}^{M \times M}$ that can be written as

$$\mathbf{D} = \mathrm{E}[\mathbf{r}^s (\mathbf{r}^s)^T] \qquad (3.69)$$

We seek the covariance of the estimates, $\mathbf{C} \in \mathbb{R}^{N \times N}$, of the N parameters, $\hat{\mathbf{x}}$, that can be defined similarly as

$$\mathbf{C} = \mathrm{E}[\hat{\mathbf{x}} \hat{\mathbf{x}}^T] \qquad (3.70)$$

However, since the data are related to the estimate by (3.64) we can express $\mathbf{C}$ as

$$\mathbf{C} = \mathrm{E}[\mathbf{H}_B \mathbf{r}^s (\mathbf{H}_B \mathbf{r}^s)^T]$$

$$= \mathbf{H}_B \mathrm{E}[\mathbf{r}^s \mathbf{r}^{s^T}] \mathbf{H}_B^T \qquad (3.71)$$

$$= \mathbf{H}_B \mathbf{D} \mathbf{H}_B^T$$

An important special case[18] is that in which the data are statistically independent and of unit variance, for then (3.71) reduces to [Crosson, 1976a]

$$\mathbf{C} = \mathbf{H}_B \mathbf{H}_B^T \qquad (3.72)$$

If we substitute for the partitioned form of $\mathbf{H}_B$ given in (3.62), we see $\mathbf{C}$ is in the form

---

[18] This "special case" is not really so special, because if $\mathbf{D}$ is known *a priori* as we are assuming anyway, an appropriate weighting matrix can always be calculated (appendix A) that will rescale the data to be in this form (see also Lawson and Hanson [1974, pp. 193-195]).

$$C = \begin{bmatrix} \mathbf{H} \\ \mathbf{S}_N^+ \mathbf{U}_N^T \end{bmatrix} [\mathbf{H}^T \ \mathbf{U}_N (\mathbf{S}_N^+)^T] \tag{3.73}$$

$$= \begin{bmatrix} \mathbf{HH}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_N^+ (\mathbf{S}_N^+)^T \end{bmatrix}$$

since

$$\mathbf{HU}_N (\mathbf{S}_N^+)^T = \mathbf{V}\Lambda^H \mathbf{U}_R^T (\mathbf{S}_N^+)^T$$

$$= \mathbf{0}$$

because $\mathbf{U}_R^T \mathbf{U}_N = \mathbf{0}$. (3.73) has an even simpler structure that we can see by recognizing that $\mathbf{HH}^T$ is of the form

$$\mathbf{HH}^T = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 & \cdots & \mathbf{0} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_{m_e} \end{bmatrix}$$

where the individual partitions $\mathbf{C}_i \in R^{4 \times 4}$ are defined as

$$\mathbf{C}_i = \mathbf{H}_i \mathbf{H}_i^T$$

Thus $\mathbf{C}$ has the block diagonal form

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 & \cdots & \mathbf{0} & \mathbf{0} \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_{m_e} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{S}_N^+ (\mathbf{S}_N^+)^T \end{bmatrix} \tag{3.74}$$

This is a significant result. It demonstrates that in the special case when $\mathbf{D}=\mathbf{I}$ random errors in the estimates of the hypocenters of individual events are not coupled. Furthermore, random errors in the estimates of the hypocenters are not coupled to errors in the estimates of the station corrections.

4. SUMMARY

In this chapter I reviewed the major methods currently used for locating earthquakes. I grouped these methods into two major groups; single and multiple event location methods. Although other methods exist, I described details of only two multiple event location schemes; the master event method and the method of joint hypocenter determination. My principle motivation for describing these two methods is that both are closely related to the progressive multiple event location procedure I have introduced here. PMEL is, in a sense, a hybrid of these two methods. PMEL is a pseudo-master event method in which the "master event" is not a single event, but an average over the entire data set which Jordan and Sverdrup [1981] call the "hypocentroid". PMEL is also essentially a JHD method since it solves essentially the same equations. Its major difference, however, is that PMEL does not suffer from the explosive growth in computer storage required by JHD. PMEL is a significant advance in earthquake location because it combines the superior averaging qualities of multiple event location with the speed and small memory requirements of single event location. Consequently, I believe PMEL makes all competitive multiple event location methods obsolete. Furthermore, with application of some of the extensions proposed in section 3.2.3 it may render the same fate to conventional single event location methods.

CHAPTER 4
CONSTRUCTING A VELOCITY MODEL

1. INTRODUCTION

The arrival time of waves generated by a point source in an elastic medium depends on the wave propagation velocity by definition. Velocity is a number we use to convert times into distances. Thus, it is immediately obvious that we ought to be able to use arrival time measurements to tell us something about the seismic velocity structure of the earth. The fundamental problem we face in doing this is that there is not "a" seismic velocity of the earth. Instead, the earth is made up of a complex assemblage of many different kinds of materials with different seismic velocities. Hence, the earth does not have a single number that quantifies its seismic velocity but a whole range of velocities that vary from point to point within it. Consequently, a complete description of the velocity structure of the earth would require us to specify velocities at an infinite number of points. This is a task that cannot be done perfectly. The reason is that the amount of arrival time data we can collect will always be finite but the velocity structure requires (in principle at least) an almost infinite number of parameters to be described exactly. The effect of this is that we can never hope to obtain a perfect image of the velocity structure, but we must be content with some blurred version of it. This obscurity of geophysical inverse problems has been widely recognized since the work of Backus and Gilbert [1967,1968,1970].

Previous workers who have used arrival time data from earthquakes (e.g. Peters[1973], Crosson [1976a], or Aki and Lee [1976]) to estimate velocity structure have evaded the issue of nonuniqueness by imposition of a special, a priori form to the velocity structure to allow them to solve the problem by least squares. This approach suffers from the problem that it is difficult to know the effect this initial, fixed parameterization has on one's solution. A major goal of

my work is to break free of the need for any such *a priori* parameterization of the velocity model. Here the velocity model is allowed to be an arbitrary function whose uniqueness can be assessed by the techniques of Backus and Gilbert [1968, 1970] and Johnson and Gilbert [1972].

This chapter is devoted to the issue of how to construct a model that fits the observed data. Arrival times are nonlinear functionals of the location of the source and the velocity model. The method I describe here for constructing a solution is an iterative procedure based on a linearization of these nonlinear equations. It shares this property with Crosson's [1976a] nonlinear least squares procedure. The scheme I use here, however, takes the view that the hypocenters and station corrections are fundamentally different entities from the velocity model. That is, a given parameter (hypocenter coordinate or station correction) requires only a single number to be specified exactly while the velocity model is an unknown function that is fundamentally ambiguous. The approach I adopt here exploits this distinction between the parameters and the velocity model through the use of the annulling transformation (Pavlis and Booker [1980], Spencer and Gubbins [1980], and Rodi et. al. [1980]). One of the most significant properties of the annulling transformation is that it permits one to estimate the velocity model and the discrete parameters in separate calculations. The end result is a procedure I have dubbed PRIMEL (progressive inversion and multiple event location). The procedure is "progressive" because different parts of the model are adjust by separate, iterative procedures. This is in contrast to the "simultaneous" procedure of Crosson [1976a] in which all parts of the model are adjusted in a single linear step.

The procedure described in this chapter has two major limitations. The first is that the procedure described uses only first arrival (P wave) data and hence estimates only the compressional

wave velocity structure. This limitation was imposed only for simplicity and extension to shear waves as well should not be a formidable problem. The second limitation is that the algorithm assumes that velocity varies only as a function of depth and that any lateral velocity variations can be accounted for by station corrections. This assumption was made primarily to make the procedure computationally tractable in spite of the fact that it is probably not very good for many regions of the earth's crust. It is important to recognize, however, that the procedure can, in principle, be directly extended to three-dimensional velocity models. These possible extensions are a significant topic for future research.

2. LINEARIZATION

The data we wish to exploit are the arrival times of body waves from a point source at a position $\mathbf{h}'$ observed by a receiver at position $\mathbf{h}_i$ (Notation here is as defined in the preceding chapter.). The travel time from $\mathbf{h}'$ to $\mathbf{h}_i$ is related to the velocity model by the relation

$$T(h_{1_i}, h_{2_i}, h_{3_i}, h_1', h_2', h_3') = \int_{\mathbf{h}' \to \mathbf{h}_i} u(h_1, h_2, h_3) ds \qquad (4.1)$$

where u is the slowness as a function of position ($u = \dfrac{1}{v}$ where v is velocity). The integral is evaluated on a unique path (ray path) chosen so as to make T stationary (Fermat's principle) and ds is an element of length along that path. Equation (4.1) is really only a statement that the total travel time can be calculated by summing incremental travel times along a ray path. A major complication is that although for a given velocity model the rays are fixed curves in space, the ray paths depend upon the velocity model in a complicated way (see e.g. Aki and Richards [1980, pp. 84-100]). Because of this, travel times are nonlinear functionals of the velocity (slowness) structure. We now seek a local linearization of these

nonlinear equations.

A small change δu in slowness will produce a change in δT in the travel time in two ways: the path of the path of integration will change by an amount proportional to δu, and the travel time along the given path will change by an amount $\int_{\mathbf{h'} \to \mathbf{h}_i} \delta u \ ds$. Fermat's principle states that the contribution of the first is of order $(\delta u)^2$ and thus is negligible [Backus and Gilbert, 1969]. This implies that small changes in the slowness, δu, do not significantly alter the ray path. We need to extend this result, however, because the primary interest here is in earthquake sources, whose actual location is not known exactly. This requires that we also permit the source location (the endpoint of the ray path at $\mathbf{h'}$) to vary. Recall, however that ray paths are always normal to wavefronts (surfaces of constant travel time) [Aki and Richards, 1980, p. 91]. Consequently, small changes in the source coordinates will induce significant changes in the travel time only when they are made in directions parallel to the ray path. Thus, those source variations that most significantly vary the travel time do not alter the ray path substantially. (An important exception to this is those points near discontinuities in the travel time curves.) Thus small changes in the travel time are related to small changes in the model by the relation

$$\delta T_i = \frac{\partial T_i}{\partial h_1'}\delta h_1' + \frac{\partial T_i}{\partial h_2'}\delta h_2' + \frac{\partial T_i}{\partial h_3'}\delta h_3' + \frac{\partial T_i}{\partial h_4'}\delta h_4' + \int_{\mathbf{h'} \to \mathbf{h}_i} \delta u \ ds \quad (4.2)$$

where the ray path in the integral is the same as in (4.1) and the partial derivatives are all calculated based on the unperturbed velocity (slowness) model. δT_i is the variation in the travel time induced by changes in the model. It is synonymous with the residual defined in equation (3.3). Furthermore, the partial derivatives in (4.2) are identical to those in equation (3.2) that are used in hypocenter location. In fact, the only difference between equations

(3.2) and (4.2) is that in (4.2) we have added a term involving varia-
tions in the velocity model. Note that the linearized equations of
(4.2) are perfectly valid for a velocity model that can vary in all
three spatial dimensions. We could proceed directly from here
using a velocity model that varies three dimensionally. Tracing rays
in a laterally inhomogeneous velocity model is a formidable numeri-
cal problem (see e.g. Pereyra et. al. [1980] or Lee and Stewart
[1981]) as is the three-dimensional velocity inversion problem [Chou
and Booker, 1979]. Consequently,for this initial investigation I made
the conventional assumption in seismology that the velocity varies
only in the vertical direction.

With this assumption the integral in (4.2) can be written in a
simpler form. At depth h_3, the ray will make an angle θ with the
positive h_3 axis. Snell's law makes

$$p = u \sin\theta$$

a constant along the ray [Aki and Richards, 1980, p.92] and as a
result the integral in (4.2) can be written as [Slotnick, 1959, pp.
201-204]

$$\int_{\mathbf{h'} \to \mathbf{h_i}} \delta u \ ds = \int_0^L G_i(h_3) \delta u(h_3) dh_3 \tag{4.3}$$

where G_i depend on the source depth h_3'

$$G_i(h_3) = \begin{cases} \dfrac{u}{(u^2 - p^2)^{1/2}} & 0 < h_3 \leq h_3' \\ \dfrac{u}{(2u^2 - p^2)^{1/2}} & h_3' < h_3 < \text{depth of ray bottom}(u = p) \\ 0 & h_3 > \text{depth of ray bottom} \end{cases} \tag{4.4}$$

L in the integral in (4.3) is chosen as some depth below the bottom
of all rays. We identify the function $G_i(h_3)$ as a **Frechet derivative**
(Frechet derivatives are the functional equivalent of partial deriva-
tives. For a further description see Parker [1977a].) which I will

also call a **data kernel.**

It is convenient to introduce the nondimensional variable r defined as

$$r = \frac{L - h_3}{L} \qquad (4.5)$$

because then (4.3) can be written as

$$\int_{\mathbf{h}' \to \mathbf{h}_i} \delta u \; ds = \int_0^1 G_i(r) \delta u(r) dr \qquad (4.6)$$

where

$$G_i(r) = -\frac{1}{L} G_i(h_3)$$

If we now substitute (4.6) back into (4.2) and identify δT_i as the residual, r_i, we obtain

$$r_i = \frac{\partial T_i}{\partial h_1'} \delta h_1' + \frac{\partial T_i}{\partial h_2'} \delta h_2' + \frac{\partial T_i}{\partial h_3'} \delta h_3' + \frac{\partial T_i}{\partial h_4'} \delta h_4' + <G_i \,|\, \delta u > \qquad (4.7)$$

where I have introduced the inner product symbolism[1]

$$<G_i \,|\, \delta u> \equiv \int_0^1 G_i(r) \delta u(r) dr \qquad (4.8)$$

as a convenient shorthand[2].

---

[1] I have chosen to use this notation due to Dirac to denote inner products instead of the symbolism $(G_i, \delta u)$ that is more conventional in the geophysics literature. The Dirac notation is superior for this work because it makes a clear distinction between operators and functions as individual entities.

[2] The inner product notation itself is useful because it makes the results described here automatically more general. Equation (4.7) is equally valid for a three dimensional velocity model except in that case the inner product takes the form of a volume integral (Wu [1977] or Chou and Booker [1979]).

Suppose we have available a set of m_e earthquakes for which we record a total of M (equation (3.24)) arrival times. For each arrival time we can write an equation of the form (4.7). We can then assemble these M equations into the form

$$r = A\delta h + \overset{\leftrightarrow}{G}|\delta u>$$

(4.9)

where r is the vector of travel time residuals, A and δh are as defined in equation (3.30) and (3.31), and

$$\overset{\leftrightarrow}{G} = \begin{bmatrix} <G_1| \\ <G_2| \\ <G_3| \\ . \\ . \\ . \\ <G_M| \end{bmatrix}$$

(4.10)

$\overset{\leftrightarrow}{G}$ is a mathematical object which is a collection of the Frechet kernels that can be visualized as shown in equation (4.10). I will use a convention in which bold, capital letters with the $\leftrightarrow$ symbol above them will denote this type of entity. Entities like $\overset{\leftrightarrow}{G}$ can be loosely considered as matrices with a finite number of rows but an infinite number of columns.

There are two practical concessions that have to be made in the face of any real data. The first is that the assumption that the velocity model varies only vertically is never totally justified. I have attempted to reduce this potential problem through the use of station corrections. This adds a term to (4.2) involving a perturbation to the station corrections. A combination of the arguments given above and those in section 3.1 of chapter 3 yield the following equivalent to (4.9) when station corrections are used

$$r^s = Bx + \overset{\leftrightarrow}{G}|\delta u>$$

(4.11)

where r^s, B, and x are as defined in equations (3.26), (3.27), and (3.28).

The second pragmatic concession that we have to make is that the Frechet derivatives, $<G_i|$, in equation (4.4) generally have to be calculated numerically. I have chosen to do this by discretizing $|\delta u>$ on a regular grid of N_g points (N_g for the calculations in chapters 6 and 7 was always between 150 and 200.). These points are assumed connected by linear velocity (note that is not a linear slowness gradient) segments for which the incremental distances, $ds = G_i(h_3)dh_3$, can be calculated analytically [Slotnick, 1959, p. 207]. Thus although it is useful for theoretical purposes to view the operator $\overset{\leftrightarrow}{G}$ as a collection of functions, as a practical matter $\overset{\leftrightarrow}{G}$ is always approximated by a $M \times N_g$ matrix.

3. CONSTRUCTING THE ANNULLED DATA SET

3.1. Introduction

The set of equations in (4.9) and (4.11) both define what Pavlis and Booker [1980] (see also Appendix B) have termed "the mixed discrete-continuous inverse problem". We used this term because the complete model consists of two quite different entities. That is, the model consists of a finite set of numbers (parameters) plus a continuous function that requires an essentially infinite set of numbers to be described exactly. We showed that whenever the number of data is greater than the total number of discrete parameters, it is possible to construct a transformation we called the "annulling transformation" We adopted this term because of its connection to the null space of the matrix associated with the discrete parameters (see chapter 2 and appendix B). This transformation is useful for mixed inverse problems because it yields a set of data that are independent (in the present case only locally independent) of the discrete parameters. In this section I will describe the application of this method to the hypocenter-velocity inversion problem. This is done for two different cases; the

problem without station corrections (equation (4.9)) and the problem with station corrections included as free parameters (equation (4.11)). The first case has essentially already been solved in chapter 3 and requires only a minor modification of the methods used in ALGORITHM PMEL. The second case requires some additional manipulations.

3.2. Case I: without station corrections

The separation procedure we want to apply here is defined in terms of two operators. These are the two operators denoted in Appendix B by the symbols $\mathbf{H}$ and $\mathbf{U}_N$. $\mathbf{H}$ is a generalized inverse used to estimate the discrete parameters by the relation (B.6) and is required to satisfy (B.7). $\mathbf{U}_N$ is an orthogonal matrix that is related to $\mathbf{H}$ by (B.14). Because of this connection, the issue of how one estimates the discrete parameters cannot be totally ignored. If station corrections are not used, Pavlis and Booker [1980] pointed out that the special form of the matrix $\mathbf{A}$ in equation (4.9) can be exploited to great advantage. This is the same property that was exploited in chapter 3 in the development of the progressive multiple event location (PMEL) procedure. That is, individual earthquake hypocenters are coupled only through the velocity model (in chapter 3 the coupling was with station corrections), because of the block diagonal form of the matrix $\mathbf{A}$. Consequently, if the form (4.9) is assumed, a sensible method for estimating the hypocenters of the m_e earthquakes is to locate them by standard least squares single event methods. In section 3.2.2 of chapter 3, I demonstrated that although single event locations adjust the hypocenters of all m_e earthquakes independently (also iteratively), the final adjustment can still be viewed as a multiplication by the $4m_e \times M$ generalized inverse matrix $\mathbf{H}$ defined in equation (3.62). Furthermore, we identify the $M \times M$ matrix $\mathbf{U}$ defined in (3.42) and (3.45) as identical to that in (B.14), because the product $\mathbf{V}\Lambda_R^H$ in (3.42) is comparable to

the matrix $\mathbf{F}_R$ in (B.14). Thus, the annulled data set for this case can be constructed from the matrix $\mathbf{U}_N^T \in \mathbb{R}^{(M-4m_s) \times M}$ defined in (3.45).

A practical difficulty, however, is that the matrix $\mathbf{U}_N^T$ can attain a large size when the number of data is large. Fortunately, the special structure of $\mathbf{U}_N^T$ can be exploited by the same method used in ALGORITHM PMEL. To see how this can be accomplished, note that the only usage (see appendix B) made of the matrix $\mathbf{U}_N^T$ is to form the following two products:

$$\mathbf{r}_N^s = \mathbf{U}_N^T \mathbf{r}^s \qquad (4.12)$$

and

$$\ddot{\mathbf{G}}_N = \mathbf{U}_N^T \ddot{\mathbf{G}} \qquad (4.13)$$

We further observe that $\mathbf{r}_N^s$ is the same as the vector defined previously in (3.55). We saw in the previous chapter that it was natural to calculate $\mathbf{r}_N^s$ in blocks defined by the special structure of $\mathbf{U}_N^T$ (see equation (3.45)). Furthermore, by utilizing Lawson and Hanson's [1974, pp. 260-269,295-300] singular value decomposition routine (SVDRS) it is possible to arrange the calculations such that $\mathbf{U}_N^T$ is never formed explicitly, but a set of products are produced by an overwritting operation (see step 2i of PMEL). The identical procedure can be used here, except in this case we form the products in (4.12) and (4.13). A sketch of an algorithm to accomplish this is given below.

ALGORITHM NULLHYP

1) For $i = 1,2,3, \ldots, m_e$

    1i) Read data and location estimate for this event.

    2i) Form r_i^s, A_i, and $\overset{\ast\ast}{G}_i$

    3i) Execute ALGORITHM SVDRS [Lawson and Hanson, 1974, pp. 260-267,295-300]. SVDRS returns:

$$r_i^s \leftarrow U_i^T r_i^s$$

$$\overset{\ast\ast}{G}_i \leftarrow U_i^T \overset{\ast\ast}{G}_i$$

    4i) Write $(r_N^s)_i$ and $(\overset{\ast\ast}{G}_N)_i$ to mass storage

2) End

The symbol $(\overset{\ast\ast}{G}_N)_i$ in step 4i) is used to symbolize the product

$$(\overset{\ast\ast}{G}_N)_i = (U_N^T)_i \overset{\ast\ast}{G}_i \tag{4.14}$$

where $\overset{\ast\ast}{G}_i$ is the collection of Frechet kernels for the ith earthquake and $(U_N^T)_i$ is as defined in equation (3.41).

The most significant feature of NULLHYP is that it makes very efficient usage of computer memory. (It shares this property with PMEL since it uses almost the same algorithm.) The minimum storage required by NULLHYP is $5n_s + n_s N_g$ where we recall N_g is the number of grid points the model is discretized at. n_s cells are required to hold the residual vector, r^s; $4n_s$ cells are required to hold the matrix of partial derivatives, A_i; and $n_s N_g$ cells are required to hold the discretized Frechet kernels, $\overset{\ast\ast}{G}_i$.

Finally, it is necessary to point out why this case (no station corrections) has any relevance. With a one-dimensional velocity model the use of station corrections is, from my experience, always required. Thus the case considered here has, in this sense, no relevance to one-dimensional velocity inversion. I have included this discussion here, however, primarily because it has great importance to future work that is anticipated with three-dimensional

velocity inversion. The equations one has to solve in the three-dimensional case are identical to (4.9) except the $<G_i|\delta u>$ term must be specified on a three dimensional grid[3]. In that case N_g can attain an enormous size so the storage reduction achieved by NULLHYP is not a moot point.

3.3. Case II: with station corrections

When station corrections are needed, I have chosen to estimate them by ALGORITHM PMEL described in the previous chapter. Other methods (e.g. JHD) could be used, but I consider PMEL a superior algorithm because of its minimal storage requirements and lack of required constraints. I noted above that in spite of the fact that PMEL uses iterative improvement techniques on several levels, the final estimate of the parameters (m_e hypocenters and n_s station corrections) returned by PMEL can still be viewed as a linear perturbation from a reference solution using the generalized inverse in equation (3.61). The matrix $\mathbf{H}_B$ used there defines half of what is needed for the application of the separation procedure (appendix B) to the system of equations in (4.11). To complete the procedure we have to find the operator $\mathbf{U}_N$ (equation (B.14)) that has the fundamental property stated in equation (B.21). The starting point for this case is equation (3.54). The annulling transformation can be had directly from the matrix $\mathbf{S}_N \in \mathrm{R}^{(M-4m_e)\times n_s}$. This can be accomplished by calculating a QR decomposition (theorem 2.1) of $\mathbf{S}_N$.

$$\mathbf{S}_N = \mathbf{QR} \tag{4.12}$$

---

[3] Chou and Booker [1979] describe a technique that actually avoids the need to specify these functions on a three-dimensional grid by reducing all terms involving these functions to a set of line integrals. This is promising because it could be utilized to further reduce computer memory requirements in a practical implementation of their procedure.

where $\mathbf{Q} \in \mathrm{R}^{(M-4m_s) \times (M-4m_s)}$ and $\mathbf{R} \in \mathrm{R}^{M-4m_s \times n_s}$ are of the form

$$\mathbf{Q} = [\mathbf{Q}_s \, \mathbf{Q}_n] \qquad (4.13)$$

and

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_s \\ \mathbf{0} \end{bmatrix} \qquad (4.14)$$

Note that the upper triangular matrix $\mathbf{R}_s \in \mathrm{R}^{n_s \times n_s}$ and the orthogonal matrix $\mathbf{Q}_s \in \mathrm{R}^{M-4m_s \times n_s}$ are the same as in equation (3.59). The more important matrix for the present is the matrix $\mathbf{Q}_N \in \mathrm{R}^{(M-4m_s) \times (M-4m_s-n_s)}$, because it leads directly to the matrix we are seeking here. To see this, observe that because of the form of $\mathbf{R}$ in (4.14), $\mathbf{Q}_N^T$ will annihilate $\mathbf{S}_N$. Consequently, the product $(\mathbf{U}_N^T)_B \in \mathrm{R}^{(M-4m_s-n_s) \times M}$ defined as

$$(\mathbf{U}_N^T)_B \equiv \mathbf{Q}_N^T \mathbf{U}_N^T \qquad (4.15)$$

is easily seen to satisfy

$$(\mathbf{U}_N^T)_B \, \mathbf{B} = 0 \in \mathrm{R}^{M_n \times (4m_s+n_s)} \qquad (4.15)$$

where

$$M_n = M - 4m_e - n_s \qquad (4.16)$$

Equation (4.15) states that $(\mathbf{U}_N^T)_B$ satisfies the equivalent of equation (B.21) so it is the annihilation operator we were seeking.

We now face the issue of how to efficiently perform the annulling transformation defined by $(\mathbf{U}_N^T)_B$. A useful way to do this is by a two step procedure virtually identical to that used in ALGORITHM PMEL. To see how this can be done, note as before that the only use we make of $(\mathbf{U}_N^T)_B$ is to form the products

$$\mathbf{n} = (\mathbf{U}_N^T)_B \, \mathbf{r}^s \qquad (4.17)$$

and

$$\mathring{\mathbf{N}} = (\mathbf{U}_N^T)_B \, \mathring{\mathbf{G}} \qquad (4.18)$$

which are analogous to equations (4.12) and (4.13). We note, however, that because $(\mathbf{U}_N^T)_B$ can be factored into the matrix product in equation (4.15), equation (4.17) and (4.18) can be rewritten as

$$\mathbf{n} = \mathbf{Q}_N^T \mathbf{U}_N^T \mathbf{r}^s \qquad (4.19)$$

$$= \mathbf{Q}_N^T \mathbf{r}_N^s$$

and

$$\mathring{\mathbf{N}} = \mathbf{Q}_N^T \mathbf{U}_N^T \mathring{\mathbf{G}} \qquad (4.20)$$

$$= \mathbf{Q}_N^T \mathring{\mathbf{G}}_N$$

where $\mathbf{r}_N^s$ and $\mathring{\mathbf{G}}_N$ are as defined in equations (4.12) and (4.13). Equations (4.19) and (4.20) illustrate what I meant above when I stated that the annulling transformation for this case can be done in two steps. The first step is the multiplication by the matrix $\mathbf{U}_N^T$ and the second step is the multiplication by $\mathbf{Q}_N^T$. In the previous section I described an algorithm (NULLHYP) for accomplishing the first step. Only a minor modification of that algorithm is needed to accomplish what we need here. A sketch of this modified version of NULLHYP is given below under the title ALGORITHM NULLWSC.

ALGORITHM NULLWSC

1) For $i = 1,2,3, \ldots, m_e$

    1i) Read data and location estimate for this event.

    2i) Form r_i^s, A_i, S_i, and $\overset{\ast}{G}_i$

    3i) Execute ALGORITHM SVDRS [Lawson and Hanson, 1974, pp. 260-267,295-300]. SVDRS returns:

$$r_i^s \leftarrow U_i^T r_i^s$$

$$S_i \leftarrow U_i^T S_i$$

$$\overset{\ast}{G}_i \leftarrow U_i^T \overset{\ast}{G}_i$$

    4i) Write $(r_N^s)_i$, $(S_N)_i$, and $(\overset{\ast}{G}_N)_i$ to mass storage

2) End

NULLHYP and NULLWSC differ only in that NULLWSC manipulates the matrix S_i in steps 2i), 3i), and 4i). This modification is necessary because the matrix Q_N^T is calculated from the matrix S_N which can be constructed by assembling the partitions $(S_N)_i$ as shown in equation (3.56).

Calculating the matrix Q_N^T and forming the products in equations (4.19) and (4.20) could be done by brute force. That is, one could assemble $S_N, r_N^s,$ and $\overset{\ast}{G}_N$; calculate the QR decomposition of S_N; and then calculate the products (4.19) and (4.20). When the number of data, M, is not extremely large this would be a reasonable way to proceed. When M is large, however, we quickly face a formidable storage problem, since the brute force approach requires storage of at least $(n_s+1)(M-4m_e)+(M-4m_e)N_g$ numbers.

A useful alternative is an approach based on sequential accumulation, similar to the approach used in PMEL. Details of how this algorithm works can be found in Lawson and Hanson's [1974, chapter 27] text, but I will highlight the key points here. Remember first that Q_N^T is derived from a QR decomposition of the matrix S_N. In addition, recall that S_N can be assembled from the

individual partitions $(\mathbf{S}_N)_i$ calculated by NULLWSC (see equation (3.56)). The basic idea of sequential accumulation is that $\mathbf{Q}_N^T$ is assembled from a sequence of matrices calculated from the individual $(\mathbf{S}_N)_i$. The net result is that $\mathbf{Q}_N^T$ is calculated as the product of m_e orthogonal matrices

$$\mathbf{Q}_N^T = \mathbf{Q}_{m_e}^T \mathbf{Q}_{m_e-1}^T \cdots \mathbf{Q}_2^T \mathbf{Q}_1^T \qquad (4.21)$$

Each $\mathbf{Q}_k^T$ is, however, calculated completely from the partition $(\mathbf{S}_N)_k$ of $\mathbf{S}_N$ and is of the special form

$$\mathbf{Q}_k = \begin{bmatrix} \mathbf{Q}_{11} & 0 & \mathbf{Q}_{13} & 0 \\ 0 & \mathbf{I} & 0 & 0 \\ \mathbf{Q}_{31} & 0 & \mathbf{Q}_{33} & 0 \\ 0 & 0 & 0 & \mathbf{I} \end{bmatrix} \qquad (4.22)$$

where $\mathbf{Q}_{11} \in R^{n_s \times n_s}$ is diagonal. $\mathbf{Q}_{31} \in R^{(m_k-4) \times n_s}$, $\mathbf{Q}_{13} \in R^{n_s \times (m_k-4)}$, and $\mathbf{Q}_{33} \in R^{(m_k-4) \times (m_k-4)}$ are calculated from a sequence of Householder transformations as described by Lawson and Hanson [1974, p. 210]. $\mathbf{Q}_k$ is "special" because it differs little from an identity matrix. It has only four partitions that differ from an identity matrix and they occupy a special position. The important point is that the structure of the $\mathbf{Q}_k$ matrices allows one to process the data from each event separately in a manner that permits a substantial storage reduction. A fixed storage of $2(n_s^2 + n_s + n_s N_g)$ cells is required. This storage scheme consists of the following elements:

(1) $2n_s^2$ cells are required to hold the matrix $\mathbf{R}_s$ and the partitions $(\mathbf{S}_N)_k$ of $\mathbf{S}_N$. The n_s^2 cells of $\mathbf{R}_s$ are used to accumulate the upper triangular matrix of the final QR factorization of $\mathbf{S}_N$ (equation (3.59) and (4.14)). The remaining n_s^2 cells are used as a scratch storage area to hold the individual partitions, $(\mathbf{S}_N)_k$, in the accumulation process.

(2) $2n_s$ cells are required to hold the vector r_Q (equation (3.60)) and the partitions $(\mathbf{r}_N^s)_k$ of the data vector $\mathbf{r}_N^s$. r_Q is accumulated in n_s cells as the algorithm proceeds. n_s additional cells are required as a work space to hold each $(\mathbf{r}_N^s)_k$ vector. These numbers are ultimately overwritten by the annulled data values n_i.

(3) $2n_s N_g$ cells are required to hold the discretized kernel functions. $n_s N_g$ cells are used to accumulate the kernels, $\overset{\leftrightarrow}{\mathbf{X}}$, that are used later to appraise the uniqueness of the station corrections (see chapter 5 and appendix B). The remaining $n_s N_g$ cells are used as a scratch area to hold the the discretized kernels $(\overset{\leftrightarrow}{\mathbf{G}}_N)_k$. This array is overwritten by the data kernels for the annulled data derived from the kth event which I write as $\overset{\leftrightarrow}{\mathbf{N}}_k$.

A sketch of how this algorithm proceeds is given below with the title ALGORITHM NULLSTA. I have omitted two details for the sake of simplicity. These are

(1) How to actually calculate the $\mathbf{Q}_k$ matrices of (4.22).

(2) Bookkeeping details involving pivot elements and some special treatment that is required for the first n_s data points.

These details can be had by consulting Lawson and Hanson's text.

ALGORITHM NULLSTA

1) For $k = 1, 2, \ldots, m_e$

   1k) Read output of NULLWSC into work areas: $(r_N^s)_k$, $(S_N)_k$, and $(\overset{*}{G}_N)_k$

   2k) Execute sequential accumulation routine (modification of SEQHT from Lawson and Hanson [1974, p. 210]) which returns:

$$\left[\frac{R_s}{(S_N)_k}\right] \leftarrow \left[\frac{(Q_{11}^T)_k R_s + (Q_{31}^T)_k (S_N)_k}{(S_N)_k}\right]$$

$$\left[\frac{r_Q}{(r_N^s)_k}\right] \leftarrow \left[\frac{r_Q}{n_k}\right] = \left[\frac{(Q_{11}^T)_k r_Q + (Q_{31}^T)_k (r_N^s)_k}{(Q_{13}^T)_k r_Q + (Q_{33}^T)_k (r_N^s)_k}\right]$$

$$\left[\frac{\overset{*}{X}}{(\overset{*}{G}_N)_k}\right] \leftarrow \left[\frac{\overset{*}{X}}{\overset{*}{N}_k}\right] = \left[\frac{(Q_{11}^T)_k \overset{*}{X} + (Q_{31}^T)_k (\overset{*}{G}_N)_k}{(Q_{13}^T)_k \overset{*}{X} + (Q_{33}^T)_k (\overset{*}{G}_N)_k}\right]$$

   3k) Write n_k and $\overset{*}{N}_k$ to mass storage.

2) End

It is quite obvious that NULLSTA is considerably more complex than the "brute force" approach of a direct QR decomposition of S_N. It has two significant advantages, however, over the brute force approach

(1) The storage required is fixed and of a modest size.

(2) Its input (blocks of data defined by individual earthquakes) interfaces nicely with NULLSTA and its output interfaces nicely with the data compression method that is described in the following section.

4. CONSTRUCTING A PERTURBATION TO THE VELOCITY MODEL

No matter how we choose to construct it, the fundamental result of the previous section is that we can obtain a set of M_n (equation (4.16)) equations of the form

$$n_i = <N_i \,|\, \delta u >\qquad (4.23)$$

These M_n equations constitute what Pavlis and Booker [1980] term the "annulled data set" They have the useful property that they are independent (here only locally independent since this is a linearization of a nonlinear problem) of the estimated hypocenters and station corrections. As a result they can be used directly to construct an estimate of a perturbation to the velocity (slowness) model, $|\delta u>$.

There are a number of methods one could use to construct a perturbation to the velocity model. Following Backus and Gilbert [1969] and Kennett [1976] I have chosen to use what Kennett calls the "flattest model" This name is appropriate because this solution is the unique one that minimizes

$$\left\|\frac{d}{dr}(\delta u\,)\right\|^2 = \int_0^1 \left[\frac{d}{dr}(\delta u\,)\right]^2 dr\qquad (4.24)$$

subject to the M_n side conditions (4.23). (i.e. the estimate is required to fit the data.) In principle, this solution is straightforward. The perturbation at a given position r_0 can be calculated as

$$\delta \hat{u}(r_0) = \sum_{i=1}^{M_n} w_i n_i\qquad (4.25)$$

where the constants w_i are calculated from the normal equations

$$\underline{N}w = h\qquad (4.26)$$

The components of the matrix $\underline{N}$ and the vector h are given by

$$\underline{N}_{ij} = <\underline{N}_i \,|\, \underline{N}_j >$$

and

$$h_i = <H(r-r_0)|\underline{N_i}>$$

where $<H(r-r_0)|$ is the Heaviside step function. The functions $<\underline{N_i}|$ are related those in (4.23) by

$$<\underline{N_i}(r)| = \int_0^r N_i(\rho)d\rho \qquad (4.27)$$

and these in turn are related to the annulled data by[4]

$$n_i = \delta u(1)\underline{N_i}(1) - <\underline{N_i}|\frac{d}{dr}(\delta u)> \qquad (4.28)$$

(see Backus and Gilbert [1969], Johnson and Gilbert [1972], and Kennett [1976])

Although in principle the solution defined in (4.25) and (4.26) seems reasonable, it almost never works in practice because the matrix $\underline{N}$ is generally singular or at least very ill conditioned. A useful method for dealing with this problem is the procedure termed "ranking and winnowing" by Gilbert [1971] and "spectral expansion" by Parker [1977a] and Oldenburg [1977]. This technique makes use of the fact that the matrix $\underline{N}$ in (4.26) is positive definite and symmetric. As a result, it can be factored as [Parker, 1977a][5]

---

[4] The solution of (4.26) inherently assumes the surface velocity is known. In that case $\delta u(1)$ in (4.28) is 0 causing the $\delta u(1)\underline{N_i}(1)$ to vanish. If the surface velocity is unknown the problem can be avoided by adding the constraint that $\sum_{i=1}^{M_n} w_i \underline{N_i}(1)=1$. This constraint causes any estimate to be independent of the surface velocity. (For further details see chapter 5 or Johnson and Gilbert [1972].) Similar constraints can also be imposed to remove the effect of unknown discontinuities in the model at a known depth (see Johnson and Gilbert's treatment of the velocity jump at the core-mantle boundary.)

[5] The actual method I have invoked to numerically perform this transformation follows a suggestion made by Parker [1977b]. This

$$\mathbf{O}^T \underline{\mathbf{N}} \mathbf{O} = \lambda$$

where $\mathbf{O} \in \mathbf{R}^{M_n \times M_n}$ is an orthogonal matrix and λ is a diagonal matrix with all positive elements. The elements of λ are the eigenvalues, λ_i, of the matrix $\underline{\mathbf{N}}$ which are assumed to be arranged such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{M_n}$ [Parker, 1977a]. The usefulness of the matrix $\mathbf{O}$ is that it can be used to construct a set of functions $<\underline{\Psi}_i|$ defined by the relation

$$<\underline{\Psi}_i| = \frac{1}{\lambda_i^{1/2}} \sum_{j=1}^{M_n} O_{ji} <\underline{N}_j| \qquad (4.29)$$

which satisfy

$$<\underline{\Psi}_i|\underline{\Psi}_j> = \delta_{ij} \qquad (4.30)$$

(i.e. the $<\underline{\Psi}_i|$ are an orthonormal set of functions [Parker, 1977a].) The analysis given here departs slightly from that in Parker's paper at this point because I am using the "flattest perturbation". It is imperative to recognize that the functions $<\underline{N}_j|$ are not the same as those in (4.23) but have been integrated (equation (4.27)). Consequently, the functions $<\underline{N}_j|$ are not related directly to $|\delta u>$, but its first derivative (equation (4.28)). Thus, the functions $<\underline{\Psi}_i|$ are related to the annulled data by the relation

$$a_i = <\Psi_i| \frac{d}{dr}(\delta u)> \qquad (4.31)$$

---

procedure works by calculating a singular value decomposition of the discrete representation of the operator $\overset{\circ}{\underline{N}}$ (This operator is made up of the kernels in (4.27).) in which the columns of $\overset{\circ}{\underline{N}}$ are weighted by an appropriate set of quadrature coefficients. This approach is equivalent theoretically to the approach described above but it is preferable numerically because it is less sensitive to computational roundoff errors [Parker, 1977b]. I have adopted Parker's [1977a] description above because it is more comprehensible.

where the numbers a_i are given by

$$a_i = \frac{1}{\lambda_i^{1/2}} \left[\sum_{j=1}^{M_n} O_{ji} \delta u (1) \underline{N}_i (1) - \sum_{j=1}^{M_n} O_{ji} n_i \right] \qquad (4.32)$$

(see comments in footnote 4 concerning the $\delta u (1) \underline{N}_i (1)$ term.) The numbers a_i are useful because we can consider an expansion of the function $|\frac{d}{dr}(\delta u)\rangle$ in terms of these orthogonal functions as

$$|\frac{d}{dr}(\delta u)\rangle = \sum_{i=1}^{M_n} a_i |\underline{\Psi}_i\rangle + |\underline{\Psi}^*\rangle$$

where $|\underline{\Psi}^*\rangle$ is an annihilator [Parker, 1977a]. If the residuals are weighted by their expected standard error (equation (3.15), the annulled data are statistically independent with unit variance (see appendix B). As a result the variance of each coefficient a_i in the expansion (4.33) is $\frac{1}{\lambda_i}$. It follows that since the λ_i are assumed to be arranged in a decreasing sequence that the expansion of $|\frac{d}{dr}(\delta u)\rangle$ in (4.33) is in terms of a set of functions whose coefficients increase in uncertainty; after we reach $|\underline{\Psi}_{M_n}\rangle$ we reach $|\underline{\Psi}^*\rangle$ whose uncertainty is total [Parker, 1977a]. In every case I have dealt with it is always the case that the $\langle\underline{\Psi}_i|$ become more oscillatory as i increases. This means that the smoothest functions are most accurately determined since their coefficients have smaller errors.

I have utilized the spectral expansion technique in a way that is different from that advocated by previous workers. Parker [1977a], Gilbert [1971], Oldenburg [1979], and others have used this technique primarily as a method for dealing with the inevitable ill conditioned nature of inner product matrices like $\underline{N}$ in (4.26). I have used it, on the other hand, primarily as a tool for data compression. This is motivated primarily by the fact the $\underline{N}$ can attain a large size

when the number of data is large (the usual case with arrival time data). On the other hand, the fact that $\underline{\mathbf{N}}$ is ill conditioned is really a statement that the data are redundant. If the data are redundant, it makes little sense to carry them all along directly and form the giant $M_n \times M_n$ matrix $\underline{\mathbf{N}}$. It would be desirable to have a method of extracting meaningful averages from blocks of data of a manageable size. These averages could then be assembled into a smaller set of normal equations. The spectral expansion technique provides a tool for doing just that. Implementing this requires a sequence of four major steps;

(1) Divide the annulled data set into groups of M_b blocks that are each of a manageable size.

(2) Apply the spectral expansion procedure to each of the M_b blocks of data. In this step, each block is treated independently.

(3) "Winnow" the data from each block. That is, keep only those a_i's and associated kernel functions that have a reasonably small uncertainty.

(4) Assemble the winnowed data from all blocks of data and estimate $|\delta u>$ from them.

The key step above is (3). It is justified on the basis of the following observation. One view of the spectral expansion technique is that it is a way of extracting a set of independent averages, a_i, of the original data and ranking them according to their relative uncertainty. Since the data are redundant, some of these averages are not very useful and we would be better off excluding them. By throwing out those a_i associated with small λ_i (eigenvalues) we are doing just that. The only question that then remains is, what is a "reasonably small uncertainty" ? That question is inseparably linked to the precision of the raw data since throwing away a_i's introduces a certain level of misfit into the observation. Parker [1977a] shows, however, that the misfit introduced by deleting a given a_i is proportional to

the size of its associated eigenvalue, λ_i. Thus an a_i associated with a small λ_i contributes little to the misfit compared its contribution to the solution or its uncertainty [Parker, 1977a]. In implementing this idea I have found that even if one takes the ultra conservative approach of only keeping a_i's whose uncertainty is within machine precision (13 significant figures on the CDC machine used here) less than 50% of the data survive the winnowing operation. I have adopted a less extreme but still conservative winnowing criteria. Only those a_i with eigenvalues larger than 10^{-4} were kept. This is still conservative because the largest eigenvalues were always of the order of 10^4 so this is only slightly different from a criteria based on machine precision. Based on this criteria, only 10%-20% of the a_i's were typically kept. This indicates that travel time data from earthquake sources are highly redundant.

The winnowing operation of step (3) above is significant because it makes step (4) tractable for large data sets. That is, winnowing, when combined with the data compression achieved by the annulling transformation, typically produces a 90% reduction in the number of "data" we need to handle in this final step. Consequently, inversion of data set of several thousand arrival times is quite conceivable. (The largest number of arrivals I have handled to date is slightly greater than 800 (chapter 7).) Once this reduction has been accomplished, a perturbation can be estimated by the same set of normal equations described above (equations (4.25) and (4.26)) except the functions $<\underline{N}_j|$ are replaced by the functions $<\underline{\Psi}_i|$ and the numbers n_i are replaced by the numbers a_i. This still requires us to invert an inner product matrix like that in (4.26). (The $<\underline{\Psi}_i|$ constructed as described above are orthogonal only to those $<\underline{\Psi}_j|$ constructed from the same block of data. There is no assurance they are orthogonal to functions from different blocks of data.) Unfortunately, this matrix is generally also ill conditioned and cannot be inverted by elementary methods. There are two

established ways of circumventing this problem;

(1) Apply the spectral expansion procedure again. In this case it is used in the manner advocated by Parker [1977a] or Gilbert [1971].

(2) Damped (Levenberg-Marquardt) inverse (see chapter 2)

The former has the advantage of consistency with the previous step, but suffers from the disadvantage of being somewhat slower to compute [Gilbert, 1971]. Because of this, for the purpose of constructing a single perturbation for a given iteration of this procedure, the damped solution is preferable. Selecting the degree of damping required for stability is facilitated by studying the trade-off curves for the velocity model (chapter 5) to which a damped solution closely related.

5. ITERATIVE METHODS

The discussion in chapter 3 and the preceding sections of this chapter described a whole host of building blocks. In this section these building blocks are combined to yield a larger scale algorithm that I have given the title PRIMEL (Progressive Inversion and Multiple Event Location). When fed a set of arrival time data this procedure will ultimately return a set of estimates for the hypocenters of all the earthquakes in the data set, estimates of station corrections, and a compressional wave velocity model specified as a continuous function of depth. The gross structure of PRIMEL is shown below.

ALGORITHM PRIMEL

1) Gather data and select an initial velocity model
2) Loop from 3) to 11) till convergence

   3) Calculate travel time table from current velocity model

   4) Execute ALGORITHM PMEL (chapter 3)

   5) Execute ALGORITHM NULLWSC (chapter 4)

   6) Execute ALGORITHM NULLSTA (chapter 4)

   7) Fragment output of NULLSTA into M_b blocks of data

   8) For $i = 1, 2, \ldots, M_b$

      1i) Calculate spectral expansion of this block of data

      2i) Winnow data

   9) Assemble winnowed data and calculate slowness perturbation, $|\delta u>$, as the flattest perturbation (minimizes equation (4.24)).

   10) $|u> \leftarrow |u> + |\delta u>$

11) End Loop
12) Error appraisal (chapter 5)

Several aspects of this algorithm require some additional comments. The first is that selecting the "initial velocity model" in step 1) is not totally arbitrary. Since this is an iterative procedure based on a linear approximation to a set of nonlinear equations, one normally wants the "initial model" to be as close as possible to the truth. There are few areas of the world were there is not at least some gross estimate of the seismic velocity for the region. These estimates are, however, almost always parameterized as a set of constant velocity layers. Such layered models are not directly suitable as initial models form PRIMEL. The reason is that perturbations calculated by PRIMEL (step 9) are always smooth (flattest perturbation) functions. Because of this, layer boundaries tend to persist indefinitely. A solution to this problem is based on the idea of resolution functions. These functions are fundamental to the analysis of errors in the velocity model because the perturbation one calculates can always be viewed as the true, unknown

perturbation smoothed by these functions [Backus and Gilbert, 1968]. How these functions are calculated and details of their interpretation are the subject of section 3 of chapter 5. For the present, it is sufficient to recognize that they indicate the degree the true velocity model is smoothed by PRIMEL to yield an estimate at a given depth. Because of this it is useful with any starting model to execute only steps 1 to 8 of PRIMEL using that initial model. The resolution functions can then be calculated from the winnowed data produced by step 8 and the starting model can then be smoothed by these functions. The resulting smoothed velocity model can then be used as a starting model. This is a useful approach because it guarantees that the final result will not contain any components that are an artifact of unresolvable structure contained in the starting model.

A second aspect of PRIMEL that requires additional comment is step 2. The statement itself is straightforward enough until one asks the question of how "convergence" is defined. Convergence of an iterative procedure usually implies two things;

(1) The model is not changed significantly by the final iteration (i.e. the perturbations are relatively small everywhere).

(2) The model yields travel times that fit the data.

The problem is that both of these criteria are rather ill defined because of the inherent nonuniqueness in inverse problems. To be more specific, criteria (1) is hazy because it depend on what perturbation you are talking about. That is, there are infinitely many perturbations we could calculate. Thus, the fact that the one I happen to choose is small could be totally irrelevant as it may indicate that the current estimate lies near a local minimum. Furthermore, criteria (2) is also difficult to specify unambiguously. The reason for this is that we cannot easily apply the usual statistical measures of misfit such as χ^2 [Wolberg, 1967, pp. 59-60] from more conventional regression analysis. The reason is that one of the entities we are

trying to estimate here (the velocity model) requires essentially an infinite number of parameters to be specified precisely. In the jargon of regression analysis this means that the number of degrees of freedom of the system is infinite [Parker, 1977a] rendering statistical tests such as χ^2 of questionable value.

In chapter 7 I describe a rather involved method for detecting convergence that essentially requires both criteria (1) and (2) to be satisfied in a prescribed way. The problem is that the method I describe there suffers from two significant problems. First, it is usually too conservative and secondly it requires too much of a subjective judgement by an analyst. In short, the method I describe there is one way of detecting convergence but that method is not entirely satisfactory. Alternatives are an important subject for future research.

Finally, it seems appropriate to stress what PRIMEL is and is not. First, I reiterate that it is not a "simultaneous" procedure in the sense of Crosson's [1976a] procedure. Following Roecker [1982] I have adopted the term "progressive" to describe it. In a "simultaneous" procedure all parts of the model are adjusted at once. In this "progressive" scheme different parts of the model that are physically distinct are adjusted in independent, iterative steps.

A second feature of PRIMEL is that it contains features that make it, in some respects, a least squares procedure while at the same time retaining a respect for the fundamental ambiguity of the velocity model. That is, hypocenters and station corrections are both estimated by least squares procedures while the velocity model is estimated as a smooth function through a solution technique similar to that of Backus and Gilbert [1969] and Johnson and Gilbert [1972]. Moreover, because of the data compression achieved by the application of the spectral expansion method PRIMEL is capable of digesting extremely large data sets; a property formerly claimed only for parameterized least squares procedures.

In summary PRIMEL is a significant advance as a method of constructing a solution to the inverse problem of using earthquake arrival time data to determine earthquake hypocenters and seismic velocity structure. One of its most significant advantages, however, remains to be discussed. That is, the estimates yielded by PRIMEL are amenable to application of the powerful error appraisal tools of Backus and Gilbert [1968, 1970] and Backus [1970a,b]. This is the topic of the next chapter.

CHAPTER 5
ERROR ASSESSMENT

1. INTRODUCTION

The material of this chapter is important because it represents one of the most significant advances of the work presented in this dissertation. Following the lead of Peters [1973], Crosson [1976], and Aki and Lee [1976] most previous workers have chosen to characterize the velocity model as a finite number of parameters. This approach can be criticized on the grounds that one cannot ascertain the affect of this arbitrary, *a priori* parameterization. In the previous chapter I described a procedure for constructing a velocity model that was not encumbered by any *a priori* specification of the form of the velocity model. Instead the velocity model was allowed to be any arbitrary piecewise continuous function of depth. The reason for doing this is that it allows me to bring the powerful error assessment techniques of Backus and Gilbert [1968, 1970] and of Backus [1970,a,b,1971] to bear upon this problem.

The material of this chapter is divided into three distinct parts. Section 2 is a review of the generalized prediction formulation of Backus [1970a,b,1971]. Sections 3 and 4 then describe details of the assessment techniques that I have applied to estimate the errors in the two different parts of the model (the velocity model and the discrete parameters). It is convenient to start with Backus' general formulation because it illustrates that the two different error assessment techniques described in sections 3 and 4 are computationally identical. Their difference lies only in the way we interpret the results.

Applying this analysis has a severe limitation that must be recognized at the outset. The equations I will work with here are produced by a local linearization of a set of nonlinear equations.

The analysis is predicated on the assumption that the reference model (the one constructed by the procedure described in the previous chapter) is sufficiently close to the truth that the linear approximation is valid.

2. GENERALIZED PREDICTION

In a series of three papers Backus [1970a,b,1971] described a very general formulation of the linear inverse problem. In this section I will review several features of this formulation that are relevant to the work presented here. My discussion will, however, be considerably less general than that of Backus, since this is a specific application of his general theory to the hypocenter-velocity inverse problem.

We start with what we have available. That is, the M residuals from a set of m_e earthquakes and m_{blasts} explosion sources that are the observed data. After applying the annulling transformations described in the previous chapter and in Appendix B we can obtain a set of $M_n = M - 4m_e - n_s$ that are related to the velocity model by[1]

$$n_i = <N_i \mid \delta u> \qquad i = 1,2,3, \ldots, M_n \qquad (5.1)$$

The general problem I wish to consider here is that of predicting a member p that is related to $|\delta u>$ by a linear relation of the form

$$p = <P \mid \delta u> \qquad (5.2)$$

I will henceforth refer to the functions $<N_i \mid$ as the **data kernels,** and I will refer to the function $<P \mid$ as the **prediction kernel.**

---

[1] If some of the explosion sources have known spatial location but an unknown origin time (a common situation) then the origin times of these sources can also be treated as free parameters. If m_{ot} of the m_{blasts} explosion have unknown origin times then we would have $M_n = M - 4m_e - n_s - m_{ot}$.

The basic idea of Backus' formulation is to estimate p with a linear combination of the M_n data. I will denote this estimate by the symbol $\hat{p}$. $\hat{p}$ is defined symbolically as

$$\hat{p} = \sum_{i=1}^{M_n} w_i n_i \tag{5.3}$$

where the numbers w_i are constants that are to be determined. By substituting for n_i from (5.1) we see that (5.3) can be written as

$$\hat{p} = \sum_{i=1}^{M_n} w_i <N_i \,|\, \delta u> \tag{5.4}$$

If I now define

$$<\hat{P}\,| = \sum_{i=1}^{M_n} w_i <N_i\,| \tag{5.5}$$

then (5.4) can be rewritten as

$$\hat{p} = <\hat{P}\,|\,\delta u> \tag{5.6}$$

Thus, it is clear that $\hat{p}$ will be a best estimate of p when the weights, w_i, are specified so that $<\hat{P}\,|$ is a best approximation to the prediction kernel, $<P\,|$. A quantitative measure of this approximation is provided by the function $<\varepsilon_p\,| = <P\,| - <\hat{P}\,|$. Normally, one could calculate a set of weights, w_i, that would minimize the size of the function $<\varepsilon_p\,|$ as measured by the norm

$$\|\varepsilon_p\|^2 = <\varepsilon_p\,|\,\varepsilon_p> \tag{5.7}$$

However, for the problem I am considering here this is not feasible because $\|\varepsilon_p\|$ is always infinite. This occurs because all the prediction kernels I will consider here are singular and not square integrable. Furthermore, the data kernels are also generally singular because they are composed of linear combinations of the Frechet kernels, $<G_i\,|$, defined in equation (4.4). The $<G_i\,|$ functions are singular at the bottoming points of refracted arrivals and are, as a

result, not square integrable either [Backus and Gilbert, 1969]. A general solution to this is a process that Backus [1970b] calls "quelling". To be consistent with the minimization criterion (4.24) used to construct the velocity model, I have used the particular quelling operation that Backus [1970b] calls "quelling by integration". The process involved in this is to integrate (5.1) and (5.2) by parts. This yields[2] (I use the nondimensional variable $r = \dfrac{L-z}{L}$ as introduced in chapter 4 here for convenience.)

$$\underline{n}_i = \delta u(1)\underline{N}_i(1) - n_i = <\underline{N}_i \mid \frac{d}{dr}(\delta u)> \qquad (5.8)$$

$$\underline{p} = \delta u(1)\underline{P}(1) - p = <\underline{P} \mid \frac{d}{dr}(\delta u)> \qquad (5.9)$$

where

$$<\underline{N}_i(r)\mid = \int_0^r N_i(\rho)d\rho \qquad (5.10)$$

$$<\underline{P}(r)\mid = \int_0^r P(\rho)d\rho \qquad (5.11)$$

If one has a reasonable estimate of the surface velocity, the first term on the left of (5.8) and (5.9) will vanish, since in that case $\delta u(1)=0$. If such an estimate is not available, one can avoid the problem by adding a constraint that

$$\sum_{i=1}^{M_n} w_i\underline{N}_i(1)=1 \qquad (5.12)$$

and any estimate of p calculated by equation (5.4) will not depend

---

[2] Equations (5.8) to (5.11) are the appropriate form for this quelling when $\mid\delta u>$ is a function only of depth. It is worth noting that a similar form can be derived when $\mid\delta u>$ is an arbitrary function of all three spatial coordinates [Chou and Booker, 1979].

on $\delta u(1)$ [3]. In either case, (5.8) and (5.9) are in the same form as (5.1) and (5.2) except that $<N_i|$ and $<P|$ are replaced by their antiderivatives and $|\delta u>$ is replaced by its derivative. The major difference is that the norm of the error function

$$<\underline{\varepsilon}_p| = <\underline{P}| - <\hat{\underline{P}}| \qquad (5.13)$$

$$= <\underline{P}| - \sum_{i=1}^{M_n} w_i <\underline{N}_i|$$

is now finite. The norm $\|\underline{\varepsilon}_p\| = \sqrt{<\underline{\varepsilon}_p|\underline{\varepsilon}_p>}$ is the root mean square misfit between the actual prediction kernel, $<\underline{P}|$, and the approximation to the prediction kernel, $<\hat{\underline{P}}|$. This misfit is a measure of the error in the estimate of $\hat{p}$ due to the fact that our knowledge of the velocity (slowness) model, $|\delta u>$, will always by imperfect. This is so because the velocity model is a function that requires (in principle at least) an infinite number of parameters to be specified exactly. Yet the amount of data we can collect is always finite. Thus, we can never know the velocity model perfectly because the data we have available will always be insufficient. Recognition of this type of error is what Parker [1977a] calls "the cardinal factor distinguishing inverse theory from conventional parameter estimation". I will henceforth refer to this type of error as **model error** because of its relation to the ambiguousness of the velocity model.

Given that $\|\underline{\varepsilon}_p\|$ is a measure of model error I would like to point out that I will interpret $\|\underline{\varepsilon}_p\|$ in two quite different ways in the following two sections. Because of that I will not be specific at this

---

[3] Johnson and Gilbert [1972] show that if a discontinuity in the model is suspected at some position in the model (in their case the core-mantle boundary) its effect can be removed by imposing a constraint similar to (5.12). An untest alternative for treating this sort of problem is to consider the depth and size of the jump as another set of free parameters that can also be removed by the annulling transformation.

point as to how $\|\underline{\varepsilon}_p\|$ should be interpreted. Instead it is sufficient for the remainder of this section to simply consider $\|\underline{\varepsilon}_p\|$ as some measure of the error in $\hat{p}$ due to the fact that the velocity model can never be known exactly.

If the measured data were perfect, it is clear that we would normally want to choose the weights,w_i, in (5.13) so that $\|\underline{\varepsilon}_p\|$ is minimized for then the model error would be as small as possible. However, in real life the data one collects will always contain random measurement errors that will propagate into errors in the estimate $\hat{p}$. The objective of the remainder of this section is to derive a relationship that will permit us to make a quantitative statement concerning the size of this error. I will henceforth refer to this type of error as the **statistical error** because it can be described only in statistical terms.

The data I am utilizing here are the measured arrival times of different phases on a seismogram (I have used only first arrival P waves in this study but other phases could also be used.). These data will always contain some error that I will write as

$$t_i = \tilde{t}_i + \Delta t_i \qquad i = 1,2,3,\ldots,M \qquad (5.14)$$

where

t_i = measured arrival time for i^{th} datum.

$\tilde{t}_i$ = actual arrival time.

Δt_i = measurement error.

We cannot know what Δt_i is (it wouldn't be an error if we did) but I assume that we know something about its statistics. Errors in picking the arrival times of impulsive P waves are statistically independent and approximately normally distributed with zero mean [Buland, 1976]. This implies that the covariance of the data is given by

$$D = E[\Delta t(\Delta t)^T] = \begin{vmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & \sigma_M^2 \end{vmatrix} \quad (5.15)$$

where[4]

$\sigma_i^2 = E[\Delta t_i^2] = $ variance of ith datum.

The annulled data, **n**, are related to the observed data by the relation

$$n = U_N^T r^s \quad (5.16)$$

$$= U_N^T(t - t' - Ss)$$

$$= U_N^T(\tilde{t} - t' - Ss) + U_N^T \Delta t$$

since r^s is defined by equation (3.25). The matrix $U_N^T \in R^{M_n \times M}$ is the matrix with orthogonal rows used to construct **n** as described in appendix B. U_N^T has the form (3.46) for the case when station corrections are neglected and the form (4.15) when station corrections are included as free parameters. I show in appendix B that the covariance of the errors in the annulled data are given by

$$C_n = E[\Delta n(\Delta n)^T] \quad (5.17)$$

$$= U_N^T D U_N$$

Our present goal is to quantify the expected statistical error in the estimate $\hat{p}$ given in equation (5.3). In particular, we seek an

---

[4] With impulsive arrivals σ_i is a constant [Buland, 1976] that can be estimated. With weaker arrivals σ_i must be guessed somewhat subjectively and the errors are no longer normally distributed but tend to be biased toward measuring the arrival time late [Anderson, 1978].

estimate of the variance of the error in $\hat{p}, \Delta\hat{p}$, defined as

$$\sigma_p^2 = E[\Delta\hat{p}^2] \tag{5.18}$$

To obtain σ_p^2 we note that (5.3) can be written as

$$\hat{p} \quad w^T n = w^T \tilde{n} + w^T \Delta n$$

where w is the vector of the M_n weights in (5.3)

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ \cdot \\ w_{M_n} \end{bmatrix} \tag{5.19}$$

$w^T \tilde{n}$ is the value one would get for the estimate $\hat{p}$ if the data were error free and $w^T \Delta n$ is the error in $\hat{p}$. Consequently, we see (5.18) can be written as

$$\sigma_p^2 = E[w^T \Delta n (w^T \Delta n)^T]$$
$$= w^T E[\Delta T(\Delta n)^T] w \tag{5.20}$$
$$= w^T C_n w$$
$$= w^T U_N^T D U_N w$$

This is the relationship we were seeking. (5.20) quantifies the effect of measurement errors in the observed data through a prediction of the expected size of the variance of statistical errors that are likely to be present in $\hat{p}$.

It is quite apparent that we cannot hope to minimize both $\|\varepsilon_p\|$ and σ_p^2 with the same set of weights, w_i. Backus [1970a], however, advocates minimizing the following linear combination of the two

$$e^2 = \sigma_p^2 + B^2 \|\varepsilon_p\|^2 \tag{5.21}$$

where B is an arbitrary number and hence B^2 can be any number

from zero to infinity. The usefulness of (5.21) is that it can be used to study how the statistical error,σ_p^2, and the model error,$\|\underline{\varepsilon}_p\|$, interact with each other. This analysis,.,which I will refer to as a trade-off analysis, consists of varying B^2 over a large range and studying how σ_p^2 and $\|\underline{\varepsilon}_p\|$ vary. I will apply this analysis to assess the errors in both the velocity model and the discrete parameters. However, the way the results can be interpreted are quite different for these two cases and so I will defer any further discussion of the principles of a trade-off analysis to the two subsequent sections. There the results can be interpreted in a better context. The major point to derive from this section is that the two methods are closely related, although they may appear quite different in detail.

3. ASSESSMENT OF ERRORS IN THE VELOCITY MODEL

In the previous chapter I described a procedure that utilized the annulled data to construct an estimate of the velocity model that fits the observed data. Having constructed such a model, it is imperative to address the question of how reliable that model is. The approach I will follow here is identical to what Johnson and Gilbert [1972] call the "objective approach". This approach is "objective" because the data are used directly to derive a quantitative statement concerning the uniqueness of the model with no *a priori* judgements about it.

If we want to assess the uniqueness of the velocity model, we must face the fact that any estimate of it is fundamentally ambiguous. This is so because the velocity model is characterized by a function that requires (in principle at least) an infinite number of parameters to be specified exactly. However, the quantity of data we have available is always finite. Thus, we cannot hope to resolve the velocity structure at an arbitrarily fine scale but we must always be content with some smoothed version of it [Johnson and Gilbert, 1972]. Backus and Gilbert [1967,1968, and 1970] developed

a procedure which, recognizing this ambiguity, concentrates not on constructing models (They prove there are always infinitely many that fit the data.); but instead concentrates on the question of how a given model is related to the truth. They do this by considering the estimate of the model at a given position, r_0, as some smoothed or locally averaged version of the actual model at that depth. This changes the emphasis from constructing models, which are fundamentally ambiguous, to studying the averaging functions which are unique. Backus and Gilbert consider the properties of averaging functions constructed by a particular application of Backus' generalized prediction procedure[5]. That is, Backus and Gilbert consider a particular type of prediction in which we choose the prediction kernel $<P| = <\delta(r - r_0)|$, where $<\delta(r - r_0)|$ is the delta function. This of interest because then (5.2) can be written as

$$\delta u(r_0) = <\delta(r - r_0) | \delta u>$$

In other words, the number p in (5.2) can, in this case, be interpreted as the actual value of $|\delta u>$ at r_0. Suppose we estimate the model at r_0 as a linear combination of the data as in (5.3). We can then interpret the number $\hat{p}$ as an estimate of $|\delta u>$ at r_0. I write this as

$$\delta \hat{u}(r_0) = \sum_{i=1}^{M_n} w_i(r_0) n_i \qquad (5.21)$$

If $<\delta u(r_0)|$ is estimated by (5.21), then the approximate prediction kernel, $<\hat{P}|$, that I defined in (5.5) can be interpreted as the averaging function I mentioned above. This interpretation is clear from

---

[5] Historically, Backus and Gilbert's "particular application" came first. Backus' general theory was an extension of his earlier work with Gilbert. I find Backus' view a useful starting point here, however, because it helps emphasize the interconnection of the material of this chapter.

(5.6) which I now write as

$$\delta \hat{u}(r_0) = <A(r,r_0)|\delta u> = \int_0^1 A(r,r_0)\delta u(r)dr \qquad (5.22)$$

where I denote $<\hat{P}|$, in this case, by the special symbol $<A(r,r_0)|$ to emphasize that it can be interpreted as an averaging function for the estimate at r_0. $<A(r,r_0)|$ is defined exactly like $<\hat{P}|$ in (5.5);

$$<A(r,r_0)| = \sum_{i=1}^{M_n} w_i(r_0)<N_i(r)| \qquad (5.23)$$

We cannot hope to make $<A(r,r_0)|$ into a true delta function [Backus and Gilbert, 1968]. Thus we must be content with averaging functions that are approximate delta functions in some sense. Ideally, we would like to choose the weights, $w_i(r_0)$, such that $<A(r,r_0)|$ is a strongly peaked function centered at r_0

The question now centers on how we should choose the $w_i(r_0)$ so that $<A(r,r_0)|$ is the peaked function we seek. Backus and Gilbert refer to this choice as a "δ-ness criteria" Conceptually the simplest of these is to make $<A(r,r_0)|$ a best approximation to a delta function as measured by the norm (5.7). As noted before, however, this is not possible here because neither the prediction kernel nor the data kernels are square integrable[6]. This problem is eliminated, as noted in the previous section, by involking what Backus [1970b] calls "quelling by integration" This operation is defined by

---

[6] It is worth noting that if the data kernels were not singular it would be possible to produce a set of averaging functions that are a best approximation to a delta function. The results are difficult to interpret, however, because $||<\delta(r-r_0)|-<A(r,r_0)|||$ is always infinite since the delta function is not square integrable. This problem can only be circumvented by selecting some quelling operation. Quelling by integration is a reasonable choice since it leads to an error assessment consistent with the minimization criteria (4.24) used to construct the velocity model.

equations (5.8) to (5.11) above. Of these, the relationship (5.9) can be made more specific by noting that

$$<\underline{P}(r)\,|=<H\,(r-r_0)\,|=\int_0^r \delta(\rho-r_0)d\rho \qquad (5.24)$$

where $<H\,(r-r_0)\,|$ is the unit step function. The error function in equation (5.13) then takes the special form

$$<\underline{\varepsilon}_{\delta u}\,(r_0)\,|=<H\,(r-r_0)\,|-<\underline{A}\,(r,r_0)\,| \qquad (5.25)$$

where

$$<\underline{A}\,(r,r_0)\,|=\sum_{i=1}^{M_n} w_i <\underline{N}_i\,| \qquad (5.26)$$

$$=\int_0^r A\,(\rho,r_0)d\rho$$

The norm of the error function in (5.25), $\|\underline{\varepsilon}_{\delta u}\,(r_0)\|$, is always finite. As I noted in the previous section $\|\underline{\varepsilon}_{\delta u}\,(r_0)\|$ is a measure of the error in the estimate of $<\delta u\,|$ at r_0 due to the fact that the velocity model cannot be known unambiguously. One of the fundamental contributions of Backus and Gilberts' work was to provide an interpretation of what this number means. $\|\underline{\varepsilon}_{\delta u}\,(r_0)\|$ is a natural measure of the width of the peak of the averaging function $<A\,(r,r_0)\,|$. This is so because if $<A\,(r,r_0)\,|$ were the boxcar function

$$<B\,(r,r_0)\,|=\begin{cases} \dfrac{1}{l} & \text{if}\,|r-r_0|<\dfrac{l}{2} \\[2ex] 0 & \text{if}\,|r-r_0|>\dfrac{l}{2} \end{cases}$$

we would find $\|\underline{\varepsilon}_{\delta u}\,(r_0)\|^2=\dfrac{l}{12}$ [Johnson and Gilbert, 1972]. Thus, the **resolution length** (also called the **spread**)

$$l \doteq 12\|\underline{\varepsilon}_{\delta u}\,(r_0)\|^2 \qquad (5.27)$$

is useful as a measure of the smoothing produced by $<A(r,r_0)|$. If the velocity model has fine scale structural detail on a length scale less than l, $<A(r,r_0)|$ will be incapable of resolving it. This is an inevitable consequence of the finite amount of data we have available and the resolution length, l, is a quantitative measure of the uniqueness of our inversion [Johnson and Gilbert, 1972]. l has the additional distinction of being easily calculated via the following simple relationship [Johnson and Gilbert, 1972]

$$l = 12[(1-r_0) - 2\mathbf{h}^T\mathbf{w} + \mathbf{w}^T\underline{\mathbf{N}}\mathbf{w}] \tag{5.28}$$

where $\mathbf{w}$ is the vector of weights as in (5.19) and I define the components of the matrix $\underline{\mathbf{N}}$ and the vector $\mathbf{h}$ by

$$\underline{N}_{ij} = <\underline{N}_i \mid \underline{N}_j> \tag{5.29}$$

and

$$h_i = <H(r-r_0) \mid \underline{N}_i> \tag{5.30}$$

As I noted in the previous section, there is another source of error in our estimates that cannot be ignored. This is the uncertainty in the estimates, which I have called the statistical error, caused by fact that the observed data are not perfectly accurate. It is measured by the expected variance of this error, which can be calculated by (5.20). The statistical error in the estimated perturbation, which I denote by the symbol $\sigma_{\delta u}$, cannot be ignored because a highly localized averaging function (small resolution length) is not very useful if it produces a large statistical error in the estimate. It seems sensible then that we should try to minimize both the resolution length, l, and the statistical error. It turns out that we cannot minimize both with the same set of weights. However, Backus and Gilbert [1970] show that we can minimize the following linear combination of the two

$$e_{\delta u}^2(r_0) = \|\underline{\varepsilon}_{\delta u}(r_0)\|^2\cos\theta + \sigma_{\delta u}^2(r_0)\sin\theta \tag{5.31}$$

where I use the symbol $\sigma_{\delta u}^2(r_0)$ to symbolize the statistical error, σ_p^2, calculated by (5.20). θ in equation (5.31) is a parameter which varies from 0 to $\frac{\pi}{2}$. When $\theta=0$, the resolution length is minimized, and when $\theta=\frac{\pi}{2}$, $\sigma_{\delta u}^2(r_0)$ is minimized [Johnson and Gilbert, 1972]. (Note that (5.21) and (5.31) are identical except in (5.21) we see $B^2=\frac{\cos\theta}{\sin\theta}=\cot\theta$ [Backus, 1970a].) Backus and Gilbert [1970] prove that as θ goes from 0 to $\frac{\pi}{2}$ the statistical error, $\sigma_{\delta u}^2(r_0)$, is a mono-tonically decreasing function of l. This says that we can always decrease the statistical error at the expense of a broadening of the averaging function $<A(r,r_0)|$ (as measured by l). For this reason they call the resulting curve the **trade-off curve** of error versus resolution length (spread). This process is analogous to the similar trade-off between resolution and variance commonly encountered in time series analysis (see for example Claerbout [1976, chapter 4]).

Constructing the family of solutions that yield these trade-off curves is relatively straightforward. The weights that minimize (5.31) are calculated by solving the following set of equations [Johnson and Gilbert, 1972]

$$[\underline{\mathbf{N}}\cos\theta+\mathbf{C}_n\sin\theta]\mathbf{w}=\mathbf{h}\cos\theta \qquad (5.32)$$

or equivalently

$$[\underline{\mathbf{N}}+\mathbf{C}_n\frac{1}{B^2}]\mathbf{w}=\mathbf{h} \qquad (5.33)$$

where $B^2=\cot\theta$ is as in (5.21) and $\mathbf{C}_n$ is as defined in (5.17). $\mathbf{w}$ found by (5.32) will be different for each value of θ. The trade-off curves are constructed by calculating $\mathbf{w}$ at selected values of r_0 for a range of θ values. Each set of $\mathbf{w}$ are determined by solving (5.32) for each value of θ. Once $\mathbf{w}$ is known, l can be calculated easily from (5.28) and $\sigma_{\delta u}^2(r_0)$ can be calculated from (5.20). As a practical matter, a

direct solution of (5.32) for each value of θ is unnecessarily inefficient. It is preferable instead to simultaneously diagonalize $\mathbf{C}_n$ and $\underline{\mathbf{N}}$ using the procedure Gilbert [1971] calls "ranking and winnowing".

Some actual timing data helps emphasize how important Gilbert's algorithm is. It requires approximately 10 cpu seconds of computer time to invert a 100 by 100 matrix on a CDC Cyber 170-750 computer. This means that a direct solution would require at least 10 cpu seconds for each point on the trade-off curves. An actual calculation I made using the diagonalizing transformations required ~20cpu seconds to perform the transformations to diagonalize $\mathbf{C}_n$ and $\underline{\mathbf{N}}$ (The algorithm to do this is slower than a direct solution of a set of linear equations like (5.32).). Calculating 13 points on the trade-off curves at 150 values of r_0 then required only 5 cpu seconds. The same calculation by a direct solution would have required at least 130 cpu seconds. This represents an improvement, for this example, of over 500%.

4. ASSESSMENT OF PARAMETER ERRORS

4.1. Introduction

Assessing the reliability of the discrete parameters (hypocenters and station corrections) presents a slightly different problem from that of assessing the reliability of the velocity model. In estimating the velocity model, its value at a given point is of less interest than its structure. Consequently, most people would be quite willing to accept some smoothed version of the true velocity model as long as the degree of smoothing was not so great as to render the result useless. A similar interpretation does not appear to be feasible for the assessment of the errors in the discrete parameters. I say this because it is difficult to decide what a "local average" is for these parameters in a manner analogous to that

discussed above for assessing the reliability of the velocity model. The concept of a local average makes perfectly good sense for a function like the velocity model because it is composed (conceptually at least) of an infinite set of points that can be ordered according to their depth. The same is not true of the discrete parameters as we can arrange them in virtually any order we feel like. Thus, it becomes difficult to say what a "local average" for any of the discrete parameters is because it is difficult to say what points are "local" to that parameter. Because of this, errors in the discrete parameters appear to be more appropriately assessed by the linear inference techniques of Backus [1970a] or the alternatives given by Parker [1977a]. The technique I describe here is based heavily on the work of Backus [1970a]. The major difference is in the way the results are interpreted.

4.2. Relation of the parameters to the velocity model

Because the data (arrival times) are nonlinear functionals of the discrete parameters and the velocity model, it is imperative that we first review how the discrete parameters are estimated. This review will lead us to a set of equations that relate the discrete parameters directly to the velocity model. These equations are the foundation of the error assessment I am presenting in this section. I will show that they follow naturally from the estimation procedure outlined in the previous chapter.

The estimation procedure described in chapter 4 is based on the linearized equations (4.11), which I repeat here for the reader's benefit.

$$\mathbf{r}^s = \mathbf{B}\mathbf{x} + \overset{\circ}{\mathbf{G}} | \delta u > \qquad (5.34)$$

At the core of that procedure are two matrix operators. The first is the matrix $\mathbf{U}_N^T \in \mathbb{R}^{M_\neg \times M}$ that is used to construct the annulled data set. The second is the generalized inverse $\mathbf{H}_B \in \mathbb{R}^{N \times M}$ that is used to

estimate the parameter correction vector[7]

$$\hat{\mathbf{x}} = \mathbf{H}_B \mathbf{r}^s \qquad (5.35)$$

The essence of the algorithm I have described in chapter 4 is that the parameters are perturbed in isolation by in iterative solution of (5.35). When that solution has converged the matrix $\mathbf{U}_N^T$ is calculated and the annulled data are used to perturb the velocity model. This sequence is repeated until the entire algorithm converges. This segregation of the adjustments of the two different parts of the model is warranted because

$$\mathbf{H}_B \mathbf{U}_N = \mathbf{0} \in \mathbf{R}^{N \times M_n} \qquad (5.36)$$

That is, the rows of $\mathbf{H}_B$ and $\mathbf{U}_N^T$ (columns of $\mathbf{U}_N$) are orthogonal. This relationship is fundamental because it indicates that $\hat{\mathbf{x}}$ and the perturbation to the velocity model are constructed from two different, orthogonal components of the data and hence are uncoupled. Furthermore, in the language of chapter 2, since the columns of $\mathbf{U}_N$ are a basis for $N(\mathbf{B}^T)$, then (5.36) is a statement that

$$R(\mathbf{H}_B^T) \subset R(\mathbf{B}) \qquad (5.37)$$

since $R(\mathbf{B})$ and $N(\mathbf{B}^T)$ are orthogonal complements. (5.37) is significant because it is identical to equation (B.7). This is a notable fact because I show in appendix B that because $\mathbf{H}_B$ satisfies (5.37) it is possible to derive the following set of N equations (see equation (B.30))

---

[7] Recall that $\hat{\mathbf{x}}$ is not actually calculated directly by a matrix multiplication like (5.35). Instead the algorithm described in chapter 3 and 4 is considerably more complex as it involves iterative solutions on several levels. Nonetheless, as I pointed out earlier (section 3.2.4 of chapter 3) the final step can still be viewed as a linear perturbation from a reference solution using (5.35) even if $\mathbf{H}_B$ is never even formed explicitly.

$$x_R = \hat{x} + \overset{\leftrightarrow}{X} \mid \delta u >$$ (5.38)

where

$$x_R = H_B B x = R_B x$$ (5.39)

$\hat{x}$ is the estimate (5.35) and the kernels of the operator $\overset{\leftrightarrow}{X}$ are as given in in (B.27)

$$\overset{\leftrightarrow}{X} = -H_B \overset{\leftrightarrow}{G}$$ (5.40)

(5.38) is the relationship we need. It provides a direct relationship between the true correction vector,x; the estimate,$\hat{x}$; and the velocity (slowness) model perturbation, $\mid \delta u >$ [8]. This relationship is not quite as direct as we might wish because of the presence of the resolution matrix R_B that relates x and x_R by (5.39). When station corrections are not used, R_B reduces to the form shown in (3.53), which in most cases is a reasonably close approximation to an identity matrix. In that case the relationship (5.38) is relatively unambiguous. However, when station corrections are used and estimated by the procedure described in chapter 3, R_B acquires the more complicated form in (3.65). This relationship is then considerably more ambiguous as this matrix is normally not even remotely close to an identity matrix. It acquires this form because the matrix B has a fundamental ambiguity caused by the fact that the average value of the station corrections cannot be determined uniquely (see chapter 3). (5.38) is a statement that we can never know x itself but only the set of number x_R. R_B is a useful tool for assessing this additional nonuniqueness in x. R_B is best interpreted somewhat

---

[8] It is imperative to recognize that both x and $\mid \delta u >$ are perturbations from a reference model. They are perturbations because (5.34) is based on a local linearization of a set of nonlinear equations. Consequently, (5.38) is generally correct only if the parameters and the velocity model are sufficiently close to the truth that the linearized equations are correct.

like the averaging functions I discussed in the previous section. That is, it provides a means of objectively assessing the nonuniqueness of **x** because we can interpret $\mathbf{R}_B$ as an additional resolution operator through which we must view our estimate of **x**. In fact, the ambiguity of **x** expressed by $\mathbf{R}_B$ in (3.65) has a quite simple and sensible interpretation for this case (see chapter 3).

4.3. Model error

Having cautioned the reader about the added uncertainty that the resolution matrix $\mathbf{R}_B$ introduces, we are now in a position to study how the more fundamental ambiguity of the velocity model can affect our estimates of the parameters. To do this I now define

$$\delta\mathbf{x} = \mathbf{x}_R - \hat{\mathbf{x}} \tag{5.41}$$

so (5.38) can be rewritten as

$$\delta\mathbf{x} = \mathbf{\ddot{X}} | \delta u > \tag{5.42}$$

Observe, however, that (5.42) represents N equations of the form

$$\delta x_i = <X_i | \delta u > \qquad i = 1, 2, \ldots, N \tag{5.43}$$

which is identical to equation (5.2) with $p = \delta x_i$ and $<P| = <X_i|$. Thus, we can assess the uncertainty in each discrete parameter using Backus' generalized prediction procedure that I outlined above. This means we again estimate each of the N numbers, δx_i, as a linear combination of the annulled data exactly as in (5.3). In this case I write this as

$$\delta\hat{x}_i = \sum_{j=1}^{M_n} w_j n_j \tag{5.44}$$

We again cannot proceed directly but are required to invoke quelling by integration (equations (5.8) to (5.11)). In this case $<P| = <X_i|$ so I define

$$<\underline{X_i}| = \int_0^1 X_i(\rho)d\rho \qquad (5.45)$$

This permits me to write (5.9) as

$$\delta \underline{x_i} = \delta u(1)\underline{X_i}(1) - \delta x_i = <\underline{X_i}| \frac{d}{dr}(\delta u)> \qquad (5.46)$$

Integration quelling is required because both the prediction kernels, $<X_i|$, and the annulled data kernels, $<N_i|$, are composed of linear combinations of the data kernels, $<G_i|$, which are singular at the bottoming points of refracted arrivals. The fundamental property of integration quelling is that is assures that the norm of the error function

$$<\underline{\varepsilon_i}| = <\underline{X_i}| - <\hat{\underline{X_i}}| \qquad (5.47)$$

$$= <\underline{X_i}| - \sum_{j=1}^{M_n} w_j <\underline{N_j}|$$

is finite (Note (5.47) is identical to (5.13) but with the symbols redefined.) I noted in section 2 that $\|\underline{\varepsilon_i}\|$ is a measure of what I have called the model error. In section 3 we saw that a quantity virtually identical to $\|\underline{\varepsilon_i}\|$ could be interpreted as a natural measure of the width of the peak of the averaging function associated with the estimate of the velocity model at a selected depth (equation (5.27)). Here, however, I will interpret $\|\underline{\varepsilon_i}\|$ differently. Following Backus, I do this by assuming that we can find a number B such that

$$\|\frac{d}{dr}(|u_{true}> - |\hat{u}>)\| \equiv \|\frac{d}{dr}(|\delta u_{true}>\| \leq B \qquad (5.48)$$

where

$|u_{true}> =$ true but unknown slowness function

$|\hat{u}> =$ final estimate of $|u_{true}>$

That is, B is an upper bound on the norm of the function that is the

derivative of the deviation of the true slowness model from the final estimate. Specifying B *a priori* in a non-arbitrary way is not trivial. Nonetheless, if we assume there is such a bound[9], then by Schwartz's inequality, the error, Δx_i, in x_i due to our insufficient knowledge of the velocity model is bounded by

$$\Delta x_i \leq \|\underline{\varepsilon}_i\| B \qquad (5.49)$$

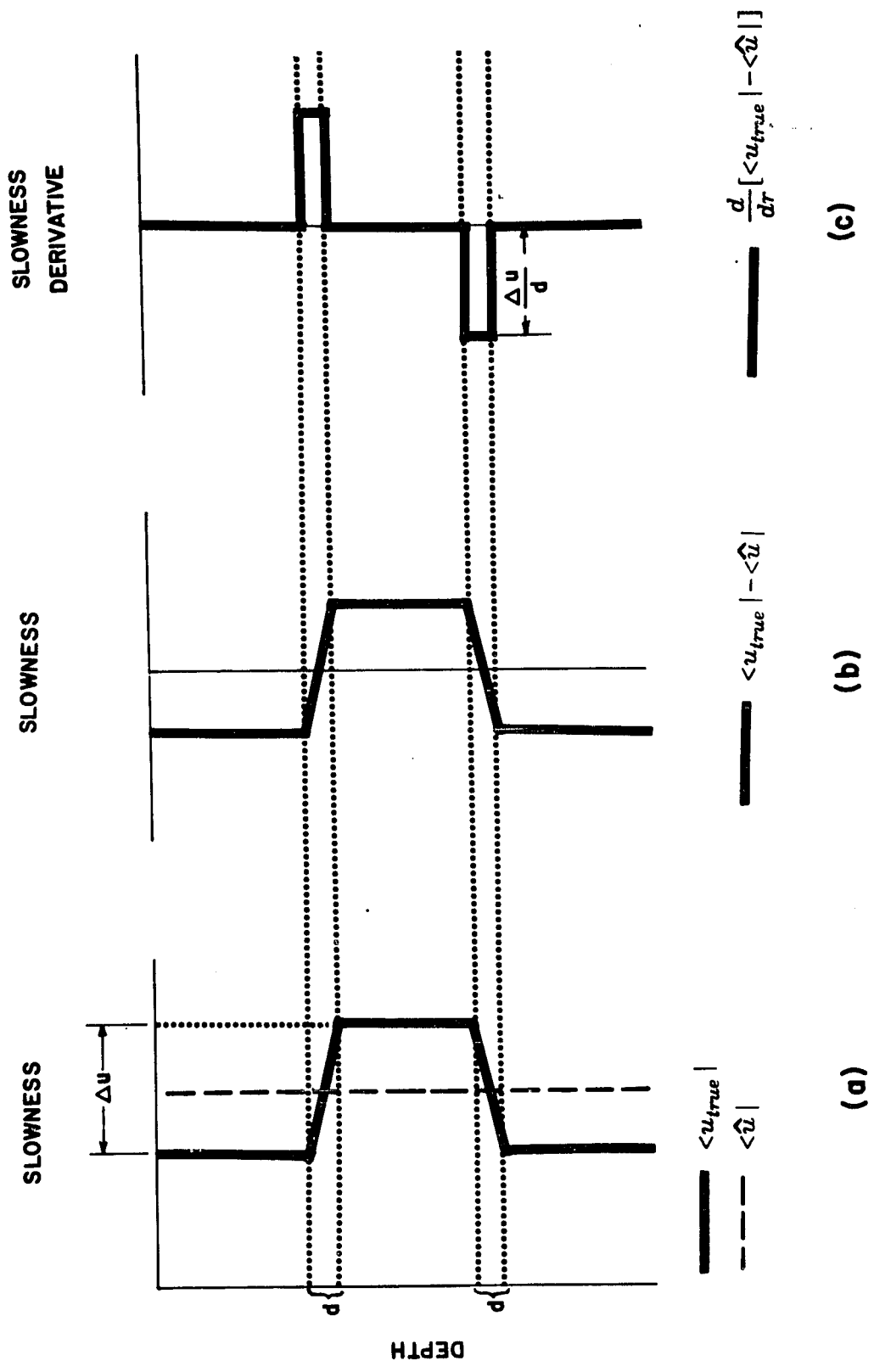## 4.4. Physical interpretation of the model bound

The bound B introduced in (5.48) plays a central role in the error analysis technique I am describing here. Consequently, it is essential to obtain a physical feeling for what this number means. I do that here by relating B to a set of two characteristic length scales that have a much simpler interpretation.

Equation (5.48) says that the function $|\delta u_{true}\rangle$ is the unknown difference between the actual slowness model and the final estimate of it. B is a bound on the norm of the derivative of this function. To interpret what this means consider the ideas presented in figure 5.1. Suppose $|u_{true}\rangle$ is as shown in figure 5.1a. That is, $|u_{true}\rangle$ is a stack of three constant velocity layers of thickness Σ. Let us assume the layer boundaries are not perfectly sharp. Assume instead that the transition region from one layer to the next occurs over a characteristic length scale d. An idealization of this is illustrated in figure 5.1a where the layers are shown joined by a constant slowness gradient transition region of thickness d. Suppose

---

[9] In chapter 7 I demonstrate that a given value of B can be interpreted physically in terms of the length scales of slowness discontinuities that could be anticipated in the real earth. I postpone that discussion to chapter 7 because the relationship I obtain for B is dependent upon the scale of the network. Since this is a data dependent property, description of this interpretation is better left to chapter 7 where it can be applied immediately to a specific example.

Figure 5.1. Prototype of an unresolved layer. (a) Slowness versus depth plot. (b) Plot of difference between true slowness versus depth function and an estimate of it. (c) Plot of derivative with respect to depth of (b).

SLOWNESS

SLOWNESS

SLOWNESS
DERIVATIVE

Δu

$<u_{true}|$

$<u_{true}|-<\hat{u}|$

$\frac{\Delta u}{d}$

$---$ $<\hat{u}|$

$<u_{true}|-<\hat{u}|$

$\frac{d}{dr}[<u_{true}|-<\hat{u}|]$

(a)

(b)

(c)

DEPTH

further that $|\hat{u}>$ is the constant slowness function shown as the dashed line in figure 5.1a. Hence, the overall picture in figure 5.1a is an idealization of an unresolved velocity variation. The error function, $|\delta u_{true}>$, in equation (5.48) is, in this case, as shown in figure 5.1b and the derivative of this error function is the two boxcar functions shown in figure 5.1c. The height of these boxcar functions is $\dfrac{\Delta u}{d}$ where Δu is the difference in slowness between the layers. The width of the boxcar is d, the thickness of the transition region. Thus, the norm of this function is $\dfrac{2\Delta u}{d}\left[\dfrac{d}{L}\right]^{1/2}$ 10.

The model in figure 5.1 is exemplary of an unresolved layer. Suppose we idealize the earth as a set of such layers with a characteristic spacing Σ as shown in figure 5.2. That is, we idealize the earth as a set of constant velocity layers with an average thickness Σ. The boundaries between the layers are presumed to not be perfectly sharp, but occur instead over a finite length scale d. One can see that this idealization is reasonable if figure 5.2 is compared with the actual sonic velocity well logs shown in figure 5.3.

---

10 The appearance of the length L, the maximum depth of the model being studied, comes from the fact that the norm of a function $f(r)$ is

$$\|f(r)\| = <f(r)|f(r)>^{1/2} = \left[\int_0^1 f^2(r)dr\right]^{1/2}$$

Because the nondimensional variable r is related to physical depth by the relation

$$r = \frac{L-z}{L}$$

we have

$$\|f(z)\| = \left[\frac{1}{L}\int_0^L f^2(z)dz\right]^{1/2}$$

SLOWNESS
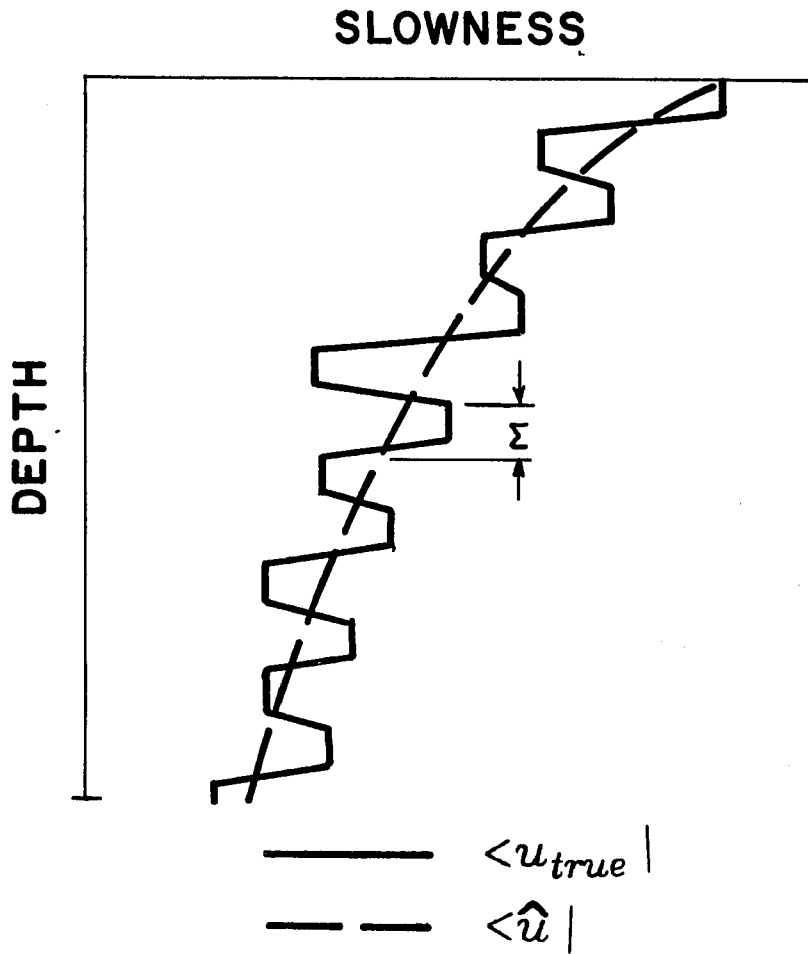
Figure 5.2. Idealized model used to interpret the bound on the norm of the derivative of the error function in equation (5.48). Truth is idealized as a set of layers idealized as in figure 5.1 and with a characteristic spacing of Σ.

Figure 5.3. Sonic velocity well logs from two different environments. (a) Typical log from a sedimentary basin [after Lindseth, 1979]. (b) Columbia river basalt. Note the similarity to the idealized model in figure 5.2.

SLOWNESS (μsec/ft)

90   65   40

SLOWNESS (μsec/ft)

120   40

DEPTH (ft)

DEPTH (ft)
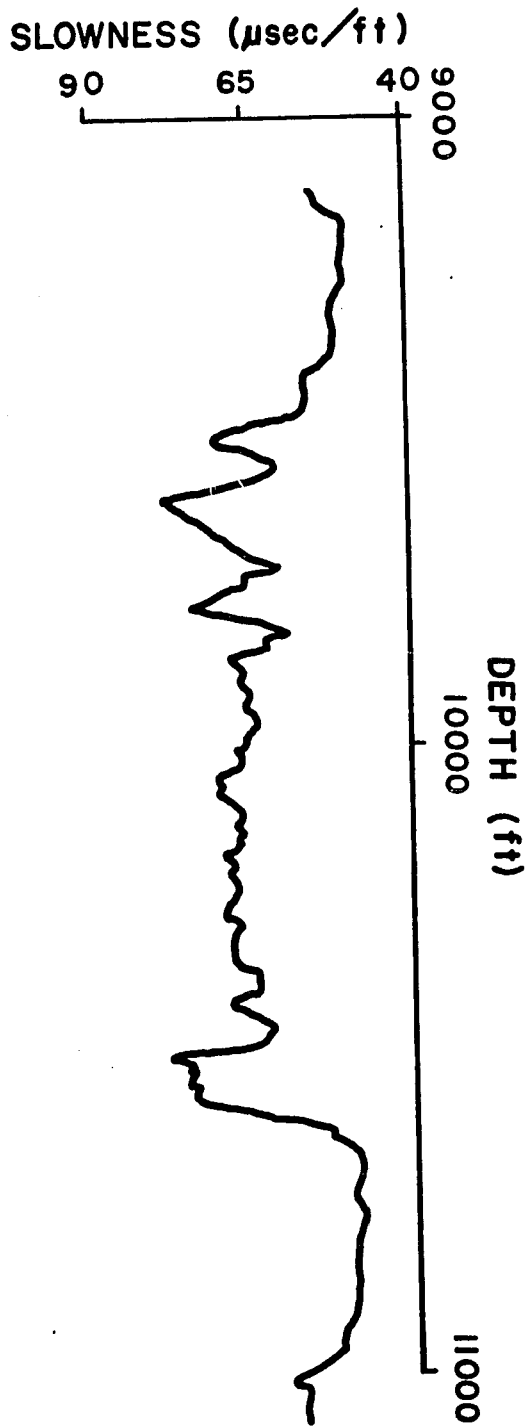
(a)
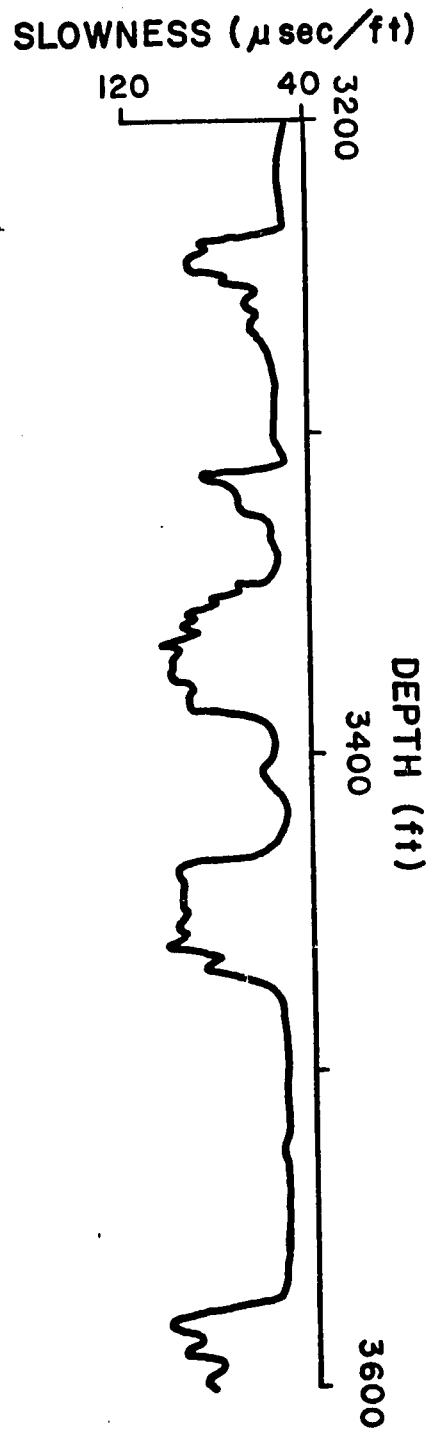
(b)

The estimate we can produce of the slowness function will always be some smoothed version of the truth similar to that shown by the dashed line in figure 5.2. For this idealized case $|\delta u_{true}>$ can be considered as the sum of a finite number of segments like that shown in figure 5.1b. The number of such segments over the depth range L of the model is $\approx \frac{L}{\Sigma}$. Hence, for this idealization

$$\|\delta u_{true}\| \approx \left[\frac{L}{\Sigma}\right] \frac{\Delta u}{d} \left[\frac{d}{L}\right]^{1/2} \tag{5.50}$$

$$= \frac{\Delta u}{\Sigma} \left[\frac{L}{d}\right]^{1/2}$$

The product $\Sigma\sqrt{d}$ will prove useful in chapter 7 when we use the analysis presented here to interpret some real data. Consequently, I define

$$\kappa^{3/2} = \Sigma\sqrt{d}$$

or

$$\kappa = [\Sigma\sqrt{d}]^{2/3} \tag{5.51}$$

I will call κ the **roughness length** because of its relation to the length scales of the variations in the velocity model. The bound B in (5.48) can be expressed in terms of κ as

$$B = \frac{\Delta u \sqrt{L}}{\kappa^{3/2}} \tag{5.52}$$

The relation in (5.50) is easier to interpret than the definition (5.49) because it involves three quantities ($\Delta u, \Sigma$, and d) that can be fairly easily related to reality and a fourth number L that is a constant for any given data set. One of these numbers, Δu, is fairly easy to define universally. The vast majority of crustal rocks have P wave velocities between 3 km/sec and 8 km/sec. As a result, it is unlikely that Δu for any given boundary is ever greater than 0.2

sec/km. A conservative guess for the average Δu is then 0.1 sec/km. The other two variables in (5.50) have a simple interpretation, but specifying them in a non-arbitrary way is not trivial. Consequently, I will defer any further discussion of them until chapter 7 where they can be better evaluated in the context of the data they are used to interpret.

κ will prove useful in chapter 7 but for the remainder of this chapter I will work with the number B instead. Since B is proportional to $\kappa^{-3/2}$, they could be used somewhat interchangeably in the analysis that follows. κ, however, appeared from an idealized model of the real earth introduced to interpret the more fundamental number B. Consequently, it is more natural to use B.

4.5. Statistical error

When the data contain no measurement errors, we would want to choose the weights, w_i, so that the norm of the error function in (5.11), $\|\underline{\varepsilon}_i\|$, is minimized. This is sensible because then the error bound given by (5.49) will be as small as possible. However, as I noted above when I considered the problem of determining the error in the velocity model estimates, minimization of $\|\underline{\varepsilon}_i\|$ is not very desirable if it produces an unacceptably large statistical error. Thus, we need to find a procedure to evaluate how the model error and statistical error interplay. To do this, however, we first have to derive a relationship that quantifies the statistical errors in the parameters analogous to equation (5.20).

We cannot procede directly from (5.20), because the complete estimate of the vector,$\mathbf{x}_R$, of the N parameters that I am using here is the sum of two terms

$$\mathbf{x}_{est} = \hat{\mathbf{x}} + \delta\hat{\mathbf{x}} \tag{5.53}$$

$\hat{\mathbf{x}}$ is the initial estimate of $\mathbf{x}$ obtained by a hypocenter location procedure (single or multiple event) calculated from equation (5.35).

$\delta\hat{x}$, the second term in (5.53), can be considered as a correction to $\hat{x}$ to account for inadequate knowledge of the velocity model. $\delta\hat{x}$ is calculated here by equation (5.44). It is useful to write the N equations of (5.44) in matrix form as

$$\delta\hat{x}=\mathbf{Wn} \tag{5.54}$$

where each equation of (5.44) is represented by one row of (5.54). The matrix $\mathbf{W}\in\mathbf{R}^{N\times M_n}$ is thus a collection of all the weights in (5.44) used in calculating $\delta\hat{x}$. The annulled data, $\mathbf{n}$, are related to the vector of residuals, $\mathbf{r}^s$, by equation (5.16). Combining (5.54) and (5.16) yields

$$\delta\hat{x}=\mathbf{WU}_N^T\mathbf{r}^s \tag{5.55}$$

The equations (5.35) and (5.55) provide a general relationship between the two estimates, $\hat{x}$ and $\delta\hat{x}$, and the initial data. The importance of these relationships is that they will provide the means for calculating the statistical error in the total estimate, $\mathbf{x}_{est}=\hat{x}+\delta\hat{x}$. Specifically, we need to calculate the covariance of the estimate $\mathbf{x}_{est}$ whose components are defined as

$$C_{ij}=\mathrm{E}[(\Delta x_{est})_i(\Delta x_{est})_j] \tag{5.56}$$

where E[ ] again denotes the expectation operator. It is preferable to write (5.56) in matrix form as

$$\mathbf{C}=\mathrm{E}[\Delta\mathbf{x}_{est}(\Delta\mathbf{x}_{est})^T] \tag{5.57}$$

Since $\mathbf{x}_{est}=\hat{x}+\delta\hat{x}$, we see that

$$\mathbf{C}=\mathrm{E}[\Delta\hat{x}(\Delta\hat{x})^T]+\mathrm{E}[(\Delta(\delta\hat{x}))(\Delta(\delta\hat{x}))^T]+2\mathrm{E}[\Delta\hat{x}(\Delta(\delta\hat{x}))^T] \tag{5.58}$$

(5.58) and the relationship (5.17) and (5.55) can now be used to calculate how random errors in the data propagate into $\mathbf{x}_{est}$. To do that, however, we have to again assume that the covariance of the data is known. We saw previously that this can indeed be estimated

and has the form of the matrix $\mathbf{D}$ in (5.15). From $\mathbf{D}$ we can calculate each term in (5.58). I show in appendix B (equation (B.39)) that the first term reduces to

$$E[\Delta\hat{x}(\Delta\hat{x})^T]=\mathbf{H}_B\,\mathbf{D}\mathbf{H}_B^T \qquad (5.59)$$

By using the same algebraic manipulations it is easy to show that the other terms reduce to

$$E[\Delta(\delta\hat{x})(\Delta(\delta\hat{x}))^T]=\mathbf{W}\mathbf{U}_N^T\mathbf{D}\mathbf{U}_N\,\mathbf{W}^T \qquad (5.60)$$

and

$$E[\Delta\hat{x}(\Delta(\delta\hat{x}))^T]=\mathbf{H}_B\,\mathbf{D}\mathbf{U}_N\,\mathbf{W}^T \qquad (5.61)$$

Combining (5.59), (5.60) and (5.61) gives

$$\mathbf{C}=\mathbf{H}_B\,\mathbf{D}\mathbf{H}_B^T+\mathbf{W}\mathbf{U}_N^T\mathbf{D}\mathbf{U}_N\,\mathbf{W}^T+2\mathbf{H}_B\,\mathbf{D}\mathbf{U}_N\,\mathbf{W}^T \qquad (5.62)$$

An important special case is that in which the data are statistically independent and of unit variance for then we would find

$$\mathbf{D}=\mathbf{I} \qquad (5.63)$$

and (5.58) reduces to

$$\mathbf{C}=\mathbf{H}_B\,\mathbf{H}_B^T+\mathbf{W}\mathbf{W}^T \qquad (5.64)$$

since $\mathbf{U}_N^T\mathbf{U}_N=\mathbf{I}\in\mathbf{R}^{M_n\times M_n}$ and $\mathbf{H}_B\,\mathbf{U}_N=0$ by (5.36). The "special case" (5.63) is important because if $\mathbf{D}$ is known *a priori*, an approximate weighting matrix can always be calculated (see Appendix A) that will rescale the data to be in this form [Lawson and Hanson, 1974, pp. 193-185]. In other words, the "general case" of (5.57) can always be converted to this "special case" by appropriate weighting. (5.63) remains, nonetheless, a useful relationship in case one chooses to use a different weighting scheme. In my application of these equations I have used weighting to yield the form (5.64) because it considerably simplifies the calculations that are involved.

Consequently, in the remainder of this chapter I will assume the form (5.64) is appropriate.

4.6. Damped solution trade-off analysis

I have now assembled all the mathematical machinery that is required for assessing the errors in the parameters. For any choice of the matrices $\mathbf{H}_B$ in (5.16) and $\mathbf{W}$ in (5.54) the predicted second moment statistics of the estimate can be calculated by (5.64) and we can bound the error due to inadequate knowledge of the velocity model by (5.49). Unfortunately, a major complication arises because (5.49) involves specification of a bound, B, on the derivative of the slowness perturbation. As I noted above, specification of this bound in a non-arbitrary fashion is difficult. Furthermore, Parker[1977b], has found that bounds of the type (5.49) tend to be overly pessimistic. He gives a number of alternative bounding criterion that could potentially be applied to this problem. However, I have elected to not pursue this approach because his alternatives appear to be computationally intractable when the effect of measurement errors is considered. I have chosen instead to adopt a general procedure discussed by Backus [1970a]. This procedure is closely related to the trade-off analysis of Backus and Gilbert [1970] that I have used for assessing the reliability of the velocity model, but the interpretation of the results is different. In this analysis the weights, w_i, in (5.44) are found such that they minimize the total prediction error for each parameter that is given by[11]

$$e_i^2 = \|\underline{\varepsilon}_i\|^2 B^2 + \sum_{j=1}^{M_n} W_{ij}^2 \qquad i = 1, 2, 3, \ldots, N \qquad (5.65)$$

where $\|\underline{\varepsilon}_i\|$ is the norm of the error function defined in (5.11) and

---

[11] The form shown here assumes a data covariance given by (5.64). A more general form could be derived by (5.63).

W_{ij} defines the matrix of weights in (5.54). The first term is (5.65) is a measure of the misfit error due to inadequate knowledge of the velocity structure, and the second term is the random error from second moment statistics of the data. For a given value of B , the matrix $\mathbf{W}$ calculated by this criterion is

$$\mathbf{W}^T = (\underline{\mathbf{N}} + \frac{1}{B^2}\mathbf{I})^{-1}\mathbf{X} \qquad (5.66)$$

where $\underline{\mathbf{N}} \in R^{M_n \times M_n}$ is as in (5.24) and I define $\mathbf{X} \in R^{M_n \times N}$ as

$$X_{ij} = <\underline{N_i} \,|\, \underline{X_j}> \qquad (5.67)$$

Because of the similarity of the form of (5.66) and the damped least squares solution (Levenberg-Marquardt inverse) defined in equation (2.38) I will refer to any solution calculated from (5.66) as a **damped solution.**

Since specifying B in a non-arbitrary way is difficult, the approach I take here is to study how the model and statistical errors compare in size as B is varied over the range one considers reasonable (Note this is equivalent to varying the measure of roughness of the true velocity model, κ, defined in (5.51).). I do this by calculating the matrix of weights, $\mathbf{W}$, by equation (5.66) for each desired value of B. Once $\mathbf{W}$ is calculated, the model errors can be bounded by (5.49) and the predicted statistical errors can be calculated by (5.64). This calculation can be repeated for any value of B one considers reasonable[12]. In this way model and statistical errors

---

[12] The actual calculation of $\mathbf{W}$ by (5.66) is much facilitated if the kernels, $<N_i|$, are orthogonalized by the method described in Appendix C or the equivalent method given by Gilbert [1971]. This is useful for two reasons. The first is that the number of calculations required is reduced substantially when $\mathbf{W}$ is calculated for many values of B because the matrix inversion in (5.66) is trivial when $(\underline{\mathbf{N}} + \frac{1}{B^2}\mathbf{I})$ is diagonal. The second reason is that there is a substantial gain in numerical stability. The matrix $\underline{\mathbf{N}}$ is always ill conditioned. Thus, direct application of (5.66) could fail for large value

can be calculated as a function of a single variable, B. The result is a set of curves that are similar to trade-off curves. However, the way one interprets these curves is different, so I shall now discuss some of their general features. Specific examples of this analysis as applied to some real data are discussed in chapter 7.

To understand the nature of the damped solutions it is useful to view them in light of two extreme solutions that they approach in the limit as $B \to 0$ and $B \to \infty$. Consider first the $B=0$ limit. As $B \to 0$ the diagonal elements of $(\underline{\mathbf{N}} + \frac{1}{B^2}\mathbf{I}) \to \infty$ so $(\underline{\mathbf{N}} + \frac{1}{B^2}\mathbf{I})^{-1} \to 0$ and hence $\mathbf{W} \to 0$ [13]. When $\mathbf{W}=0$ we see that (5.49) reduces to

$$\Delta x_i \leq B \, \|\underline{X_i}\| \tag{5.68}$$

since in this case $<\hat{\underline{X}}_i| = 0$. Similarly, (5.64) reduces to

$$\mathbf{C} = \mathbf{H}_B \mathbf{H}_B^T \tag{5.69}$$

This represents an extreme solution that minimizes the statistical error at the expense of maximizing the model error. Note that $\mathbf{H}_B \mathbf{H}_B^T$ represents the smallest possible statistical error[14]. Its size is

of B. This problem can be remedied when $\underline{\mathbf{N}}$ is diagonal by deleting equations associated with small diagonal entries. This yields a solution analogous to the pseudoinverse discussed in chapter 2.

[13] If the constraint (5.12) is imposed $\mathbf{W}$ does not approach zero but instead each component approaches $W_{ij} = \dfrac{N_j(1)}{\sum\limits_{i=1}^{M_n} \underline{N_i}^2(1)}$. This fact alters only the size of the intrinsic error.

[14] The fact the $\mathbf{C}$ does not reduce to zero in this limit is a bit unusual and is a major difference between this analysis and that proposed by Backus [1970a]. The difference arises because Backus considered only parameters that were related to the model by linear functions of the form (using my notation) $x_i = <X_i|\delta u>$. In the analysis presented here the parameters are related to the model by linear functions of the form $x_i = \hat{x}_i + <X_i|\delta u>$. As we saw above the $\mathbf{H}_B \mathbf{H}_B^T$ term arises from $\hat{x}_i$.

independently adjustable by the choice of the matrix $\mathbf{H}_B$ but it will never be zero unless $\mathbf{H}_B$ is zero. Because of this, I will refer to $\mathbf{H}_B \mathbf{H}_B^T$ as the **intrinsic statistical error**. On the other hand, the model error in (5.68) is the largest that is possible for any reasonable solution and is thus the most pessimistic possible bound on the model error. It is of some interest to plot both these error terms as shown in figure 5.4. In that figure I have sketched a plot of the statistical error and model error for this extreme solution as a function of B. $\|X_i\|$ is a constant, so $\|X_i\|B$ plots as a straight line with slope $\|X_i\|$. The statistical error, $\sqrt{C_{ii}}$, is also a constant and hence plots as a horizontal line[15]. A major point of interest from figure 5.4 is that this solution allows us to divide this graph into two regions. These regions are labeled $\boxed{1}$ and $\boxed{2}$ in figure 5.4. The boundary between these two regions is marked by the vertical line that passes through the intersection of the model and statistical error lines. Region $\boxed{1}$ is that area of the graph to the left of this line. It represents a region where the model error is always small compared to the statistical error. This is the case, because $\|X_i\|B$ is the most pessimistic bound on the model error that is possible. Consequently, the actually model error can be expected to be considerably less. Region $\boxed{2}$, on the other hand, is a region where the model errors dominate the errors for this estimate. The errors predicted for this region are again presumably somewhat pessimistic because the misfit error is as large as is reasonable. One suspects that this error can be reduced considerably as we shall see shortly is the case. It is interesting to note, however, that figure 5.4 is actually

---

[15] I have chosen to use the **standard error** defined as $\sqrt{C_{ii}}$ as a measure of the statistical error in the ith parameter rather than a more complete description using error ellipsoids as is frequently used in modern hypocenter location error analysis (see e.g. Klein [1978]). This was done primarily to reduce the required computational burden.
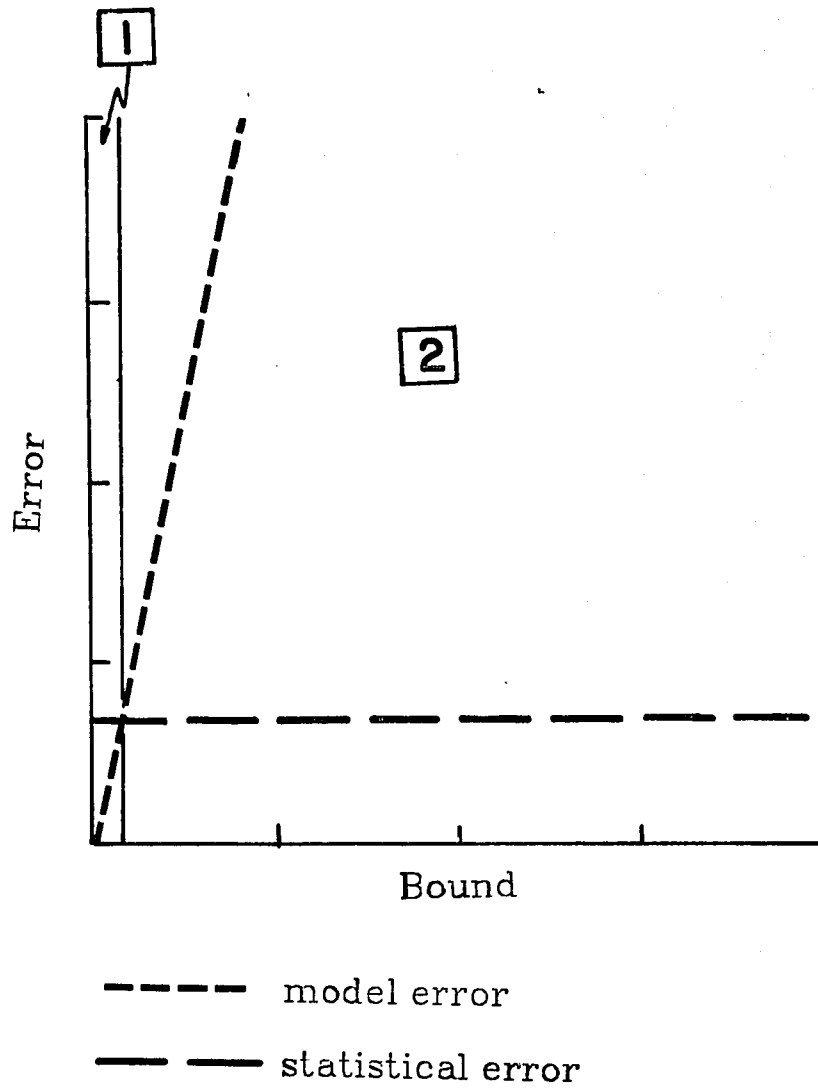
Figure 5.4. Model and statistical errors for $B = 0$ solution.

the proper representation of the error bounds for the way that virtually all hypocenters are currently estimated. That is, hypocenter location procedures inherently assume the velocity model is known exactly and estimate the hypocenter coordinates by setting $W=0$ and declaring that the error is given by $\mathbf{H}_B \mathbf{H}_B^T$. The analysis here shows that this error estimate is overly optimistic as it ignores the contribution from the model error.

The second extreme solution that we need to consider is that for the case as $B \to \infty$. As $B \to \infty$, $\frac{1}{B} \to 0$ and (5.66) reduces to

$$\mathbf{W}^T = \underline{\mathbf{N}}^{-1} \mathbf{X} \tag{5.70}$$

where $\underline{\mathbf{N}}$ and $\mathbf{X}$ are as defined previously in (5.29) and (5.67). This solution is the same as the unique solution that minimizes the root mean square misfit of the error function $<\underline{\varepsilon}_i|$ defined in (5.47) [Johnson and Gilbert, 1972][16]. Because of this, I will denote quantities associated with this estimate by the label $b.f.$, which is an abbreviation for "best fit" This particular solution is the opposite of the $B=0$ case discussed above. This solution makes the model error as small as possible from the available data, but ignores statistical errors. This estimate would be optimum if the observed arrival times were measured with no error. However, because the observed arrival times always contain random errors, the statistical errors calculated for this estimate are generally very large. Thus, (5.70) produces a solution where the model error is made as small as

---

[16] As a practical matter $\underline{\mathbf{N}}$ is nearly always sufficiently ill conditioned to be numerically singular. Because of this, in my implementation of this procedure I have actually used a pseudoinverse solution using a tolerance parameter set to discard eigenvalues of $\underline{\mathbf{N}}$ that were not significant to machine precision. This provides a stable means of calculating $\mathbf{W}$ in the case when B is large but has the disadvantage of causing the results to sometimes depend on the size of the tolerance parameter (see chapter 7).

possible at the expense of increasing the statistical error substantially. In figure 5.5 I have drawn a sketch of how the errors for this estimate might plot in the graph introduced in figure 5.4. The point to observe is that the model error again plots as a straight line passing through the origin. However, this line has a smaller slope than the corresponding line in figure 5.4 because the slope of the line is given by $\|\varepsilon_i\|_{b.f.}$ and $\|\varepsilon_i\|_{b.f.} < \|X_i\|$. Similarly, the statistical error again plots as a horizontal line that now lies considerably higher on the error axis since the error for this estimate is always larger than the intrinsic statistical error. As before, it is useful to divide the error plane into two regions defined by the point where the model and statistical error lines intersect. I call the region to the left of a vertical line through this point region ③ and the region to the right of this line region ④. Region ③ is similar to region ① in figure 5.4. In region ① statistical errors dominate. Similarly, region ④ is similar to region ② in figure 5.4. However, region ④ has a more fundamental importance. The model error sketched in figure 5.5 is the smallest that is possible with the available data[17]. As a practical matter, this means that if we believe B is sufficiently large that the errors lie in region ④, then any estimate of x_i by (5.53) will be dominated by model errors. This fact occurs because the model error for this best fit solution is as small as possible, but it is still larger than the statistical error. This problem can be alleviated only by using more data or modifying the bounding criterion on the model error.

I introduced the two extreme solutions illustrated in the graphs in figures 5.4 and 5.5 because they are fundamental as endpoints in the family of damped solutions given by (5.66). To show the

---

[17] This statement should really be qualified by the added clause "using the bounding criterion (5.49)" The model error could also potentially be reduced further be using a less pessimistic bounding criterion such as those given by Parker [1977b] or Jackson [1979].
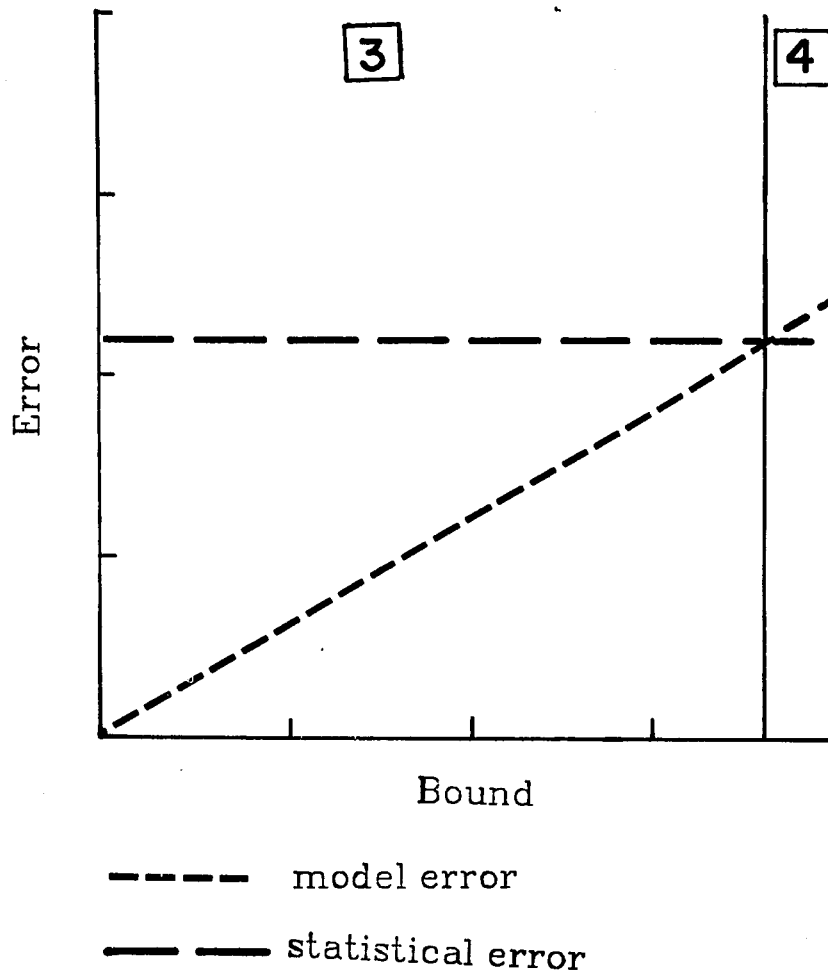
Figure 5.5. Model and statistical errors for $B \to \infty$ (best fit) solution.

relationship of these extreme solutions to the damped solutions, it is useful to combine them all on a single graph like those in figure 5.4 and figure 5.5. Combining them on a linear-linear graph as I have sketched there presents a display problem, however. The reason for this is that I have invoked a considerable amount of artistic license in figures 5.4 and 5.5. In every real case I have considered, the extreme misfit errors, $\|X_i\|$ and $\|\varepsilon_i\|_{b.f.}$, and the extreme statistical errors, $\mathbf{H}_B \mathbf{H}_B^T$ and $\mathbf{C}_{b.f.}$, differ by by many orders of magnitude. As a result, it is preferable to combine the two solutions on a log-log plot as sketched in figure 5.6. This changes only the appearance of the resulting display. The statistical errors still plot as horizontal lines. The model errors, however, both plot in this display as lines with a slope of unity but separated vertically by $\log\|X_i\| - \log\|\varepsilon_i\|_{b.f.}$. Regions $\boxed{1}$ and $\boxed{4}$ from figures 5.4 and 5.5 map unchanged into figure 5.6. This combined display, however, leads to the definition of a new region of this graph that is labeled $\boxed{5}$ in figure 5.6. Regions $\boxed{1}$ and $\boxed{4}$ are characterized by the fact that if B lies in these regions one of the extreme solutions, $B=0$ or $B=\infty$, provide the optimum estimate for the ith parameter since in these regions either the statistical errors or model errors always dominate. The region label $\boxed{5}$ is, on the other hand, characterized by the fact that it is a grey region between these two extremes. In this area both model and statistical errors are potentially important. It is in this region that the damped solution plays a useful role because it recognizes the existence of both types of error and finds the optimum solution that minimizes the total error.

To help clarify the nature of the damped solutions, I turn finally to a real example. Figure 5.7 is a log-log plot, similar to figure 5.6, of the errors calculated for the x-component of the hypocenter of one event from a data set collected by a local network near Cape Mendocino in northern California (The reader is referred chapter 7 for a more complete discussion of the data that was used here.).
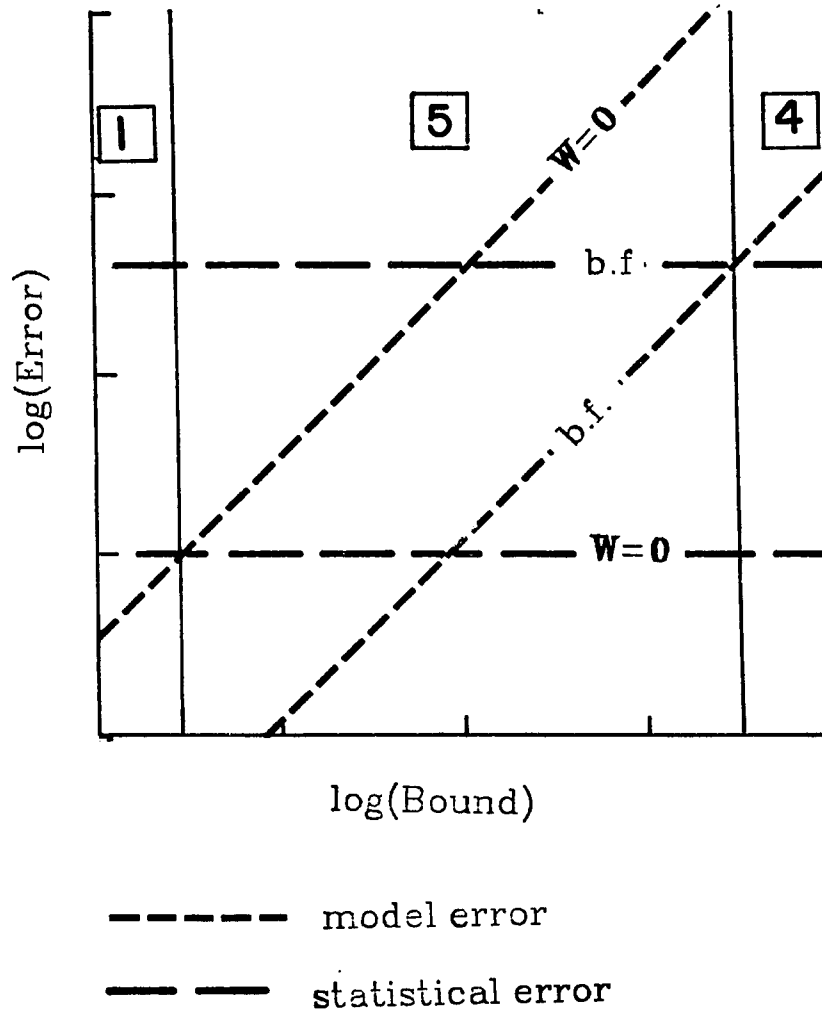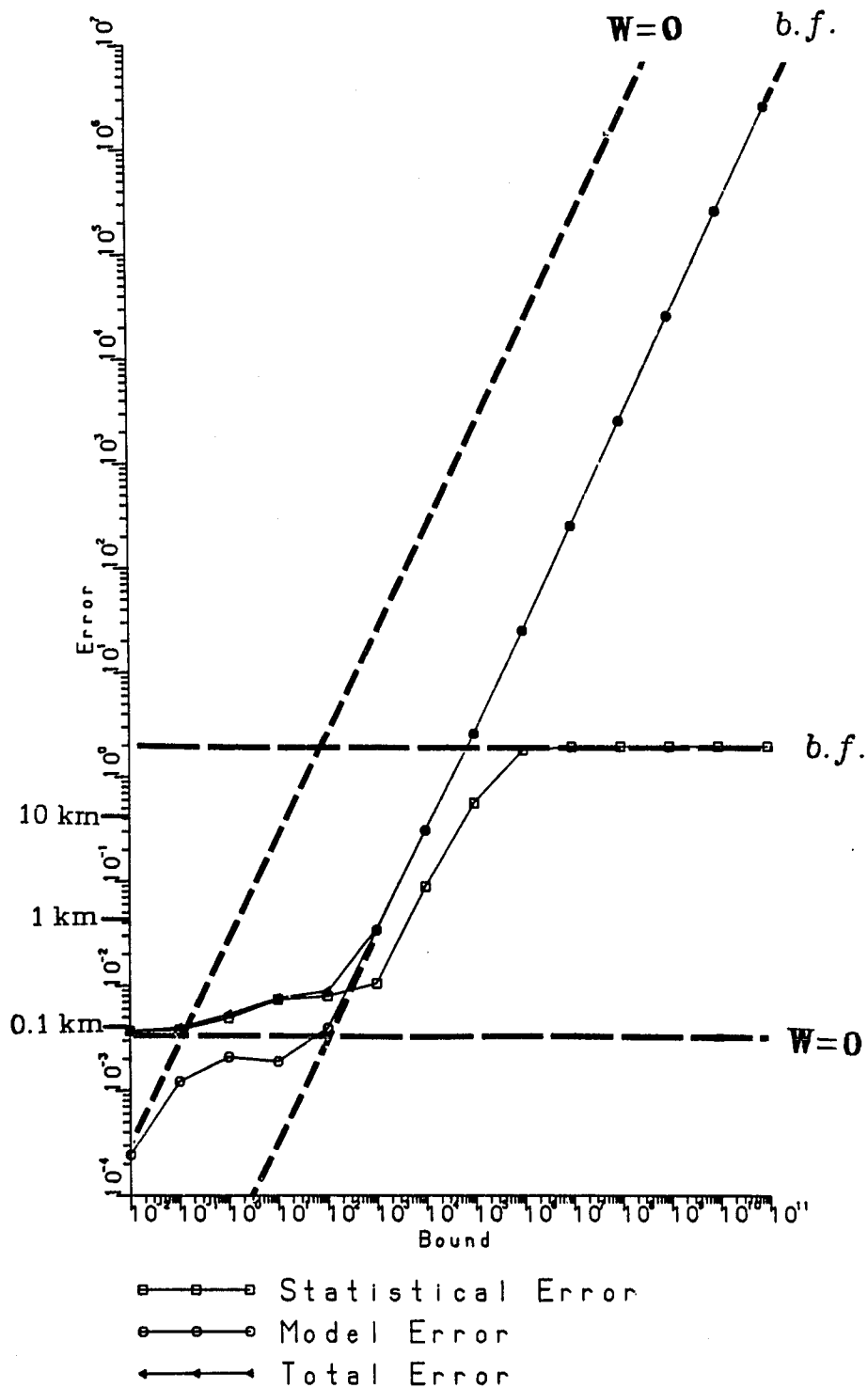
Figure 5.6. Log-log plot of both extremal solutions.

Figure 5.7. Example parameter tradeoff curves. Plot is for the x (north) component of the hypocenter of event 1482 from Humboldt Bay network data (origin time: 75 225 17 34 36.10). Both axes are in nondimensional units. Three representative lengths in physical units are noted on the error axes.

x component

The errors from the extreme solutions produce the characteristic set of four intersecting lines as in figure 5.6. Of major interest here, however, are the other three labeled curves. These curves define the errors in the family of all optimum solutions calculated by (5.66) for this parameter. The curves are calculated simply by calculating the weights for each given value of B by (5.66). The statistical errors can then be calculated by (5.64), and the model error can be bounded by (5.49). The "total error" is then calculated from

$$e_{total} = [\|\underline{\varepsilon}_i\|^2 B^2 + C_{ii}]^{1/2} \qquad (5.71)$$

The three curves in curves in figure 5.7 are produced by plotting each of these error terms ($\|\underline{\varepsilon}_i\|B$, C_{ii}, and e_{total}.) as a function of B, the bound on the norm of the derivative of the deviation of the true model from the estimate. The advantage of this analysis is that it allows us to assess the relative importance of the model and statistical errors over a large range of bound values. For this particular example the errors are characterized by two regions. For bounds less than 10^2 the errors are dominated by statistical errors but for bounds greater than 10^2 model errors become increasingly dominant. As a practical matter this says that if we believe $B < 10^2$ then the standard error as calculated by a typical earthquake location routine is a reasonably accurate measure of the true error.

Two features of figure 5.7 are universal. The first is that the total error curve is always asymptotic to the $\|\underline{\varepsilon}\|_{b.f.}$ line as $B \to \infty$. The second is that the total error curve is also asymptotic to the horizontal line defining the intrinsic statistical error as $B \to 0$. This is the major reason I spent the time to introduce to these extremal solutions since the damped solution reduces to them in these limits. Everything else (including the scale) is highly dependent upon the actual data. As a result, I will defer any further discussion of interpretation of this kind of analysis to chapter 7 where it can be evaluated in a better context.

CHAPTER 6
TESTS WITH SYNTHETIC DATA

1. INTRODUCTION

The first step to verify the validity of a procedure like this one is to test it on artificial data generated from known models. This type of experiment serves two purposes. The first is the very pragmatic need to test our computer programs. The second purpose of these tests was to attempt to answer four key questions;

(1) What are the limitations of linearization?

(2) What are the limits imposed by random error in the measurements of arrival times.

(3) What type and distribution of sources provides the most information from the fewest data?

(4) Is the method capable of resolving low velocity zones and what distribution of sources is required to resolve them.

This study has focused on question (1) because I believe it is the single most important question to ask about this problem. The reason I state that with such dogmatism is that the relationship of the data (arrival times of direct P and S waves) to the model (elastic wave velocity structure and the hypocenters of earthquakes) is nonlinear. This inversion method rests on a local linearization to the nonlinear equations relating the observed data to the model (see chapter 4). It shares this limitation with all other existing schemes for solving this problem (e.g. Crosson [1976a] or Aki and Lee [1976]) The question of how good the linear approximation is, is fundamentally important, since the only general method of assessing errors in inverse problems is valid only for linear problems (see chapter 5). The linear assessment theory can be extended to nonlinear problems provided the problem is sufficiently well behaved that we

can somehow find an estimate of the true earth model that is "linearly close" to the truth [Backus and Gilbert, 1968]. We shall discuss what is meant by "linearly close" at some length in the following paragraphs, but for the moment the point the reader must realize is that unless we can be confident that we can obtain a good estimate in this sense, we cannot say anything quantitative about our estimates. I have found that estimates that are good enough to permit a linear assessment are possible but not easy. Several pitfalls that must be avoided have been disclosed. In addition, these experiments have gone a good distance toward answering the other three questions outlined above.

2. LIMITS OF LINEARIZATION: THEORY

Synthetic data permits one to analyze the results of inversion of that data in some ways that are not possible with real data. The reason is simply that with artificial data one always knowns the true model from which the data were derived. In this section I define explicitly what is meant by the statement that two models are linearly close and then describe a method of analysis that can be used to determine if two models are linearly close. This analysis is inherently useless for real data, since it requires a priori knowledge of the true model. However, it is an important method of analysis for artificially generated data since it will help us determine how far we can push the linearization approximation.

I have described the theory for the method of inversion I have used here in the previous chapters. The theory developed there is based on the linearization given by equation (4.2). A fundamental feature of this method is a transformation of the arrival time data that produces a new "annulled" set of data that are locally independent of the earthquake hypocenters and depend only on the velocity model. The result we need for the present is that if we record a

total of M arrivals from m_e earthquakes we can obtain a set of $M_n = M - 4m_e - n_s$ (see discussion preceding equation (5.1)) linear equations of the form

$$n_i = <N_i \,|\, \delta u>$$ (6.1)

(see section 3 of chapter 4). These M_n equations constitute what Pavlis and Booker [1980] have called the "annulled data set" The functions $<N_i\,|$ are as defined in equation (4.20) or (B.23), and the inner product symbolism is defined here by equation (4.8).

To review, the basic approach here is to estimate the model perturbation at some depth z_0 as

$$\delta \hat{u}(z_0) = \sum_{i=1}^{M_n} w_i n_i$$ (6.2)

In particular the weights w_i are calculated to minimize

$$\| <H(z_0 - z)\,| - \sum_{i=1}^{M_n} w_i <\underline{N_i}\,|\ \|$$ (6.3)

where $<H(z_0 - z)\,|$ is the Heaviside function. $<\underline{N_i}\,|$ is defined by equation (5.10). In this way we can construct an estimate of the slowness perturbation function $|\delta u>$, which I denote by the symbol $|\delta \hat{u}>$. The slowness perturbation function we are trying to estimate is

$$|\delta u> = |u_{true}> - |\hat{u}>$$ (6.4)

where $|\hat{u}>$ is some current estimate of the true slowness function, $|u_{true}>$. With real data the function $|\delta u>$ is not known since $|u_{true}>$ is what we are trying to determine. When $|u_{true}>$ is known, as it is here, we can compare the estimate (6.2) to the following

$$\delta u_{smoothed}(z_0) = \sum_{i=1}^{M_n} w_i <N_i\,|\, \delta u>$$ (6.5)

The reader will note that $\sum_{i=1}^{M_n} w_i < N_i \mid$ is the resolution function for the estimate at depth z_0 by equation (5.23). Hence, $\delta u_{smoothed}$ is the true perturbation smoothed by this resolution function. If this problem were linear and the data had no errors, we see (6.5) and (6.2) would always give the same result in view of equation (6.1). However, because this problem is nonlinear, (6.5) and (6.2) will generally not be the same, even with error free data. As a result, the difference

$$e_{nonlinear}(z_0) = \delta u_{smoothed}(z_0) - \delta \hat{u}(z_0) \qquad (6.6)$$

is a measure of the affect of nonlinearity on our solution at z_0. The importance of the affect of nonlinearity on our solution depends upon how $e_{nonlinear}(z_0)$ compares to the potential error due to measurement error. That is, $e_{nonlinear}(z_0)$ being nonzero is irrelevant if $e_{nonlinear}(z_0)$ is negligible compared to the expected error due to noise. Consequently, we need to analyze how $e_{nonlinear}$ trades off with the other two major variables of inversion, resolution and variance. This analysis is greatly facilitated if we perform the transformation termed "ranking and winnowing" by Gilbert [1971] and "spectral expansion" by Parker [1977a] (see also section 4 of chapter 4). In this way we can convert the M_n equations of (6.1) into M_n equations we can write as

$$a_i = < \underline{\Psi}_i \mid \frac{d}{dr}(\delta u) > \qquad (6.7)$$

which have the properties

$$< \underline{\Psi}_i \mid \underline{\Psi}_j > = \delta_{ij} \qquad (6.8)$$

and

$$E[a_i a_j] = \frac{1}{\lambda_i} \delta_{ij} \qquad (6.9)$$

That is the functions $<\underline{\Psi}_i|$ form an orthonormal set and the transformed data, a_i are statistically independent.

With this simplification we can examine the sequence of partial sums

$$l_K(z_0) = 12(z_0 - \sum_{i=1}^{K} b_i^2) \tag{6.10}$$

$$\sigma_K^2(z_0) = \sum_{i=1}^{l} \frac{b_i^2}{\lambda_i} \tag{6.11}$$

and

$$e_{nonlinear_K}(z_0) = \sum_{i=1}^{K} b_i <\underline{\Psi}_i | \frac{d}{dr}(\delta u)> - \sum_{i=1}^{K} b_i a_i \tag{6.12}$$

where

$$b_i = <H(z_0 - z)|\underline{\Psi}_i> \tag{6.13}$$

$l_K(z_0)$ is the resolution length at depth z_0, a measure of the smoothing performed by the resolution function. $\sigma_K^2(z_0)$ is the predicted variance from statistical error. $e_{nonlinear_K}(z_0)$ is as in (6.6).

The actual analysis used is this: We compare the error of our estimate caused by nonlinearity, (6.12), with the expected statistical error, (6.11), as we increase the expansion order, K. If $e_{nonlinear_K}$ is small compared to σ_K for the range of K giving reasonably small statistical error, then we can say model $|\hat{u}>$ is linearly close to the truth. "Linearly close" in this context means that the true slowness perturbation when smoothed by all acceptable resolution functions, is the same as our estimate of this perturbation to within the precision imposed by measurement errors. If a reference model, $|\hat{u}>$, is "close to" the true model, $|u_{true}>$, in this sense, then this is an

indication that $|\hat{u}>$ is sufficiently close to the truth that a linear assessment will yield results that are correct. This analysis is used below in two different contexts. In section 2.3 $|\hat{u}>$ is chosen arbitrarily to study the range over which linearity is valid. In section 2.4, $|\hat{u}>$ is determined by ALGORITHM PRIMEL introduced in chapter 4. The analysis I just described can be used to determine if the model the procedure has found is the global or a local minimum.

3. NUMERICAL TESTS

3.1. The Data

A wise experimentalist designs his experiments to address a single problem and tries as hard as possible to isolate his study from contamination by random uncontrolled variables. The same wisdom applies to numerical experiments such as the one reported here. The emphasis in the present study was to determine the limitations of a linear approximation to this, a nonlinear problem. Consequently, all the results shown in this section were produced from basically the same data. By that I mean that the geometry of the sources and receivers is the same throughout. The only variable is the velocity model used to produce the synthetic data. Figure 6.1 shows a map view of the source and receiver geometry used in this study. Figure 6.2 is a cross sectional view from the left of figure 6.1. The configuration of the 18 stations of this network is that of the local earthquake network in the area of Humboldt Bay in northern California (see chapter 7). These stations are shown by the triangles in figures 6.1 and 6.2. A total of ten hypothetical sources were used in this study. The locations of these sources are shown in figures 6.1 and 6.2 by asterisks. Finally, 10% of the arrivals were randomly eliminated. This was done to loosely simulate a normal
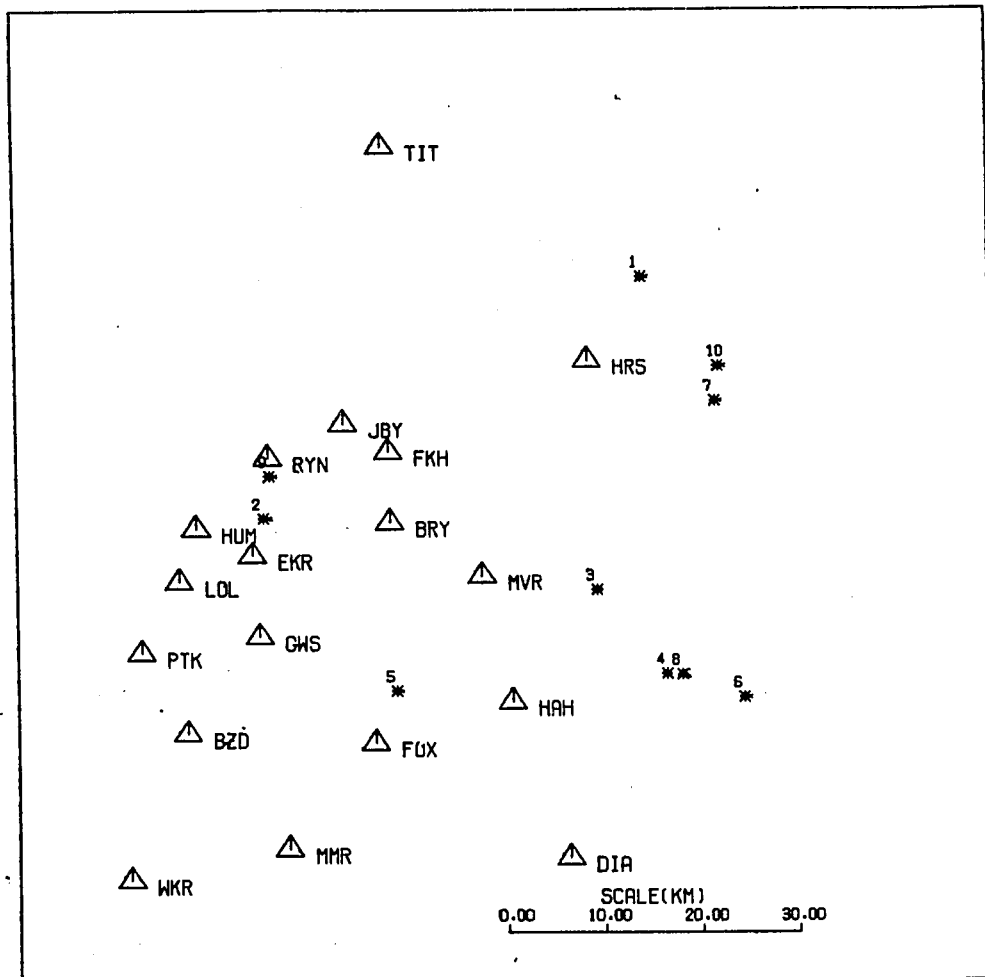
Figure 6.1. Map view of source and receiver geometry of artificial data used in this study. Stations are labeled as triangles. Epicenters of sources are labeled as asterisks.
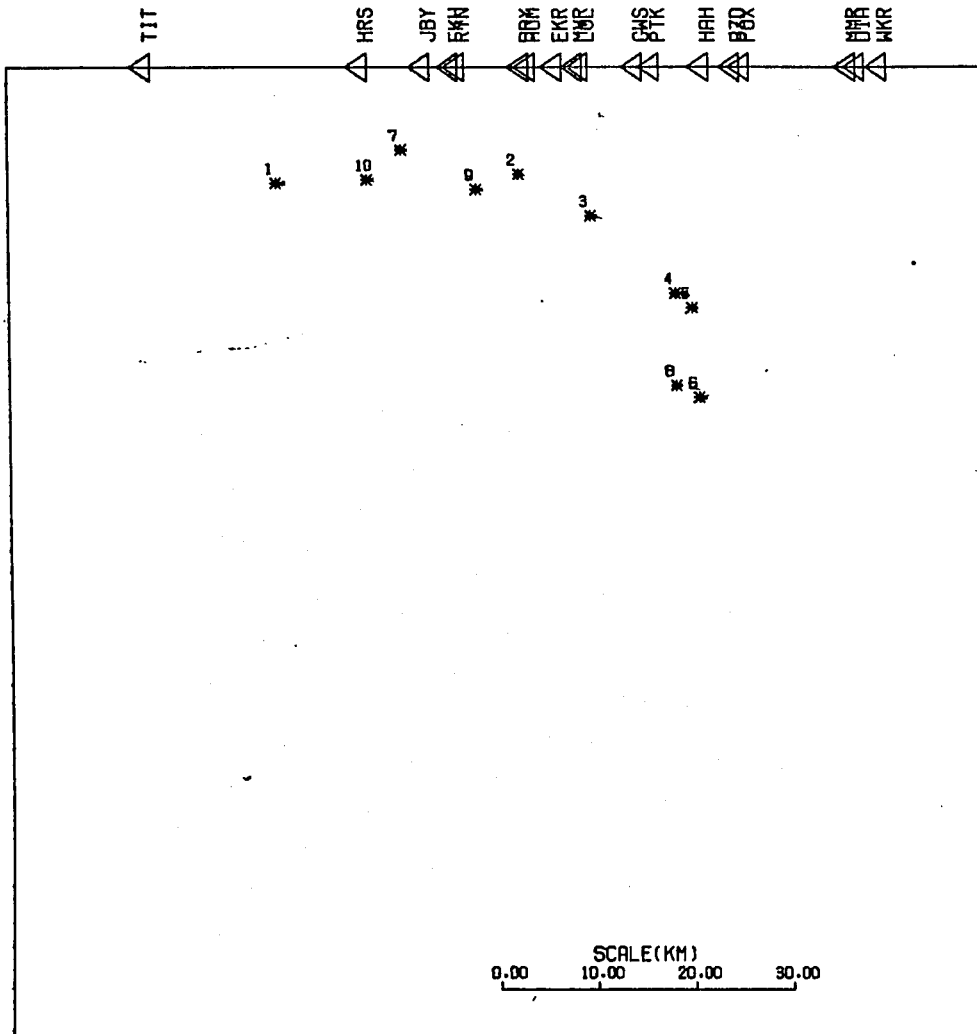
Figure 6.2. Depth distribution of sources for artificial data used in this study. View is a cross section from the left of figure 6.1. Sources are labeled with asterisks.

feature of real data, since only a certain percentage of stations normally record a single event. A summary of which stations record a given events is shown in figure 6.3.

Station	Event Number									
	1	2	3	4	5	6	7	8	9	10
BRY		**	**	**	**		**	**		**
DIA	**	**	**	**	**	**	**	**	**	**
EKR	**	**	**	**	**	**		**		**
FKH	**	**	**	**	**		**	**	**	**
FOX	**	**	**	**	**	**	**	**	**	
GWS	**	**	**		**	**	**	**	**	**
HAH	**	**	**	**	**	**	**	**	**	**
HRS	**	**		**	**		**	**	**	**
JBY	**		**	**	**	**	**	**	**	**
LOL	**	**	**	**	**	**	**	**		**
MMR	**	**	**	**	**	**	**	**		**
MVR	**	**	**	**	**	**	**	**	**	
PTK	**	**	**	**	**	**	**	**	**	**
RYN	**	**		**	**	**	**	**	**	**
TIT		**	**	**		**	**	**	**	**
WKR	**	**	**	**	**	**	**	**	**	**
BZD	**	**	**	**	**	**	**	**	**	**

Figure 6.3. Histogram showing what stations recorded arrivals for each event in synthetic data set. Gaps indicate no arrival at that station from that event.

3.2. Limits of linearization: numerical results

The first set of tests I conducted were designed to use the theory sketched above to ascertain how good one's initial guess needed to be to the truth to permit a linear assessment. This was done with the faint hope that perhaps the problem was sufficiently well behaved that we could avoid the inevitable problems involved in an iteration scheme if we could somehow make a good enough initial guess. Unfortunately, it was found that the initial guess, in general, would have to be extremely good and iteration is usually unavoidable.

Travel times were calculated for the data set described above for two different models. These two models are shown as the solid lines in figures 6.4 and 6.5. No random errors were added, so the only errors in the data are from computational rounding error. Tests using a linear model indicate rounding errors are never greater than $\sim$0.005 seconds and are usually less than $\sim$0.001 seconds. This is at least an order of magnitude smaller than typical picking errors for impulsive P wave arrivals so these data are essentially perfect. This is not a moot point because when we calculate $e_{nonlinear}$ by equation (6.12) we can then be assured $e_{nonlinear}$ is not contaminated by noise. This does not affect our analysis for calculating the variance estimates in equation (6.11) since σ_l^2 is determined completely by the data kernels to within a scale factor. With real data the scale factor is determined by estimates of the variance for each datum. Here we have set the scaling so the variance estimates we calculate are those that could be expected if all data points had some constant standard error. We have calculated results for both models assuming a constant picking error with an expected variance of 0.10 seconds. 0.10 seconds was used because we considered that as nominal for poorer quality data like helicorder records. Impulsive arrivals can often be picked with a

precision ~0.05 seconds on film records [Lee and Stewart, 1981] and perhaps half that again on digital records. However, since the scaling of the predicted variance is linear, our results are easily converted to any picking error desired.

Hypocenters were initially located by a standard, nonlinear, least squares procedure (LQUAKE2 written by Robert S. Crosson) using travel times calculated from the linear model shown as the dashed line in figures 6.4 and 6.5. Figures 6.6 to 6.8 all show results of the calculations described above for expansion order, l, of 15. Fifteenth order was chosen because higher order expansions produce large increases in variance with little gain in resolving power. This is roughly equivalent to choosing an optimum smoothing from a set of trade-off curves [Parker, 1977a].

The model in figure 6.4 is identical to the model used by Pavlis and Booker [1980]. The gradient in this model is absurdly large but I used it primarily for continuity with this earlier work. The results of the nonlinearity tests for this model are shown in figure 6.6. We see that although the initial model appears to be quite close to the true model, there is a significant error due to nonlinearity. However, in this case, the nonlinear error is not really significant compared to the noise. Unfortunately, that is not as nice as it may sound since the variance here is sufficiently large that the initial and true model are virtually indistinguishable statistically with the assumed picking error. This is basically a proof of something we already knew. The initial model used was a very good estimate of the true model.

The model in figure 6.5 was chosen for continuity with the results for the model from figure 6.4. The idea was to see what would happen with a very smooth model like the one in figure 6.4, but with a sufficiently large sinusoidal component to produce a triplication in the travel time curves. Furthermore, by using the same
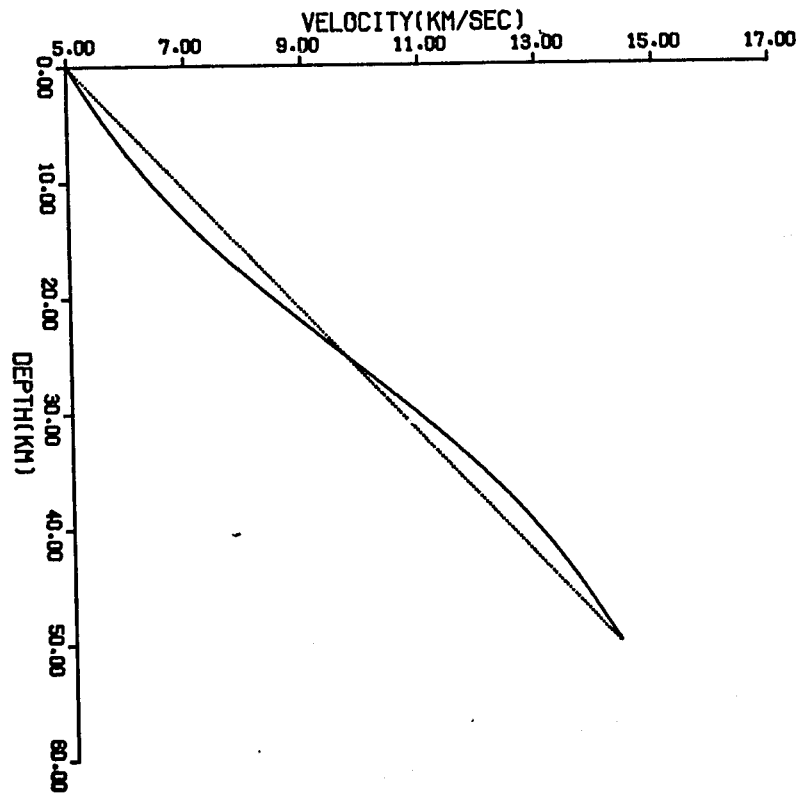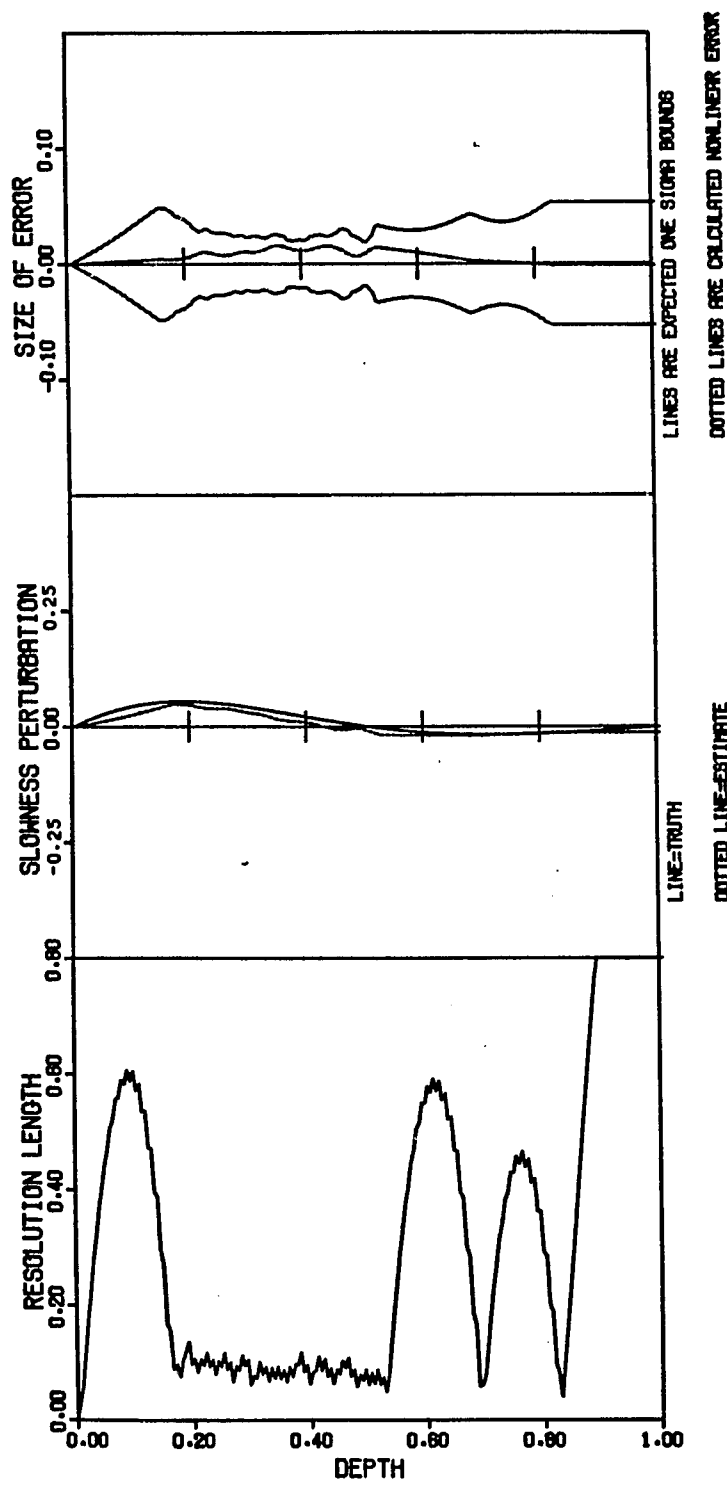
Figure 6.4. True and reference velocity models for first test of limits of linearity study. Solid line is the velocity model used to calculate travel times for artificial data. Hypocenters were located using travel times calculated from linear model shown by dotted line.
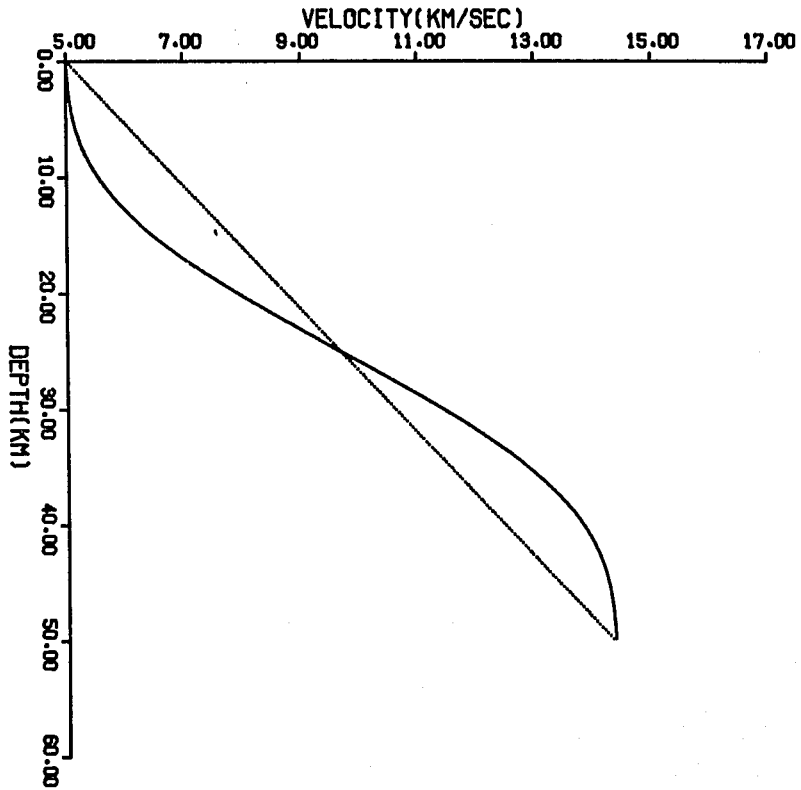
Figure 6.5. True and reference model for second test of limits of linearity study. Travel times for artificial data were calculated from the model drawn with a solid line. Linear velocity model shown as a dotted line is the reference model.

Figure 6.6. Results of nonlinear error analysis using models shown in figure 6.4. Left panel gives resolution length measure from integration quelling as a function of depth. Center panel shows true slowness perturbation and Backus-Gilbert estimate of that perturbation. Solid line is the true perturbation and the dotted line is the estimate of it. Right panel shows errors due to nonlinearity compared to measurement errors. Dotted line is the error from nonlinearity. Solid lines are predicted one standard error bounds assuming a constant 0.10 second picking error in all arrival time measurements. All quantities shown are in nondimensional units. Scales used are the following: length = 60 kilometers, slowness=0.2 sec/km.

SIZE OF ERROR

LINES ARE EXPECTED ONE SIGMA BOUNDS

DOTTED LINES ARE CALCULATED NONLINEAR ERROR

SLOWNESS PERTURBATION

LINE=TRUTH

DOTTED LINE=ESTIMATE

RESOLUTION LENGTH

DEPTH

initial model, the procedure was required to produce a perturbation of identical form, but much larger amplitude from the earlier study of the model in figure 6.4. The results for this case are shown in figure 6.7. Note that in this case the affect of nonlinearity cannot be ignored. The error due to nonlinearity is at least three times the noise level. Moreover, a linear assessment would clearly give wrong results here. Not only is the nonlinear error large but the resolution length and variances calculated here are very different from those in figure 6.6.

A question which arose was whether the bad results discussed above where caused by nonlinearity in the velocity inversion or by nonlinearity due to source mislocation. To test this, the sources were fixed at their true location and the same analysis described above was repeated. These results are plotted in figure 6.8. Observe that this has helped considerably but has not totally solved the problem. Hence, it appears that both source and velocity errors have contributed to the large nonlinear error here but the source mislocation appears to cause the most severe problems. This clearly demonstrates that an iterative method is required in this case.

3.3. Convergence Study

Chronologically the work described in the previous section was the earliest test I made. The conclusions derived from it led directly to the development of PRIMEL introduced in chapter 4. This algorithm was tested for a number of examples. Figure 6.9 shows the results obtained from five iterations of this procedure on the same data that was analyzed in figure 6.7. The final model almost exactly reproduces the true model and is indeed linearly close to the truth. The failure at the bottom of the model is of no consequence as the resolution blows up there because the deepest
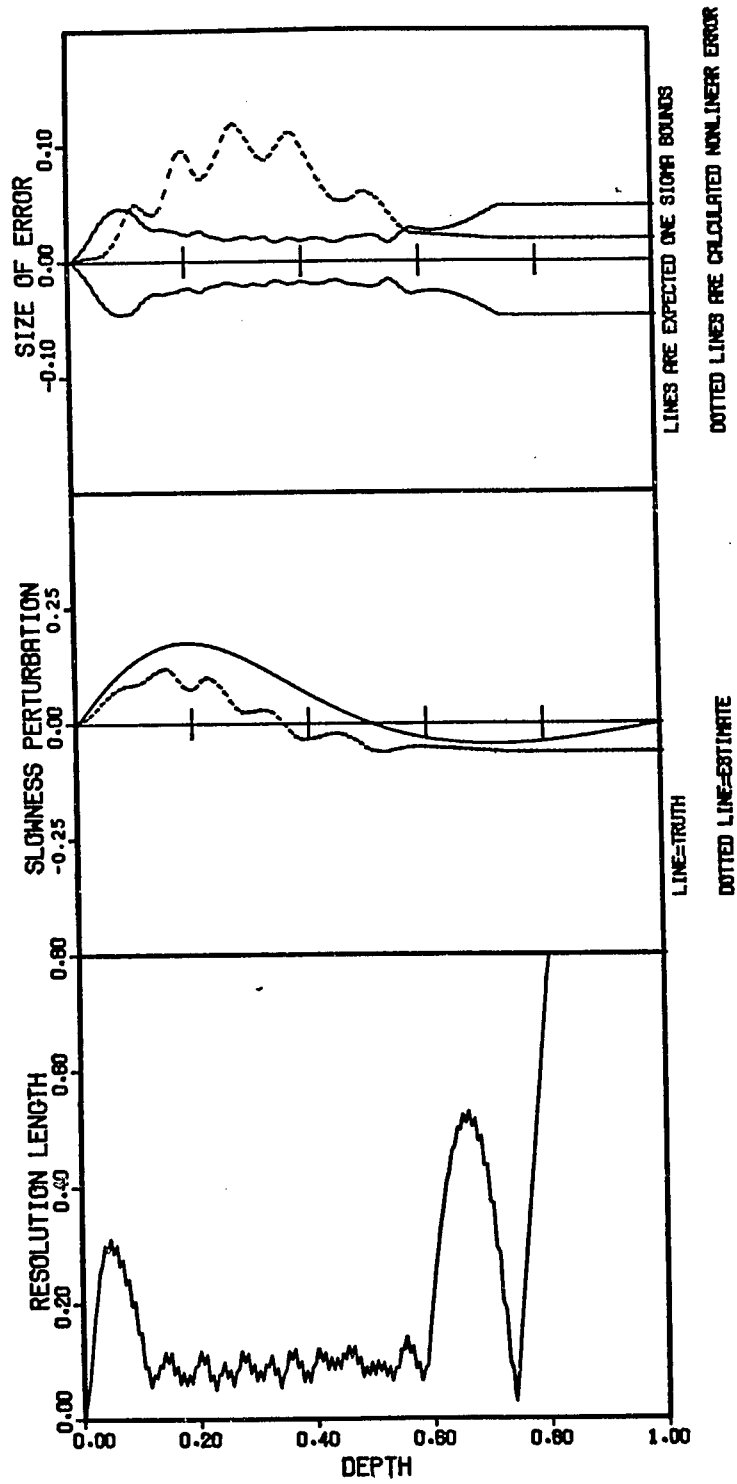
Figure 6.7. Same as figure 6.6 but using data calculated from the model in figure 6.5. (See figure 6.6 caption for further description.) Hypocenters for this data were located using travel times from the linear model in figure 6.5. All quantities shown are in nondimensional units. Scales used are the following: length = 60 kilometers, slowness=0.2 sec/km.
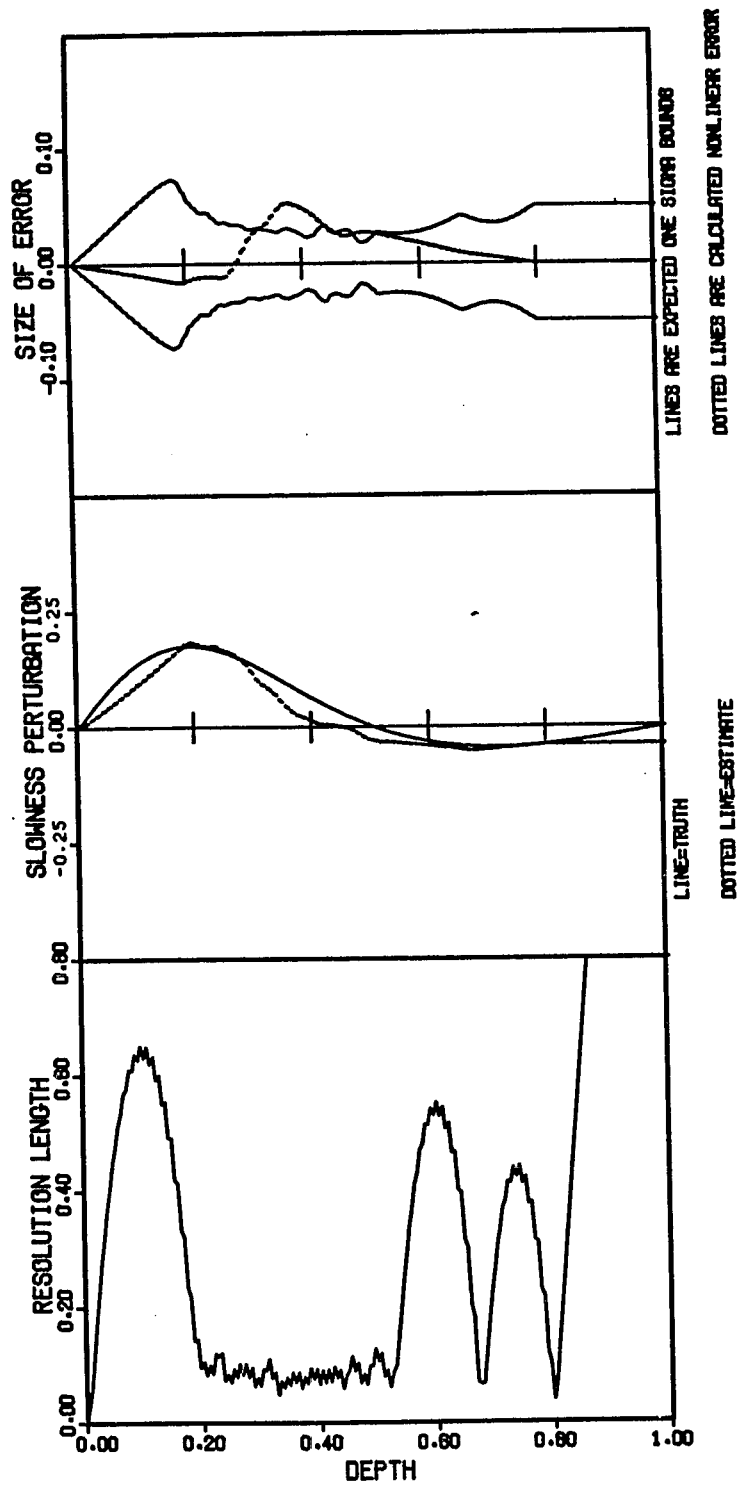
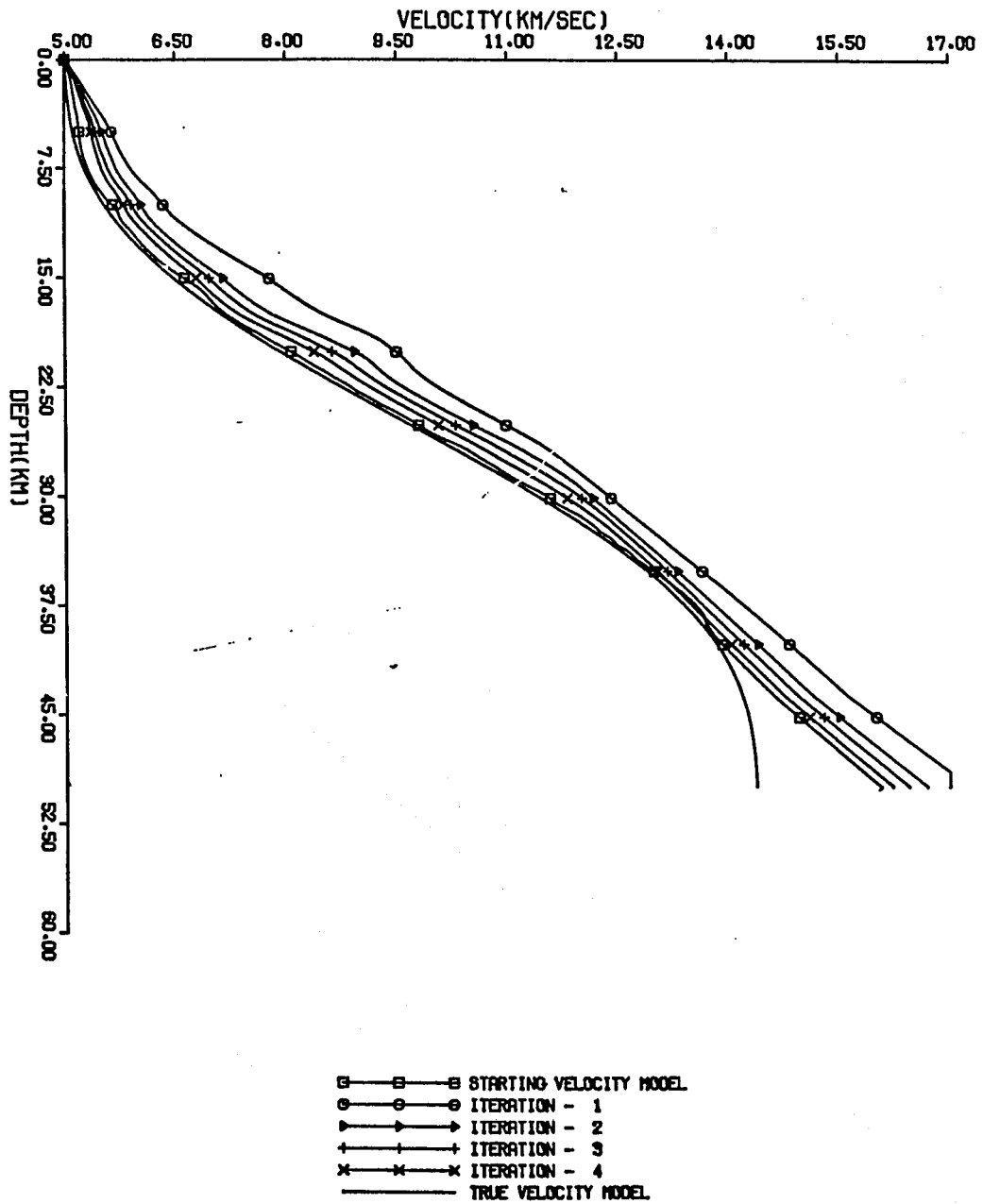**Figure 6.8.** Same as figure 6.7 but here hypocenters were fixed at their true positions. (See figure 6.6 caption for further description.) All quantities shown are in nondimensional units. Scales used are the following: length = 60 kilometers, slowness=0.2 sec/km.

Figure 6.9. Convergence history of iterative procedure using data from the model in figure 6.5.

source in this data set is at 39.2 kilometers. Good results like this are typical for data derived from very smooth models. Additional tests not shown suggest that generally the estimate one obtains is reasonably independent of the starting model when the true model is very smooth.

If all my results were as good as that in figure 6.9 it would be very pleasing. However, there is always fear in an iterative procedure like this that the procedure may converge to a solution that apparently fits the data but is not linearly close to the truth. Two such potential pitfalls have, in fact, been found. The first is illustrated by the results plotted in figure 6.10. This model was chosen to study the effect and resolvability of low velocity zones. The results were not encouraging. The solution has clearly converged since the third and fourth iteration velocity models are virtually indistinguishable. This final velocity model, however, is not linearly close to the true model. This apparently happened because of the combination of two deficiencies of this data set. The shallowest source was at a depth of 8.5 kilometers. The resolution of the inversion above that depth is very poor. As a result the procedure was incapable of producing a reasonable estimate of the model near the surface. The second problem with this data is that almost all the rays joining sources and receivers are direct rays that travel upward from the source position. Hence, most ray paths were close to straight lines joining source and receiver. This caused the errors to tend to average to small numbers because the estimated model is too low near the surface and too high at depth. The procedure found a local minimum solution that is not linearly close to the truth.
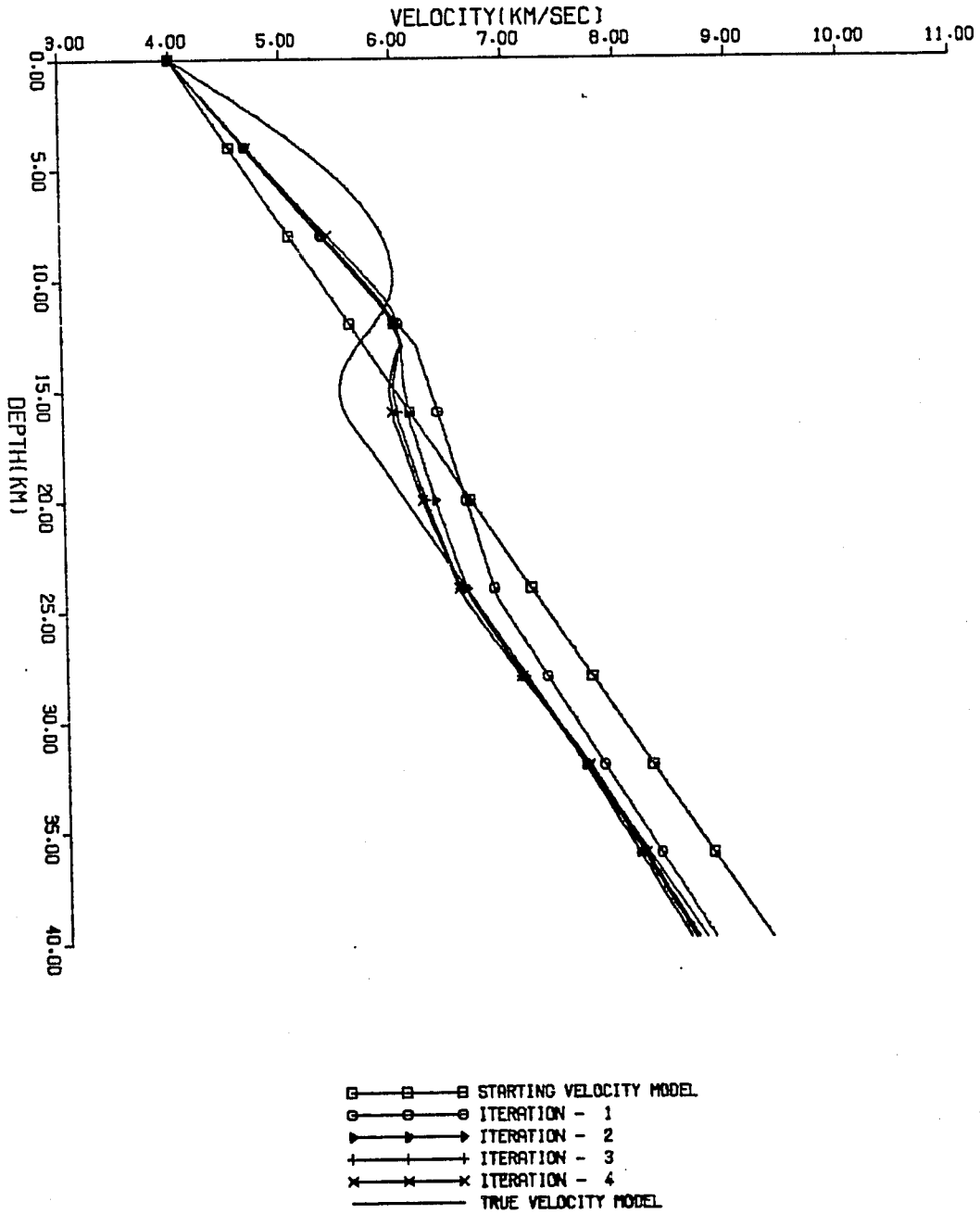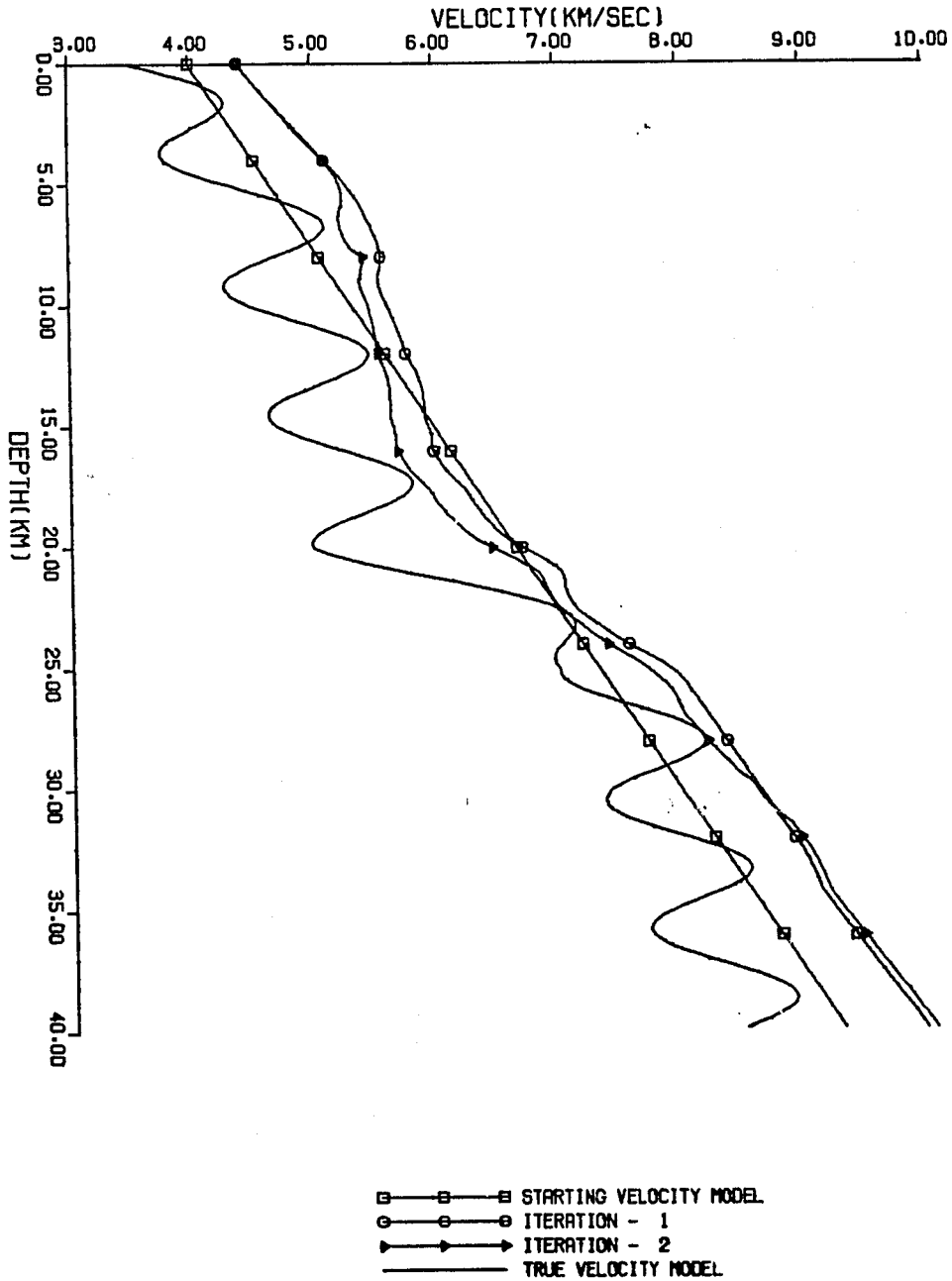
Figure 6.10. Convergence history for low velocity zone model.

A second convergence problem is revealed by the results illustrated in figures 6.11 and 6.12. The model used to generate the data in this case is the wiggly curve with no labels. This model was designed to determine if the algorithm could recover the low frequency information contained in the true model without being affected by the unresolvable high frequency information. The results in figure 6.11 were obtained with the surface velocity fixed at 3.5 km/sec, the true value. Here the inversion has indeed recovered the low frequency information in the true model but the velocity estimates tend to be too large. This is compensated by errors in the origin times of all events ~+0.2 seconds. The net affect is a model that fits the data but has velocities that are too high. It is also interesting to note that the final model tends to follow the peaks of the sinusoidal oscillations in the model that form alternating high and low velocity zones. Many of the arrivals in the artificial data from this model were refractions off the high velocity zones. This could be an alternate explanation for offset of the estimated model as the refractions produce arrivals with apparent velocities at the peaks of the sinusoidal oscillations in the model.

The offset problem is intensified in a similar study summarized in figure 6.12. The difference here is that the surface velocity was free to vary. The model labeled with the small triangles was found to fit the data. This model is similar to the final model in figure 6.11, except it is shifted to the right ~ 1 km/sec on the velocity scale. Moreover, in this case the errors in the origin times are even larger, ~ +0.5 seconds. These results reveal a second convergence problem with this procedure. Earthquake data consisting of only P wave arrival times, measure only relative, not absolute travel times. As a result, there appears to be a considerable trade-off between the origin times of the earthquakes in the data set and the dc level of the velocity model. The procedure can converge to a local
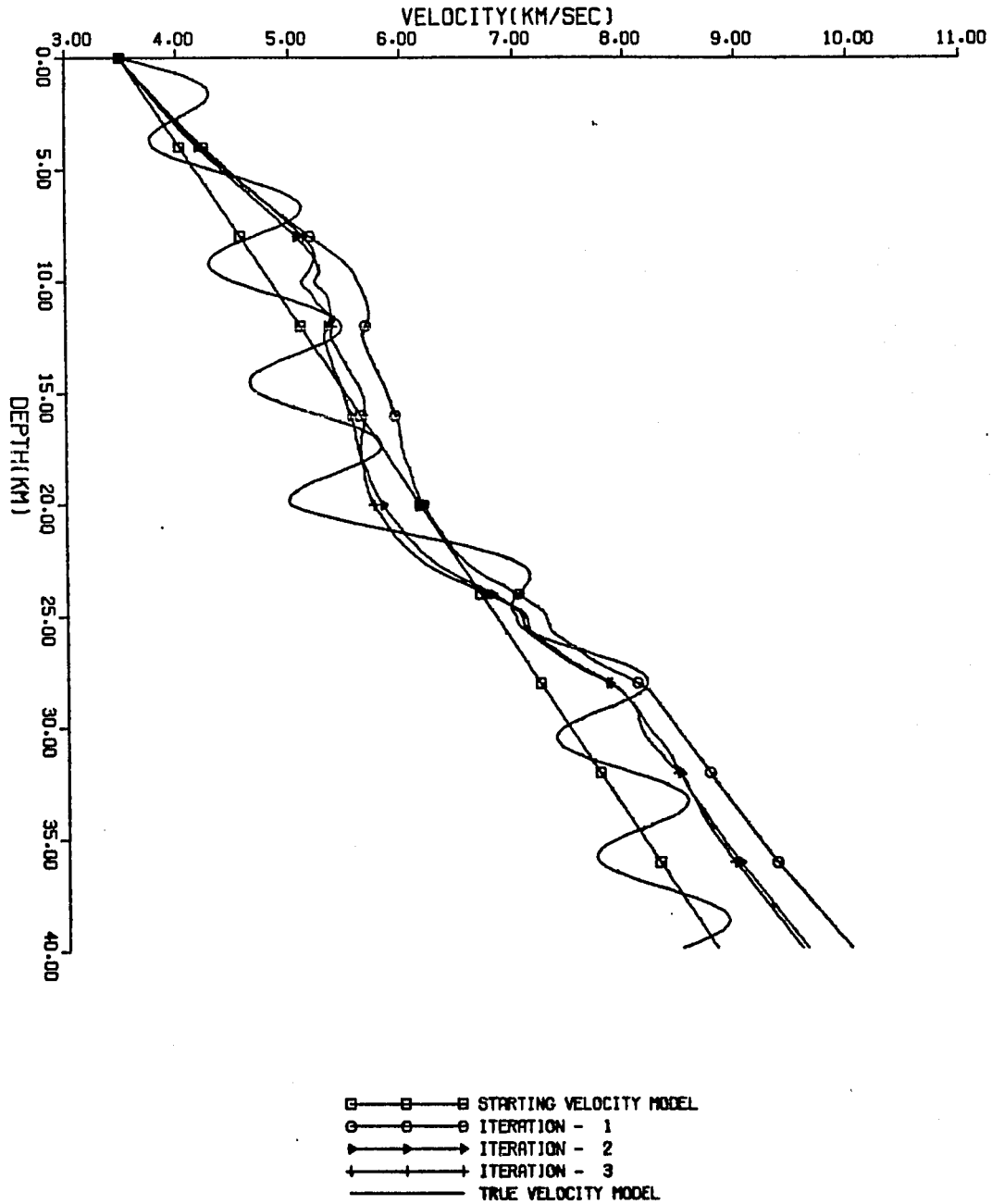
Figure 6.11. Convergence history plot. Surface velocity was fixed at 3.5 km/sec here.

Figure 6.12. Convergence history plot. Data here are the same as in figure 6.11. The difference here is that the surface velocity was allowed to vary as a free parameter.

minimum in which the velocities estimated are everywhere too large or too small. Residuals for the local minimum are small because origin times of the earthquakes in the data set are systematically estimated late or early.

4. CONCLUSIONS

The experiments described here have produced several important conclusions regarding inversion of local earthquake data for seismic velocity structure. The first major result is that the problem appears to be sufficiently nonlinear that an iterative solution method is unavoidable to obtain results that are quantitatively correct. The convergence properties of the iterative procedure developed in chapter 4 (PRIMEL) were studied using synthetic data. These studies indicate the algorithm converges to an estimate that is linearly close to the truth whenever the resolution of the data is sufficient to resolve the structure of the true model. Additional convergence studies, however, have uncovered two pathologic cases that must be avoided when dealing with real data. Both can lead to convergence to a local minimum solution that is not linearly close to the truth. The first is caused by the inability of earthquake data to resolve velocity structure above the shallowest source of the data set. If no shallow sources are present in the data set, errors in estimating the near surface velocity can cause errors of opposite sign to be introduced at depth to produce an apparent fit to the data. This problem appears to be exceptionally bad when the data set is dominated by direct, rather than refracted, ray paths. The second problem uncovered is that there exists a major trade-off between the dc level of the velocity model and the origin times of the earthquakes in the data set. Both problems can be avoided. The best solution is to add refraction data whenever it is available. Refraction data is capable of resolving near surface structure which

should avoid the downward error propagation problem. In addition refraction measurements provide absolute, rather than relative, travel times. This should eliminate the origin time problem. An alternate solution to the origin time problem may be to include S wave data and invert simultaneously for the shear wave structure. This could solve the problem since S wave data in conjunction with P wave data are well known to provide good constraints on the origin times of earthquakes.

CHAPTER 7
TESTS WITH REAL DATA

1. INTRODUCTION

Tests with synthetic data, like those described in the previous chapter, are essential for developmental purposes. Such studies serve much the same purpose as testing of a piece of field equipment in the controlled environment of a laboratory. Once the piece of equipment is found to function in the manner it was designed for on the bench, the next step is to test it in a real world situation. By analogy, the work described in the previous chapter served a purpose similar to laboratory testing and this chapter represents the next logical step; application of the procedure to the real world. Real data contain two features that were not correctly represented or ignored by the tests with synthetic data described above. These are:

(1) Measurement errors in the observed arrival times[1]

(2) Lateral velocity variations.

The effects of (1) can be appraised unambiguously by the methods described in chapter 5. (2) is a more subtle matter and concern for this problem is one of the dominant themes of this chapter. A basic assumption of the procedure I have applied here is that any lateral velocity variations in the true earth structure can be absorbed (within the precision of measurement) by station corrections. As a practical matter, this is generally true only if lateral velocity variations are restricted to the near surface [Crosson, 1976a], because

---

[1] The synthetic data used above actually contained uncertainties because of computational roundoff errors. These errors were, however, at least an order of magnitude smaller than those that are typical of real arrival time measurements.

then station corrections are analogous to static corrections used in reflection seismology (see Dobrin [1976, pp. 211-222]).

Two sets of data are studied here. The first of these is from a local network operating in the region near Humboldt Bay in northern California. This data set was selected primarily because Knapp [1976] had used these data in an earlier study using the least squares procedure of Crosson [1976a]. The major goal of the work with this data set was to compare the procedure I have described here with Crosson's. The results I obtained were reasonably consistent with Knapp's. Unfortunately, however, I was led to conclude that no one-dimensional velocity model could be found that fit these data to within the precision of measurement. I interpret this as an implication that lateral velocity variations are significant within this region and that a one-dimensional velocity model is inadequate to describe the velocity structure of the Humboldt Bay region.

Because of the complications introduced by lateral velocity variations with the Humboldt Bay data, a second set of data was studied from a local network operated by the United States Geological Survey in the Coso Range in east-central California. This area was chosen because previous studies by Walter and Weaver [1980a,b] indicate the velocity structure of the upper crust there was relatively simple. Results from analysis of these data are in sharp contrast to the data from Humboldt Bay. The approximation of a one-dimensional velocity model appears to be adequate and the application of the error appraisal techniques described in chapter 5 is justified.
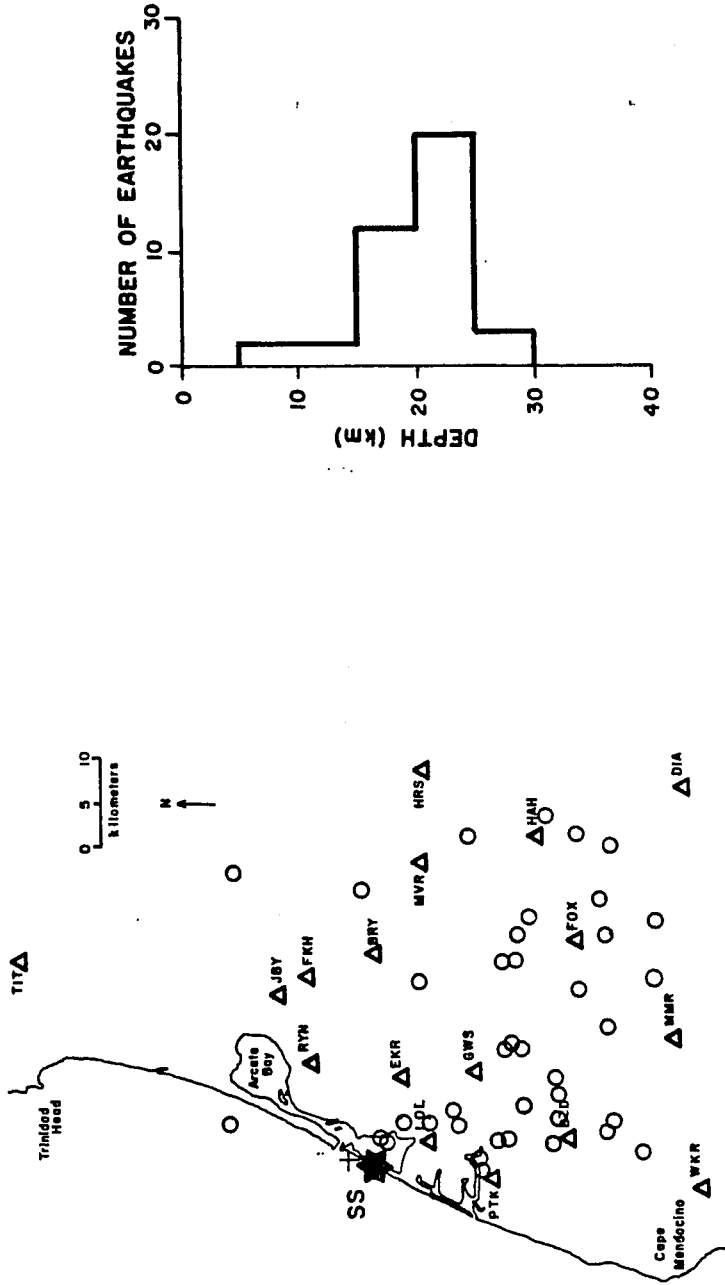
2. HUMBOLDT BAY SEISMIC NETWORK STUDY

2.1. The Data

The Humboldt Bay seismic network was installed in August, 1974, to monitor local seismicity [Knapp, 1976]. This network of sixteen station extends north from Cape Mendocino to Trinidad Head and east from the Pacific coast to the Coast Range Mountains as shown in figure 7.1. The aperture of this array is approximately 70 km. in the north-south direction (50 km. if TIT is excluded) and approximately 50 km. in the east-west direction.

The earthquake arrival time data I have used in this study is the same as the "combined data" used by Knapp [1976] with the exception of one event (event number 1164), which was missing from the data file I used. The epicenters of these earthquakes are shown in figure 7.2a and the depth distribution of these sources is shown in figure 7.2b. Although the areal distribution of these sources is reasonably good, the depth distribution is far from ideal. Most of the earthquakes in this data set occur in the depth range of 15 to 25 kilometers. The worst deficiency of these data is that there are no sources shallower than 8 kilometers. Previous experience with synthetic data had taught that this sort of deficiency can have a detrimental effect on any solution (see chapter 6), so I elected to supplement Knapp's data with data from a large explosive source set off on the south spit of Humboldt Bay (The details of the experiment are given by Smith and Knapp [1978].). The location of this explosion is shown in figure 7.2a where it is labeled SS.
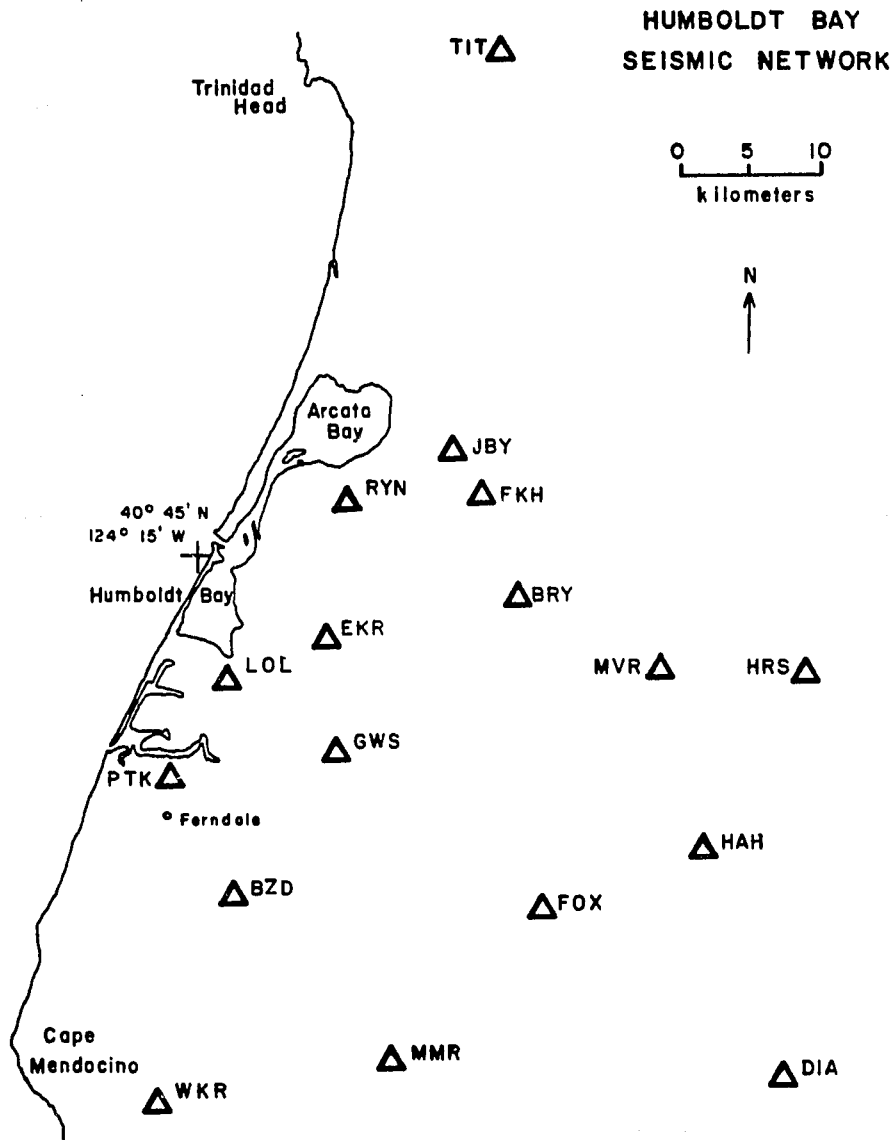
Figure 7.1. Humboldt Bay seismic network station geometry. Triangles are seismic stations. [after Knapp, 1976].

Figure 7.2. Hypocenters of sources used in Humboldt Bay network study. (a) Epicenters of sources [after Knapp, 1976]. SS marks the location of the south spit explosion. (b) Depth distribution of earthquakes. Depths used here were from final locations in this study.
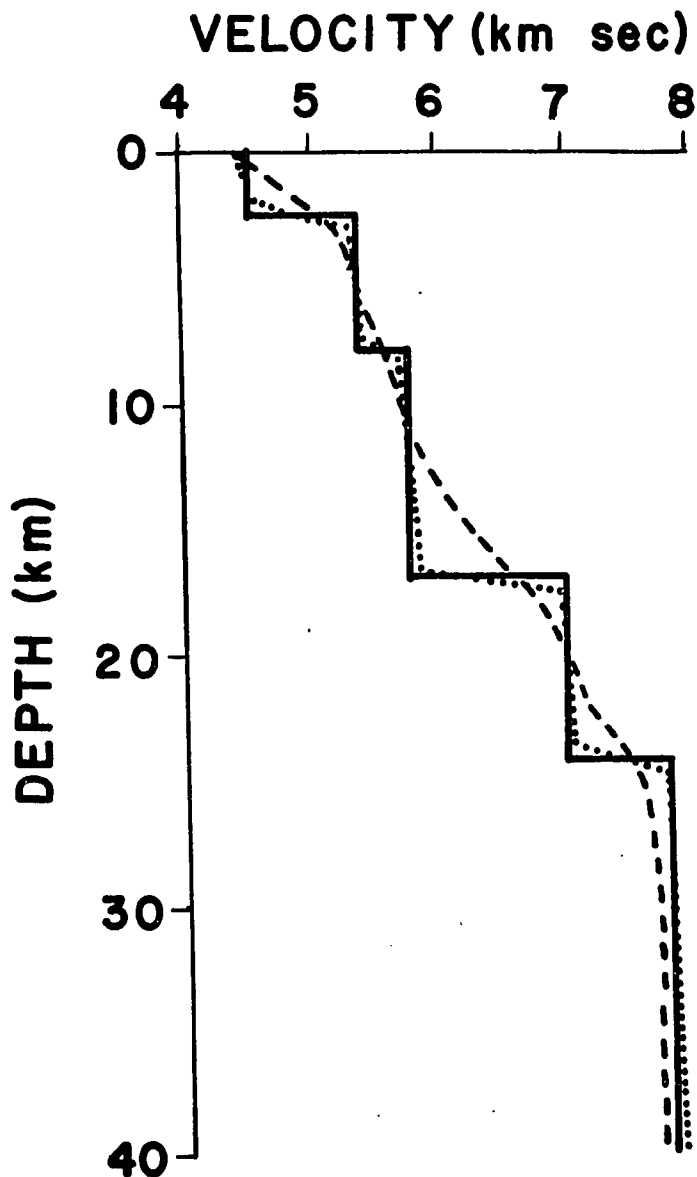
2.2. Model Construction

The procedure I have described here for constructing a velocity model is iterative and hence requires an initial guess for the velocity model. The initial model for this study was constructed from the velocity model used by Smith and Knapp [1978] and Knapp and Smith [1979] as illustrated in figure 7.3. Smith and Knapp's model was not used directly because layer boundaries in a starting model will persist indefinitely with this procedure. The reason for this is that the perturbations that one calculates are always smooth but the layer boundaries are always sharp and hence tend to persist.

I first attempted to relocate the earthquakes in this data set using travel times calculated from this initial model but with all station corrections set to zero. The location procedure, however, diverged with several of these events. Consequently, I adopted the station corrections given by Smith and Knapp[1978] with a minor modification. I chose to subtract the large positive mean value in these station corrections (0.79 seconds) to yield a set of initial station corrections that had a mean of zero. With these station corrections I was then able to successfully locate all the earthquakes in this data set.

From this initial model the procedure converged to the model shown in figure 7.4 in six iterations[2]. In all six iterations no iterative improvement of the station corrections was required and as a result the final station corrections estimated (table 7.1) differ only

----

[2] It is worth noting in retrospect that the solution appears to actually have converged in only two iterations with the most significant change occurring in the first step. Chronologically, this work occurred during a period in which I was investigating alternative convergence criteria. The criteria used at this time was overly conservative, which lead me to perform more iterations than were really necessary.

Figure 7.3. Construction of initial velocity model for Humboldt data. Smith and Knapp [1978] layered model was used as a guide to construct the model made up of the 9 linear segments (dotted lines) that is essentially equivalent to Smith and Knapp's layered model. This model was used to calculate the Frechet derivatives for this data set which were then used to calculate the resolution functions associated with these data (damped solution with $B = 0.10$ was used). The linear segment model was then smoothed by these functions to produce the model labeled by the dashed line.

Figure 7.4. Humboldt velocity models. (a) Velocity models and associated errors. Solid, bolder line is model from iteration 6. Model shown by short dashes is the iteration 6 model with the final perturbation estimate added (perturbation shown is a damped solution with $B=0.10$ and surface velocity free to vary.). Thinner lines to each side are 95% confidence intervals of the estimated errors in this perturbation. Note the perturbed model lies within these errors which is indicative of convergence. (b) Resolution lengths as a function of depth for the perturbation shown in (a). (c) Comparison of iteration 6 model with layered model from Smith and Knapp [1978] in relation to estimated error bars. Thinner lines here are again the 95% confidence intervals as in (a).
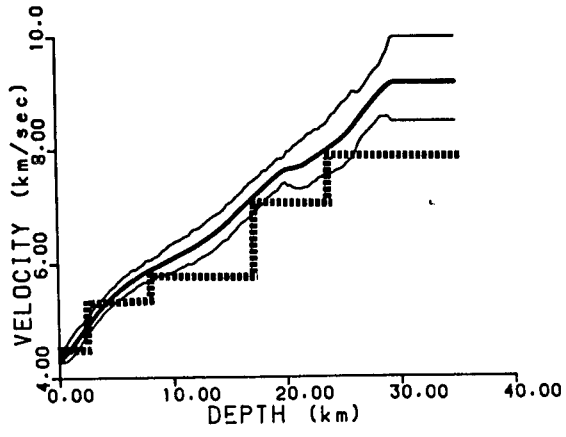
(c)
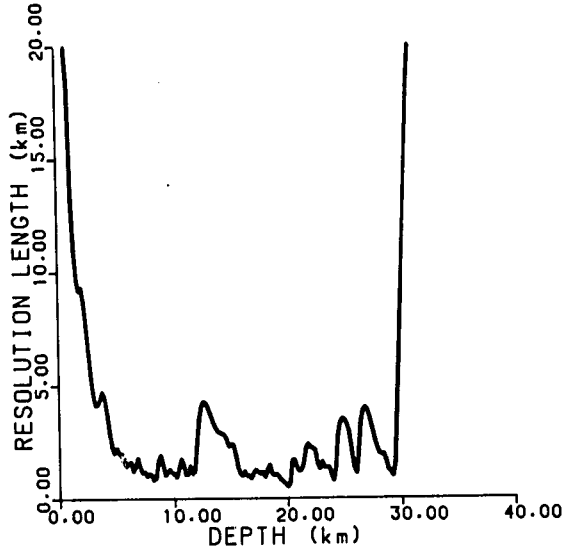
(b)
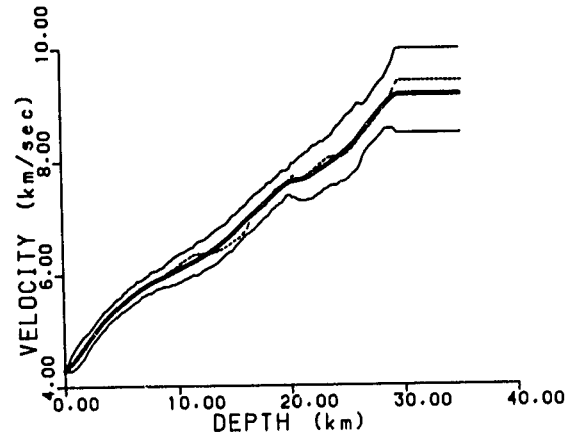
(a)

slightly from those quoted by Smith and Knapp [1978]. Each adjustment of the velocity model was made under the following conditions:

(1) The surface velocity was allowed to vary as a free parameter.

(2) The south spit explosion origin time was considered a free parameter (shot point correction).

(3) Velocity model perturbations were calculated by a damped solution with $B = 0.10$.

(4) After each perturbation the velocity model was smoothed by the resolution functions appropriate for that solution to eliminate spurious high frequency variations introduced into the perturbation by noise.

Convergence of this solution is based on the analysis summarized graphically in figure 7.5. That figure is a contour plot of the quantity

$$P/E(z_0,B) = \frac{\delta\hat{u}(z_0,B)}{\sigma_{\delta u}(z_0,B)} \qquad (7.1)$$

where $\delta\hat{u}$ is the slowness perturbation estimate at the depth z_0 and $\sigma_{\delta u}$ is the expected error in that estimate. ($\delta\hat{u}$ is calculated by equation (5.21) and $\sigma_{\delta u}$ is calculated by equation (5.20).) P/E measures the size of a given perturbation relative to its expected error. It is a function of both the depth z_0 and the size of the number B because both $\delta\hat{u}$ and $\sigma_{\delta u}$ are complicated functions of z_0 and B. B is a quantity that can be interpreted in terms of the degree of "roughness" that is permitted in the final solution (chapter 5). That is, when B is small, $\delta\hat{u}$ will be a very smooth function of depth; but when B is large $\delta\hat{u}$ may have very rapid variations. Solutions with a constant value of B are fundamentally related because they are constructed from the same matrix inverse (equation (5.33)) and
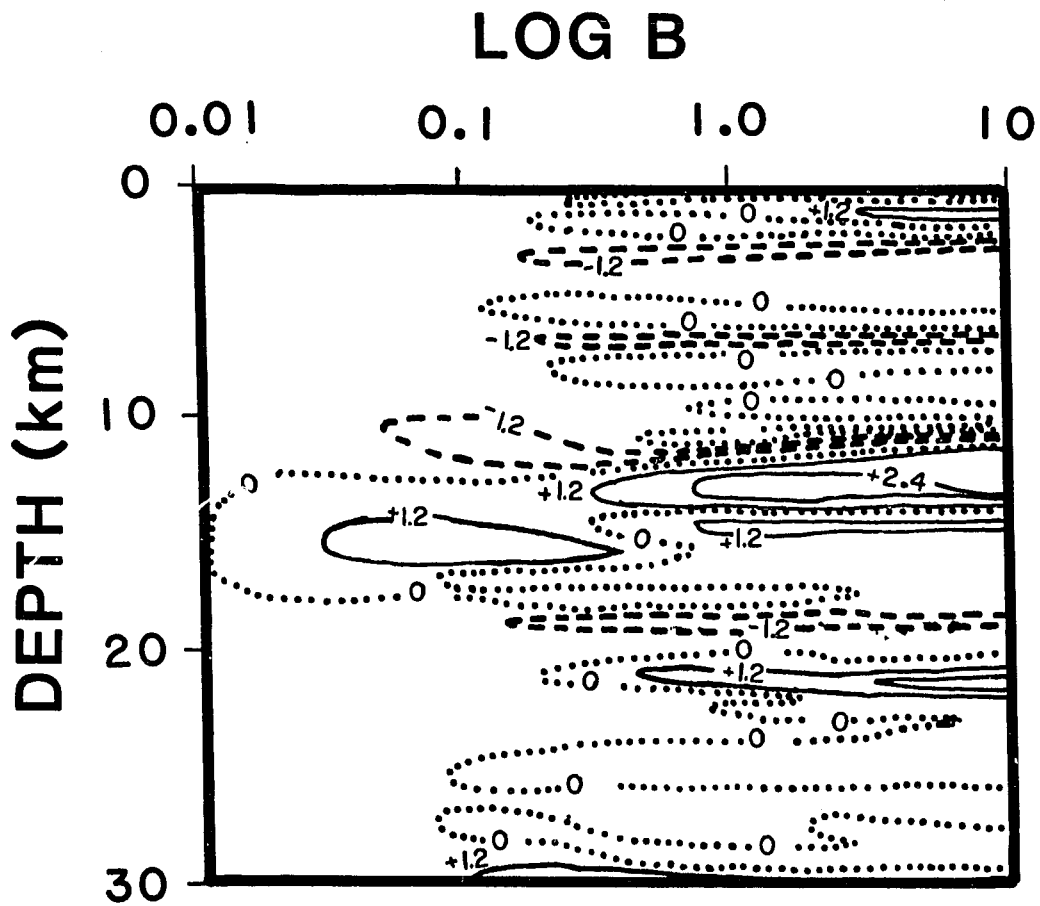
Figure 7.5. P/E contour plot for all final perturbations estimated from Humboldt data. *B* values quoted are nondimensional.

Table 7.1
Station Corrections for Humboldt Bay

Station Name	Station Corrections (seconds)			
	Knapp [1976][†]	Smith and Knapp [1978][‡]	Initial values for this study[§]	Final values for this study
TIT	0.23	0.46	-0.33	-0.36
PTK	1.15	1.14	0.35	0.35
HRS	0.76	0.71	-0.08	0.01
BZD	1.25	1.25	0.46	0.39
RYN	0.45	0.53	-0.26	-0.40
HUM	****	0.52	-0.27	-0.29
BRY	0.62	0.58	-0.21	-0.30
EKR	0.55	0.50	-0.29	-0.36
LOL	1.00	0.98	0.19	0.15
GWS	1.05	1.02	0.23	0.13
HAH	0.88	0.75	-0.04	-0.03
WKR	0.85	1.07	0.28	0.12
FOX	1.04	0.97	0.18	0.11
DIA	0.77	0.70	-0.09	0.01
MVR	0.41	0.45	-0.34	-0.51
MMR	0.93	0.97	0.18	0.10
JBY	0.36	****	-0.44[*]	-0.49
FKH	****	****	-0.44[#]	-0.27

[†]   obtained from earthquake data used in this study with the non-linear least squares procedure of Crosson [1976a]

[‡]   obtained by modifying Knapp's station corrections using data from the south spit explosion

[§]   Obtained by subtracting 0.79 seconds from Smith and Knapp's corrections. Consequently, these corrections have a mean of zero.

[*]   Not given by Smith and Knapp. Knapp's "geologic correction" used.

[#]   Correction not quoted by any source. Value guessed from proximity to station JBY.

hence are relatively easy to construct. Figure 7.5 shows that when B is small (a smooth solution) P/E is relatively small (less than 2.0) at all depths. This is a demonstration that the more slowly varying part of the velocity model is adequately fit by the model shown in figure 7.4. On the other hand, as B increases the solution is permitted to become more jagged (higher resolution) at the cost of a steady increase in uncertainty due the effect of measurement errors in the data. Thus, although P/E shows some fairly large peaks for higher values of B, that fact is beside the point because the statistical errors of the estimated perturbations rapidly become very large. Furthermore, the rapid variations of P/E for larger values of B (right hand side of figure 7.5) occur on length scales shorter than the resolution of the data, indicating a loss of precision due to data errors. B was set to 0.10 for all six iterations of PRIMEL that yielded the model in figure 7.4. Figure 7.4 shows that this solution lies right at the boundary of the region where data precision is lost, which verifies that it is a solution that optimizes resolution[3]. We see from figure 7.5 that P/E is relatively small for all values of B smaller than 0.1 indicating the more slowly varying portion of the model that we have hope of recovering is adequately fit by the model from iteration 6. This is a sign that the solution has converged because larger values of B yield estimates with an unacceptably large statistical uncertainty.

As I noted above, the major purpose of this study was to compare the results of this approach with those Knapp[1976] obtained using Crosson's nonlinear least squares procedure. The velocity

---

[3] $B = 0.10$ was selected from an inspection of a set of trade-off curves that are not presented here. This value was chosen to give the best combination of resolution and statistical errors for these data. Hence, it should have this optimization quality since it was designed that way.

model determined by Smith and Knapp [1978] is compared to the final model found here in figure 7.4c. The station corrections I obtained are also compared to their station corrections in table 7.1. The station corrections found here are similar[4] as a whole, although a few of them differ fairly substantially. The most significant difference, however, is in the velocity models. The model found here has significantly higher velocities at all but the shallowest depths. The reason for these differences cannot be stated unequivocally, but it appears to be one of three possibilities:

(1) Smith and Knapp's velocity model was constructed in two somewhat discordant steps. The bottom three layers are based on Knapp's [1976] work. The top two layers were derived from an analysis of the data from the south spit explosion and the two models were directly pieced together. In that analysis they applied Knapp's station corrections to the observed travel times and then used the standard technique of fitting two lines to these corrected travel times to obtain the two surface layer velocities. They then obtained the station corrections listed in table 7.1 by forcing all the observations for this shot to lie on these two lines. My results were obtained in a much more direct fashion and the differences may be a result of this fact.

---

[4] These stations corrections are similar in a relative but not absolute sense. The large difference in the absolute size of these parameters is due to the fact that the station corrections found here have a mean that is fairly small whereas those quoted by Knapp [1976] and Smith and Knapp [1978] have a substantial positive mean value. For the purpose of earthquake location, the issue of which is strictly correct is irrelevant because the difference only effects the origin times of the earthquakes. In this case, however, it appears that Smith and Knapp's station corrections are more correct in an absolute sense because the station corrections I found here require an enormous shot point correction of 1.45 seconds to fit the data from the south spit explosion.

(2) Knapp [1982] has investigated the convergence properties of Crosson's procedure using synthetic data generated from a continuous velocity model. He found that if the near surface velocities were fixed at a velocity that was too high, Crosson's procedure would often converge to velocities that were too low at depth. At the time of Knapp's [1976] study the data from the south spit was not available. Because of this he correctly concluded that the near surface resolution of the velocity model was poor and elected to fix the velocity of the near surface (top ten kilometers) layer at 5.0 km/sec. His later work with the south spit data [Smith and Knapp, 1978] showed that this velocity was too high. This suggests that the differences between the models in figure 7.4c could be due to this convergence problem with Crosson's procedure.

(3) The model I have calculated here may suffer from the dc offset convergence problem noted above in chapter 6. That is, the velocity model I have calculated may be too large at all depths because the procedure converged to a local minimum.

Which of these is true could be answered by an additional test. One could either repeat Knapp's study using Crosson's procedure on this data with the south spit explosion data added, or the procedure I have developed could be used with the south spit explosion data excluded. I felt this exercise would, however, be somewhat pointless. I will now show that neither model fits the data adequately suggesting a one dimensional model may simply not be appropriate for this region.

2.3. Data Misfit

The analysis presented in figure 7.5 demonstrates that PRIMEL has converged. Any further attempt to perturb the velocity model is pointless because the size of any reasonable perturbation one can construct is not significant compared to the noise introduced by measurement errors. Unfortunately, however, the velocity model shown in figure 7.4 and the station corrections given in table 7.1 are not capable of fitting the data to within the precision of measurement. The first indication of this is that the root mean squared value of the final residuals is 0.052 seconds. This is essentially identical to the value of 0.05 seconds quoted for this statistic by Knapp [1976] for his final model. Since the nominal measurement error of these data is of the order of 0.05 seconds, one might be led to believe that this number is better than it really is. A more rigorous method to appraise the misfit of the data is provided by standard techniques from regression analysis. One can view the estimation of the discrete parameters in this formulation as a regression procedure (albeit nonlinear) that occurs in two stages. The first stage is the process of single event hypocenter location; the second stage is the process of estimating the station corrections (steps 5 and 6 of ALGORITHM PMEL). I have found it useful to examine the results from each of these stages separately.

Earthquake location is often viewed as a regression analysis (e.g. Buland [1976] or Peters [1973]). A commonly used location statistic is the quantity

$$\frac{SSWRES}{NDGF} = \frac{\sum_{j=1}^{m_i} (r_w)_j^2}{m_i - 4} \tag{7.2}$$

where m_i is the number of arrival times measured for the ith earthquake and r_w is a "weighted residual" defined previously in

equation (3.11). The quantity $m_i - 4$ is usually called the "number of degrees of freedom" (NDGF), because it can be shown [Wolberg, 1967, pp. 59-60] that the sum of the squares of the weighted residuals (SSWRES= $\sum_{j=1}^{m_i} (r_w)_j^2)$) has a χ^2 distribution with $m_i - 4$ degrees of freedom[5]. A χ^2 distribution has the property that its mean is equal to the number of degrees of freedom. Because of this property, it is conventional in earthquake location to divide SSWRES by NDGF. This yields a number of order one that can be used to infer the relative quality of any given set of earthquake locations that is less influenced by the differences in the number of arrival times recorded for that event. A justification for this is demonstrated in figure 7.6. There I show the probability density function for the random variable $\frac{SSWRES}{NDGF}$ when the number of degrees of freedom is 4, 8, and 12. We see that the probability density of $\frac{SSWRES}{NDGF}$ is not highly dependent upon the number of degrees of freedom in SSWRES. The reason for this digression into the statistical properties of $\frac{SSWRES}{NDGF}$ is that it justifies presenting the results of the locations of the 39 earthquakes used in this study in the form shown in figure 7.7. The idea is that $\frac{SSWRES}{NDGF}$ from separate earthquake locations can be viewed as a random variable that should have the probability density shown in figure 7.6. Figure 7.7 is a histogram showing the number of earthquakes in this data set that had $\frac{SSWRES}{NDGF}$ within a specified range. NDGF varies in these data from 3 to 12 but

[5] This statement is generally true only if the measurement errors are normally distributed and the weights are set by the variance of this distribution as in equation (3.15).
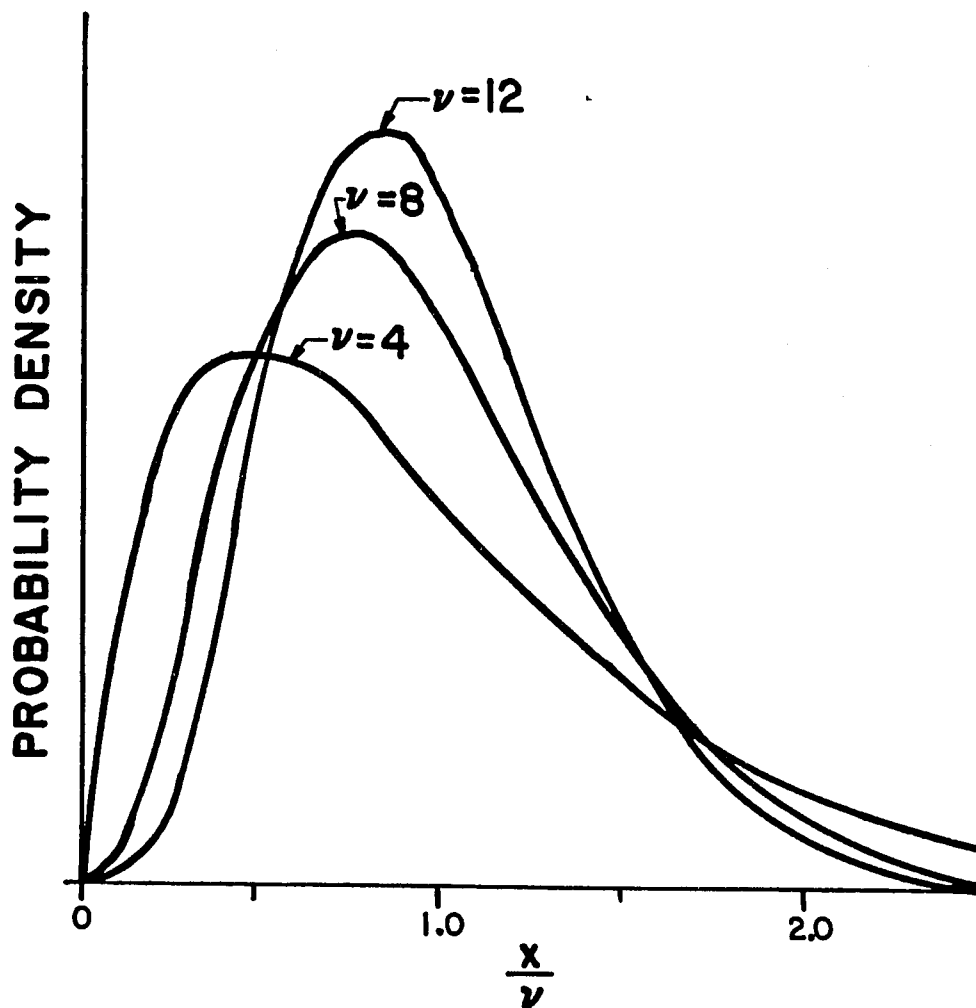
Figure 7.6. Probability density functions for $\dfrac{SSWRES}{NDGF}$ when SSWRES has a χ^2 distribution. If x is a random variable from a χ^2 distribution with ν degrees of freedom, then the random variable $\dfrac{x}{\nu}$ will have the distrubition shown above (Distributions shown are those for $\nu=4,8,$ and $12.$). The quantity $\dfrac{x}{\nu}$ is equivalent to $\dfrac{SSWRES}{NDGF}$ when the residuals have a normal distribution and weighting by the standard errors is used.

Figure 7.7. Location summary of Humboldt earthquakes. Plot is a histogram for 37 of the 39 earthquakes used in this study showing the number of events that yielded a given value of the location statistic $\frac{SSWRES}{NDGF}$. $\frac{SSWRES}{NDGF}$ should have the probability density shown in figure 7.6. Two events not shown have $\frac{SSWRES}{NDGF}$ of 4.9 and 8.2.
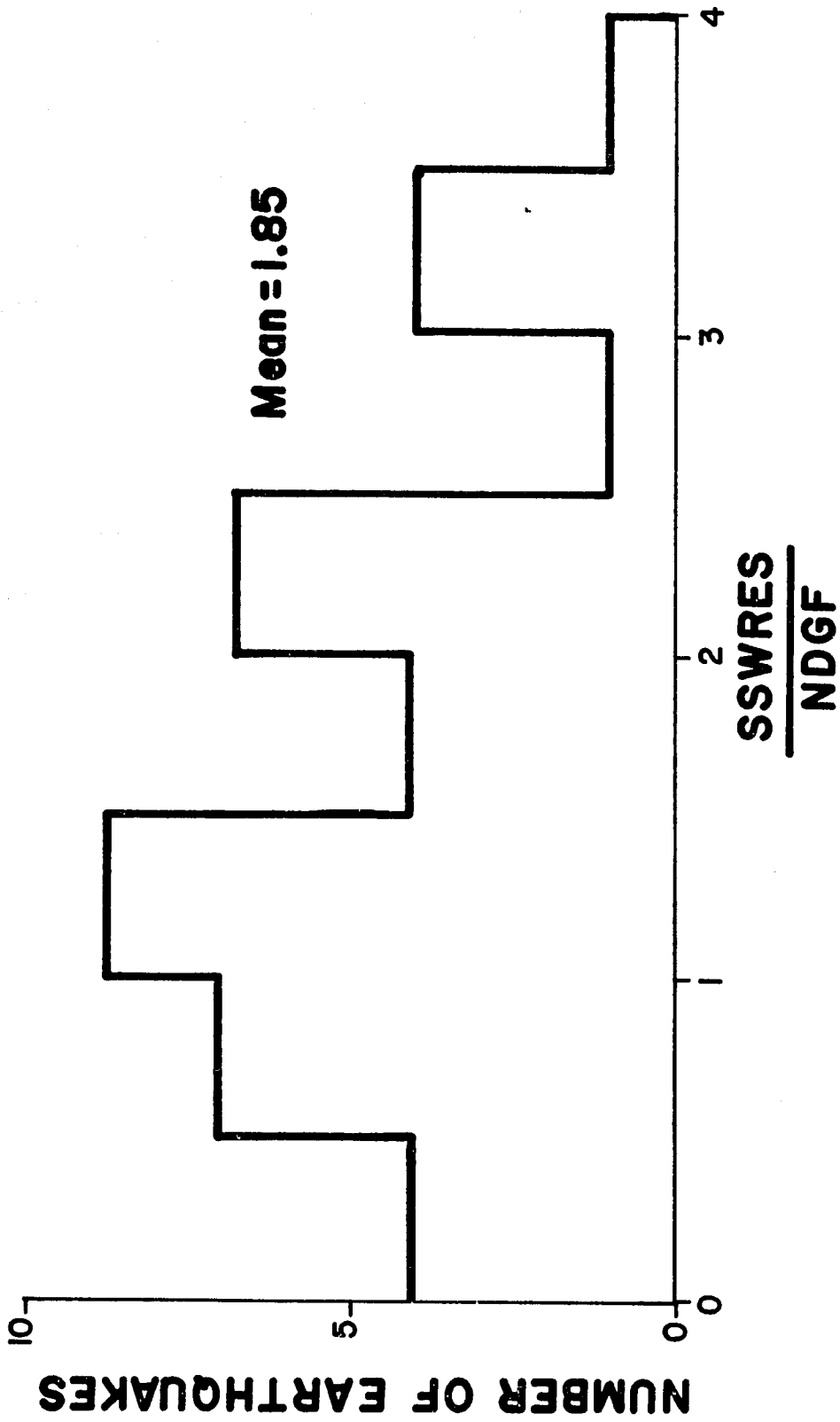
figure 7.6 shows that $\dfrac{SSWRES}{NDGF}$ should in all cases have a distribution that is peaked near unity. The actual data have a wide range in $\dfrac{SSWRES}{NDGF}$ with the center of the distribution centered closer to two than one. This is an indication of a poor fit to the observed data.

A second testimony that condemns the one dimension model assumption is given by the same statistical test for the second regression stage of this procedure; the estimation of station corrections by ALGORITHM PMEL. PMEL uses the $M_s = M - 4m_e - m_{ot}$ (M is the total number of arrival times, m_e is the number of earthquake sources, and m_{ot} is the number of explosions with origin times as a free parameter.) components of the residuals that have not been exploited in the m_e earthquake locations and the m_{ot} origin time estimates. Estimation of the station corrections can then be viewed as a regression procedure that uses these M_s numbers. Thus, SSWRES after the station corrections are estimated, should have a χ^2 distribution with $M_s - n_s$ degrees of freedom ($n_s =$ number of station corrections), and $\dfrac{SSWRES}{NDGF}$ should again be of order one. $\dfrac{SSWRES}{NDGF}$ for these data was found to be 1.8, which is essentially the same as the mean of the distribution in figure 7.7 and is much larger than unity.

Both of the tests described above indicate that this model provides a poor fit to the data. The scatter in the final residuals is about 35% larger than is anticipated from the expected errors in measuring the arrival times. This could be due to one of four possibilities:

(1) The errors (and thus the weights) assigned to these data are wrong.

(2) The velocity structure is indeed one-dimensional but has significant unresolved features that cause a poor fit to the observations.

(3) The procedure has converged to a local and not a global minimum.

(4) The true velocity structure in this region has significant lateral velocity variations and no one-dimensional velocity model can possibly fit the observations because they are too inconsistent.

(1) is conceivable but not likely. This data was recorded on analog 16mm film records [Knapp, 1976]. The errors assigned to these data (0.05 seconds) is nominal for such records [Lee and Stewart, 1981]. The records were, however, processed using a computer assisted picking procedure on the Varian computer operated by the Geophysics Program at the University of Washington. The precision of measuring arrival time of impulsive arrivals with this system is from 0.02 to 0.03 seconds [Malone, personal communication]. This suggests that the errors which were assigned to these data were, if anything, too large. Thus, I consider (1) unlikely. (2) is unlikely as the resolution of the velocity structure here is reasonably good (see figure 7.5b). Furthermore, a more complete analysis using trade-off curves (not shown for the sake of brevity) indicates that unresolved velocity structure should not significantly effect this result. (3) is always a fear with a nonlinear problem like this one and cannot be totally excluded. I conclude, however, that (4) is the major cause for the poor fit obtained to this data. I reached this verdict based on two additional observations. The first is the enormous range in the station corrections found here (0.9 seconds), which is a significant fraction of the average travel time for this data set (~6

seconds). The second is the tectonic setting of the region. Cape Mendocino is the site of the triple junction that marks the northern termination of the San Andreas fault system [Smith and Knapp, 1981] The geology of the Humboldt Bay area is complicated because of the complex history of deformation of rocks of this region [Ogle, 1953]. There are huge variations in sediment thickness in the region [Ogle, 1953] and the entire region is underlain by the complex rock assemblage known as the Franciscan formation [Bailey et. al., 1962]. Furthermore, the crust offshore thins seaward from continental to relatively normal oceanic crust [Knapp and Smith, 1979] which is by definition a lateral velocity variation. Thus, there is little doubt that significant lateral velocity variations exist and it is perhaps surprising that we were able to obtain as good a fit to these data as we have. Because of this fact, any further error analysis applied to this data set is of somewhat dubious value as one of its basic assumptions breaks down here (a one-dimensional velocity model). Consequently, I have elected not to discuss the application of the error analysis techniques described in chapter 5 to these data for the sake of brevity. I will concentrate instead on the application of these techniques to the second set of data I have studied from the Coso Range. There lateral velocity variations appear to be relatively unimportant, so believing the results of such an error analysis does not require an act of faith.

3. COSO DATA

3.1. Introduction

The procedure I have utilized in this study inherently assumes that the velocity structure varies only as a function of depth and that any lateral velocity variations can be adequately approximated by station corrections. As I noted above, this assumption is

probably a poor one for the data from Humboldt Bay. Consequently, it was apparent that a further test of the procedure was desirable with a set of data for which these assumptions were valid. I chose to analyze a rather extensive data set collected by a local network operated in the Coso Range of east-central California (figure 7.8) by the United States Geological Survey (USGS). This decision was based on an earlier study by Walter and Weaver [1980b]. They studied travel time residuals from both earthquake and explosion sources and concluded that there was no evidence for significant lateral velocity variations for this regions which could not be accounted for by near surface effects [Weaver, personal communication].

3.2. The data

The Coso Range lies in the southeastern portion of Inyo County, California (figure 7.8). The region has been investigated extensively in recent years because within it is an area of hydrothermally altered rock and fumarolic activity known as the Coso geothermal area [Bacon and Duffield, 1980]. As part of that investigation the USGS installed a network of sixteen seismograph stations in September, 1975 (figure 7.8) to monitor local seismicity [Walter and Weaver, 1980]. The area is very active seismically. Over 4000 earthquakes were located by the USGS from data collected by this network in the two year period from the installation of the network until September, 1977 [Walter and Weaver, 1980]. This represents an impractically large data set to use directly for an inversion. Consequently, a considerably smaller subset of the total data was adopted. The selection criteria used were the following:

REGIONAL MAP
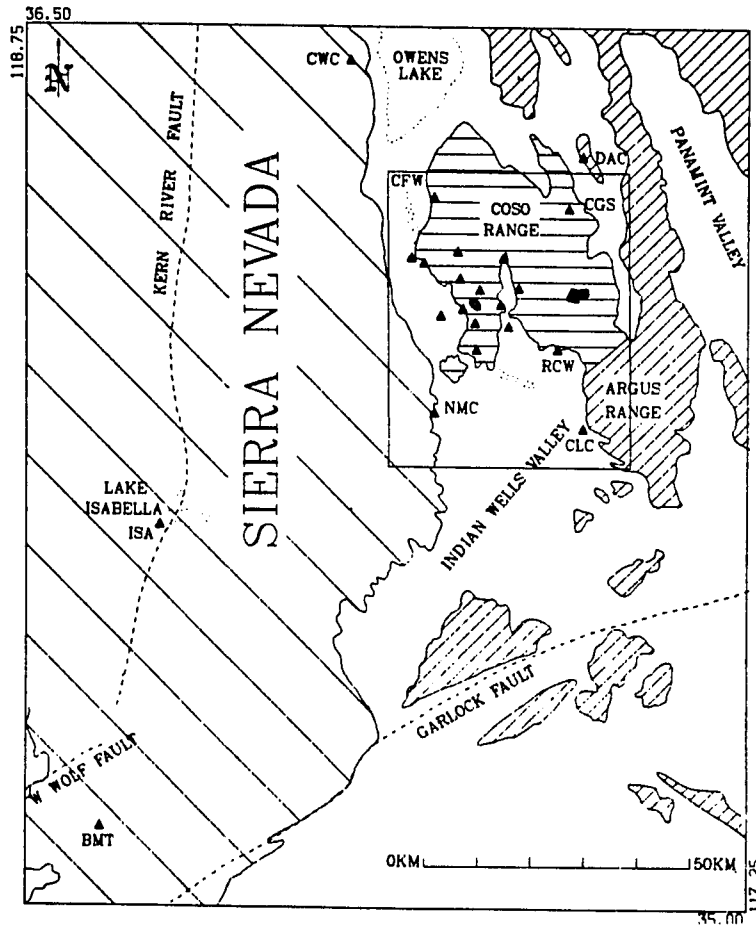
Figure 7.8. Regional map showing location of Coso seismic network. (from Walter and Weaver [1980a])

(1) Duration (coda) magnitude > 1.5

(2) Number of stations recording the earthquake > 10

(3) 36° 55' < Latitude of epicenter < 36° 7.5'

(4) 117° 40' < Longitude of epicenter < 117° 55'

(5) 0°<φ<120° where φ is the largest azimuthal gap in station coverage

107 earthquakes from the full data set satisfied these five criteria. The depth distribution of these events, however, was not ideal as the distribution was peaked between a depth of 5 to 7 km in depth. Consequently, I hand selected 59 of these to yield a more uniform depth distribution. The epicenters of these 59 events are shown in figure 7.9 and the depth distribution is shown in figure 7.10. In addition, the hypocenters of all of these events are summarized in tabular form in appendix C. We see from figure 7.9 that the epicenters of these events are reasonably well distributed areally through the central part of the network. Note, however, that over half of these sources lie within three clusters near station DKN. Furthermore, all but one of the sources deeper than 5 kilometers lie within one of these clusters. This is significant to this work because it tended to make this data set excessively redundant (Two earthquakes with the same spatial location but difference origin times will produce completely redundant data.). This is, in one respect desirable. The data compression techniques described in chapter 4 can be used to great advantage to reduce the size of the matrices one has to manipulate in the inversion procedure. It may be undesirable, however, in another respect. If hidden lateral velocity structure exists, it could bias the estimated station corrections toward those corrections that are appropriate only for the clusters and produce a misleadingly good fit to the data.
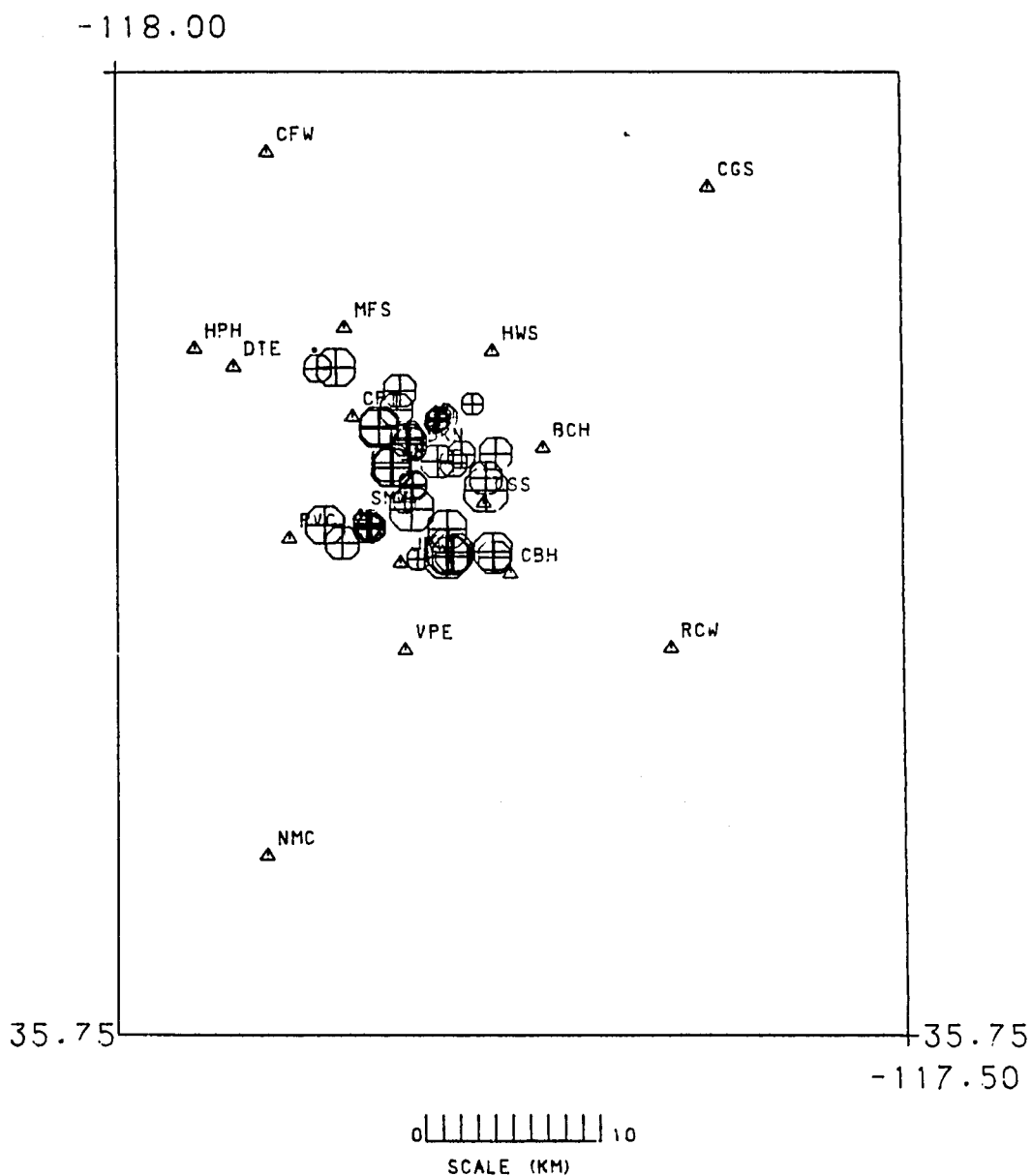
Figure 7.9. Epicenters of 59 local earthquakes used in this study. The triangles are seismic stations and the octagonal symbols indicate the epicenters of these earthquakes calculated from final model. Symbol sizes are scaled linearly with depth with the largest symbols corresponding to the shallowest earthquakes.
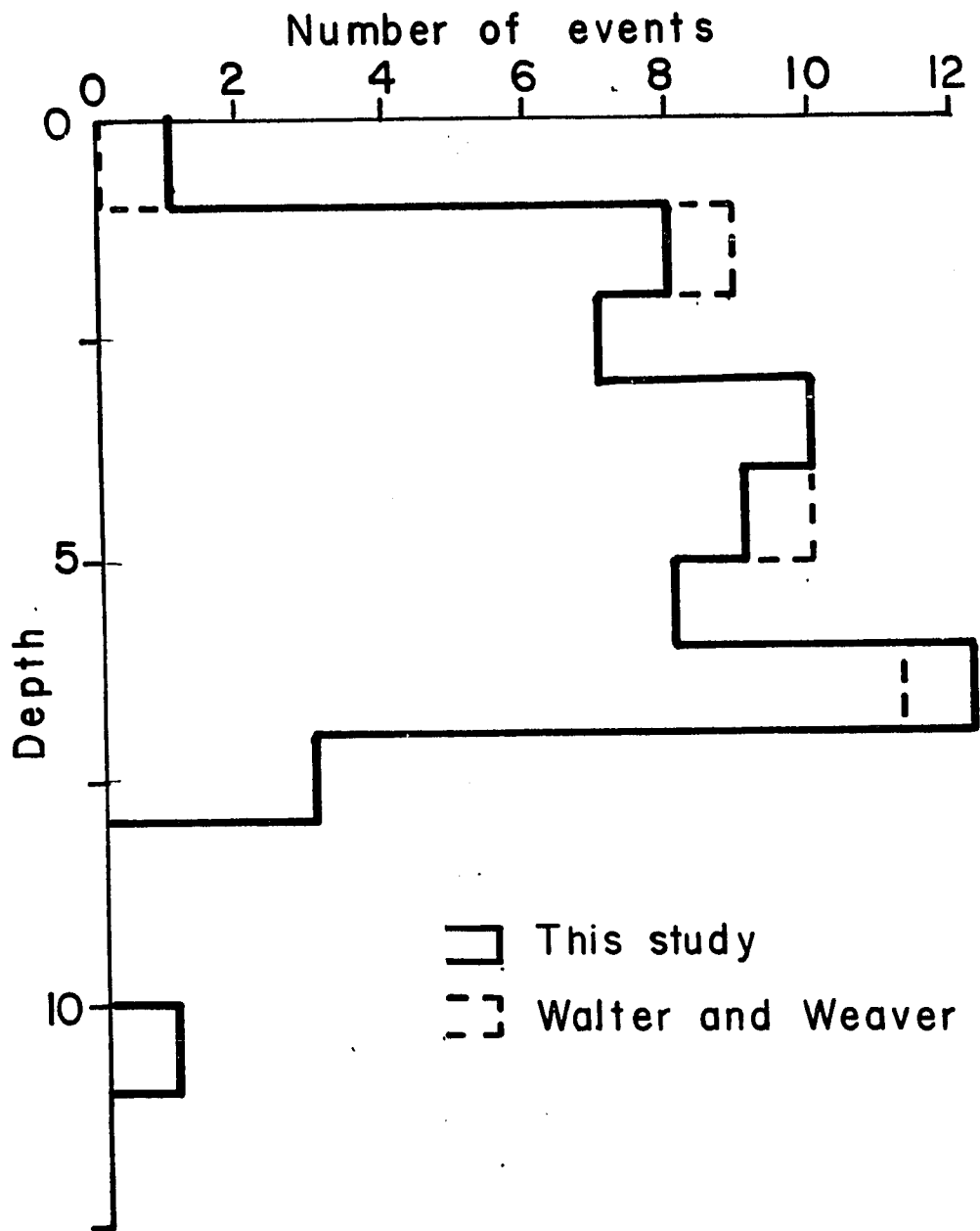
Figure 7.10. Depth distribution of 59 local earthquakes used in this study. Solid line is the depth distribution for the final locations from this study and the dotted lines denote the depth distribution from locations previously calculated by Walter and Weaver [1980a].

In October, 1976, the USGS conducted a fairly extensive refraction experiment in the Coso range [Walter and Weaver, 1980b]. Most of the data from their experiment was also utilized in this study. For this experiment twenty temporary stations were installed across the Coso range along a line trending in a northeast-southwest direction. Explosive charges were detonated at both ends of this line and at a third point near the center of the line close to station C09. These three shot locations are shown in figure 7.11 along with the locations of the temporary stations. For the remainder of this chapter I will refer to these shots by the abbreviations given in figure 7.11. SW will refer to the shot on the southwest end of the profile, NE will refer to the shot on the northeast end of the profile, and C will refer to the shot at the center of the profile. Additional data was also included from a large explosion set off by a quarry operated by Kerr-McGee in the Argus Range just off the eastern edge of the map in figure 7.11. Walter and Weaver [1980b] constructed record sections for all four of these shots. These are reproduced here for the readers convenience in figures 7.12 to 7.15. Data from two of the temporary stations (C01 and C15) along the NE-SW profile and all the temporary stations installed for the quarry shot were not used. They were not used because they only recorded arrivals from a single source. When stations record data record data from only a single source they contain no usable information about the velocity model since estimating a correction for such a station uses all the available information. Consequently, there was no point in including this type of data in this analysis.

The total number of arrival times in this data set was 828. 726 were from earthquake sources and 102 were from explosion sources from the refraction experiment. Figure 7.16 shows the pattern of arrivals recorded for each station. The major point to glean from figure 7.16 is that although the permanent stations recorded arrival

STATION MAP
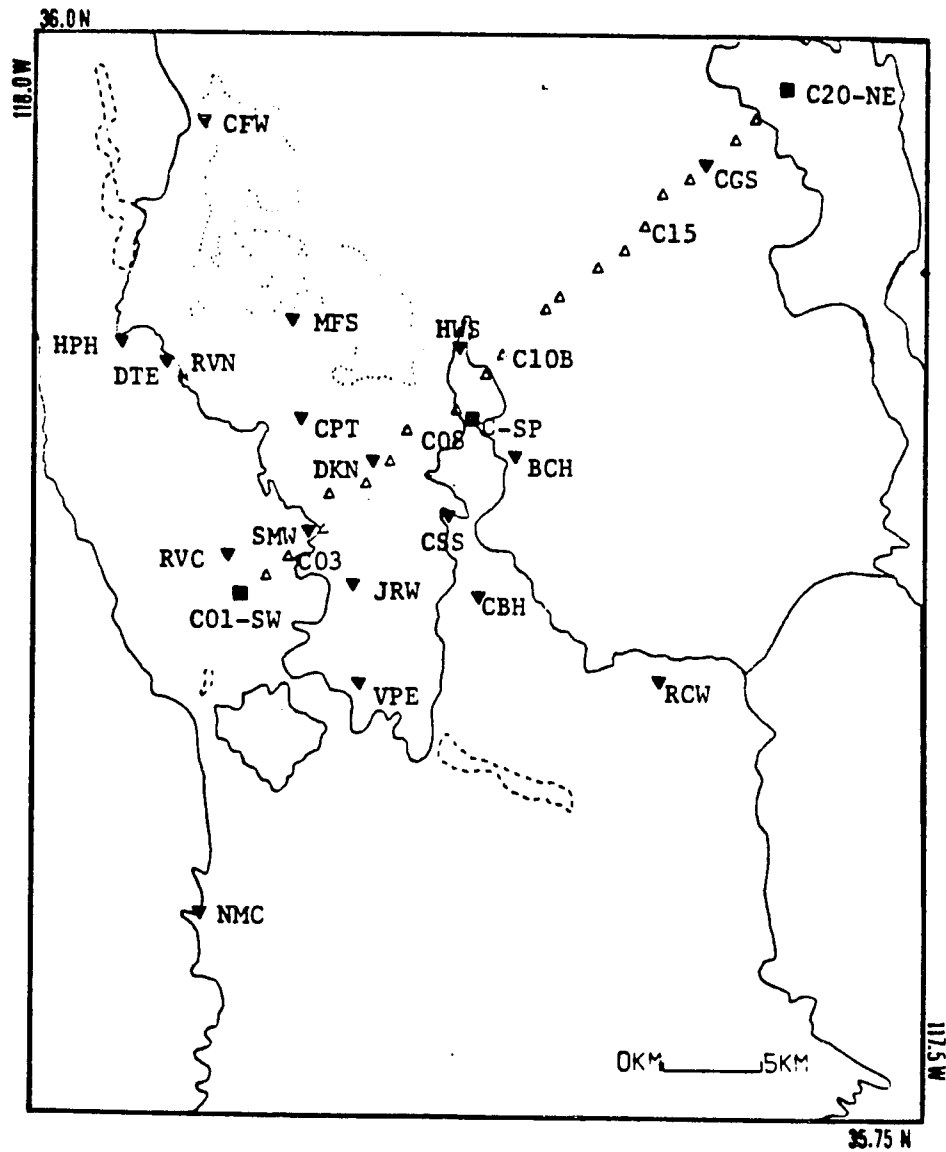
Figure 7.11. Station and shot point locations in relation to local geography for the Coso seismic network. Shot points are indicated by solid squares, permanent stations are indicated by solid inverted triangles, and the temporary stations used for the NE-SW profile are indicated by the open triangles. The Kerr-McGee quarry is located just off the eastern edge of this map. (from Walter and Weaver [1980b])

SOUTHWEST SHOT NE PROFILE
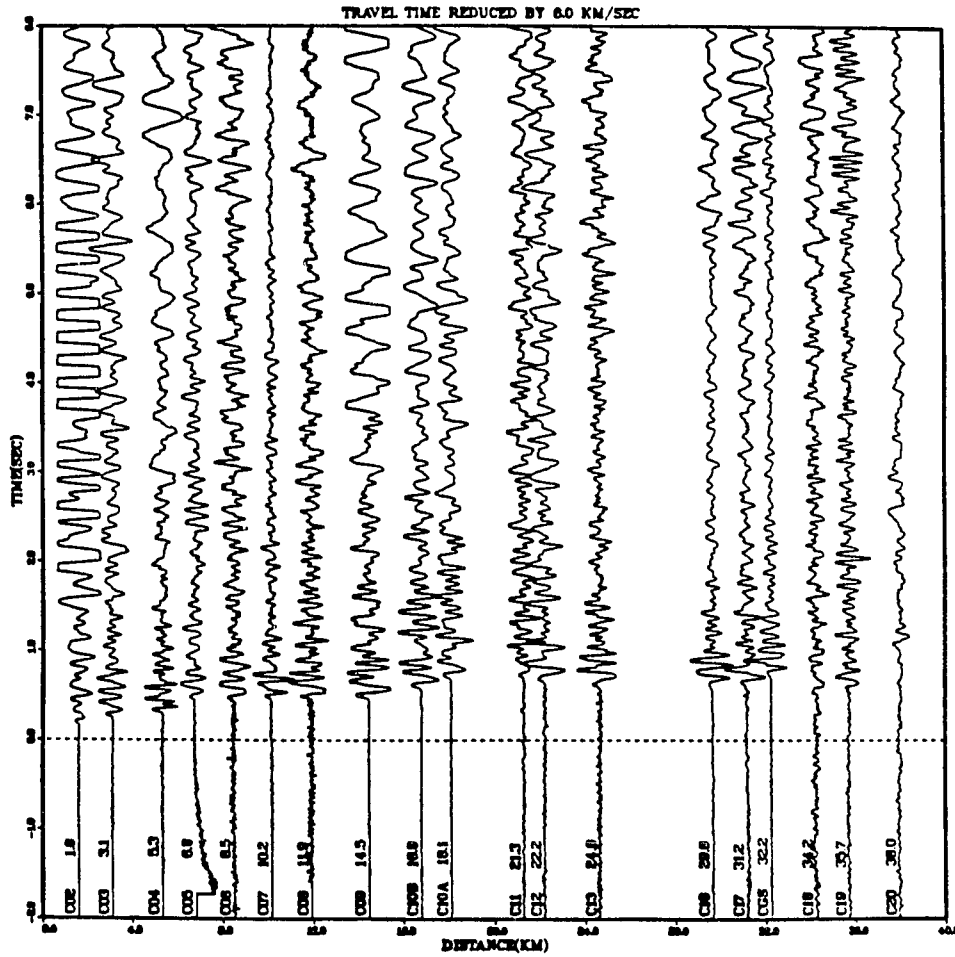
Figure 7.12. Record section showing data recorded by profile stations from SW shot point. (from Walter and Weaver [1980b])

CENTER SHOT PROFILE
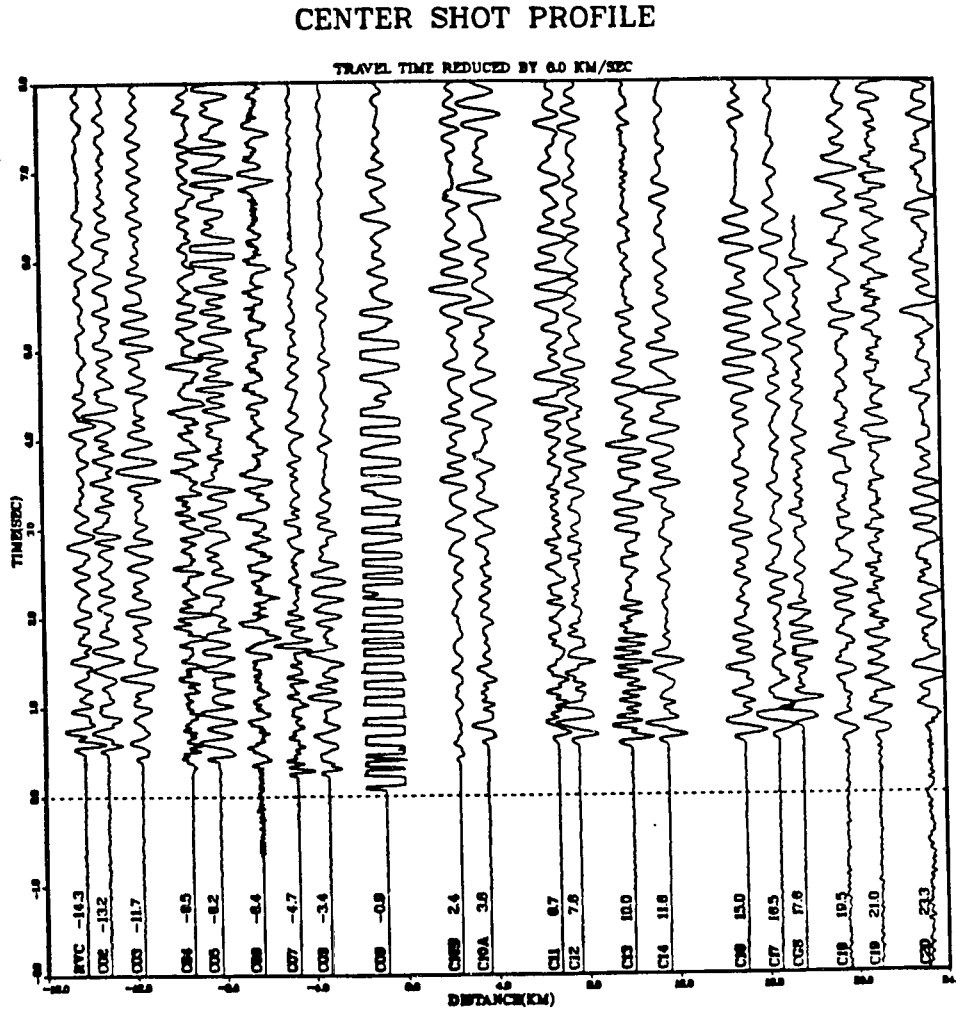
Figure 7.13. Record section showing data recorded by profile stations from the C shot point. Note the marked difference in arrival times for stations northeast (to the right) and southwest (to the left) of the shot point. (from Walter and Weaver [1980b])

NORTHEAST SHOT SW PROFILE
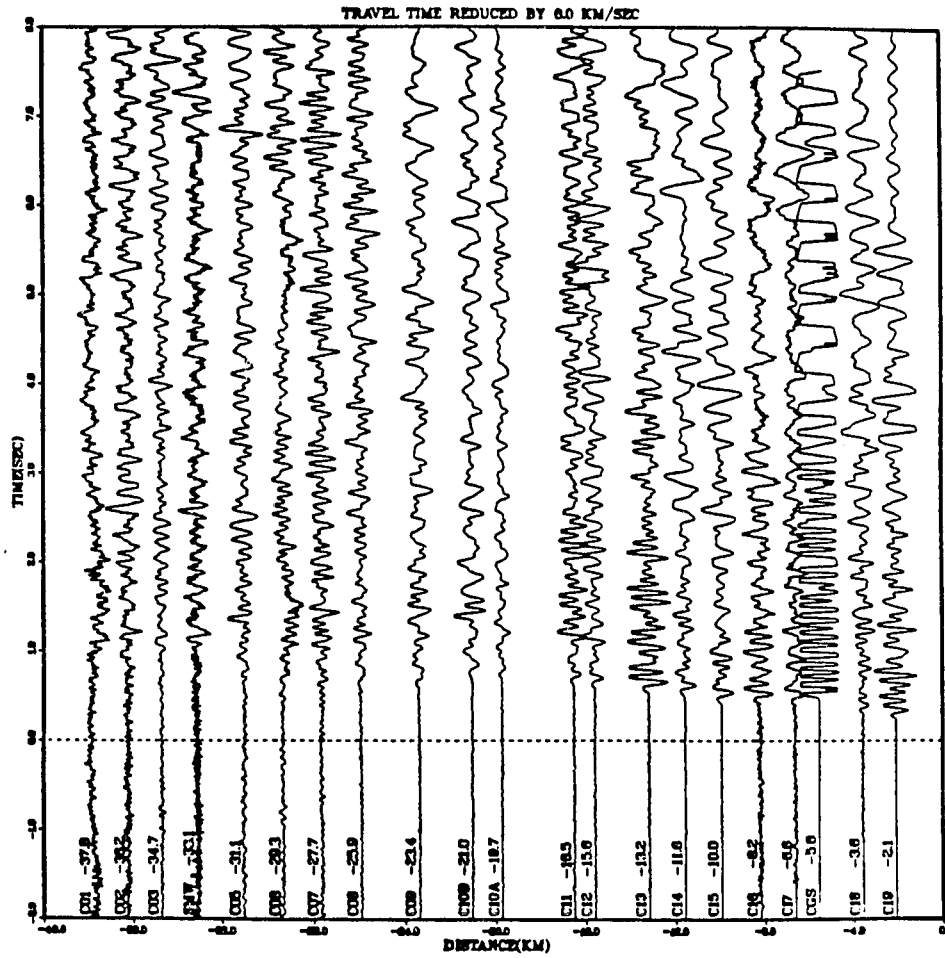
Figure 7.14. Record section showing data recorded by profile stations from the NE shot point (from Walter and Weaver [1980b]).

QUARRY BLAST REFRACTION PROFILE
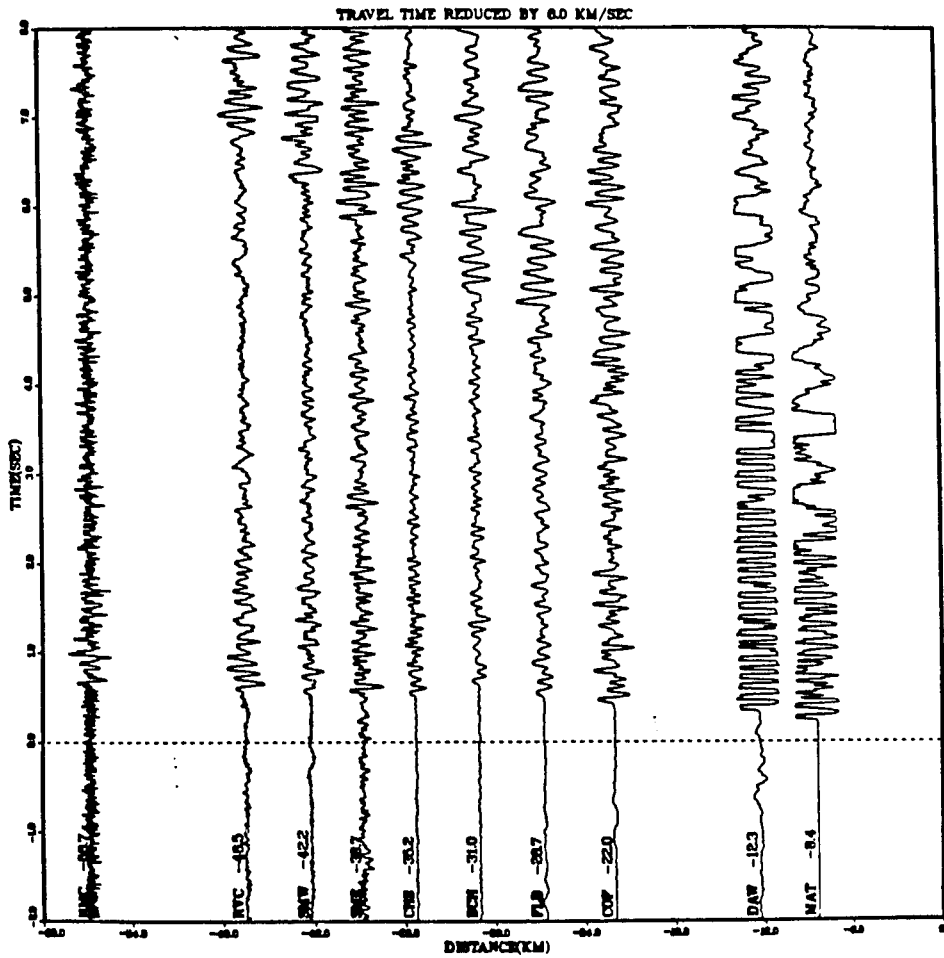
Figure 7.15. Record section showing data recorded for Kerr-McGee quarry blast. Stations SME, CHS, FLB, COF, DAW, and MAT were temporary stations installed to observe this shot. Data from these stations was not used in the velocity inversion because they recorded no other events (from Walter and Weaver [1980b]).

times from both earthquakes and explosion, the temporary stations (C02, C03, etc.) recorded arrival times only from the explosions. This is significant because we will see later that the errors in the station correction estimates for the temporary stations are much larger than those for the permanent stations.

Table 7.2
Coso data weighting scheme

HYPO71 weight number	HYPO71 weight	Assigned Standard error (sec)
0	1.00	0.05
1	0.75	0.10
2	0.50	0.15
3	0.25	0.25
4	0.00	∞

Because the original film records had been archived by the USGS, the original seismogram were not available. Consequently, I was unable to make estimates of the actual picking errors for these data directly. Instead, I was forced to rely on the weights listed with the phase data from the USGS. The associations I made are given in table 7.2.

Similarly, the arrival times listed by Walter and Weaver [1980b] for the refraction data were used directly. Because Walter and Weaver's record sections were available, however, I assigned uncertainties to these arrival times based on a subjective judgement of their quality. Unfortunately, the scheme I used did not permit a standard error any smaller than 0.05 seconds. In retrospect, this is almost certainly an overestimate of the true measurement errors
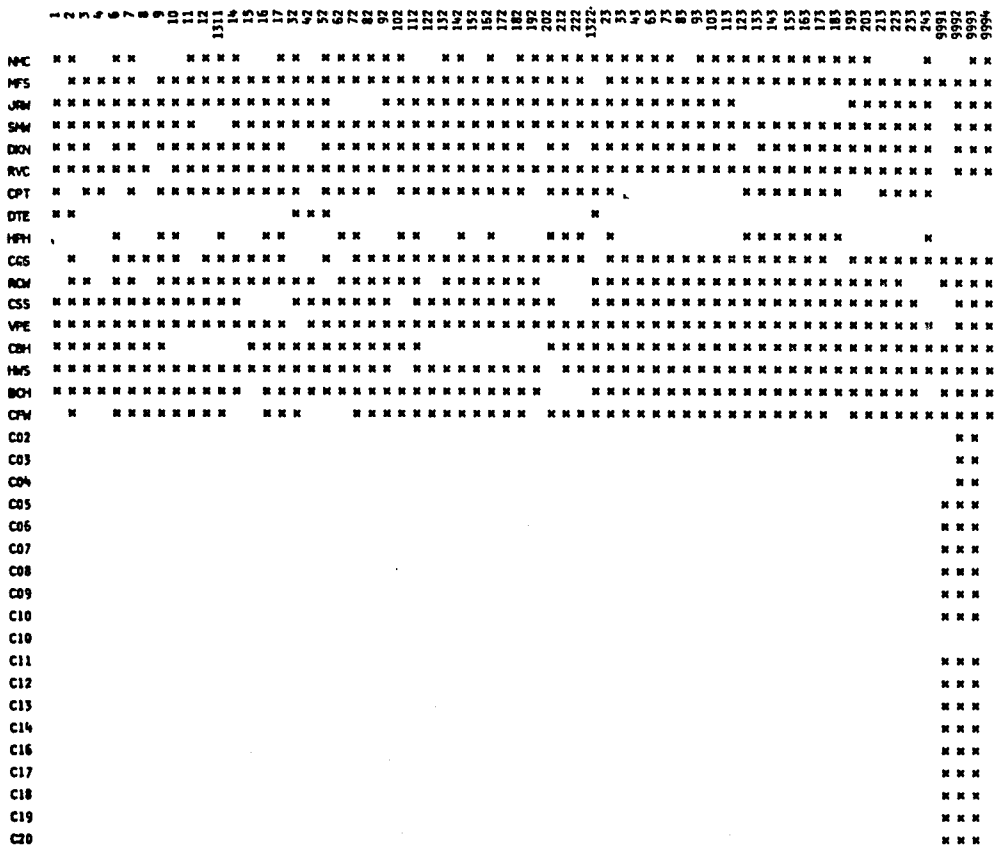
Figure 7.16. Diagram showing what stations recorded arrival times for each event used in this study. Asterisks denote an arrival recorded at a given station-source pair. The event numbers that identify each event are tabulated with the final hypocenters of all these events in Appendix D. Event 9991 is 9991 is the NE shotpoint, 9992 is the C shotpoint, 9993 is the SW shotpoint, and 9994 is the quarry blast.

for some of these data. 0.05 seconds is a reasonable estimate for the error in reading arrival times for an impulsive arrival from a film record [Lee and Stewart, 1981]. Walter and Weaver's picks, however, were made from digital records, which probably have a precision somewhat better than ~ 0.02 seconds [Weaver, personal communication]. This has two possible consequences that should be noted.

(1) Statistical errors may be overestimated.

(2) Arrival times from shots have been given less weight than they deserve. This effectively down-weights the contribution of the refraction data.

3.3. Model Construction

Walter and Weaver [1982] had previously analyzed the refraction data utilized in this study. They interpreted these data by conventional refraction analysis techniques in terms of the four layer model shown in figure 7.17. I adopted a simplified version of Walter and Weavers' model as a starting model for subsequent inversion. This model consisted of the three linear segments shown in figure 7.17. The most notable difference between my model and that found by Walter and Weaver is that I have chosen to ignore the thin, low velocity, surface layer. Walter and Weaver opted to include this near surface layer based on an analysis of data from a string of geophones that were deployed near each of the three shots on the NE-SW profile. The shot points, however, were chosen primarily by the pragmatic consideration of were holes could be drilled to plant the explosives [Weaver, personal communication]. This meant that the shot points were intentionally chosen where near surface sediments were present. Thus, I would argue that a surface velocity of 3.5 km/sec, although known to a high precision locally, is not necessarily a representative average for the entire region. Because the
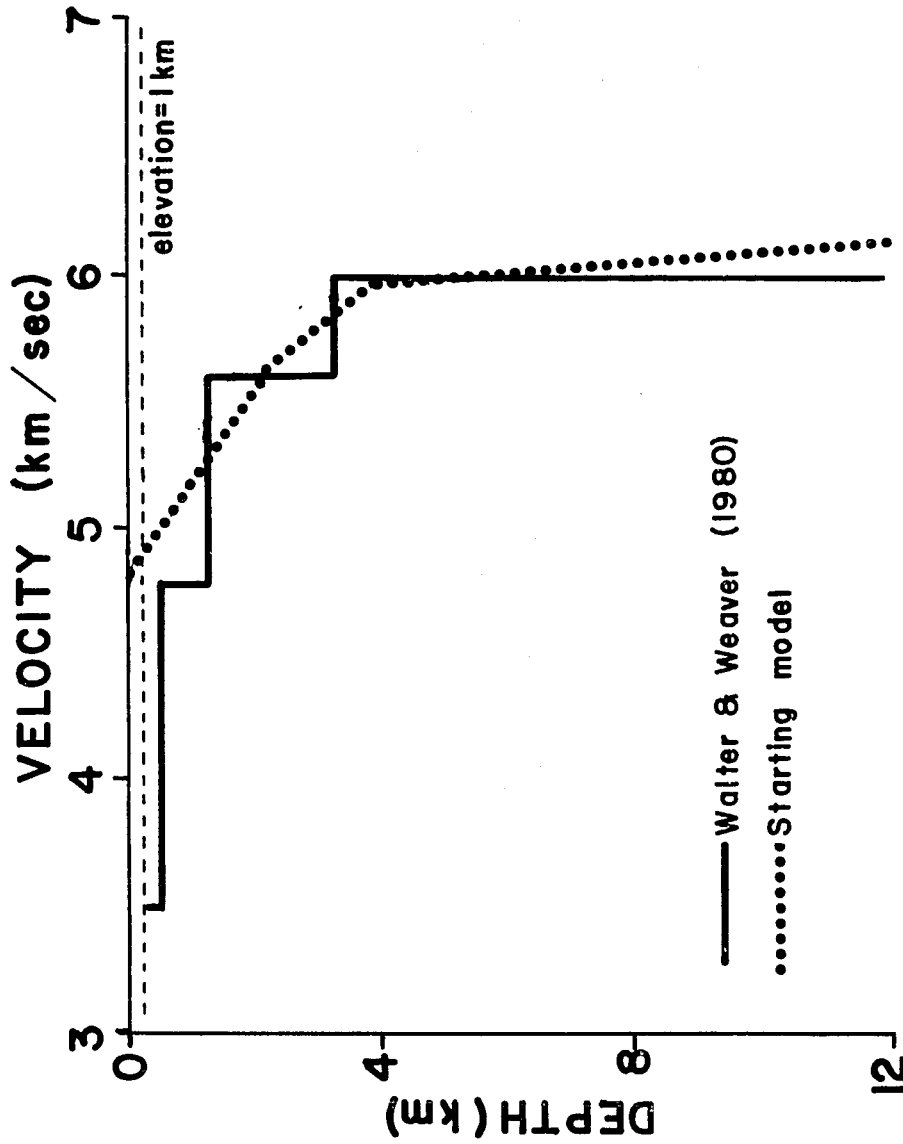
Figure 7.17. Coso velocity models. The layered model is from Walter and Weaver [1980a,b]. Model labeled "starting model" was constructed of three linear segments from Walter and Weavers' layered model.

sediment cover that does exist is known to be quite thin in this region [Duffield et. al., 1980], I believe it is preferable to ignore the existence of any sediments at the surface and absorb their effect in station and shot point corrections.

A second, minor difference I have introduced is a change in the reference elevation. Walter and Weaver's velocity model was based on a reference elevation of 1000 meters [Weaver, personal communication]. I chose, instead, to use 1250 meters as a reference elevation because the mean elevation of the permanent stations in the Coso network is 1264 meters.

The sequence of steps that led to a final result was not direct. That is, the final results were not derived by n steps of the inversion program from the "initial model" in figure 7.17. The major reason for this was that during the course of working with this data set, I uncovered an inadequacy in my original method of calculating travel times. The original procedure followed a practice that is common in earthquake location of ignoring elevation effects completely. That is, travel times were calculated as if all stations were at the reference elevation. Elevation corrections are then assumed to be absorbed in the station corrections. This is an adequate approximation when the topography is not too great and the sources are all relatively deep as is illustrated in figure 7.18a. That approximation can be very poor, however, when sources are shallow as shown in figure 7.18b. At Coso there is a substantial amount of topographic relief (The stations range in elevation from 884 meters to 1975 meters.) and the seismicity is shallow. As a result, the errors introduced in this way were substantial, particularly for station close to the explosion sources. To correct this problem, the procedure was modified to use a more accurate slant elevation correction with the reduction to the reference elevation based on a the velocity at $z = 0$.
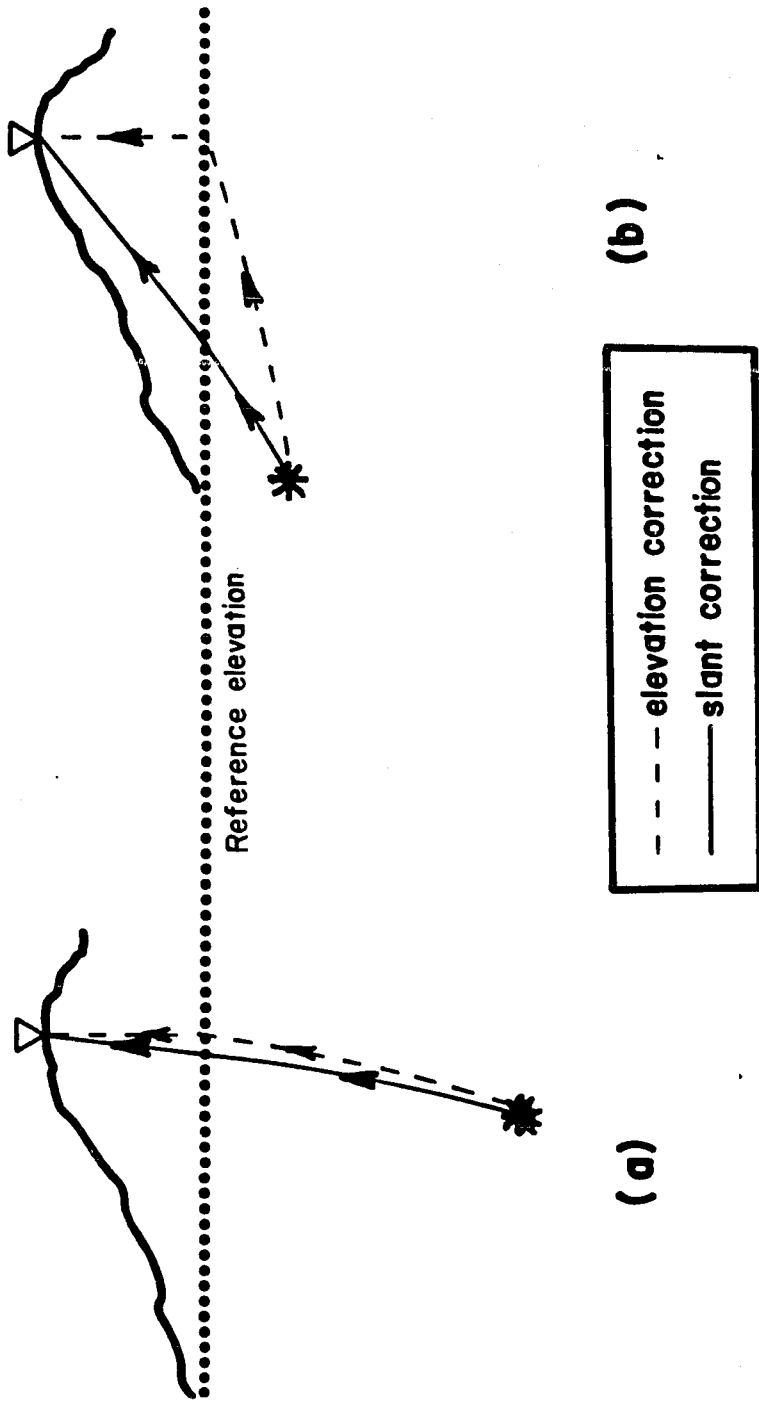
Reference elevation

elevation correction

slant correction

(b)

(a)

Figure 7.18. Cartoon showing differences in slant elevation correc-
tions for (a) rays with steep (deep sources) and (b) rays with low
(shallow sources) angles of incidence.

After the procedure had been modified to correct for the elevation correction problem I described above, the "final model" shown in figure 7.20 was then obtained by the following set of steps:

(1) Prior to the discovery of the elevation correction problem, the procedure had converged to the model labeled "initial model 2" in figure 7.19. That model was used as a starting model with all station corrections set to an initial value of zero.

(2) The progressive multiple event location algorithm (ALGORITHM PMEL described in chapter 3) was applied using travel times calculated from "initial model 2" This procedure converged after three adjustments of the station corrections.

(3) The velocity model was then perturbed by the methods described in chapter 4. This yielded the model labeled "perturbation" in figure 7.19.

(4) The velocity model labeled "perturbation" in figure 7.19 was smoothed by the resolution functions calculated for this data from "initial model 2" This yielded the model labeled "smoothed perturbation" in figure 7.19.

(5) Previous experience had taught that a low velocity zone like the one in figure 7.19 would have an undesirable effect if it were used as the starting model for a subsequent iteration. Low velocity zones introduce shadow zones into the travel time curves that make calculating the ray paths joining some sources and receivers difficult or impossible. The effect of this in my implementation was that an excessive amount of data was discarded, because the ray paths could not be found or were considered questionable. Consequently, I generated a new starting model by piecing together the two models shown in figure 7.19 that are labeled "initial model 2" and "smoothed perturbation". The "smoothed perturbation" model was used from the surface
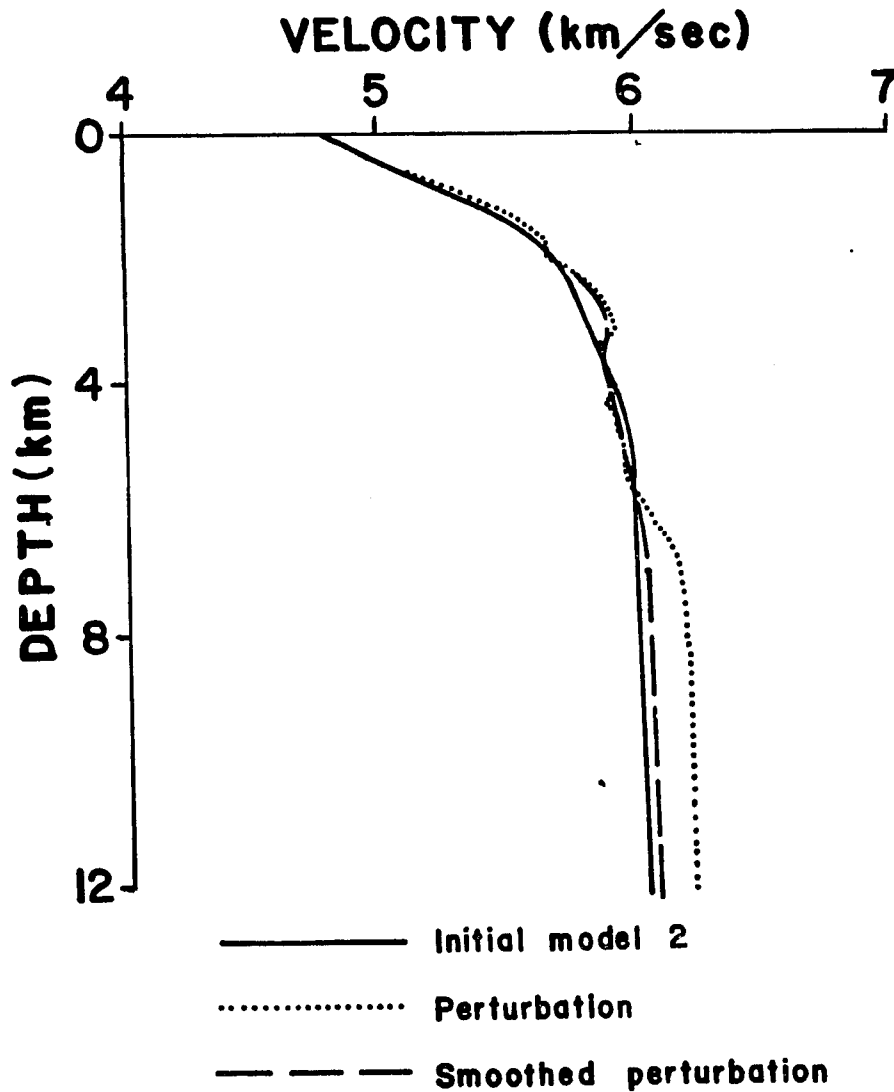
Figure 7.19. Model perturbations. The model labeled "initial model 2" was used as a second starting model after initial run failed because of problems caused by elevation corrections (see text). The velocity model labeled "perturbation" was constructed directly from the residuals calculated from initial model 2. The velocity model labeled "smoothed perturbation" was obtained by smoothing the "perturbation" model by the resolution functions derived from initial model 2.
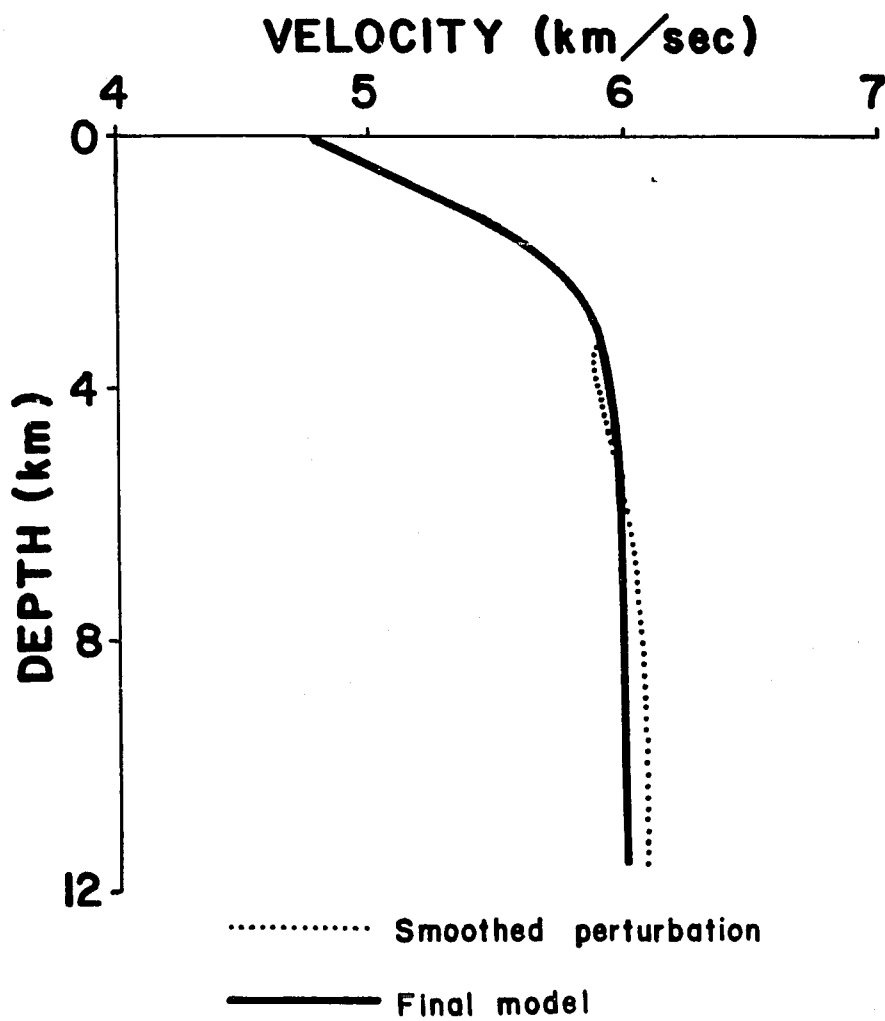
Figure 7.20. Final model compared to smoothed perturbation model that was shown previously in figure 7.19.

down to the top of the low velocity zone. Below 6 kilometers velocities from "initial model 2" were used[6]. These two sections were then joined by a linear velocity gradient segment and the resulting model was again smoothed by the resolution functions calculated from this model. The resulting model is shown in figure 7.20 (labeled "final model") where it is compared with the "smoothed perturbation" model shown previously in figure 7.19. (This model is the same as that shown in figure 7.17.)

(6) PMEL was again applied. Only one adjustment of the station corrections was required in this case.

(7) The velocity model was again perturbed but the calculated perturbation was found to be insignificant compared to the noise (see figure 7.21) indicating convergence.

Two lines of evidence indicate that the model labeled "final model" in figure 7.17 and 7.19 is a good approximation to the truth;

(1) Any reasonable perturbation that can be calculated from the residuals for this model is not significant compared to the statistical noise introduced by measurement error.

(2) This model produces travel times that fit the observed data very well.

---

[6] I used velocities from initial model 2 in this rather *ad hoc* method of constructing a velocity model, because previous studies by Eaton [1966] and Walter and Weaver [1982] found a velocity of 6.0 km/sec for the Mesozoic basement rocks that underlie this area. The smoothed perturbation model in figure 7.19 has velocities ~6.1 km/sec for the deeper structure. Since there is very little control on the velocity below 5 km depth (see section 3.5 which follows) I judged the 6.0 km/sec velocity in initial model 2 more appropriate.
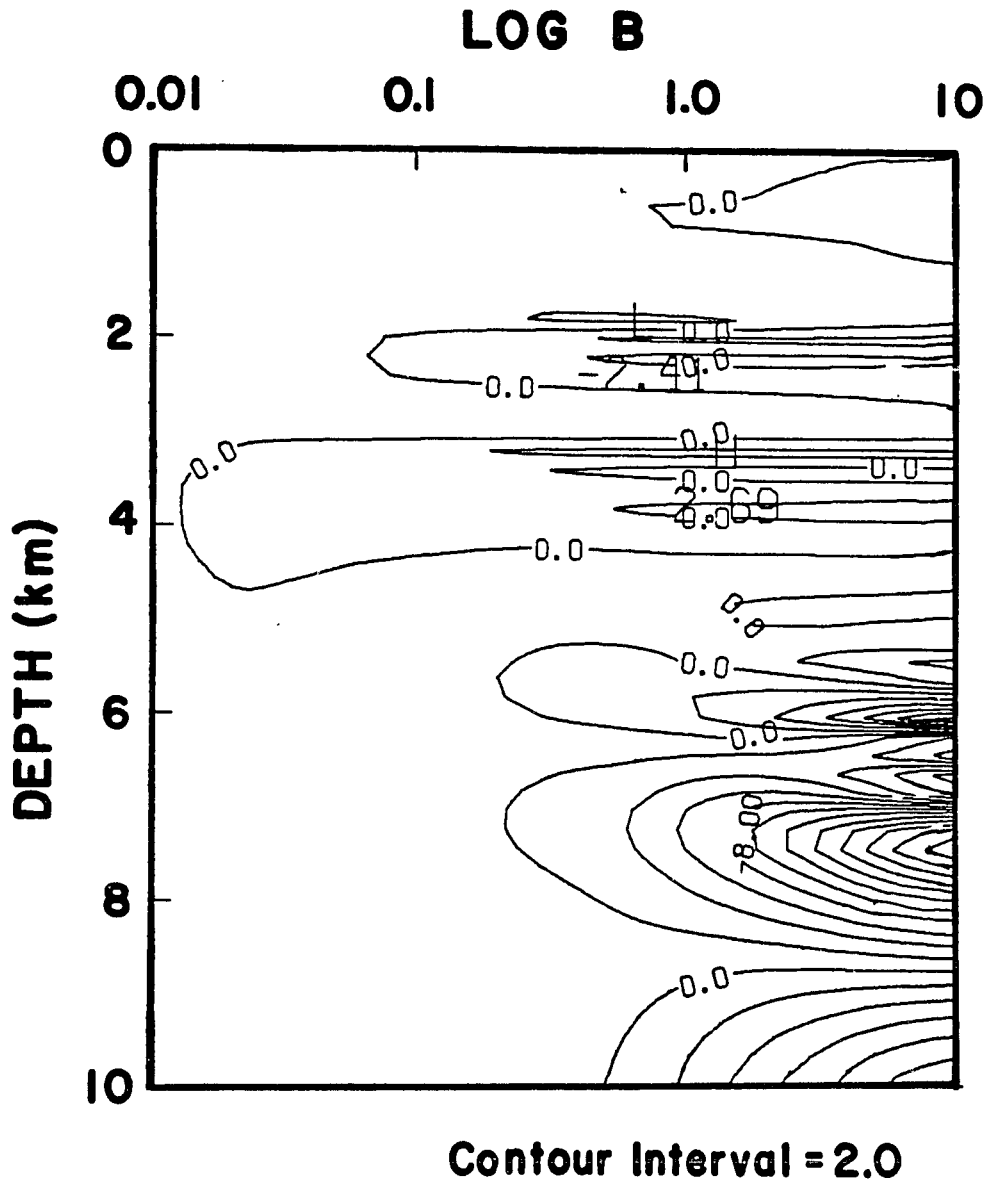
Figure 7.21. P/E ratio contour plot for possible perturbations from the "final model" shown in figure 7.20.

I will conclude this section with a brief discussion justifying (1). I will discuss (2) extensively in the section that follows.

Figure 7.21 is a contour plot similar to figure 7.5 of the ratio P/E defined above in equation (7.1). The interpretation here is similar to that of figure 7.5. The small values of P/E when B is smaller than 0.1 indicates that the more slowly varying portion of the velocity model is adequately fit by this "final model" Although P/E becomes locally quite large as B increases from 0.1, that fact is irrelevant because statistical errors increase extremely fast as B increases. In fact, when $B > 1.0$ the statistical errors begin to become of the order of the total size of the model (i.e. the error bars of the perturbation become ~ 6.0 km/sec.). The implication of this is that the "final model" in figure 7.20 is as good an approximation to the truth as we can hope for, since any reasonable perturbation we can construct from it is not significant compared to the statistical noise.

3.4. Data Misfit

Study of the statistical significance of the family of solutions presented in figure 7.21 is a useful analysis that gives some indication of the stability of our final solution. Application of that analysis to this example shows that the solution is stable in the sense that any reasonable (By "reasonable" I mean a solution that has an acceptably small statistical error.) perturbation we can calculate will be small compared to measurement errors. This analysis is, however, incomplete. It is conceivable that one could find a model that was stable in the sense of the above analysis that did not fit the observed arrival times to within measurement error (Recall this is precisely what happened in the work described previously with the data from Humboldt Bay.). In this section I will demonstrate that the model shown in figure 7.20 yields travel times that fit the

observed data very well. This is important because it demonstrates that the approximation I have used for the velocity model (velocity varying only in the vertical direction supplemented with station corrections) is adequate to fit this data.

The fit of the arrival time data for the 59 local earthquakes used in this study is extremely good. This is demonstrated dramatically in figure 7.22, which is a histogram comparable to figure 7.7. As I noted above, the quantity SSWRES/NDFR, (defined in equation (7.2) when presented this way, should have a χ^2 distribution. The mean of this distribution is a measure of the fit of the model to the observed data. We see in figure 7.22 that the mean SSWRES/NDFR for these 59 earthquakes is considerably less than unity, indicating a good fit of this model to the observations. This is in marked contrast to the comparable results shown in figure 7.7 from the work with the data from Humboldt Bay. Figure 7.7 was the primary basis for my conclusion that the Humboldt Bay data could not be adequately fit with a one dimensional velocity model. Here I make the opposite conclusion. The approximation of a one dimensional velocity model appears to be more than adequate.

With earthquake arrival time data, a presentation like that in figure 7.22 is about the best we can do. Arrival times from explosion data, however, are less ambiguous, because the hypocenter of the source is known. As a result, it is useful to study this data more carefully to search for any systematic patterns in the observation that might be indicative of hidden lateral velocity variations that cannot be corrected for with station corrections. My conclusion from this study is that there is no evidence for any significant lateral velocity variations at Coso except for the region near the earth's surface where the effects of these variations are absorbed in station and shot point corrections. The evidence for this is shown in three different forms in figures 7.23 to 7.35. The first group of these
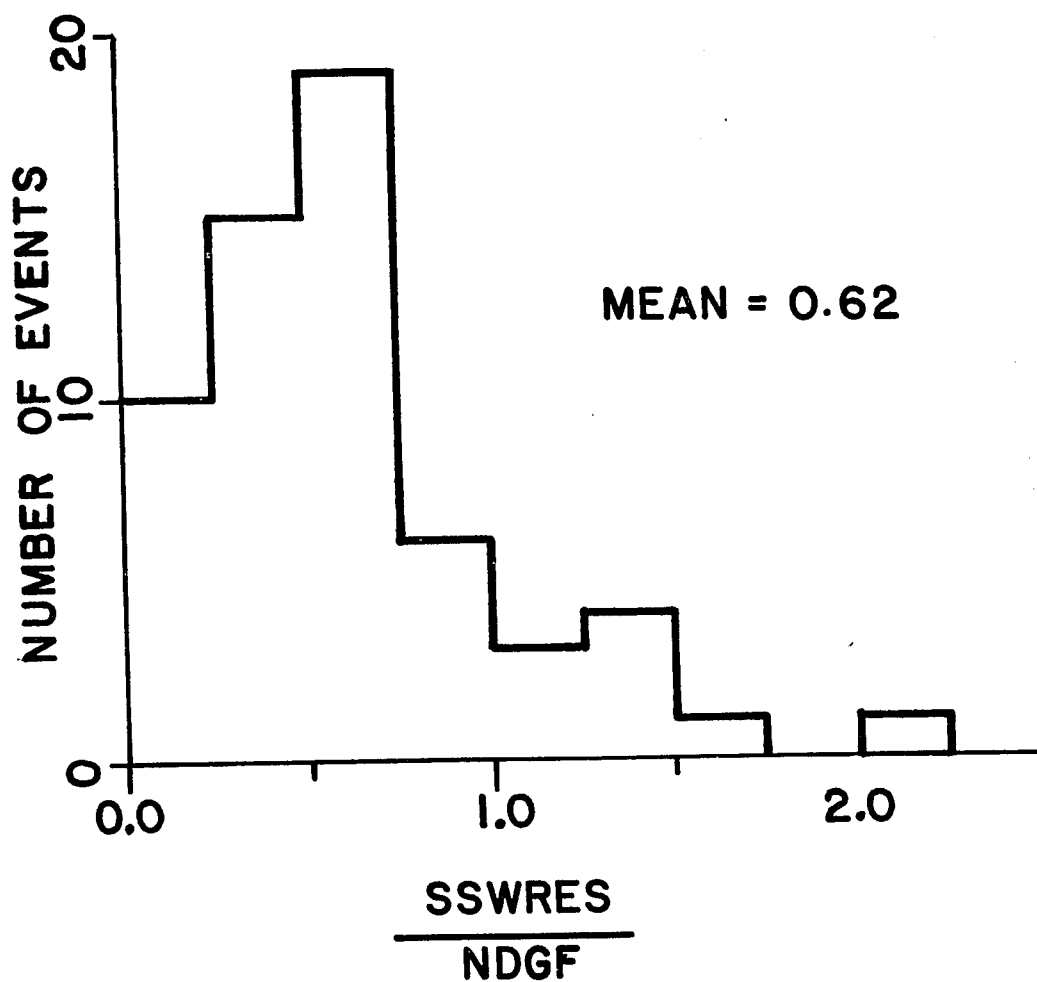
Figure 7.22. Misfit of local earthquake data. This plot is a histogram showing the number of events for which the location statistic SSWRES/NDGF (sum of squared weighted residuals divided by the number of degrees of freedom) falls within a specified interval. Comparison with figure 7.7 demonstrates dramatically the exceptionally good fit of the results of this study.

Figure 7.23. Reduced travel times for stations on the NE-SW profile from the SW shot point. Travel times have been reduced by 6.0 km/sec. Triangles are the measured travel times. The circles are the measured travel times with elevation, station, and shot point corrections applied. The error bars drawn on these circles are the standard errors that were assigned to that datum in the inversion. The curve in this figure is the travel time curve calculated from the model in figure 7.19. A noteable feature of the raw data for this shot is the prominent delay in stations in the distance range from 20km to 25km. These stations lie at the point were the profile crosses the northern end of the sediment filled Coso Basin and the fan deposits of the Coso Formation (see Figure 7.41).

SW SHOT
Temporary Stations
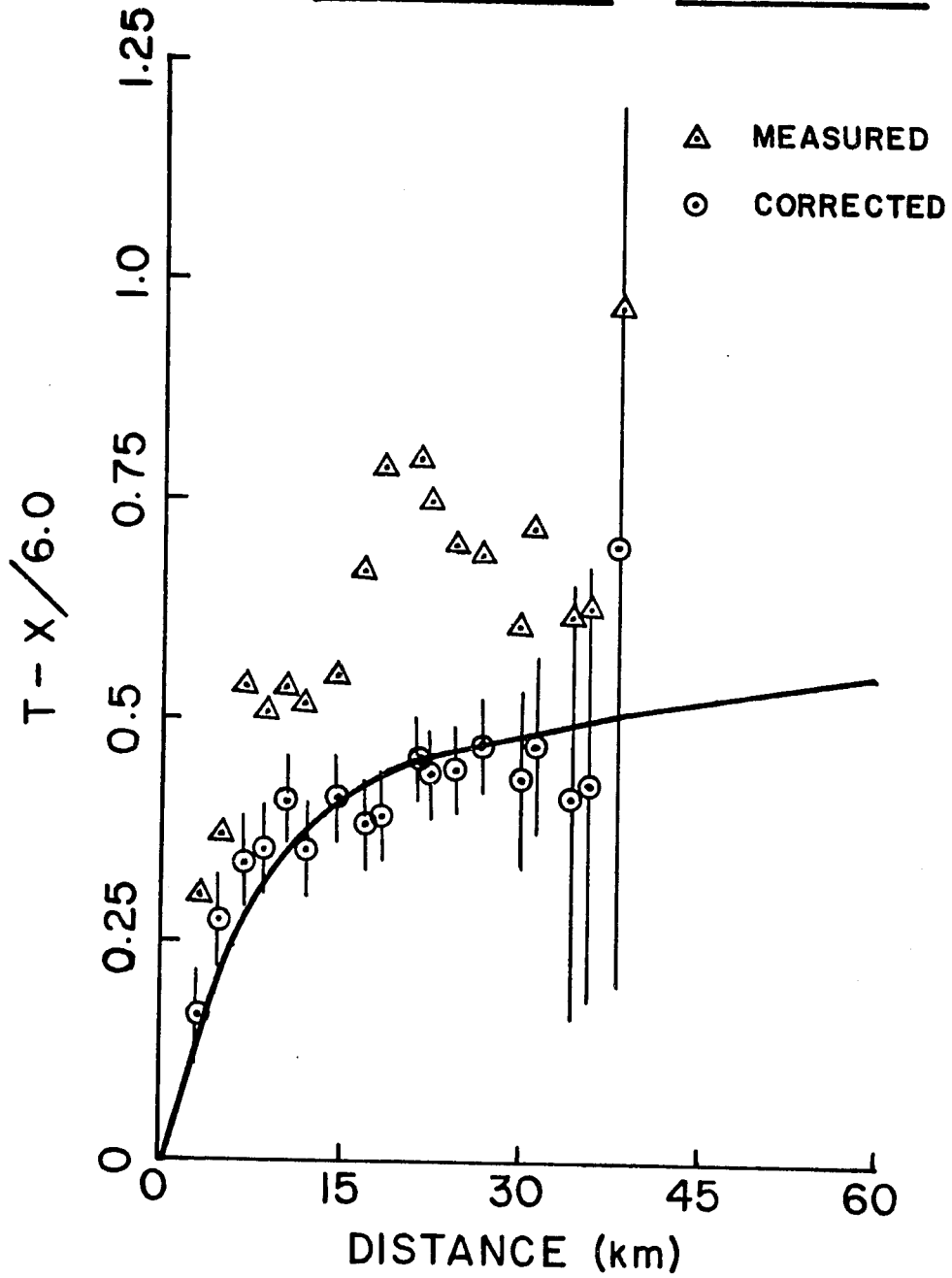
Figure 7.24. Reduced travel times for permanent stations from the SW shot. Here the squares are the measured travel times and the hexagons are the travel times with elevation, station, and shot point corrections applied. Other features are as described in figure 7.23.

SW SHOT
Permanent Stations
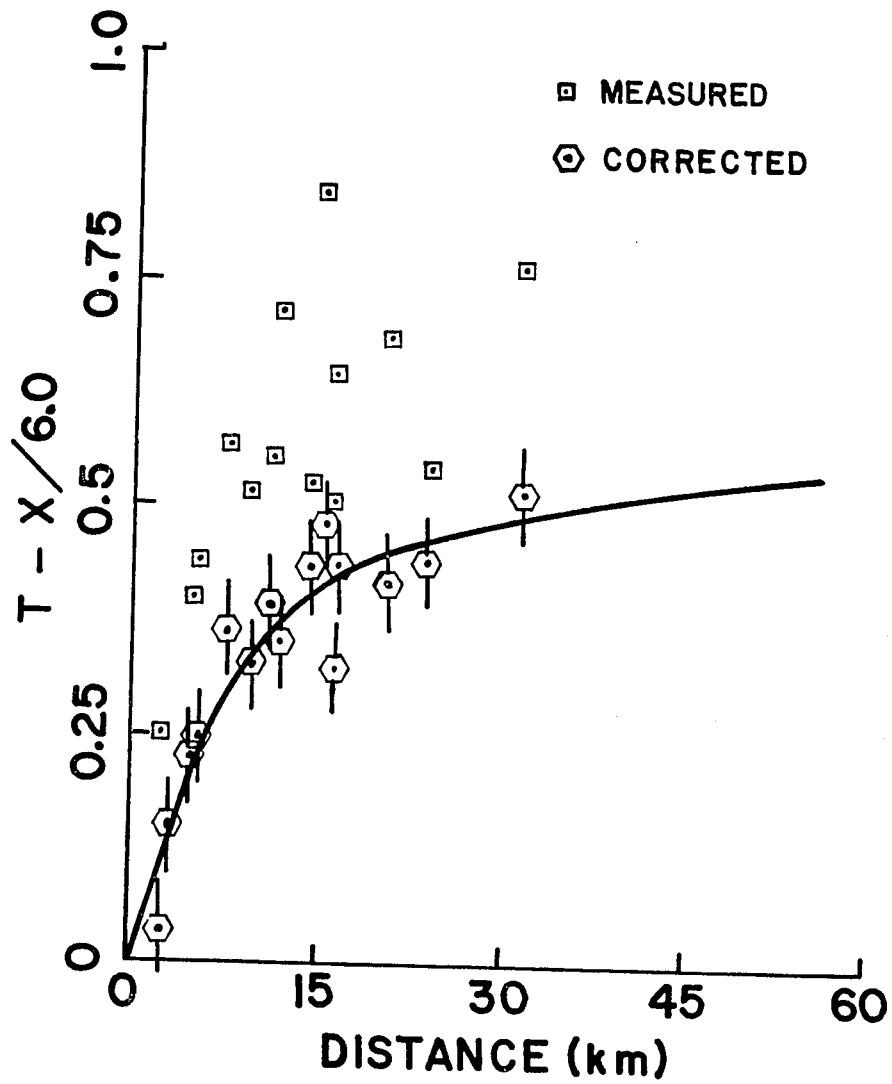
Figure 7.25. Reduced travel times for profile stations from the C shot point. The triangles and stars are the measured travel times and the circles and crosses are travel times corrected for elevation, station, and shot point delays. Seperate symbols have been used to denote those stations that are southwest of the shot to emphasize how different they are when compared to arrival at station northeast of the shot. This difference is due to the fact that the shot was located near the western edge of the Coso Basin. The Coso Basin is bounded on the west by a normal fault with the Coso Basin on the downthrown side [Duffield et. al., 1980]. This delayed arrivals northeast of the shot relative to those southwest of the shot.

C SHOT
Temporary Stations

STATIONS NE OF SHOT

△ MEASURED

⊙ CORRECTED

STATIONS SW OF SHOT

✡ MEASURED

∗ CORRECTED

T–X/60

1.0

0.75

0.5

0.25

0

0   15   30   45   60

DISTANCE (km)
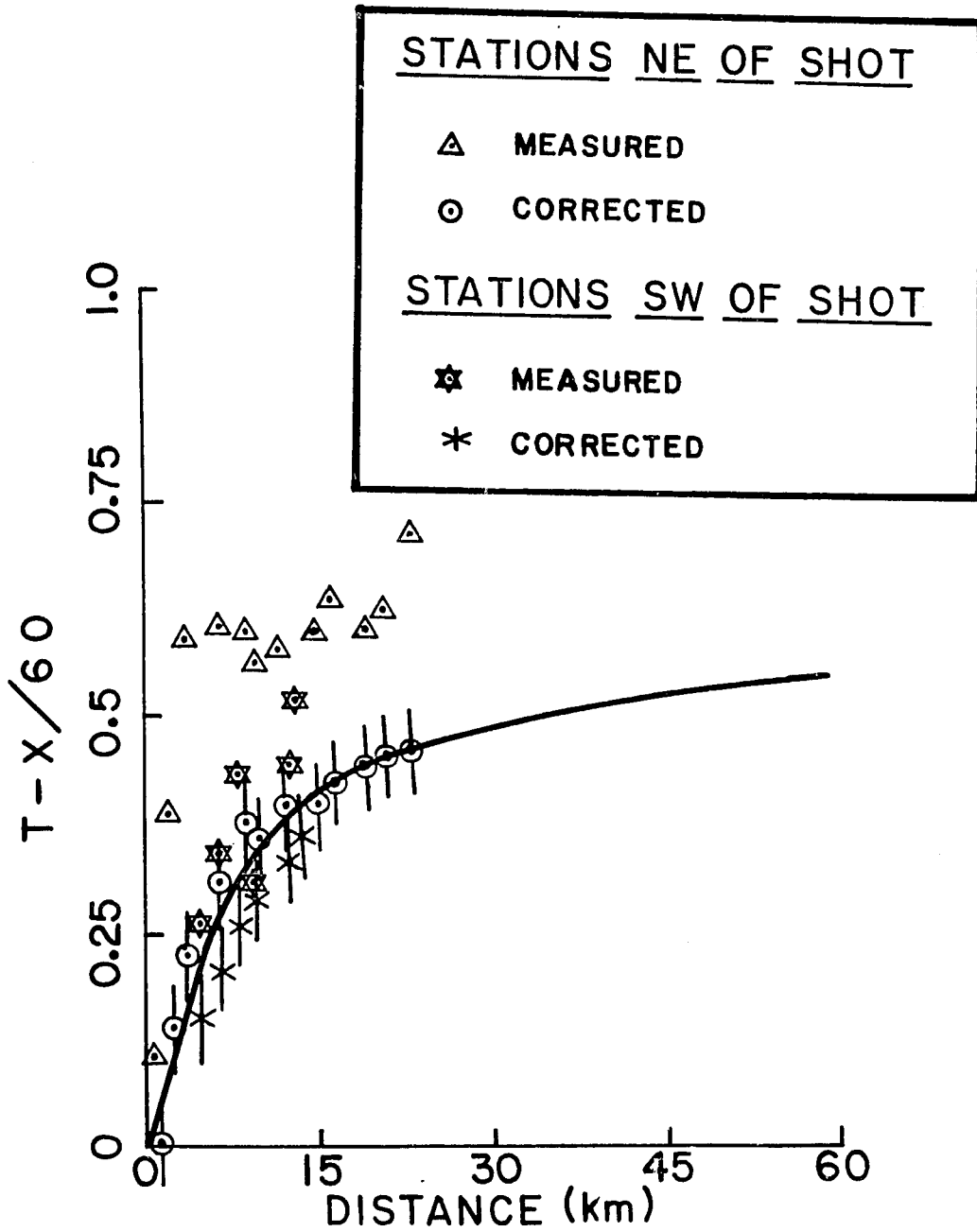
C SHOT
Permanent Stations
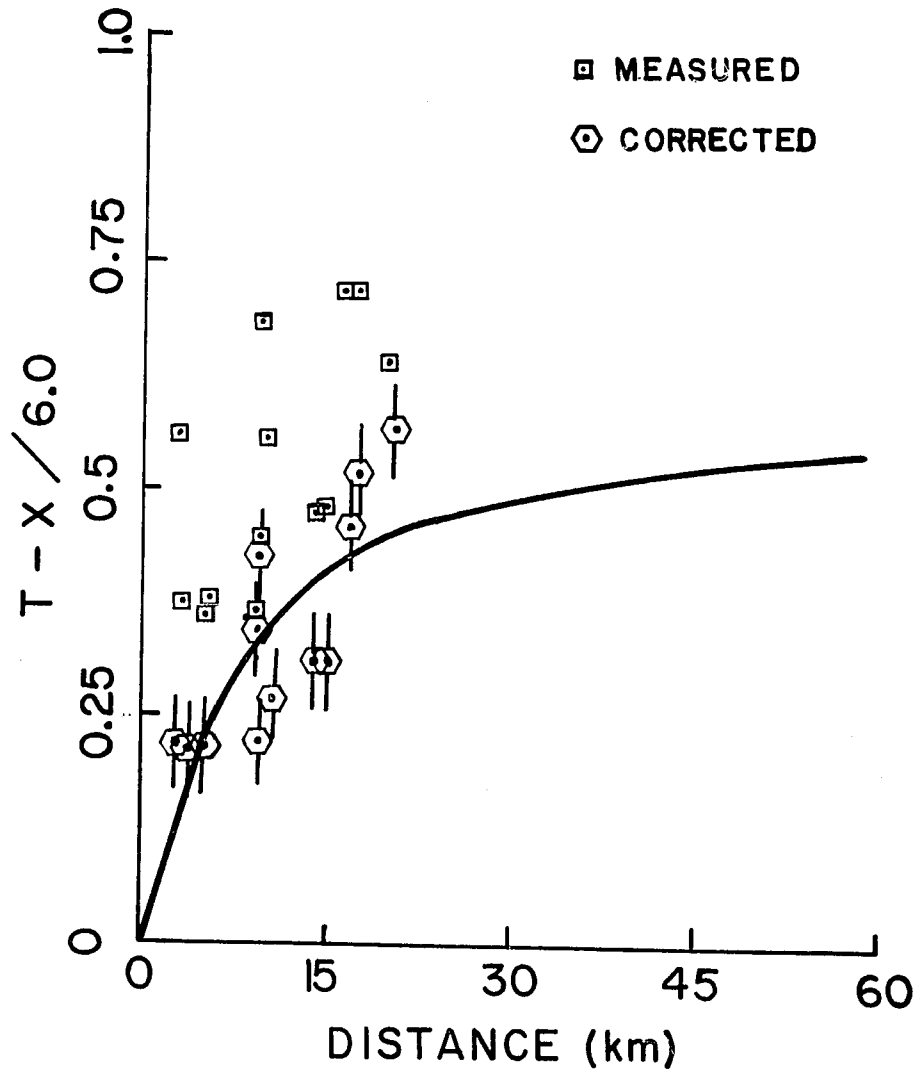
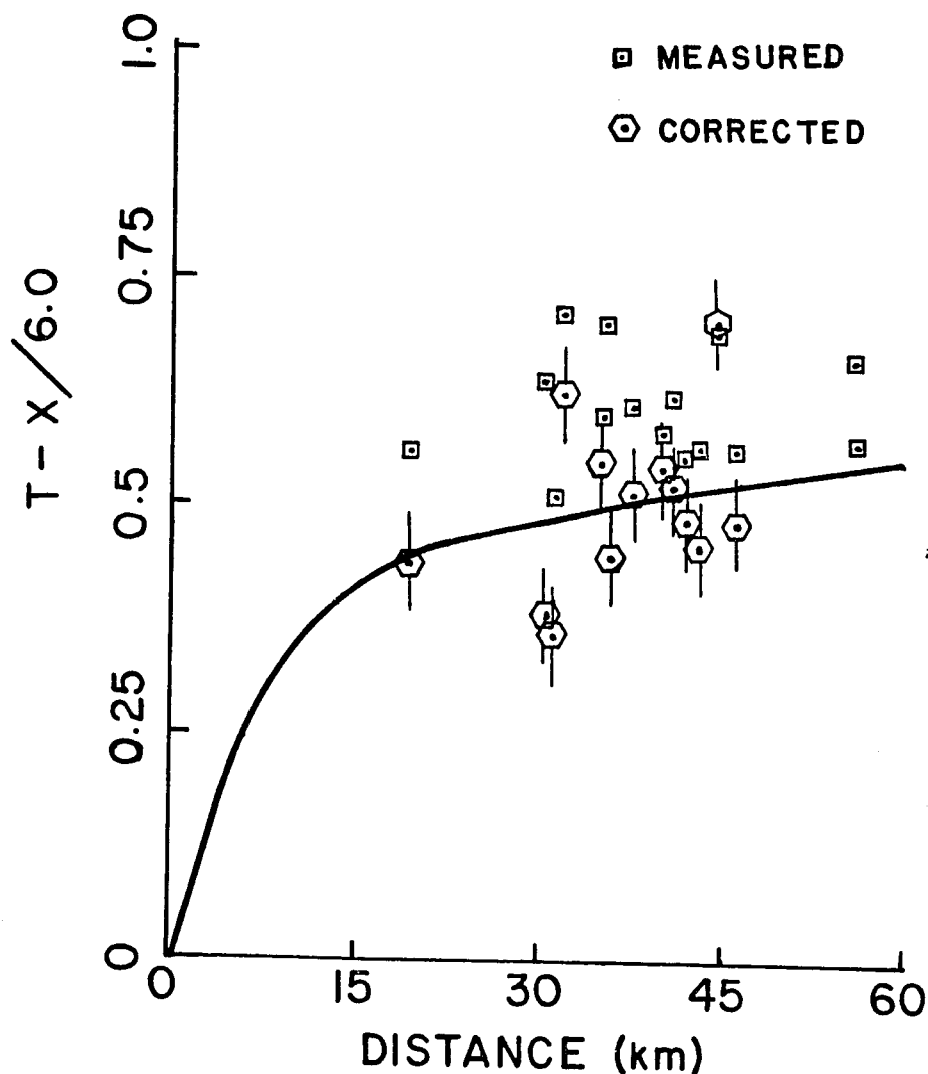**Figure 7.26.** Reduced travel times for permanent stations from C shot. Symbols are as defined in figure 7.24.
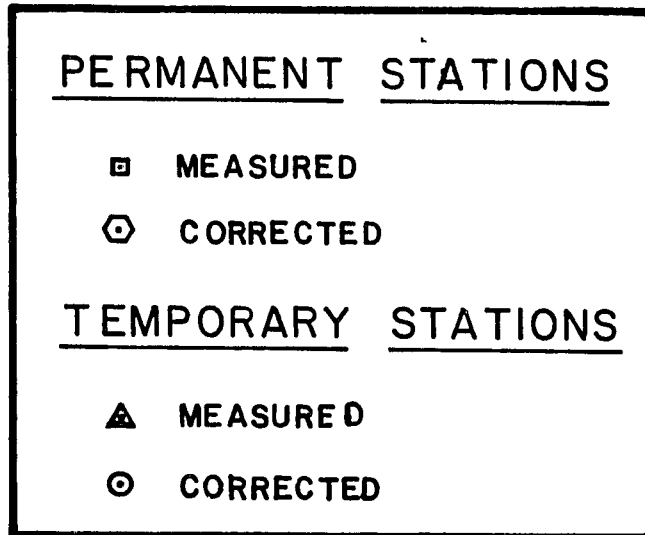
Figure 7.27. Reduced travel times for NE shot (all data). Symbols are as described in figure 7.23 and 7.24. Large error bars for these data are due to a misfire of the explosives for this shot [Weaver, personal communication].

NE SHOT
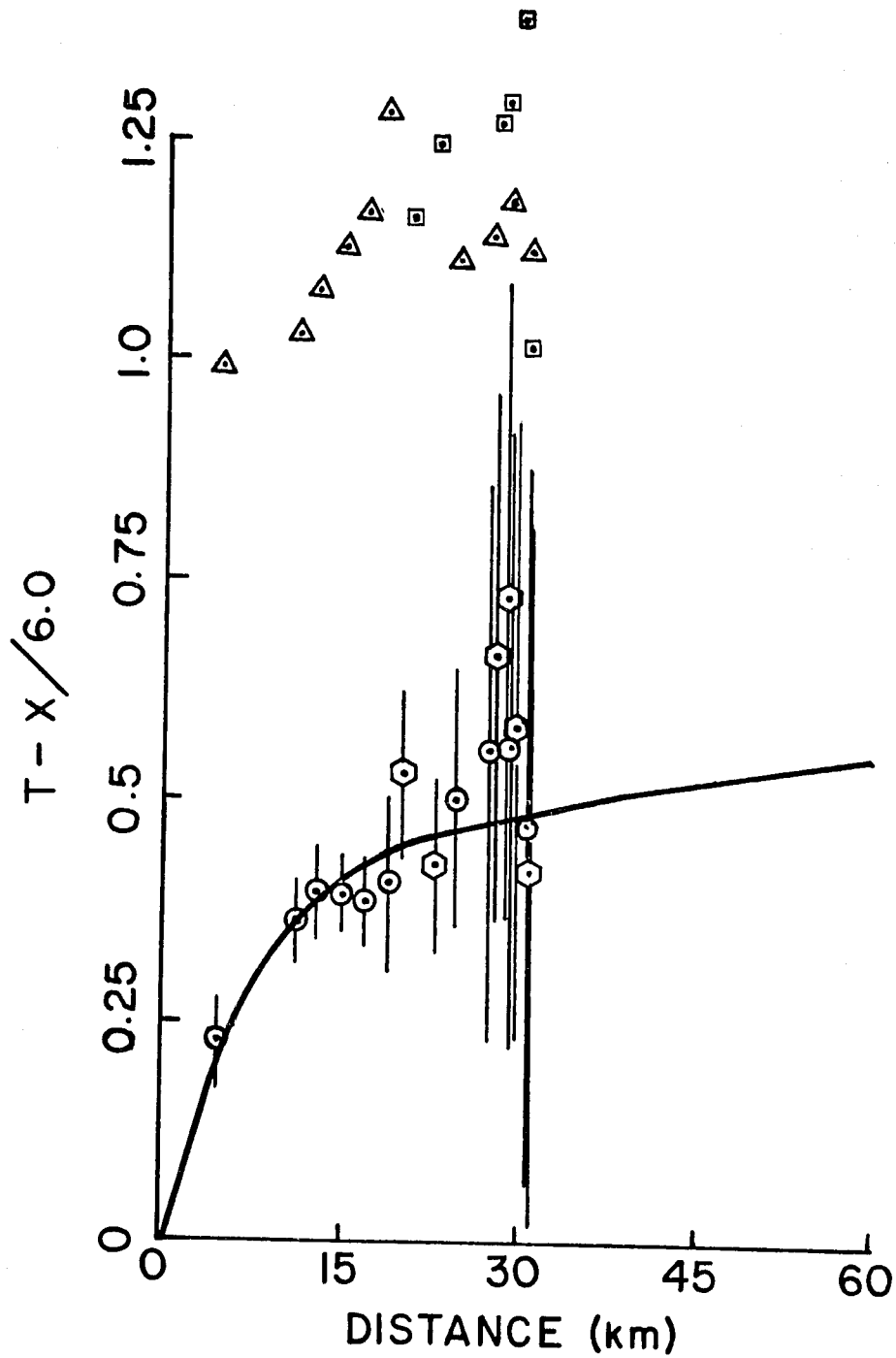
Kerr-McGee quarry blast

Figure 7.28. Reduced travel times for permanent stations from quarry blast. Symbols are as described in figure 7.24. Note that there is a considerably larger scatter in the corrected data than from the other shots. This is probably due to the fact that these data sample a different region of the crust than the remainder of this data set since most of these rays bottom near the eastern boundary of the Coso Range.

Figure 7.29. Residuals for profile stations from SW shot point. Tri- angles are the residuals one obtains when only the shot point correction and elevation corrections are applied. Circles are the final residuals with station corrections also applied. The bold trum- pet shaped lines denote standard error estimates assigned to the data as a function of distance for weighting purposes. These error bars are referenced to zero to permit a comparison of the size of a given residual or group of residuals in the context of the error to be anticipated in that datum.

SW shot point residuals

DISTANCE (km)

RESIDUAL (sec)

△--△ raw residuals

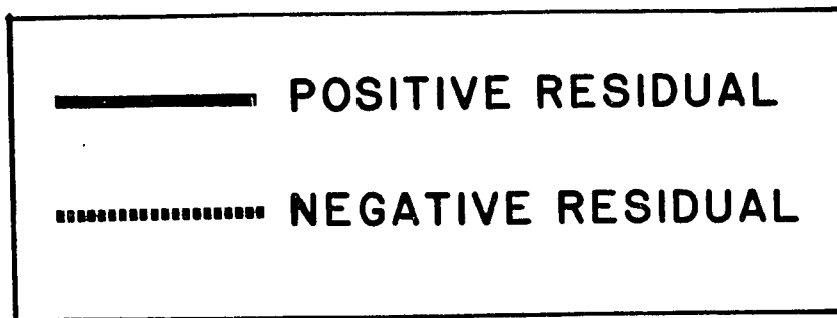⊙——⊙ residuals (station correction used)
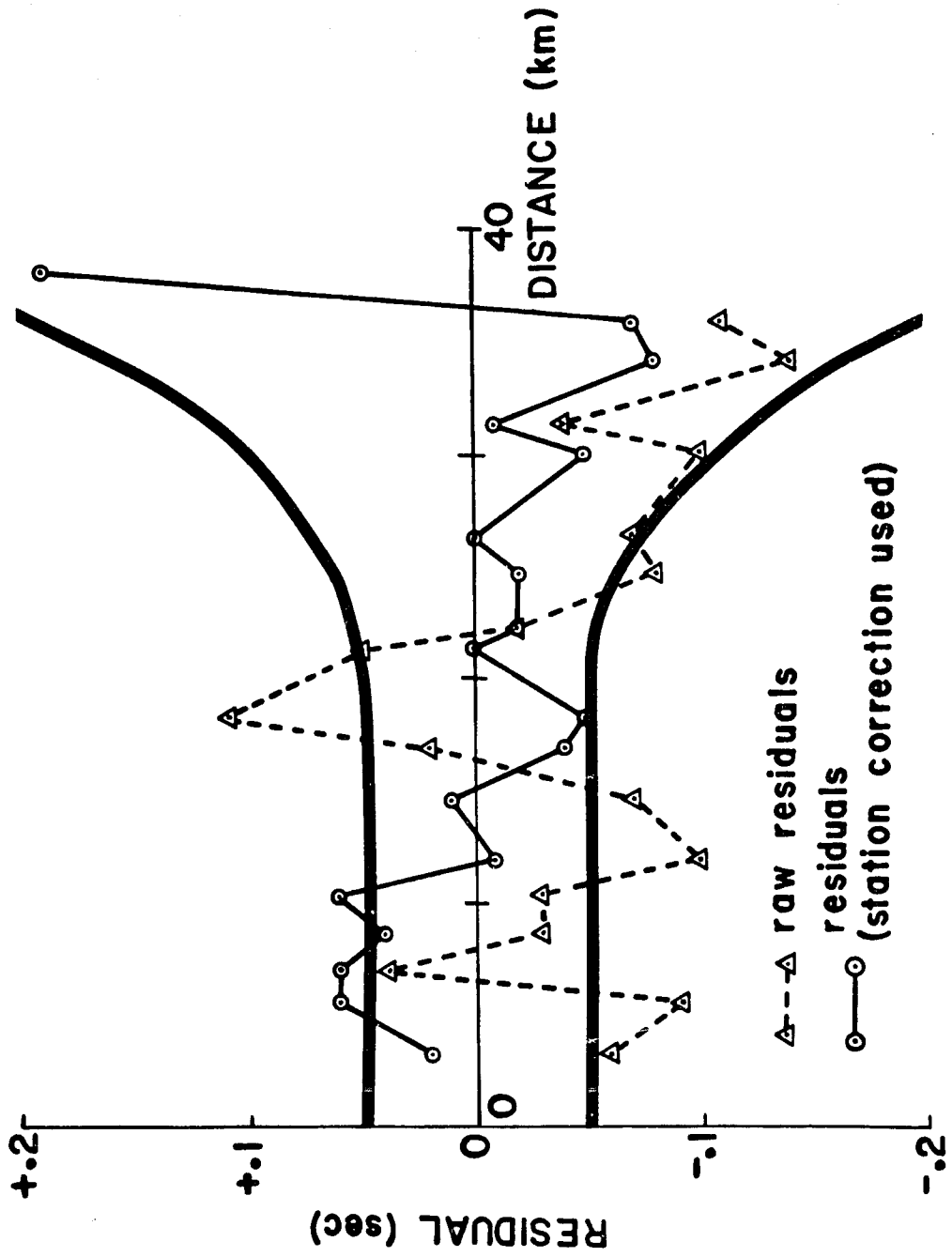
Figure 7.30. Residuals for profile stations from C shot point. Symbols are as described in figure 7.29. The orientation of the distance axis here is consistent with that in figure 7.29 and the record sections in figures 7.12 to 7.14 (The left is the southwest end of the profile and the right is the northeast end of the profile.).
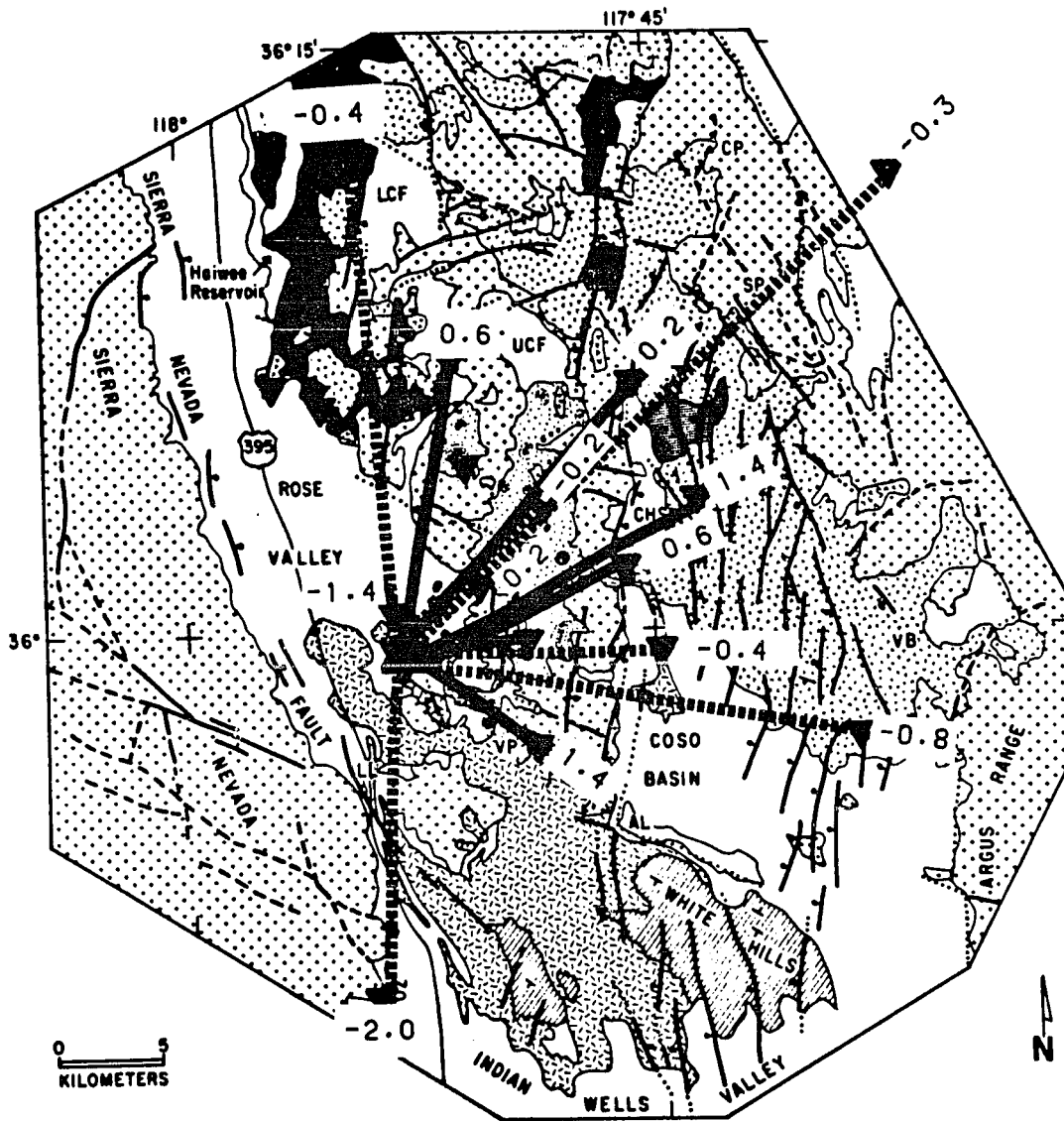
C shot point residuals

△ - - △ raw residuals
residuals
(station correction used)

DISTANCE (km)
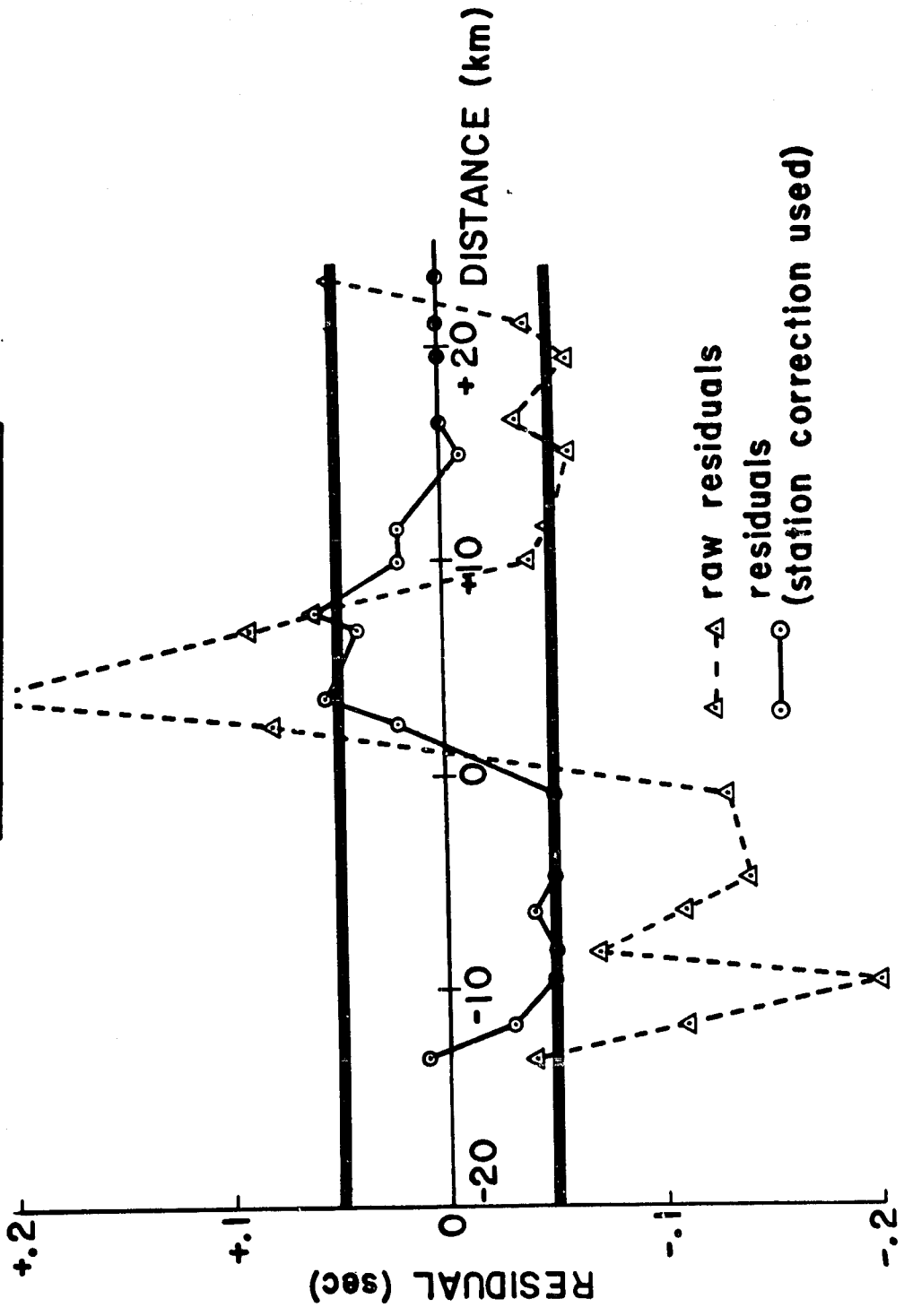
RESIDUAL (sec)

NE shot point residuals
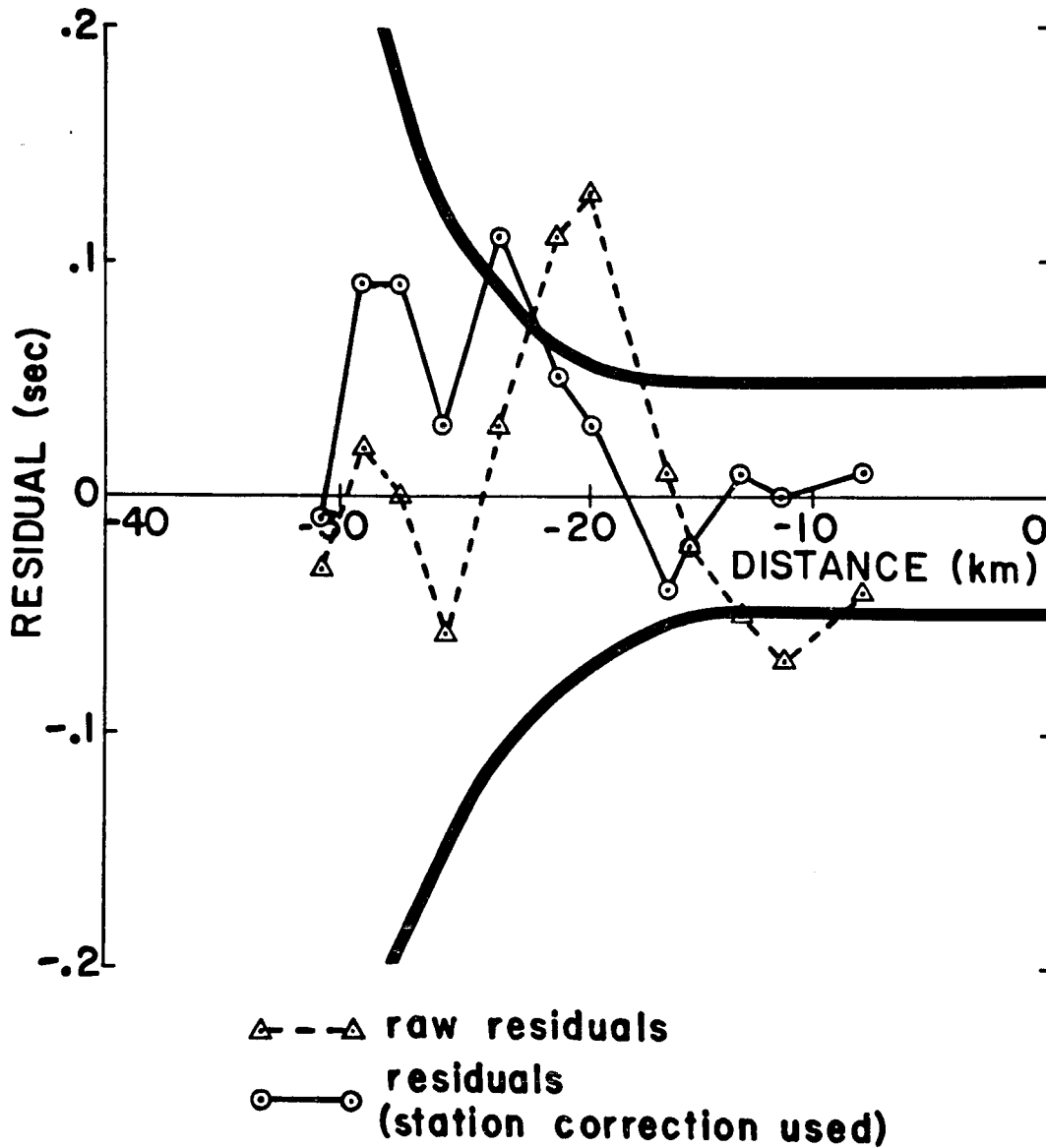
Figure 7.31. Residuals for profile stations from NE shot point. Symbols are as described in figure 7.29. The distance axis is again oriented consistently with that in figure 7.29 and 7.30.

```
┌─────────────────────────────────────────┐
│                                         │
│   ━━━━━━━━━  POSITIVE RESIDUAL          │
│                                         │
│   ·········  NEGATIVE RESIDUAL          │
│                                         │
└─────────────────────────────────────────┘
```

Figure 7.32. Weighted residuals for permanent stations from SW shot. This figure is a map view of the Coso network overlaid on a simplified version of Duffield and Bacon's [1980] regional geologic map (simplification from Duffield et. al. [1980]). Lines shown here connect shot and station location. This path is equivalent to the surface projection of the ray that joins the source and receiver positions. This is useful to clarify the relation of these ray paths to surface geology. Paths with positive residuals are shown as solid lines. Paths with negative residuals are shown as dotted lines.

117° 45'

36° 15'

118°

-0.4

-0.3

LCF

Haiwee
Reservoir

CP

SP

UCF

0.6

0.2

395

ROSE

0.2

0.2

VALLEY

-1.4

0.6

1.4

CHS

0.6

VP

-0.4

VB

36°

SIERRA

NEVADA

SIERRA

NEVADA

FAULT

COSO

BASIN

-0.8

AL

1.4

ARGUS RANGE

WHITE

HILLS

-2.0

INDIAN

VALLEY

WELLS

N

0   5
KILOMETERS

EXPLANATION
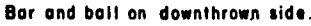
Quaternary alluvial, fluvial,
playa, and wind-blown deposits

Pliocene volcanic rocks and
intercalated Coso Formation

Pleistocene basalt

Pre-Cenozoic granitic and
metamorphic rocks

Pleistocene rhyolite-domes and
flows, concentric dots-pyroclastic
deposits, scattered dots

Fault, dotted where concealed and dashed
where uncertain.
Bar and ball on downthrown side.

Pleistocene sedimentary rocks
of the White Hills

Generalized attitude of sedimentary
rocks and Pliocene basalt flows

POSITIVE RESIDUAL

NEGATIVE RESIDUAL
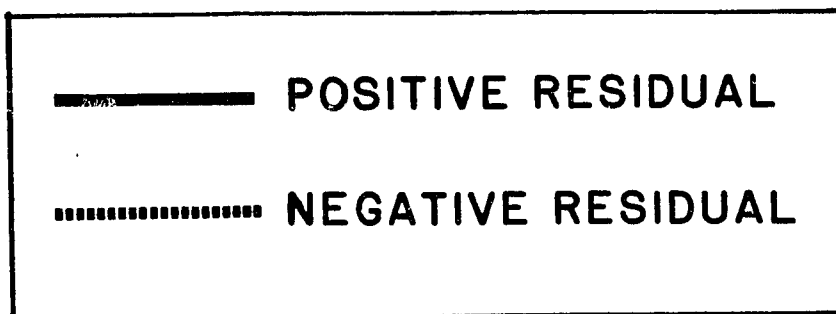
Figure 7.33. Weighted residuals for permanent stations from C shot. See figure 7.32 caption for further description.

EXPLANATION

☐ Quaternary alluvial, fluvial, playa, and wind-blown deposits

▨ Pleistocene basalt

▨ Pleistocene rhyolite-domes and flows, concentric dots-pyroclastic deposits, scattered dots

▨ Pleistocene sedimentary rocks of the White Hills

◼ Pliocene volcanic rocks and intercalated Coso Formation

☐ Pre-Cenozoic granitic and metamorphic rocks

····· ━━━•━ ━ ━ ━
Fault, dotted where concealed and dashed where uncertain.
Bar and ball on downthrown side.

⊥
Generalized attitude of sedimentary rocks and Pliocene basalt flows

Figure 7.34. Weighted residuals for permanent stations from NE shot. See figure 7.32 caption for further description.

EXPLANATION

Quaternary alluvial, fluvial, playa, and wind-blown deposits

Pleistocene basalt

Pleistocene rhyolite-domes and flows, concentric dots-pyroclastic deposits, scattered dots

Pleistocene sedimentary rocks of the White Hills

Pliocene volcanic rocks and intercalated Coso Formation

Pre-Cenozoic granitic and metamorphic rocks

Fault, dotted where concealed and dashed where uncertain.
Bar and ball on downthrown side.

Generalized attitude of sedimentary rocks and Pliocene basalt flows

Figure 7.35. Weighted residuals for permanent stations from quarry shot. See figure 7.32 caption for further description.

EXPLANATION

▭ Quaternary alluvial, fluvial, playa, and wind-blown deposits

▨ Pleistocene basalt

▨ Pleistocene rhyolite-domes and flows, concentric dots-pyroclastic deposits, scattered dots

▨ Pleistocene sedimentary rocks of the White Hills

◼ Pliocene volcanic rocks and intercalated Coso Formation

▨ Pre-Cenozoic granitic and metamorphic rocks

····· ━━━ ─ ─ ─
Fault, dotted where concealed and dashed where uncertain.
Bar and ball on downthrown side.

⊥
Generalized attitude of sedimentary rocks and Pliocene basalt flows

is figures 7.23 to 7.28. There I show the actual, measured travel times and the "corrected" travel times (The corrected travel time is the measured travel time minus an elevation, station (geologic), and a shot point correction.) for all four shots on a reduced travel time plot (reducing velocity = 6.0 km/sec). The theoretical travel time curve for the model shown in figure 7.20 is also shown in each of these figures for a reference. The success of the fit of the corrected data to the theoretical travel time curve here is striking. The scatter in the raw travel time data is fairly substantial. Furthermore, the overall trend of the raw data looks quite different for the four separate shot points (Note especially the substantial delay in the arrival times from the NE shot.). The scatter in the corrected data, by comparison, is enormously less and represents a major success for the application of this procedure. Closer examination of the results for the profile stations, however, reveals some noticeable trends that deserve further comment.

The final residuals observed at all the profile stations for the three shots fired on this line are presented in figure 7.29 to 7.31. The data from the NE shot (figure 7.31) is not very good, but there are not any obvious trend in these data. The same is not true of the SW and C shots (figure 7.29 and figure 7.30). These shots show fairly substantial residuals at the stations closest to the shot. The anomalous arrivals at stations close to the shot are associated with ray paths that do not penetrate very deeply into the earth (All of these arrivals bottom within the top kilometer.). Consequently, the waves that arrive at these stations have passed at low angles through the complex structures that surface geology shows are present within the upper few hundred meters of the Coso Range (see geologic map in figure 7.41). Station corrections cannot completely compensate for these complexities at stations close to the shots, which suggests that the near surface seismic velocity

structure has significant lateral velocity variations.

Two trends are also apparent in the residuals from the explosions recorded at the permanent stations. This is demonstrated in figure 7.32 to 7.35 in the form of a map view in which lines have been drawn connecting each source-receiver pair. These lines represent the surface projection of the ray path that a given residual (note the numbers shown are weighted residuals) is associated with. The first trend to note is the strong pattern that can be seen in the residuals from the C shot (figure 7.33). All stations in the quadrant southwest of the shot have substantial negative residuals while all other stations show positive residuals. Furthermore, the same pattern is apparent in the residuals from the profile stations for this shown shown in figure 7.30. This residual pattern has a fairly simple explanation. The C shot was fired near the western edge of the Coso Basin north of Coso Hot Springs. The Coso Basin is bounded on the west by a major normal fault with the downthrown block to the east as shown in figure 7.36 from Duffield et. al. [1980]. Because the C shot was positioned near this boundary, the waves that traveled to the southwest of the shot had to travel only a short distance through the fan deposits within this basin. This led to arrival times that were early compared to other stations where the waves had to travel larger distances through this low velocity material.

A second pattern can be seen most clearly in the residuals from the SW shot (figures 7.29 and 7.32) and more weakly in the residuals from the C (figure 7.30 and 7.33) and the NE shot (figure 7.31 and 7.34). These data suggest a pattern in which stations to the west tend to have arrivals that are relatively late (positive residuals) compared to stations further east. This pattern is consistent with the regional dip in the structure from the east to the west that has been produced within the eastern sections of the Coso Range by a

series of north-south normal faults as illustrated in figure 7.36. This trend apparently does not continue to the east as a similar trend is not apparent in the data from the quarry blast shown in figure 7.35.

To conclude this section, I would like to point out that although this detailed study of the residuals from the four shot points reveals some significant patterns that are indicative of lateral velocity variations, I would argue that this fact is of little consequence to the analysis of earthquake data. I make this claim primarily because earthquake and explosion sources have a fundamental difference. The explosions always occur very close to the earth's surface. As a result, the waves they produce spend a much higher percentage of the total travel time passing through the complex near surface structure than the waves produced by earthquake sources. Geologic mapping of this area indicates that the overall structure of the Coso Range is quite simple. The range is underlain primarily by Mesozoic plutons and subordinate metamorphic rocks that are part of the Sierra Nevada batholith. Pliocene and Pleistocene volcanic rocks bury this basement complex over most of the range, but this cover is generally less than a few hundred meters thick [Duffield et. al., 1980]. Because the data from explosion sources is more sensitive to this near surface structure, it is not surprising that there is more scatter in these data than is apparent in the earthquake data[7]. This fact is of little consequence, however, to the problem of obtaining precise earthquake locations, because the seismicity is concentrated below the complex, near surface structure. That the

---

[7] The scatter in the refraction data is significantly different from the earthquake data. $\frac{SSWRES}{NDGF}$ of all the refraction data is 1.14. An F-test [Hoel, 1971, pp. 269-273] shows that this is significantly different from the value of $\frac{SSWRES}{NDGF}$ for all of the earthquake data.
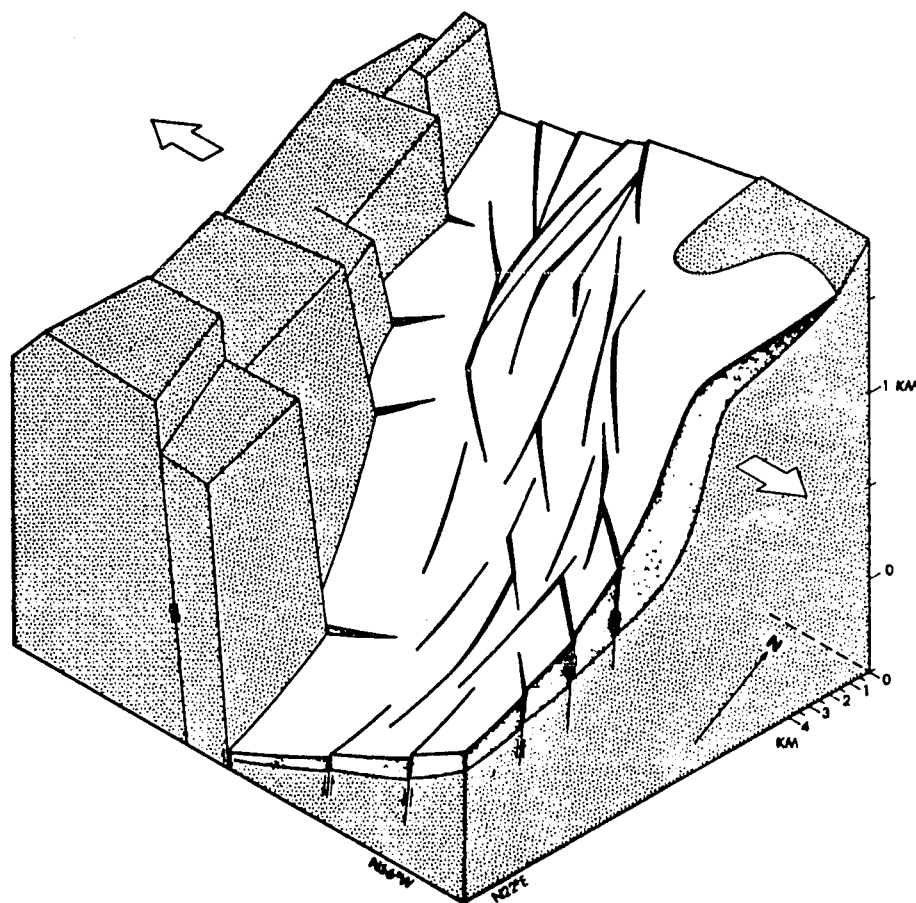
Figure 7.36. Block diagram showing structure of eastern Coso Range. The major northwest trending fault zone through the Airport Lake area (figure 7.41) forms the south edge of the diagram. The western edge passes through the major horst on which the Pleisocene rhyolite domes of the Coso volcanic field have been emplaced. The C shot was fired in a position near the northern edge of this figure in the fan sediments on the immediately east of this major horst (from Duffield et. al. [1980]).

near surface complexities of the velocity structure at Coso can be
adequately approximated by station corrections is undeniable in
light of the excellent fit of this model to the earthquake data used
in this study (figure 7.14). Thus, I conclude that the approximation
of a one-dimensional velocity model is a very good one for this
region.

3.5. Error Appraisal

3.5.1. Introduction

The purpose of the previous section was to demonstrate that
the model found here fits the observations to within the expected
measurement errors. This step was necessary to justify that these
data are adequately fit by a one dimensional velocity model. Having
demonstrated that, I address in this section the more basic ques-
tion of how this model (the total model consists of a velocity model,
a set of station corrections, 4 shot point corrections, and the hypo-
centers of the 59 earthquakes.) relates to reality. This section is an
application to these data of the error assessment techniques
described in chapter 5. It is broken into three subsections
corresponding to the three distinct parts of the model; the velocity
model, station (shot) corrections, and the hypocenters of the earth-
quakes.

3.5.2. Velocity model

An important result of the application of the annulling transfor-
mation is that it allows one to directly apply the powerful error
assessment techniques of Backus and Gilbert [1968,1970] to
appraise the uniqueness of the estimated velocity model. This is
the technique I referred to earlier (see section 3 of chapter 5 which
serves as a companion to this section) as the "trade-off curves of

error versus resolution length" This technique recognizes that the solution of any inverse problem is fundamentally ambiguous because of two sources of uncertainty. The first of these is the ubiquitous problem of the propagation of random measurement errors into anything estimated from data (statistical errors). The second is unique to inverse problems. The velocity model is fundamentally ambiguous because it requires (in principle at least) an infinite number of parameters to be specified exactly. The quantity of data we have available is, however, always finite so we cannot hope to resolve the velocity model on an arbitrarily fine scale but we must always be content with some smoothed version of it. One measure of this smoothing is the resolution length defined in equation (5.27). Backus and Gilbert [1970] show that these two sources of uncertainty are mutually exclusive. That is, statistical error can only be reduced at the expense of a degradation in resolution (increase in resolution length) and vice versa. There is a trade-off between the resolution length and the statistical errors. This trade-off can be studied by plotting the size of the resolution length as a function of the statistical error (or vice versa) for the family of solutions obtained as the number B in equation (5.33) varies. The resulting curves, which are referred to as trade-off curves, are strictly decreasing functions that are unique properties of the data [Backus and Gilbert, 1970]. These curves are generally different for estimates of the velocity at different depths. Johnson and Gilbert [1972] show how variations in the form of these trade-off curves with depth can be used to objectively assess the nonuniqueness of the velocity model.

The trade-off curves associated with the velocity model estimates from these data are presented in figures 7.37 and 7.38. The trade-off curves for all depths are shown in the form of a contour plot similar to those given by Johnson and Gilbert [1972]. The
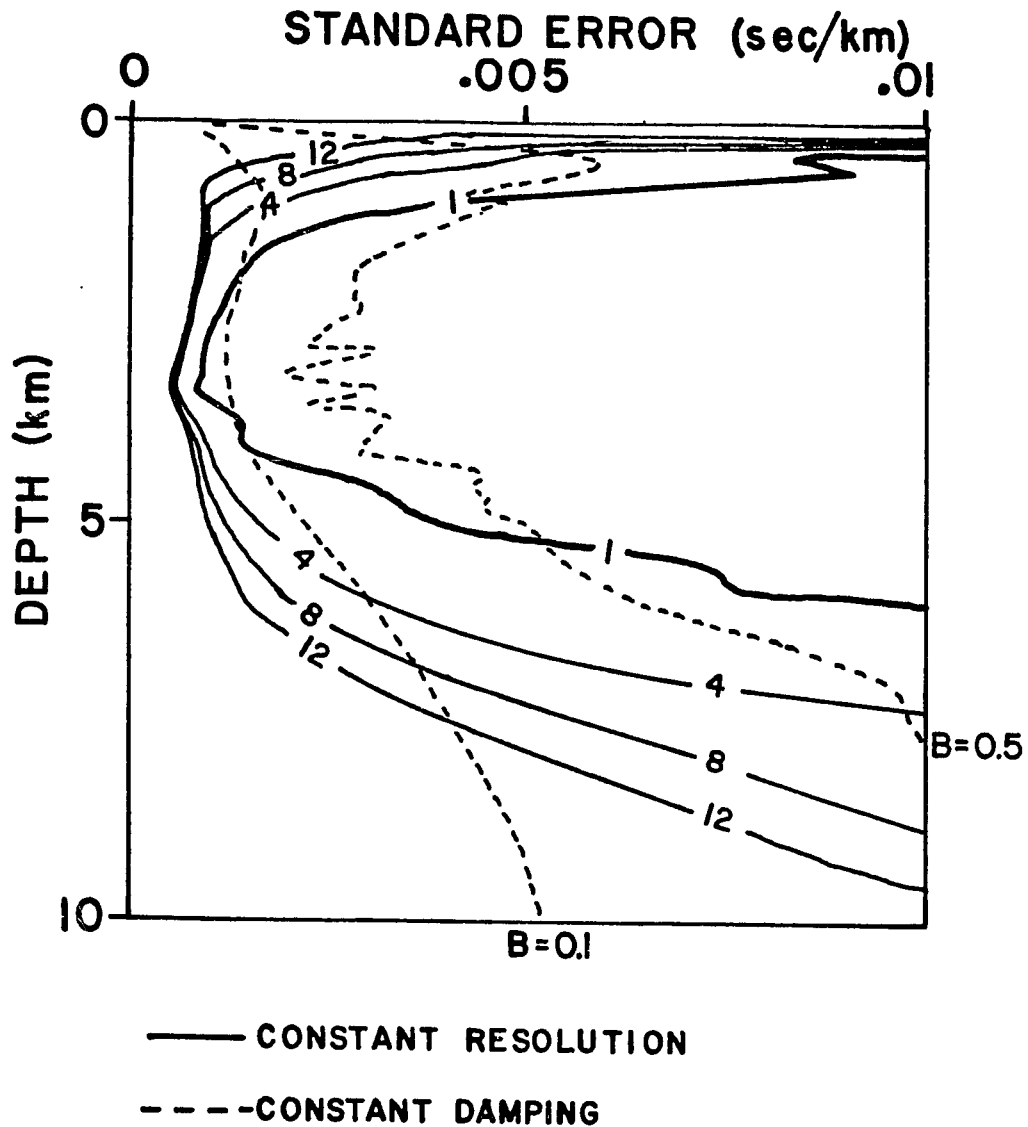
Figure 7.37. Trade-off curves for Coso velocity model (surface velocity free). Horizontal axis is the predicted standard error for the slowness perturbation. Solid lines are contours of constant resolution length. Dashed lines show standard errors (and resolution lengths) associated with a slowness perturbation constructed using a constant value of the bound B (equivalent to a constant damping: see text).
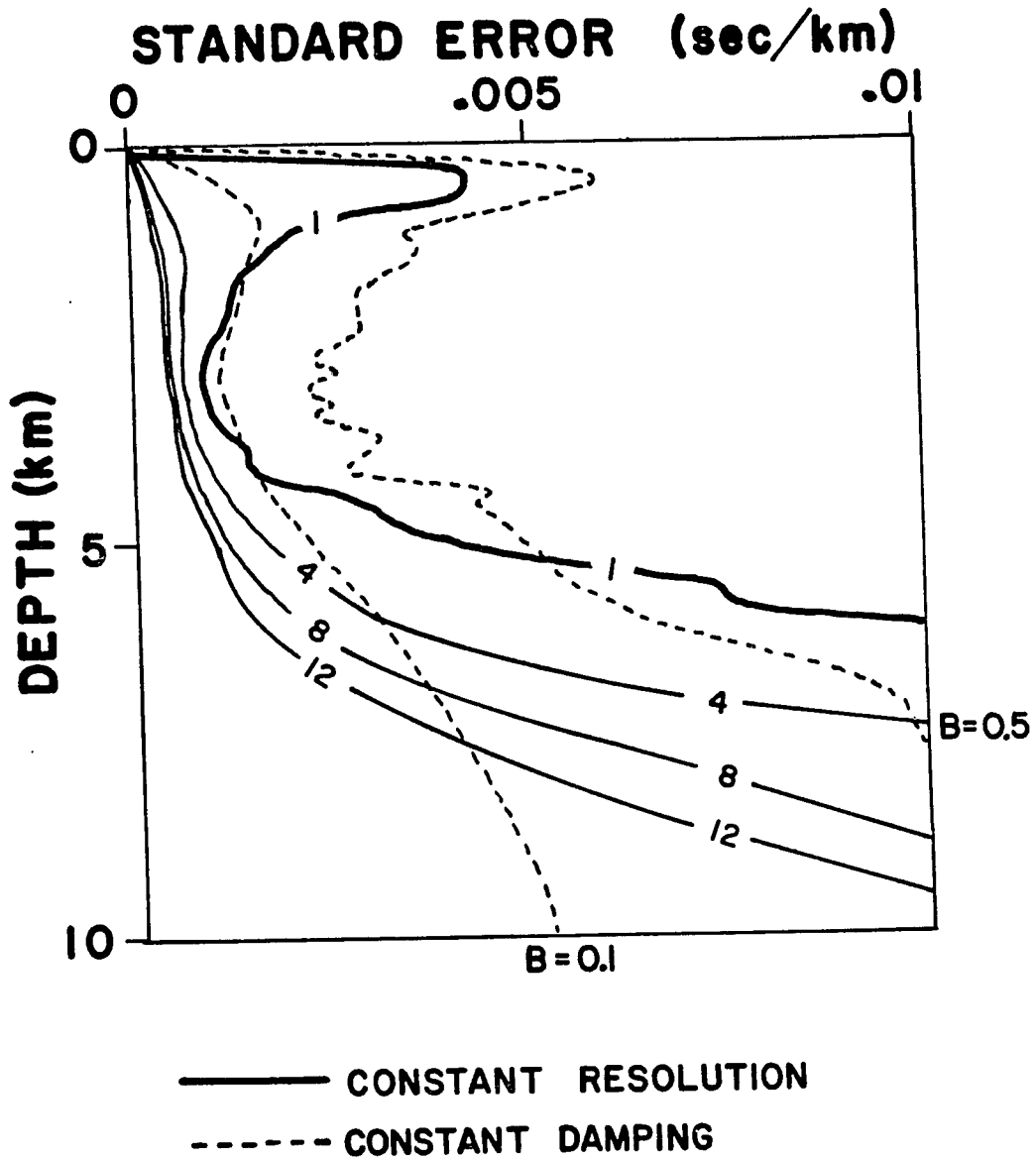
Figure 7.38. Trade-off curves for Coso velocity model (surface velocity fixed). Plot is as explained in caption of figure 7.37.

vertical axis is the depth and the horizontal axis gives the statistical error in the estimated slowness perturbation. The trade-off curve at each depth is the plot of resolution length as a function of statistical error. Consequently, the resolution length can be considered as a function of both the depth and the statistical error that can be contoured as shown in figures 7.37 and 7.38. The two differ in that figure 7.37 shows the results for the case when the surface velocity is free to vary (assumed unknown) and figure 7.38 shows the same results for the case when the surface velocity is fixed (assumed known exactly). From this we see that the only significant effect that occurs if the surface velocity is fixed, is a substantial improvement in the resolution of the near surface velocities. This is to be expected since fixing the surface velocity is appropriate only when it is already known. This represents an addition piece of information that manifests itself here in a dramatic improvement in the resolution of near surface velocities.

The dashed lines in figure 7.37 and 7.38 are also significant. They identify the uncertainties that are present in a set of estimates related by the number B in equation (5.33). These sibling estimates are important because they represent an estimate for the entire slowness perturbation function, $|\delta \hat{u}>$, that is relatively easy to calculate[8]. Figure 7.39 shows an example of the usefulness of these curves. There a velocity model perturbation estimated from the final residuals using $B = 0.10$ is shown in the context of its associated uncertainties. The statistical errors used to construct

---

[8] The only quantities that differ in equation (5.33) when B is fixed are the numbers h_i on the right hand side. The h_i depend only on the depth the model is to be estimated at (equation (5.30)). Thus, an estimate of a perturbation to the entire velocity model can be had from a single inversion of the matrix on the left hand side of (5.33).

Figure 7.39. Coso velocity model uncertainties. (a) Reference model from figure 7.20 is shown with a perturbation calculated from travel time residuals derived from this velocity model. Perturbation shown is a constant B (constant damping) solution using $B = 0.10$ (see figure 7.37) and is shown in relation to the predicted statistical errors for this perturbation. (b) Resolution length as a function of depth for perturbation shown in (a).

RESOLUTION LENGTH (km)

DEPTH (km)

(b)

VELOCITY (km/sec)

DEPTH (km)

——— Model

········ Perturbation

——— 95% Confidence Limits

(a)
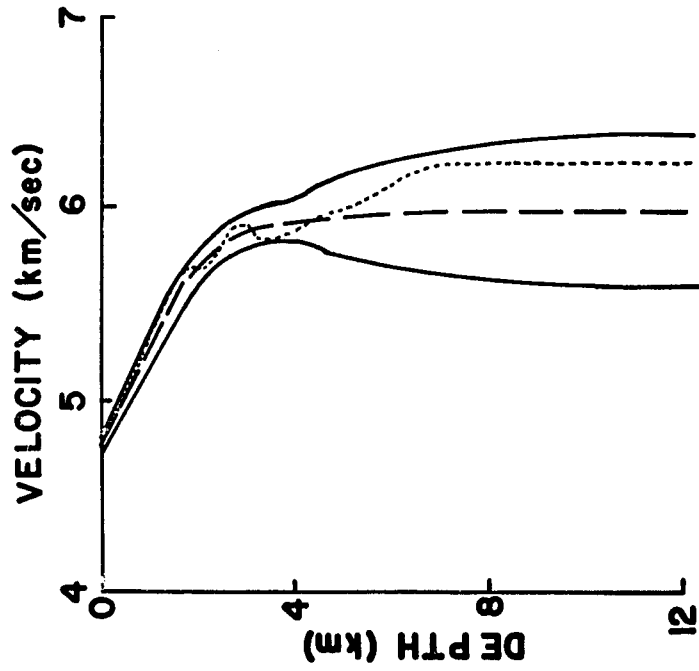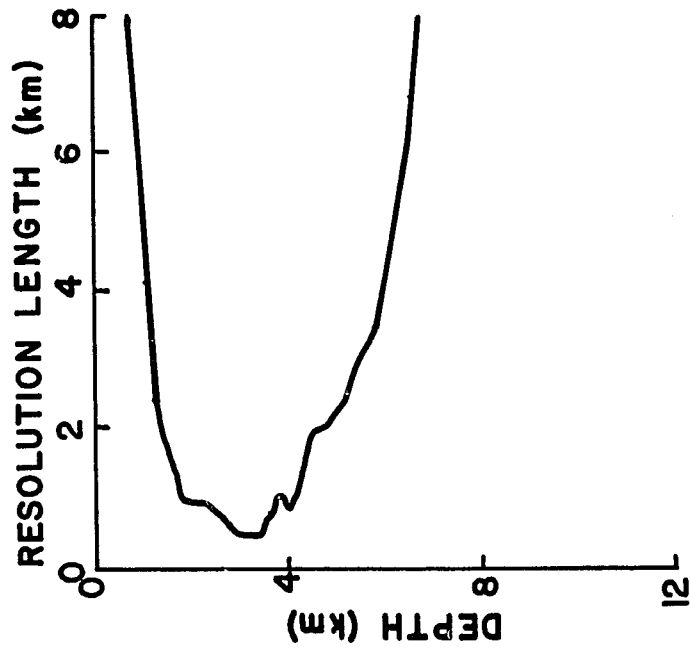
the 95% confidence limits in figure 7.39a can be obtained directly from figure 7.37 by following the leftmost dashed line. Similarly, the resolution length shown in figure 7.39b can be had by noting how that dashed line crosses the resolution length contours. The solution shown in figure 7.39 is of special interest because it is optimum in the sense that larger values of B lead to a rapid increase in statistical errors and smaller values of B lead to a rapid degradation in resolution. Figure 7.39 also demonstrates one of the most useful applications of trade-off curves. The low velocity zone that is introduced by the perturbation here is seen to be fairly well resolved but of questionable significance. That is, the width of the feature is of the order of the resolution that we can expect in this depth range, but its significance is questionable because its size is of the order of the noise. Thus, this feature is intriguing but we cannot be certain that it is real.

Overall, the trade-off curves in figure 7.37 and 7.38 show a pattern in which the resolution of the velocity model is quite good (resolution of 1 km or less is achievable with an acceptable statistical uncertainty) down to a depth of slightly greater than 4 kilometers. Below that depth, one sees that resolution degrades rapidly to the point that no reasonable estimate of the velocity model is possible for depths greater than about 6 kilometers. This result was quite surprising considering that over half of the earthquakes in this data set are deeper than 4 kilometers (see figure 7.10). Since the deepest refraction arrival bottomed at a depth of 4.1 kilometers, I was concerned that the results were being overly dominated by the refraction data. Consequently, I repeated the trade-off curve calculation with all the refraction data excluded. The results of this calculation are shown in figure 7.40 where they are compared directly with a duplicate of figure 7.37. Here we see that removing the refraction data does not change the overall picture

Figure 7.40. Effect of refraction data on resolution of Coso velocity model. Left figure shows trade-off curves for a velocity model estimate derived only from earthquake arrival time data. Right figure shows trade-off curves using all data (same as figure 7.37). Dashed lines in both figures are constant damping curves with B of same size as those shown in figure 7.37.

TRADEOFF CURVES
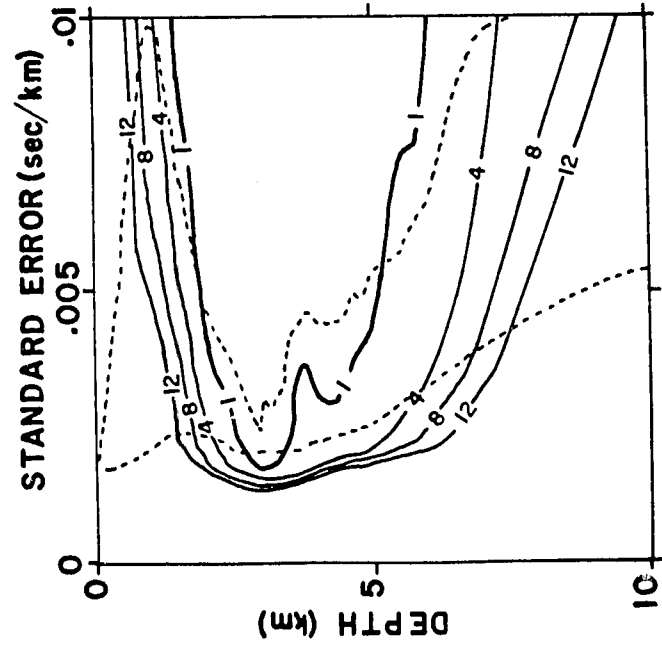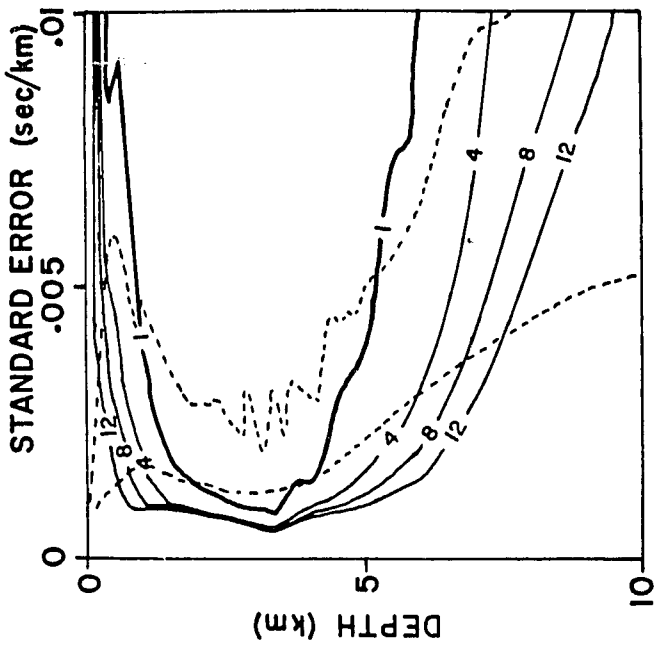
Refraction Data Included

Refraction Data Excluded

——— CONSTANT RESOLUTION

— — —CONSTANT DAMPING

substantially. The only major differences are an overall increase in statistical errors at all depths and a decrease in resolution of the near surface velocities. The loss of resolution at depth still occurs at about the same depth, indicating that this must be a geometric effect due to the small aperture (50 kilometers maximum) of the Coso network.

3.5.3. Station and Shot Point Corrections

The station correction estimates for this network are listed in Table 7.3 and the four shot point correction estimates are listed in Table 7.4. In addition, these station corrections are displayed in figure 7.41 overlaid on Duffield and Bacon's [1980] geologic map for the region. We see that these station corrections correlate very well with surface geology. Station that lie on or near the major basins in the area (Rose Valley and Coso Basin) and those stations located on the fan sediments of the Coso formation [Duffield et. al. 1980] tend to show positive station corrections. This is consistent with what we expect as the lower velocity sediments at the surface in these areas should cause arrivals to be observed relatively late (a positive correction).

When presented with a set of numbers like this, a normal question to ask is how believable they are. In this case I can show quite convincingly that they are known very well. The way I can do this is through the trade-off analysis described in section 4 of chapter 5. There I showed that estimation of discrete parameters like these from arrival time data is fundamentally ambiguous for the same reason that estimation of the velocity model is ambiguous. That is, these estimates have uncertainties due to both measurement errors in the data, and the fact that the velocity model cannot be determined uniquely. In chapter 5 I introduced a new method of analysis that can be used to appraise the relative importance of

EXPLANATION

Quaternary alluvial, fluvial, playa, and wind-blown deposits

Pleistocene basalt

Pleistocene rhyolite-domes and flows, concentric dots-pyroclastic deposits, scattered dots

Pleistocene sedimentary rocks of the White Hills

Pliocene volcanic rocks and intercalated Coso Formation

Pre-Cenozoic granitic and metamorphic rocks

Fault, dotted where concealed and dashed where uncertain.
Bar and ball on downthrown side

Generalized attitude of sedimentary rocks and Pliocene basalt flows

Figure 7.41. Station corrections for Coso network. Corrections for all stations used in this study are shown color coded by the size of the estimated correction. These have been overlaid on a a simplified version of Duffield and Bacons's geologic map (simplification from Duffield et. al. [1980]) to show how strongly these station corrections correlate with surface geology. Station names can be had by comparison with figure 7.11. The exact numerical values of these station correction estimates are listed in Table 7.3.
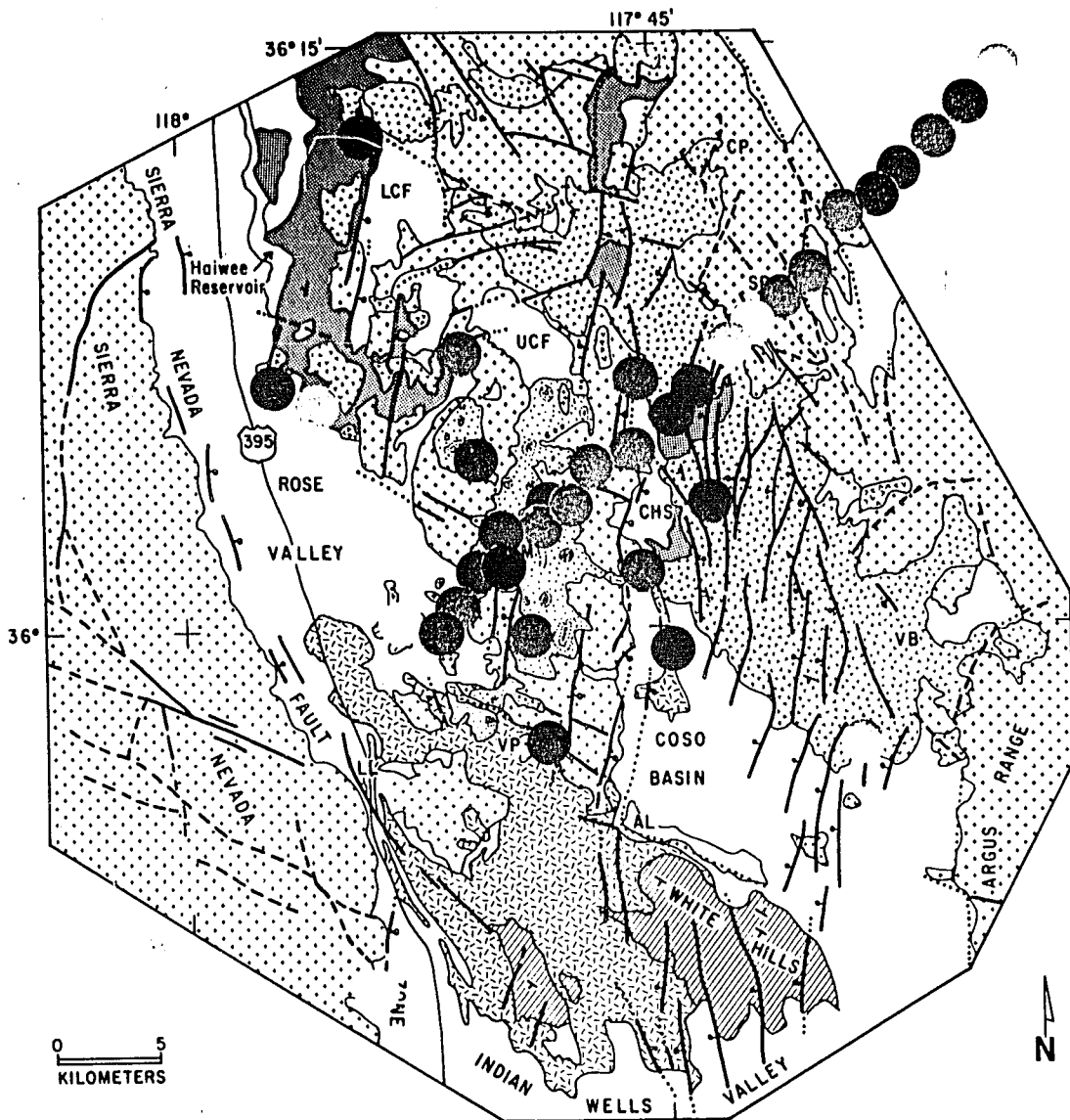
STATION CORRECTION (sec)

⬤   s < -0.10

◕   -0.10 ≤ s < -0.05

◔   -0.05 ≤ s < 0.0

     0.0 ≤ s < 0.05

◌   0.05 ≤ s < 0.10

⬤   0.10 ≤ s

these two sources of error. This technique utilizes a set of trade-off curves that show how these two sources of error interplay. Figure 7.42 is an example of the application of this analysis for the estimated station correction at station JRW. Figure 7.42 is similar to figure 5.6 and 5.7 described above. It shows the size of the model, statistical, and total errors (see chapter 5) plotted as a function of the number B (equation (5.49)). Figure 7.42 differs from figure 5.6 and 5.7, however, in that the number B in figure 7.42 is interpreted in terms of the roughness length,κ, defined in equation (5.51). κ is useful because it can be interpreted in terms of the scale lengths of the hypothetical, unresolved velocity structure. (I also show the results in terms of B to facilitate comparison with the velocity model trade-off curves.) Figure 7.42 is significant because it shows that statistical errors are essentially constant and several orders of magnitude larger then the model errors for most of the range of κ shown. This fact implies two things:

(1) Model errors can be essentially neglected for the estimate of this station correction. This is so because one has to assume κ is unreasonably small before model errors become significant in comparison to statistical errors. To see this, note that $\kappa = 10^{-4}$ kilometers implies that there are large velocity variations occurring over distances $d \approx 1$ millimeter and $\Sigma \approx 1$ meter (see definition of κ in (5.51)). Thus 10^{-4} is a lower limit on a reasonable value for κ, because smaller values of κ are rejected as physically unrealistic. Whenever, $\kappa > 10^{-4}$ model errors are small compared to statistical errors so model errors can be ignored for all intents and purposes.

(2) Because the statistical error curve is essentially flat out to very small κ, a statistical error estimate based only on the covariances calculated by PMEL (I called this the "intrinsic statistical error" in chapter 5) gives a reasonable error estimate for this

EXPLANATION

———— TOTAL ERROR

———————— STATISTICAL ERROR

·························· MODEL ERROR
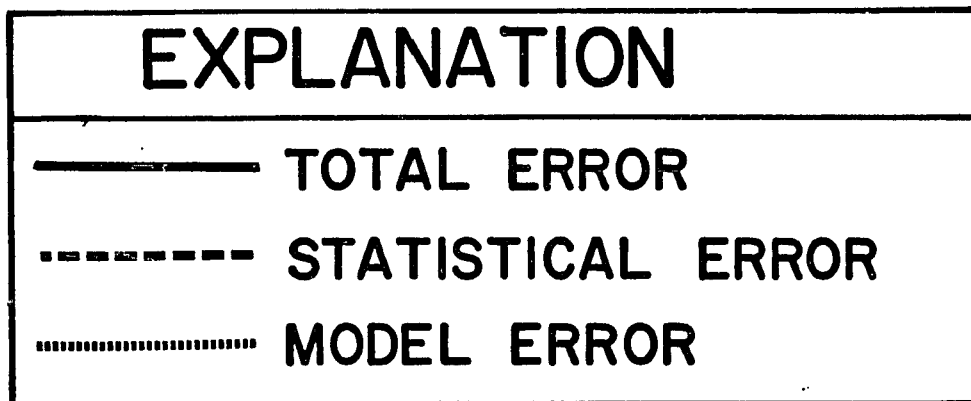
Figure 7.42. Trade-off curves for station correction estimate for station JRW. Both axes are logarithmic. The horizontal axis is tabulated in terms of both the nondimensional number B and the roughness length κ. B is useful for comparison of this figure to the trade-off curves for the velocity model in figures 7.36 and 7.37. κ is an easier number to interpret physically (see text). B and κ are related by equation (5.52) and for this figure the number Δu was set to a conservative value of 0.1 sec/km. A smaller, less conservative estimate of Δu would shift the κ scale to the left as can be seen from equation (5.52).

B

ERROR (sec)

ROUGHNESS LENGTH (km)
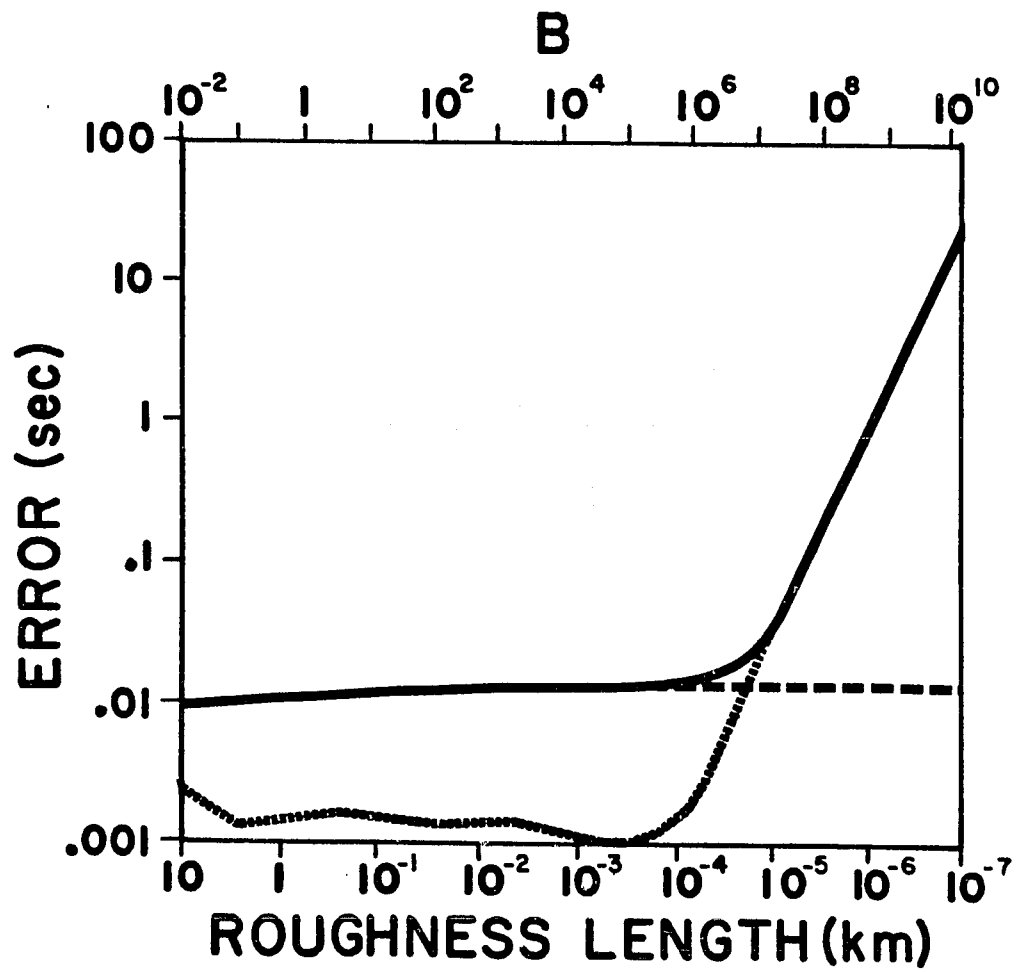
STATION: JRW

station correction.

The above conclusions were stated only for station JRW. These conclusions, however, extend directly to the other station corrections estimated from these data because the trade-off curves for all of these estimates are nearly indistinguishable. The only major difference is in the size of the intrinsic statistical error in the station corrections for the temporary and permanent stations. Standard errors for the permanent stations were all of the order of 0.01 seconds while those for the temporary stations were all of the order of 0.04 seconds (table 7.3). The reason for this is that station corrections for the temporary stations were based only on arrivals from the three profile shots while the station corrections for the permanent stations were based on arrivals from both shots and earthquakes. The station corrections for the temporary stations were thus obtained as averages of fewer numbers, so they should be expected to have a greater uncertainty. The final station corrections for all the Coso stations are

3.5.4. Hypocenters

I have assessed the reliability of the hypocenters of the 59 earthquakes used in this study by the same techniques I used above for the station corrections. It is impractical to present all these results since it would require 236 figures like 7.42 to do so. Fortunately, it is not at all necessary to do this anyway because all 59 events can be placed in two categories based on the form of their trade-off curves. A typical example of each category is shown in figures 7.43 and 7.44. The trade-off curves for the hypocenter of the earthquake (event number 1 (see appendix C)) shown in figure 7.43 are typical of the first category. The curves for the epicentral coordinates for this event look similar to figure 7.42. There is a difference, however, in the form of these curves for the depth and

Table 7.3
Coso Network Station Corrections

Station	Station Correction (sec)	Standard Error (sec)
NMC	0.01	0.008
MFS	-0.09	0.008
JRW	-0.03	0.009
SMW	-0.04	0.008
DKN	-0.03	0.009
RVC	0.00	0.008
CPT	-0.02	0.010
DTE	0.09	0.025
HPH	0.16	0.011
CGS	-0.03	0.010
RCW	0.10	0.010
CSS	-0.04	0.010
VPE	-0.03	0.009
CBH	0.19	0.011
HWS	-0.04	0.011
BCH	0.14	0.011
CFW	-0.14	0.10
C02	-0.04	0.051
C03	-0.08	0.037
C04	-0.14	0.037
C05	-0.02	0.037
C06	-0.07	0.037
C07	-0.09	0.037
C08	-0.10	0.047
C09	-0.09	0.035
C10A	0.61	0.035
C10B	0.16	0.035
C11	0.05	0.032
C12	0.00	0.032
C13	-0.06	0.033
C14	-0.06	0.034
C16	-0.06	0.036
C17	-0.03	0.044
C18	-0.07	0.048
C19	-0.04	0.048
C20	0.05	0.048

EXPLANATION

—————— TOTAL ERROR

-------- STATISTICAL ERROR

···················· MODEL ERROR

Figure 7.43. Trade-off curves of hypocenter coordinates for event number 1 (see Appendix C). Trade-off curves shown are all similar to that described in figure 7.42 except here the parameters of interest are the hypocenter coordinates of an earthquake. Coordinates given are local cartestian coordinates.
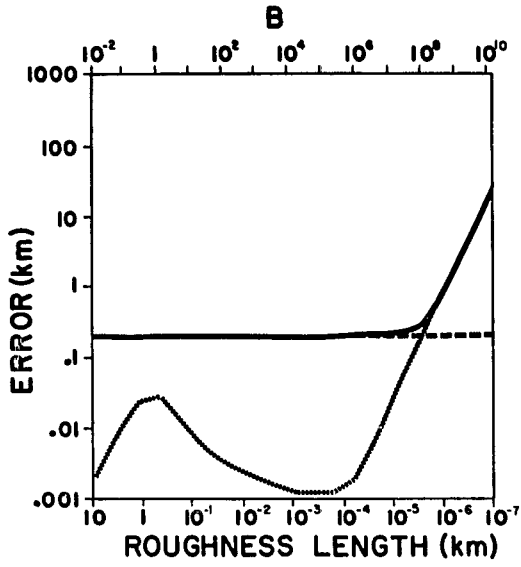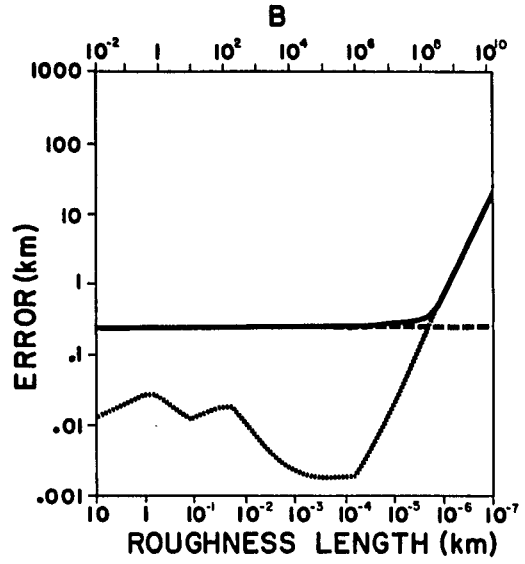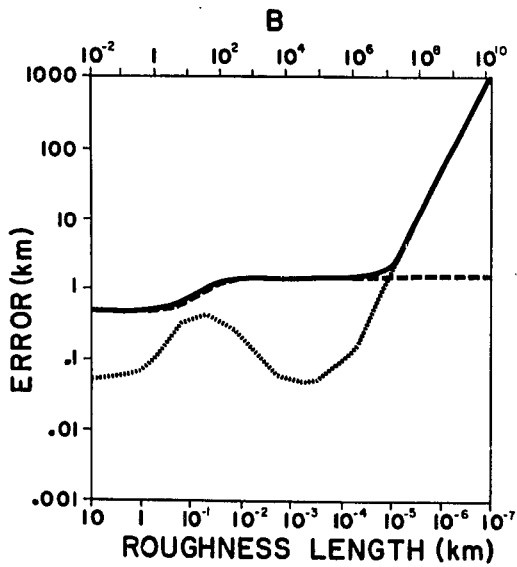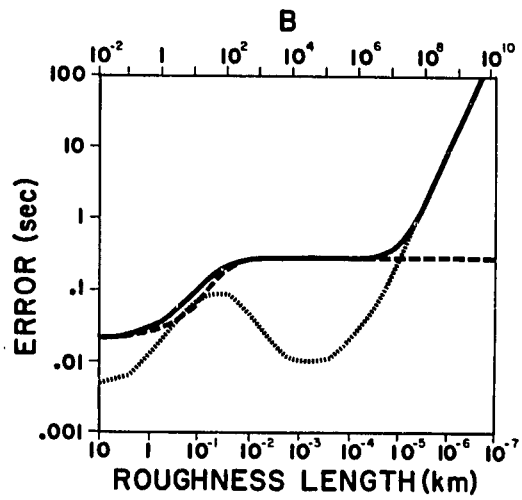
EXPLANATION

———— TOTAL ERROR

------ STATISTICAL ERROR

············ MODEL ERROR

Figure 7.44. Trade-off curves of hypocenter coordinates for event number 223 (see Appendix C). Trade-off curves shown are all similar to that described in figure 7.42 except here the parameters of interest are the hypocenter coordinates of an earthquake. Coordinates given are local cartestian coordinates.

B

ERROR (km)

ROUGHNESS LENGTH (km)

N-S

B

ERROR (km)

ROUGHNESS LENGTH (km)

E-W

B

ERROR (km)

ROUGHNESS LENGTH (km)

DEPTH

B

ERROR (sec)

ROUGHNESS LENGTH (km)
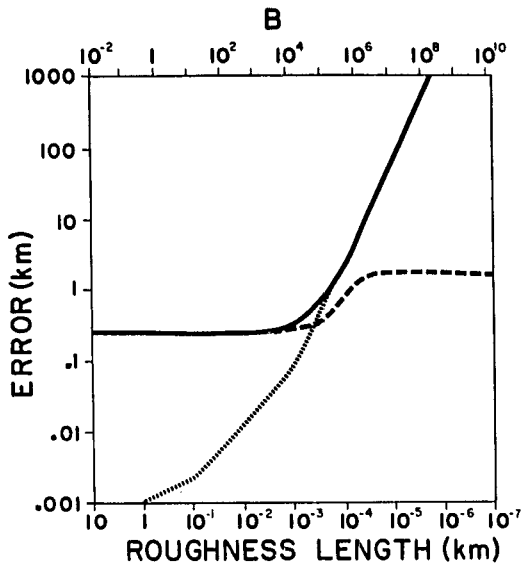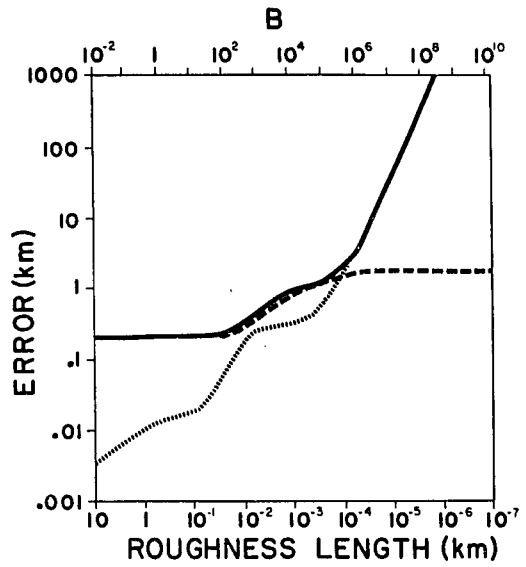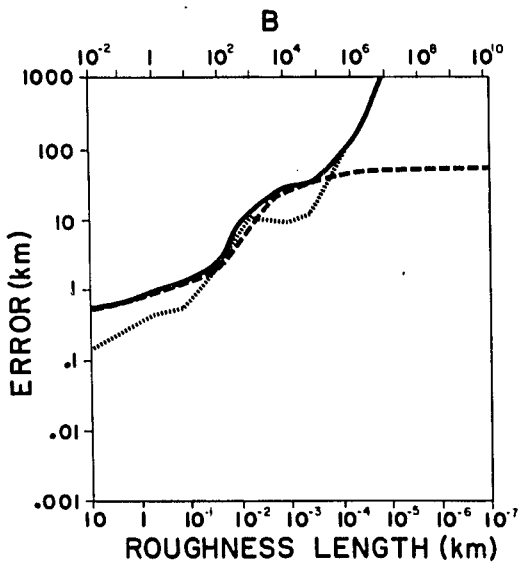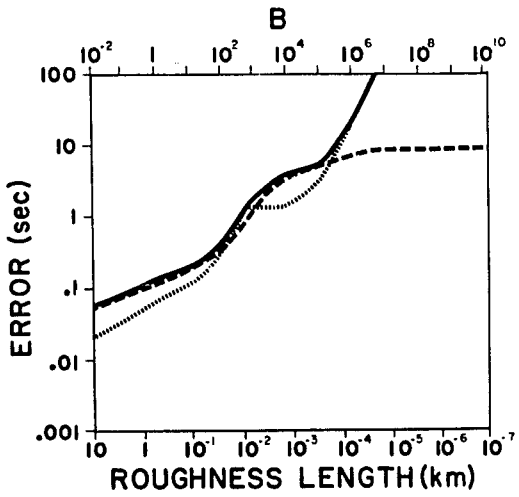
ORIGIN TIME

origin time of this event. Both show a general rise in the size of the statistical error at $\kappa \approx$ 1 km. Note that for the depth coordinate, however, the model error always remains below the intrinsic statistical error. The same is not true for the origin time of this earthquake, and it is clear that the origin time is the most poorly constrained coordinate of this hypocenter. The fact that the origin time is poorly known is of little consequence, however, since one rarely cares whether the origin time is specified to a high accuracy. As a result, the overall conclusion for events like this one is similar to that made from the same analysis applied to the station corrections estimates. That is, the velocity model estimated here is sufficiently good that model errors can be essentially neglected for the spatial coordinates of the hypocenter. Furthermore, for this event conventional hypocenter error calculations (standard errors or error ellipses) will give an adequate representation of the errors in these locations. These error estimates are most reliable for the epicenter coordinates and are less reliable for the depth and origin time.

The situation for the second category of earthquake locations is not as clear-cut. These hypocenter, which are typified here by event 63 (appendix C) whose trade-off curves are displayed in figure 7.44, have a larger propensity for errors due to the nonuniqueness of the velocity model. This can be seen by comparing figure 7.44 and 7.43. Overall, figure 7.45 looks similar to figure 7.43. The most significant difference, however, is that for the hypocenter whose errors are displayed in figure 7.44, the crossover point where model errors start to dominate over statistical errors occurs at a much larger value of the roughness length, κ. In this case, model errors become significant in the epicenter coordinates for $\kappa \approx 1$ to 10 meters. These length scales are still relatively small, but are only slightly smaller than those that are conceivable for the real earth.

This suggests that the epicenter of this event, at least, is probably not unduely influenced by model errors. This epicenter is, however, more prone to such errors than its sibling shown in figure 7.43. Unfortunately, things are not so nice for the depth and origin time estimated for this earthquake. The trade-off curves for both coordinates are virtually identical, and model errors cannot be totally neglected throughout the entire range of roughness lengths plotted in figure 7.44. Consequently, the primary conclusion for this hypocenter location is the the epicenter for this event is reasonably well determined, but the same cannot be said for the depth and origin time of this event. Both of the latter may be strongly influenced my model errors, and a correct appraisal of the errors in these coordinates cannot neglect model errors.

The type of earthquakes that fall into the two categories typified by the two sample sets of trade-off curves in figure 7.43 and 7.44 show a definite pattern. All earthquakes whose locations have trade-off curves like figure 7.43 are relatively shallow, while those whose locations have trade-off curves like figure 7.44 are relatively deep. This is consistent with the error assessment results for the velocity model, because we saw that the velocity structure was well resolved only down to a depth slightly below 4 kilometers. Thus, it is only to be expected that sources located below this depth would have a greater tendency to contain measurement errors from unresolved velocity structure.

Table 7.4
Coso Shot Point Corrections

Shot Name	Shot Point Correction
SW	0.24
C	0.20
NE	0.96
Quarry	0.14

4. CONCLUSIONS

The inversion technique (PRIMEL) described in the previous chapters was successfully applied here to data from two different seismic networks. The first of these was from the local network near Humboldt Bay in northern California. Results I obtained were comparable to those obtained previously by Knapp [1976] and Smith and Knapp [1978] who analyzed the same data set by the least squares procedure of Crosson [1976a]. There was a fairly substantial difference, however, in the velocity model's estimated by the two procedures whose cause remains somewhat of an enigma. In any case, the key result is that neither Knapp's results or those that I found fit the data to within the precision of measurement. I concluded from this that no one-dimensional velocity model is capable of fitting these data because of lateral velocity variations that must exist in this region. In defense of Knapp and Smith's work, however, it should be pointed out that the principle objective of their work was to obtain good earthquake locations to study local seismicity, not to obtain a velocity model closest to the real earth. Within the constraints of present day hypocenter determination procedures with one dimensional models, their station corrections

and velocity model do appear to fulfill this objective.

Application of PRIMEL to a second set of data from a local network operated by the USGS in the Coso Range was more successful. This success was due primarily to the simple velocity structure that typifies this region. The Coso Range is dominated by the plutonic basement rocks of the Sierra Nevada batholith that have been covered to depths of from ten to a few hundred meters by volcanic and sedimentary rocks associated with the Plio-Pleistocene volcanism that produced the Coso volcanic field [Duffield et. al., 1980]. The results I have obtained here agree well with that interpretation for three reasons;

(1) The station corrections that I found correlate well with surface geology. This indicates they are associated with near surface velocity variations due to difference in the type of material that covers the basement rocks.

(2) The velocity model I obtained is simple. Velociy increases steadily from a surface velocity of 4.8 km/sec to a velocity of 6.0 km/sec at a depth of 4 kilometers (figure 7.20). This model is consistent with velocities characteristic of the plutonic basement rocks if one interprets the decrease in velocity near the surface as a result of opening of cracks in these rocks.

(3) The fit of this model to the observations is extremely good.

Application of the error assessment techniques described in chapter 5 yielded three significant conclusions;

(1) The velocity model found is well resolved only to a depth of slightly greater than 4.0 kilometers. This result was surprising since nearly half the earthquakes in this data set occurred below this depth. The phenomenon appears to be a geometric effect due to the small aperture of the Coso network.

(2) Station correction estimates are not substantially influenced by possible unresolved structure in the velocity model. Statistical error estimates for these parameters provide a realistic assessment of their uncertainties.

(3) Hypocenters for earthquakes that were shallow (depth less than about 6 kilometers) are not substantially influenced by possible unresolved velocity structure. Deeper events, by contrast, showed a greater propensity for errors due to inadequacies in the velocity model. This result is not surprising in light of the poor resolution of the deeper velocity structure. Because of this, the usual statistical measures of the errors in the locations of the shallower events provide a reasonable estimate of the total error in these hypocenters. The same methods may not, on the other hand, give correct results for the deeper events.

In summary the results from Coso represent a major success of this procedure and clearly demonstrate its usefulness in a real world situation.

CHAPTER 8
SUMMARY AND CONCLUSIONS

In this dissertation I have described theoretically and success-fully implemented, a new method of exploiting arrival time data from earthquake and explosion sources. This procedure estimates earthquake hypocenters and a velocity model which fit the observed data. The theoretical foundation is laid to permit velocity models that vary arbitrarily in three spatial dimensions. In the actual implementation, however, the velocity model is assumed to vary only in the vertical direction. As a result, to be of use with real data, it is necessary to include station corrections to account for near surface lateral velocity variations. This leads to a new multiple event location procedure (PMEL) that is a significant development in its own right.

PMEL is a significant advance over the method of joint hypo-center determination (JHD), which is its only major competition. The primary reason for this is that JHD requires iterative solution of a set of equations where the number of rows and columns of a matrix that must be inverted grows with each earthquake added to the data set. This can lead to extreme storage problems with large data sets and has been a severe limitation on the applicability of the JHD method. PMEL, on the other hand, operates within a small, fixed storage area that is attainable even on a micro-computer. Consequently, PMEL has the ability to digest essentially unlimited data sets. This fact makes conventional multiple event locations obsolete. Furthermore, with further development it may render the same fate to customary single event location methods.

PMEL's is one component of a larger scale solution construction algorithm I have dubbed PRIMEL (Progressive Inversion and Multiple Event Location). This procedure is "progressive" because different parts of the total model we want to estimate (hypocenters, station corrections, and the velocity model) are determined in separate stages. This is in contrast to the "simultaneous" method of Peters and Crosson [Peters and Crosson, 1972; Peters, 1973; and Crosson, 1976a] where all the components of the model are adjusted at the same time. The progressive approach advocated here is superior to a simultaneous procedure for four reasons.

(1) Hypocenters, station corrections, and the velocity model are three quite different entities. The progressive approach has the advantage that it allows one greater freedom in how the different parts of the model are treated.

(2) The progressive approach has significant advantages from the practical standpoint of its implementation into a computational procedure. The simultaneous method, as formulated by Peters and Crosson, has the unfortunate property of requiring the solution of a matrix equation that grows in the same way as the matrix used in JHD. As a result, in that formulation one is virtually forced to specify the velocity model in terms of a small number of discrete parameters because of limitations in the amount of storage available on typical computing machinery. The annulling transformation releases us from the prison of parameterized velocity models by removing the walls raised by finite computer storage. PRIMEL, in a sense, also parameterizes the velocity model, since it specifies the velocity model on a discrete grid of points. The point, however, is that the storage reduction produced by the annulling transformation allow one to grid the model on a scale that is considerably smaller than the data are capable of resolving. This means the

velocity model can be considered as an arbitrary function for all intents and purposes.

(3) Because PRIMEL specifies the velocity model as an arbitrary function, the uniqueness of the velocity model can be appraised (within the limits of linearization anyway) by the methods of Backus and Gilbert [1968, 1970] and Johnson and Gilbert [1972].

(4) If we recognize that the velocity model is fundamentally ambiguous, then we must also admit that the hypocenters and station corrections we estimate are influenced by errors we may have committed in estimating the velocity model. Because PRIMEL allows the velocity model to be an arbitrary function, the importance of this source of error can be appraised through an application of Backus' [1970a,b] method of generalized prediction. That analysis is significant because this study is the first to ever quantify such errors in hypocenter and station correction estimates.

Finally, the work presented here is not just some grandiose theory. The methods describe above are a practical procedure that works. This was clearly demonstrated by its successful application to inversion of both synthetic and real data.

LIST OF REFERENCES

1. AKI, K. AND W. H. K. LEE , "Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes, 1, A homogeneous initial model," *J. Geophys. Res.* **81**, pp.4381-4399 (1976).

2. AKI, K. AND P. RICHARDS, *Quantitative Seismology: Theory and Methods*, Freeman, San Francisco, California (1980).

3. ANDERSON, K. A., *Automatic Processing of Local Earthquake Data*, Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts (1978).

4. BACKUS, G. E. AND F. GILBERT, "Numerical application of a formalism for geophysical inverse problems," *Geophys. J. R. Astr. Soc.* **13**, pp.247-276 (1967).

5. BACKUS, G. E. AND F. GILBERT, "The resolving power of gross earth data," *Geophys. J. R. Astr. Soc.* **16**, pp.169-205 (1968).

6. BACKUS, G. E. AND F. GILBERT, "Constructing P-velocity models to fit restricted sets of travel-time data," *Bull. Seism. Soc. Amer.* **59**, pp.1407-1414 (1969).

7. BACKUS, G. E. AND F. GILBERT, "Uniqueness in the inversion of inaccurate gross earth data," *Phil. Trans. Roy. Soc. London, Ser. A* **266**, pp.123-192 (1970).

8. BACKUS, G. E., "Inference from inadequate and inaccurate data, I," *Proc. Nat. Acad. of Sciences* **65**, pp.1-7 (1970a).

9. BACKUS, G. E., "Inference from inadequate and inaccurate data, II," *Proc. Nat. Acad. of Sciences* **65**, pp.281-287 (1970b).

10. BACKUS, G. E., "Inference from inadequate and inaccurate data, III," *Proc. Nat. Acad. of Sciences* **67**, pp.282-289 (1971).

11. BACON, C. R. AND W. A. DUFFIELD, "Coso geothermal area," *J. Geophys. Res.* **85**, p.2379 (1980).

12. BAILEY, E. H., W. P. IRWIN, AND D. L. JONES, "Franciscan and related rocks, and their significance in the geology of western California," *Bulletin 183, Calif. Div. of Mines and Geology* (1964).

13. BEN-ISRAEL, A. AND T. N. E. GREVILLE, *Generalized Inverses: Theory and Applications,* Wiley, New York (1974).

14. BOLT, B. A., "The revision of earthquake epicenter, focal depths, and origin times using a high-speed computer," *Geophys. J. R. Astr. Soc.* **3**, pp.433-440 (1960).

15. BOLT, B. A., "Earthquake location for small networks using the generalized inverse matrix," *Bull. Seism. Soc. Am.* **60**, pp.1823-1828 (1970).

16. BULAND, R., "The mechanics of locating earthquakes," *Bull. Seism. Soc. Am.* **66**, pp.173-187 (1976).

17. BULLEN, K. E., *An Introduction to the Theory of Seismology,* Cambridge Univ. Press, Cambridge, Great Britain (1965).

18. BUSINGER, P. A. AND G. H. GOLUB, "Linear least squares solutions by Householder transformations," *Numer. Math.* **7**, pp.269-276 (1965).

19. CHOU, C. W. AND J. R. BOOKER, "A Backus-Gilbert approach to inversion of travel-time data for three-dimensional velocity structure," *Geophys. J. R. Astr. Soc.* **59**, pp.325-344 (1979).

20. CLAERBOUT, J. F., *Fundamentals of Data Processing with Application to Petroleum Prospecting,* McGraw-Hill, New York (1976).

21. CROSSON, R. S., "Crustal structure modeling of earthquake data, 1, Simultaneous least squares estimation of hypocenter and velocity parameters," *J. Geophys. Res.* **81**, pp.3036-3046 (1976a).

22. CROSSON, R. S., "Crustal structure modeling of earthquake data, 2, velocity structure of the Puget Sound region, Washington," *J. Geophys. Res.* **81**, pp.3047-3054 (1976b).

23. DEWEY, J. W., "Seismicity and tectonics of western Venezuela," *Bull. Seism. Soc. Am.* **62**, pp.1711-1751 (1972).

24. DOBRIN, M. B., *Introduction to Geophysical Prospecting*, McGraw-Hill, New York (1976). 3rd Edition,

25. DONGARRA, J. J., C. B. MOLER, J. R. BUNCH, AND G. W. STEWART, *LINPACK User's Guide*, Society of Industrial and Applied Mathematics, Philadelphia, Pennsylvania (1979).

26. DOUGLAS, A., "Joint epicentre determination," *Nature* **215**, pp.47-48 (1967).

27. DUFFIELD, W. A. AND C. R. BACON, "Geologic map of the Coso volcanic field and adjacent areas, Inyo County, California," Misc. Geol. Invest. Map I-1200, U.S. Geol. Surv., Menlo Park, California (1980). Scale 1:50,000

28. DUFFIELD, W. A., C. R. BACON, AND G. B. DALRYMPLE, "Late Cenozoic volcanism, geochronology, and structure of the Coso Range, Inyo County, California," *J. Geophys. Res.* **85**, pp.2381-2404 (1980).

29. EATON, J. P., "Crustal structure in northern and central California from seismic evidence, geology of Northern California," *Bull. 190, Calif. Div. Mines and Geol.* , pp.419-426 (1966).

30. ENGDAHL, E. R. AND W. H. K. LEE , "Relocation of local earthquakes by seismic ray tracing," *J. Geophys. Res.* **81**, pp.4400-4406 (1976).

31. EVERDEN, J. F., "Identification of earthquakes and explosion by use of teleseismic data," *Bull. Seism. Soc. Am.* **59**, pp.1365-1398 (1969).

32. FLINN, E. A., "Local earthquake location with an electronic computer," *Bull. Seism. Soc. Am.* **50**, pp.467-470 (1960).

33. FRANCIS, J. G., "The QR transformation I, II," *Computer J.* 4, pp.265-271, 332-345 (1961, 1962).

34. GEIGER, L., "Herdbestimmung bei Erdbeden aus den Ankunfzeitzen," *K. Gessell. Wiss. Goett.* 4, pp.331-349 (1910).

35. GILBERT, F., "Ranking and winnowing gross earth data for inversion and resolution," *Geophys. J. R. Astr. Soc.* 23, pp.125-128 (1971).

36. GOLUB, G. H. AND R. J. PLEMMONS, "Large-scale geodetic least-squares adjustment by dissection and orthogoan decomposition," in *Large Scale Matrix Problems*, ed. H. Schneider, Elsevier, New York (1981).

37. HERRIN, E., J. TAGGART, AND C. F. BROWN JR., "Machine computation of earthquake hypocenters," *J. Grad. Res. Ctr, Sourthern Methodist Univ.* 30, pp.79-106 (1962).

38. HOEL, P. G., *Introduction to Mathematical Statistics*, Wiley, New York (1971).

39. JACKSON, D. D., "Interpretation of inaccurate, insufficient, and inconsistent data," *Geohys. J. R. Astr. Soc.* 28, pp.97-109 (1972).

40. JACKSON, D. D., "The use of a priori data to resolve non-uniqueness in linear inversion," *Geophys. J. R. Astr. Soc.* 57, pp.137-157 (1979).

41. JOHNSON, L. E. AND F. GILBERT, "Inversion and inference for teleseismic ray data," *Methods in Computational Physics* 12, pp.231-266, Academic Press, New York (1972).

42. JORDAN, T. H., "Structural geology of the Earth's interior," *Proc. Nat. Acad. Sci.* 76, pp.4192-4200 (1979).

43. JORDAN, T. H., "Earth structure from seismological observations," pp. 1-40 in *Physics of the Earth's Interior: Proceedings of the International School of Physics "Enrico Fermi", Course LXXVIII*, ed. E. Boschi, North Holland, Amsterdam, Netherlands (1980).

44. JORDAN, T. H. AND K. A. SVERDRUP, "Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific," *Bull. Seism. Soc. Amer.* **71**, pp.1105-1130 (1981).

45. KLEIN, F. W., "Hypocenter location program -- HYPOINVERSE: Users guide to versions 1, 2, 3, and 4.," *U.S. Geol. Surv. Open-file Rept. 78-694* (1978).

46. KNAPP, J. S., *Velocity changes associated with the Ferndale earthquake*, M.S. Thesis, University of Washington, Seattle, Washington (1976).

47. KNAPP, J. S. AND S. W. SMITH, *Seismic velocity structure of the Humboldt Bay Region, California*, Unpublished final report to: Pacific Gas & Electric Co., San Francisco, California (1979).

48. KNAPP, J. S., *Seismicity, crustal structure and tectonics at the northern termination of the San Andreas Fault*, Ph.D. Dissertation, University of Washington, Seattle, Washington (1982).

49. LANCZOS, C., *Linear Differential Operators*, Van Nostrand, New York (1961). Chapter 3,

50. LAWSON, C. H. AND R. J. HANSON, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, New Jersey (1974).

51. LEE, W.H.K. AND J. C. LAHR, "HYPO71, A computer program for determining hypocenter, magnitude, and first motion pattern of local earthquakes," U.S. Geol. Survey Open-File Report (1972).

52. LEE, W.H.K. AND S. W. STEWART, "Principles and applications of microearthquake networks," in *Advances in Geophysics, Vol. 23*, Academic Press, New York (1981).

53. LINDSETH, R. O., "Synthetic sonic logs -- a process for stratigraphic interpretation," *Geophysics* **44**, pp.3-26 (1979).

54. MOORE, E. H., "On the reciprocal of the general algebraic matrix," *Bulletin, AMS* **26**, pp.394-395 (1920).

55. NORDQUIST, J. M., "A special-purpose program for earthquake location with an electronic computer," *Bull. Seism. Soc. Am.* **52**, pp.431-437 (1962).

56. OGLE, B. A., "Geology of the Eel River Valley Area, Humboldt County, California," *Bulletin 164, Calif. Div. of Mines and Geology* (1953).

57. OLDENBURG, D. W., "One-dimensional inversion of natural source magnetotelluric observations," *Geophysics* **44**, pp.1218-1244 (1979).

58. PARKER, R. L., "Understanding inverse theory," *Ann. Rev. Earth Planet. Sci.* **5**, pp.35-64 (1977a).

59. PARKER, R. L., "Linear inference and underparameterized models," *Rev. of Geophys. and Space Physics* **15**, pp.446-456 (1977b).

60. PAVLIS, G. L. AND J. R. BOOKER, "The mixed discrete continuous inverse problem: application to the simultaneous determination of earthquake hypocenters and velocity structure," *J. of Geophys. Res.* **85**, pp.4801-4810 (1980).

61. PENROSE, R., "A generalized inverse for matrices," *Proc. Cambridge Phil. Soc.* **51**, pp.403-413 (1955).

62. PEREYRA, V., W.H.K. LEE, AND H. B. KELLER, "Solving two-point seismic-ray tracing problems in a heterogeneous medium. Part I. A general adaptive finite difference method," *Bull. Seis. Soc. Am.* **70**, pp.79-101 (1980).

63. PETERS, D. C. AND R. S. CROSSON, "Application of prediction analysis to hypocenter determination using a local array," *Bull. Seis. Soc. Amer, Abst.* **62**, p.775 (1972).

64. PETERS, D. C., *Hypocenter location and crustal structure inversion of seismic array travel-times*, Ph.D. Dissertation, University of Washington, Seattle, Washington (1973).

65. RODI, W. L., T. H. JORDAN, J. F. MASSO, AND J. M.SAVINO, "Determination of the three-dimensional structure of eastern Washington from the joint inversion of gravity and earthquake travel-time data," Report SSS-R-80-4516, Systems, Science and Software, La Jolla (1980).

66. ROECKER, S. W., "Velocity structure of the Pamir-Hindu Kush region: possible evidence of subducted crust," *J. Geophys. Res.* **87**, pp.945-959 (1982).

67. ROGERS, D. B. AND R. S. CROSSON, *Hypocenter locations for seismic arrays via GENSURF and LQUAKE2*, University of Washington, Seattle, Washington (1976). (unpublished computer software manual)

68. ROHAY, A. C., *Crust and upper mantle structure in the North Cascades Region, Washington*, Ph.D. Dissertation, University of Washington, Seattle, Washington (1982).

69. SLOTNICK, M. M., "Lesson in Seismic Computing," in *Soc. of Exploration Geophysics*, ed. Richard A. Geyer, Tulsa, Oklahoma (1959).

70. SMITH, S. W. AND J. S. KNAPP, *Seismic velocity structure, Humboldt Bay, California*, Unpublished report to: Pacific Gas & Electric Co., San Francisco, California (1978).

71. SPENCER, C. AND D. GUBBINS, "Travel-time inversion for simultaneous earthquake location and velocity structure determination in laterally varying media," *Geophys. J. R. Astr. Soc.* **63**, pp.95-116 (1980).

72. SPIETH, M. A., *Two detailed seismic studies in central California; Part I: Earthquake clustering and crustal structure studies of the San Andreas fault near San Juan Bautista, Part II: Seismic velocity structure along the Sierra foothills near Oroville, California*, Ph.D. Dissertation, Stanford University, Stanford, California (1981).

73. STEWART, G. W., *Introduction to matrix computations*, Academic Press, New York (1973).

74. THURBER, C. H. AND W. L. ELLSWORTH, "Rapid solution of ray tracing problems in heterogeneous media," *Bull. Seism. Soc. Amer.* **70**, pp.79-90 (1980).

75. THURBER, C. H., *Earth structure and earthquake locations in the Coyote Lake area, central California*, Ph.D. Dissertation, Massachusetts Institute of Technology (1981).

76. WALTER, A. W. AND C. S. WEAVER, "Seismicity of the Coso Range, California," *J. Geophys. Res.* **85**, pp.2441-2458 (1980a).

77. WALTER, A. W. AND C. S. WEAVER, "Seismic refraction data for shots recorded in the Coso Range, California, October 1976," U.S. Geol. Surv. Open-File Rept. 80-186 (1980b).

78. WIGGINS, R. A., "The general linear inverse problem: implication of surface waves and free oscillations for earth structure," *Rev. of Geophys. and Space Physics* **10**, pp.251-285 (1972).

79. WOLBERG, J. R., *Prediction Analysis*, Van Nostrand, Princeton, New Jersey (1967).

80. WU, J. C., *Inversion of travel-time data for seismic velocity structure in three dimensions*, Ph.D. Dissertation, University of Washington, Seattle, Washington (1977).

81. YELIN, T. S. AND R. S. CROSSON, "A note on the south Puget Sound basin magnitude 4.6 earthquake of March 11, 1978 and its aftershocks," *Bull. Seism. Soc. Am.*, (in press) (1982).

APPENDIX A
WEIGHTED LEAST SQUARES

It was noted in chapter 3 that all modern hypocenter location algorithms are actually weighted least squares procedures. I will show here that weighted least squares is the same as normal least squares after a simple change of variables.

We wish to consider the problem of finding a weighted least squares solution to the general problem

$$\mathbf{Ax=b} \tag{A.1}$$

where $\mathbf{x} \in R^n$, $\mathbf{b} \in R^m$, and $\mathbf{A} \in R^{m \times n}$. A weighted least squares solution of (A.1) is defined as follows

$$\mathbf{x}_w = \min_{\mathbf{x} \in R^n} \left[\|\mathbf{b} - \mathbf{Ax}\|_w \right] \tag{A.2}$$

where the norm here is the **ellipsoidal** or **weighted Euclidean** norm defined as

$$\|\mathbf{y}\| = \mathbf{y}^T \mathbf{D} \mathbf{y} \tag{A.3}$$

and $\mathbf{D}$ is some positive definite matrix[1]. We require the following lemma [Ben-Israel and Greville, 1974, p. 124]

Lemma A.1

Every positive definite matrix, $\mathbf{D} \in R^{m \times m}$, has a unique, positive definite **square root** defined as $\mathbf{D}^{1/2} = \mathbf{W}$ where $\mathbf{D} = \mathbf{W}^T \mathbf{W}$.

$\mathbf{W}$ is useful because if we multiple (A.1) by $\mathbf{W}$ we obtain

$$\mathbf{A}_w \mathbf{x} = \mathbf{b}_w \tag{A.4}$$

where I define

$$\mathbf{A}_w = \mathbf{WA} \quad \text{and} \quad \mathbf{b}_w = \mathbf{Wb} \tag{A.5}$$

---

[1] For the hypocenter location problem $\mathbf{D}$ is always a diagonal matrix with $D_{ii} = W_{ii}^2$ where $\mathbf{W}$ is defined by equation (3.10).

This shows that the transformation (A.5) defines a simple change of variables that produces a set of equations, (A.4), whose form is identical to equation (A.1). A least squares solution of (A.4) solves the minimization problem (A.2) [Ben-Israel and Greville, 1974, pp. 121-123]. The practical consequence of this is that the more general minimization problem involving ellipsoidal norms can always be made equivalent to a simple least squares problem by the change of variables defined by (A.5). Hence, after this change of variables, weighted least squares problems can be treated exactly like simple least squares, since

$$\|\mathbf{b}-\mathbf{Ax}\|_w = \|\mathbf{b}_w - \mathbf{A}_w \mathbf{x}\|_2^2 \qquad (A.6)$$

An important application of weighted least squares is the case when $\mathbf{D}$ is calculated as

$$\mathbf{D}=\mathbf{C}^{-1} \qquad (A.7)$$

$\mathbf{C}^{-1}$ is the inverse of the covariance matrix of the data where the covariance matrix, $\mathbf{C}$, is defined as

$$\mathbf{C}=E[\mathbf{b}\mathbf{b}^T] \qquad (A.8)$$

and E[] denotes the expectation value operator. This is a useful choice for the weighting matrix because then the covariance of the weighted data, $\mathbf{b}_w$, is given simply by

$$\mathbf{C}_w = E[\mathbf{b}_w \mathbf{b}_w^T] \qquad (A.9)$$

$$= E[\mathbf{C}^{-1/2}\mathbf{b}\mathbf{b}^T(\mathbf{C}^{-1/2})^T]$$

$$= \mathbf{C}^{-1/2}\mathbf{C}(\mathbf{C}^{-1/2})^T$$

$$= \mathbf{I}$$

where $\mathbf{C}^{-1/2}$ denotes the square root of $\mathbf{C}^{-1}$ as in Lemma A.1.

APPENDIX B
THE MIXED LINEAR INVERSE PROBLEM

1. PROBLEM STATEMENT

In a previous paper Pavlis and Booker [1980] discussed a method of solution of a general type of linear inverse problem that they called mixed. These are problems for which the earth model from which an observed set of data are derived have a natural specification in terms of some piecewise continuous function and a finite set of parameters. This appendix is essentially part 1 of that paper. I have repeated it here for completeness and to avoid any potential confusion that might arise from differences in notation necessitated by differences in the set of possible symbols available on the typesetter used to prepare this dissertation. In addition, the results given here are also more general than those given by Pavlis and Booker [1981]. Their work centered around the Moore-Penrose inverse. That restriction is relaxed here to include a larger class of generalized inverse solutions.

The problem I wish to consider here is this. Suppose one has a set of d measured data, b_i. Assume the data are related by known linear functionals to an unknown earth model that consists of some piecewise continuous function and p independent variables that I will call parameters. I denote the continuous part of the model by $|M>$ and the parametric part of the model as the p dimensional vector, $\mathbf{x}$. The inverse problem is to estimate both $|M>$ and $\mathbf{x}$, and to assess the relative reliability of these estimates.

Given the above definitions the functional relation of each datum to the model can be written as

$$b_i = \sum_{j=1}^{p} A_{ij} x_j + <G_i | M> \qquad i=1,2,\ldots,d \qquad (B.1)$$

where $A_{ij} = \dfrac{\partial b_i}{\partial x_j}$ and $<G_i|$ is a Frechet kernel. $<G_i|M>$ is an inner product that is normally expressible in the form

$$<G_i|M> = \int_V G_i(\mathbf{r})M(\mathbf{r})\,dV \qquad (B.2)$$

where V is the domain on which $|M>$ is defined. It is desirable to write the d equations (B.1) in matrix form as

$$\mathbf{b} = \mathbf{Ax} + \overset{*}{\mathbf{G}}|M> \qquad (B.3)$$

where $\mathbf{b}$ is the data vector, $\mathbf{A}$ is the $d \times p$ matrix of partial derivatives with elements defined in equation (B.1), and

$$\overset{*}{\mathbf{G}} = \begin{bmatrix} <G_1| \\ <G_2| \\ <G_3| \\ \cdot \\ \cdot \\ \cdot \\ <G_d| \end{bmatrix} \qquad (B.4)$$

$\overset{*}{\mathbf{G}}$ is a mathematical object like equation (4.10).

2. THE OVERDETERMINED PROBLEM

The unknown earth model has infinitely many unknowns. The number of data, d, is, on the other hand, always finite. As a result, the system of equations (B.3) is grossly undetermined and will always have infinitely many solutions. However, whenever $d > p$, the equations of (B.3) contains the paradox that although the system as a whole is grossly underdetermined, the matrix $\mathbf{A}$, when considered in isolation, is an overdetermined matrix. Before considering the solution of the problem as a whole, it is useful to consider some properties of overdetermined systems of equations. In particular, it is of interest to see what happens if we solve (B.3) ignoring the terms

involving $|M>$. In other words, we seek solutions of the equation

$$b = Ax \qquad (B.5)$$

When $d > p$, it is generally impossible to satisfy all d equations of (B.5) exactly. Indeed, we should not expect to be able to satisfy all d equations because (B.5) ignores the $<G_i|M>$ term and statistical errors in the data. For any estimate $\hat{x}$ of x we expect to have a nonzero residual vector, r, given by

$$r = b - A\hat{x} \qquad (B.6)$$

Normally, we can never make the residual vector vanish and we must be content with some approximate solution. All procedures I am aware of for solving overdetermined systems of linear equations calculate $\hat{x}$ using a generalized inverse. Any solution of this type can be expressed as

$$\hat{x} = Hb \qquad (B.6)$$

where $H \in R^{p \times d}$ is a **generalized inverse**. The set of all possible generalized inverses is obviously infinite since in a sense H can by any $p \times d$ matrix. The only restriction on H that is required here is a restriction on its range and null space. The restriction required is this. If $R(A)$ denotes the range of the matrix A in equation (B.3) (see chapter 2) then I require

$$R(H^T) \subset R(A) \qquad (B.7)$$

This means that the subspace of the data space, R^d, spanned by the rows of the generalized inverse matrix, H, is a subset of the subspace $R(A)$ of R^d. Since $R(A)$ is defined by the set of all possible linear combinations of the columns of the matrix A, this is equivalent to saying that the rows of H can all be expressed as linear combinations of the columns of the matrix A. Hence when

(B.7) is satisfied $\mathbf{H}$ can be written as

$$\mathbf{H=KA}^T \tag{B.8}$$

where $\mathbf{K} \in R^{p \times p}$ is arbitrary[1].

The representation (B.8) is inconvenient for the purpose of this work because the columns of $\mathbf{A}$ are not always the most convenient set of basis vectors for $R(\mathbf{A})$. A more convenient representation can be had through the singular value decomposition introduced in chapter 2. The singular value decomposition theorem guarantees that we can write $\mathbf{A}$ as

$$\mathbf{A=U\Lambda V}^T \tag{B.9}$$

where $\mathbf{U} \in R^{d \times d}, \mathbf{\Lambda} \in R^{d \times p}$, and $\mathbf{V} \in R^{p \times p}$ are as in theorem 2.2. From (B.9) we can also write $\mathbf{A}^T$ as

$$\mathbf{A}^T = \mathbf{V\Lambda}^T \mathbf{U}^T \tag{B.10}$$

so (B.8) can be expressed as

$$\mathbf{H=FU} \tag{B.11}$$

where I define $\mathbf{F} \in R^{p \times d}$ as

$$\mathbf{F=KV\Lambda}^T \tag{B.12}$$

Observe that Λ is a diagonal matrix with $rank(\mathbf{A}) = r \leq p$, nonzero elements (see equation (2.21) and (2.22)). As a result only the first r columns of the matrix $\mathbf{F}$ (defined in (B.12)) are nonzero and it is natural to partition $\mathbf{F}$ as

$$\mathbf{F} = [\mathbf{F}_R \, 0] \tag{B.13}$$

---

[1] As an example, if $\mathbf{K} = (\mathbf{A}^T \mathbf{A})^{-1}$, (B.8) yields the inverse used to find a least squares solution to (B.5).

where $\mathbf{F}_R \in \mathbb{R}^{p \times r}$. This leads to a conformal partitioning of the matrix $\mathbf{U}^T$ that allows us to write (B.11) in the partitioned form[2]

$$\mathbf{H} = [\mathbf{F}_R \, 0] \begin{bmatrix} \mathbf{U}_R^T \\ \mathbf{U}_N^T \end{bmatrix} \qquad (\text{B.14})$$

where $\mathbf{U}_R^T \in \mathbb{R}^{r \times d}$ and $\mathbf{U}_N^T \in \mathbb{R}^{(d-r) \times d}$. An important observation to make from (B.14) is that $\mathbf{H}$ can be written simply as

$$\mathbf{H} = \mathbf{F}_R \mathbf{U}_R^T \qquad (\text{B.15})$$

because the rightmost $d-r$ columns of $\mathbf{F}$ are zero. Thus, $\mathbf{H}$ is constructed as the p linear combinations, $\mathbf{F}_R$, of the r vectors that are the rows of the matrix $\mathbf{U}_R^T$. This is important because it demonstrates a fundamental property of all generalized inverses that satisfy the requirement (B.7). This property can be seen by substituting (B.14) into (B.6) to produce

$$\hat{\mathbf{x}} = [\mathbf{F}_R \, 0] \begin{bmatrix} \mathbf{U}_R^T \\ \mathbf{U}_N^T \end{bmatrix} [\mathbf{b}] = \mathbf{F}_R \mathbf{U}_R^T \mathbf{b} \qquad (\text{B.16})$$

We saw in chapter 2 that orthogonal matrices reorient Euclidean space without distortion. I pointed out there that the matrix $\mathbf{U}^T$ produces a special reorientation of the vector space, $\mathbb{R}^d$, in which the data vector, $\mathbf{b}$, lies. $\mathbf{U}^T$ reorients the coordinate axes of $\mathbb{R}^d$ such that the first r components of the vector $\mathbf{U}^T \mathbf{b}$ (These are denoted $\mathbf{U}_R^T \mathbf{b}$ above.) lie in the subspace $R(\mathbf{A})$ of $\mathbb{R}^d$. The remaining $d-r$ components of $\mathbf{U}^T \mathbf{b}$ (These are denoted $\mathbf{U}_N^T \mathbf{b}$ above.) lie in the null space of $\mathbf{A}^T$, $N(\mathbf{A}^T)$. The point is that any generalized inverse that

---

[2] For the sake of continuity of notation, it should be noted that the partitioning of $\mathbf{U}$ in (B.14) is identical to that used in equation (2.21). However, I have dropped the subscripts on N here for simplicity since there is no need to distinguish them here.

satisfies property (B.7) uses only those components of the data vector, **b**, that lie in the range of the matrix **A**, $R(\mathbf{A})$. Thus, an estimate of **x**, $\hat{\mathbf{x}}$, calculated by (B.6) only utilizes r independent components of the data. The remaining $d-r$ components (those that lie in $N(\mathbf{A}^T)$) are not utilized in calculating $\hat{\mathbf{x}}$. This is the foundation of the separation procedure that I now describe.

3. SEPARATION OF DISCRETE AND CONTINUOUS COMPONENTS OF THE DATA

The estimate $\hat{\mathbf{x}}$ in equation (B.6) is a perfectly valid estimate of the parameter vector **x**. We could simply substitute $\hat{\mathbf{x}}$ into (B.3) and we would obtain

$$\mathbf{r}=\overset{\circ}{\mathbf{G}}|M>\qquad(B.17)$$

The residuals in (B.17) could then be used to estimate $|M>$. This is, however, a highly suspicious procedure since such a solution is equivalent to the prejudice that $\hat{\mathbf{x}}$ is a perfect estimate of **x**. Since $\hat{\mathbf{x}}$ was obtained by ignoring the term involving $|M>$, there is no reason to believe $\hat{\mathbf{x}}$ is any better than the infinity of other possible estimates one could make of **x**. Let us instead admit that $\hat{\mathbf{x}}$ has some unknown error $\delta\mathbf{x}$. I denote this by writing

$$\mathbf{x}=\hat{\mathbf{x}}+\delta\mathbf{x}\qquad(B.18)$$

where $\delta\mathbf{x}$ is an unknown error in the estimate $\hat{\mathbf{x}}$. By a simple substitution of (B.18) into (B.3) one gets

$$\mathbf{b}=\mathbf{A}\hat{\mathbf{x}}+\mathbf{A}\delta\mathbf{x}+\overset{\circ}{\mathbf{G}}|M>\qquad(B.19)$$

The fundamental step in the separation procedure I am advocating here is to apply the reorienting matrix $\mathbf{U}^T$ introduced in the previous section in equation (B.10). This reorientation can be instituted by multiplying (B.19) by $\mathbf{U}^T$ to yield[3]

---

[3] There is a well disguised limitation that is intrinsic to the use of orthogonal matrices. Orthogonal matrices are important because

$$\begin{bmatrix} \mathbf{U}_R^T \mathbf{b} \\ \mathbf{U}_N^T \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_R^T \mathbf{A} \\ 0 \end{bmatrix} [\hat{\mathbf{x}}] + \begin{bmatrix} \mathbf{U}_R^T \mathbf{A} \\ 0 \end{bmatrix} [\delta \mathbf{x}] + \begin{bmatrix} \mathbf{U}_R^T \mathbf{G} \\ \mathbf{U}_N^T \mathbf{G} \end{bmatrix} | M > \qquad (B.20)$$

where the partitioning of $\mathbf{U}^T$ is as defined in equation (B.14). Equation (B.20) is the fundamental result of this appendix. The partitioning I have used here shows that (B.20) consists of two quite different types of equations. Of these, the most essential for the purpose of this work are the $d-r$ equations of the lower partition of (B.20). The form of these equations follows directly from the manner in which I defined the partitioning of the matrix $\mathbf{U}^T$. The partition $\mathbf{U}_N^T$ was defined by the relation

$$\mathbf{U}_N^T \mathbf{A} = 0 \in \mathrm{R}^{(d-r) \times p} \qquad (B.21)$$

Because $\mathbf{U}_N^T$ annihilates $\mathbf{A}$, the lower partition of (B.20) represents a set of equations that depend **only** on the continuous part of the model, $|M>$. Because of their connection with the null space of $\mathbf{A}^T$, I will refer to these $d-r$ equations as the annulled data set. Similarly, I will refer to the $d-r$ numbers $\mathbf{U}_N^T \mathbf{b}$ as the annulled data.

The annulled data are fundamental to the work presented in this dissertation. The desire to exploit the special form of the annulled data set was the reason for the restriction (B.7) on the generalized inverse matrix, $\mathbf{H}$, used to calculate $\hat{\mathbf{x}}$. When $\hat{\mathbf{x}}$ is calculated using a generalized inverse satisfying (B.7) the annulled data

---

they reorient Euclidean space without distortion. As a practical matter this means that the use of orthogonal matrices here limits the application of this method to the class of weighted Euclidean norm problems (see Appendix A). A more general treatment for arbitrary normed vector spaces would require the use of metric projectors [Ben-Israel and Greville, 1974, pp. 128-137]. I have elected to not inflict this upon the reader as it would introduce unnecessary complications that would probably be of no practical value anyway.

are unbiased by $\hat{x}$ because we see from (B.16) that they are ignored in calculating $\hat{x}$. Thus, the annulled data are properties of the original data[4] and are not coupled to the estimate $\hat{x}$.

For convenience, I now define

$$n = U_N^T b \tag{B.22}$$

and

$$\hat{N} = U_N^T \hat{G} \tag{B.23}$$

The convenience is that the equations of the annulled data set can then be written simply as

$$n = \hat{N} | M > \tag{B.24}$$

The fundamental property of the annulled data is that they can be used in isolation to estimate $|M>$. There are a number of ways of estimating a function $|M>$ from a finite number of data of the form (B.24) (see Parker[1977a] for a clear review). In chapter 4 I describe the method I have applied to the hypocenter-velocity structure inverse problem that is the major topic of this dissertation. I emphasize here that other methods for constructing an estimate of $|M>$ from (B.24) are possible but this is a separate issue

---

[4] A note of caution is in order here. Although theoretically $r = rank\,(\mathbf{A})$ is well defined, in practice it is not necessarily so clear cut. As a practical matter the determination of the true rank is not trivial because of computational round-off errors. Consequently, r should normally be considered the pseudorank [Lawson and Hanson, 1974, pp. 77-78] that I discussed in chapter 2. The key to a proper application of the procedure outlined here is consistency. The generalized inverse $\mathbf{H}$ that I define here always has a well defined rank. As long as the annulled data set are defined consistently with $\mathbf{H}$ (i.e. r is determined by $\mathbf{H}$ not $\mathbf{A}$) then no problems should ensue.

from the separation procedure itself.

Given that $|M>$ can be estimated from the annulled data, the question that remains is how to refine our estimates of the discrete parameters, $\mathbf{x}$. The key to this is the r equations defined by the upper partition of (B.20). These can be manipulated to yield a direct relationship between $\mathbf{x}$ and $|M>$. That is the subject of the remainder of this section.

The agent to accomplish this task is the matrix $\mathbf{F}_R$ defined in equation (B.12) and (B.13). If the upper partition of (B.20) is multiplied by $\mathbf{F}_R$, the result is

$$\mathbf{Hb} = \mathbf{HA}\hat{\mathbf{x}} + \mathbf{HA}\delta\mathbf{x} + \mathbf{H}\overset{\circ}{\mathbf{G}}|M> \qquad (B.25)$$

since $\mathbf{H}$ is defined by (B.15). $\hat{\mathbf{x}}$ is related to the data vector by equation (B.6) so (B.25) can be written (after some rearrangement) as

$$\mathbf{HA}\delta\mathbf{x} = \mathbf{Hb} - \mathbf{HAHb} + \overset{\circ}{\mathbf{X}}|M> \qquad (B.26)$$

where $\overset{\circ}{\mathbf{X}}$ is defined as

$$\overset{\circ}{\mathbf{X}} = -\mathbf{H}\overset{\circ}{\mathbf{G}} \qquad (B.27)$$

or equivalently

$$<X_i| = -\sum_{j=1}^{d} H_{ij} <G_j|$$

(The minus sign is introduced in definition (B.27) for convenience to simplify the final result.) (B.26) is close to the form we seek. To see how close we are, consider the special case when $\mathbf{A}$ is full rank and $\mathbf{H}$ is the least squares inverse, $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$. In this special case we see $\mathbf{HA} = \mathbf{I}$ and (B.26) reduces to the simple form

$$\delta\mathbf{x} = \overset{\circ}{\mathbf{X}}|M> \qquad (B.28)$$

For this special case, we can now obtain the relationship we seek by

a simple substitution into the definition (B.18)

$$\mathbf{x} = \hat{\mathbf{x}} + \overset{\star}{\mathbf{X}} |M>$$ (B.29)

A similar relationship can be derived for the more general case when $\mathbf{H}$ is any arbitrary generalized inverse that fits the requirement (B.7). This can be had by rewritting the definition (B.18) as

$$\mathbf{HAx} = \mathbf{HA\hat{x}} + \mathbf{HA}\delta\mathbf{x}$$

and substituting for $\mathbf{HA}\delta\mathbf{x}$ using (B.26). This yields

$$\mathbf{HAx} = \mathbf{HA\hat{x}} + \mathbf{Hb} - \mathbf{HAHb} + \overset{\star}{\mathbf{X}} |M>$$

Because $\hat{\mathbf{x}} = \mathbf{Hb}$, this reduces to

$$\mathbf{Rx} = \hat{\mathbf{x}} + \overset{\star}{\mathbf{X}} |M>$$ (B.30)

where the matrix $\mathbf{R}$ is defined as

$$\mathbf{R} = \mathbf{HA}$$ (B.31)

and is usually called the resolution matrix [Wiggins, 1972].

Equation (B.29) and (B.30) are of the same general form. $\hat{\mathbf{x}}$ is an initial estimate of $\mathbf{x}$ obtained by ignoring $|M>$. $\overset{\star}{\mathbf{X}} |M>$ is the correction that must be made to $\hat{\mathbf{x}}$ because $\hat{\mathbf{x}}$ ignored $|M>$. Equation (B.29) and (B.31) are identical when $\mathbf{A}$ is full rank and $\mathbf{H}$ is the least squares inverse, for then $\mathbf{R}$ reduces to an identity matrix. Otherwise, $\mathbf{R}$ is only an approximate identity matrix. When this is the case, it is a symptom of an additional deficiency of the data. The data are inadequate to define a distinct relation between x and $|M>$ as is the case in (B.29), but instead are only able to define the set of r independent averages of the true $\mathbf{x}$ defined by the vector $\mathbf{Rx}$. When this deficiency arises, one should always calculate $\mathbf{R}$. It allows a means of objectively assessing this added deficiency because we can interpret $\mathbf{R}$ as an additional resolution operator through which

we must view our final estimate of $\mathbf{x}$. Thus, the best we can ever hope to estimate from the data is the set of averages $\mathbf{Rx}$. I define the vector $\mathbf{x}_R$ as

$$\mathbf{x}_R = \mathbf{Rx} \qquad\qquad (B.32)$$

with the understanding that $\mathbf{x}_R = \mathbf{x}$ in the special case when $\mathbf{A}$ is full rank and $\mathbf{H} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$.

The separation procedure that is the topic of this appendix is now completed. The original set of equations, (B.3), have been split into two different types of equations. The $d-p$ equations of the annulled data set, (B.24), depend only on $|M>$. The p equations of (B.24), on the other hand, relate $\mathbf{x}_R$ directly to $|M>$. We see that the basic result of this splitting procedure is a change in the unknown part of the model. That is, in the original system of equations the unknown quantities are the p numbers, $\mathbf{x}$, and the unknown function $|M>$. However, in the separated problem $|M>$ is the only unknown because if we knew $|M>$ exactly, we would know $\mathbf{x}_R$ exactly by equation (B.31). This reduction in the number of unknowns was not a gift. The price we paid was the removal of the r components of the data $\mathbf{U}_R^T\mathbf{b}$ in equation (B.16) from the original data. However, by constructing the annulled data set we can recover the remaining $d-r$ pieces of independent information that remain in the data. A fundamental fact is that the annulled data are also independent of $\hat{\mathbf{x}}$ because they are orthogonal to the components of the data used by the estimate $\hat{\mathbf{x}}$. Consequently, the annulled data can be used in isolation to estimate $|M>$. Unfortunately, the opposite is not true. That is, the estimates of the parameters remain forever coupled to $|M>$ as is clear from (B.30). A proper scheme for estimating $\mathbf{x}$ must account for this, as errors in the estimate of $|M>$ can clearly propagate into estimates of $\mathbf{x}$. How one chooses to do this is again somewhat a matter of choice. To emphasize this I will not consider

this question here, but refer the reader to chapter 5 where I apply a particular method similar to that given by Backus [1970a]. I stress the point, however, that other schemes are feasible and are a separate issue from the results given here.

4. COVARIANCE OF THE SEPARATED DATA

The two fundamental results from the previous section were equations (B.24) and (B.30). They are the culmination of the manipulations given there. I have stressed that how one actually uses these equations to subsequently estimate the function $|M>$ and the parameters x is largely a matter of choice. However, there remains one topic that must be addressed by any reasonable procedure for making such estimates. This is the question of how random errors in the measurements of the data values, b_i, propagate into these estimates. To answer that question we have to know how such random errors propagate into the separated equations (B.24) and (B.30). That is topic of this section. The major result is that data errors introduce uncertainties into both sets of equations since both (B.24) and (B.30) have a term that is constructed as a linear combination of the data (equations (B.22) and (B.6)).

The problem we face here is that the data values, b_i, in real life will always be contaminated by random measurement errors. That is, each measured datum has an unknown error, δb_i, which is related to it by the relation

$$b_i = \widetilde{b}_i + \Delta b_i \qquad (B.33)$$

where $\widetilde{b}_i$ is the unknown, correct value for the ith datum. We don't know what Δb_i is, but I assume we know something about its statistics. Specifically, I assume Δb_i comes from a population with zero mean and that we know the covariance of the data defined as

$$D_{ij} = E[\Delta b_i \Delta b_j] \qquad (B.34)$$

where E[] denotes the expectation value operator. It is preferable to write (B.34) in matrix form as

$$\mathbf{D} = E[\Delta \mathbf{b}(\Delta \mathbf{b})^T] \qquad (B.35)$$

where $\mathbf{D}$ is called the **covariance matrix**.

Given that we know the covariance of the data we can utilize this knowledge to calculate the uncertainty that the statistical errors introduce into (B.24) and (B.30). I begin with the annulled data defined in (B.24). If we substitute equation (B.33) into the defining relation for $\mathbf{n}$ (equation (B.22)), then we see

$$\mathbf{n} = \mathbf{U}_N^T \tilde{\mathbf{b}} + \mathbf{U}_N^T \Delta \mathbf{b} \qquad (B.36)$$

$$= \tilde{\mathbf{n}} + \Delta \mathbf{n}$$

where I define $\tilde{\mathbf{n}} = \mathbf{U}_N^T \tilde{\mathbf{b}}$ and $\Delta \mathbf{n} = \mathbf{U}_N^T \Delta \mathbf{b}$. $\tilde{\mathbf{n}}$, like $\tilde{\mathbf{b}}$, is the true but unknown annulled data vector and $\Delta \mathbf{n}$ denotes the unknown error in $\mathbf{n}$. Clearly, if $\Delta \mathbf{b}$ has zero mean then $\Delta \mathbf{n}$ will also. Hence, $\mathbf{n}$ is unbiased. Furthermore, the covariance of the errors, $\Delta \mathbf{n}$, in $\mathbf{n}$ is given by

$$\mathbf{C}_n = E[\Delta \mathbf{n}(\Delta \mathbf{n})^T]$$

$$= E[\mathbf{U}_N^T \Delta \mathbf{b}(\mathbf{U}_N^T \Delta \mathbf{b})^T] \qquad (B.37)$$

$$= \mathbf{U}_N^T E[\Delta \mathbf{b}(\Delta \mathbf{b})^T] \mathbf{U}_N$$

$$= \mathbf{U}_N^T \mathbf{D} \mathbf{U}_N$$

which is the relationship we seek.

Data errors also introduce an uncertainty into the equations (B.30). This happens because $\hat{\mathbf{x}}$ is calculated by equation (B.6). We now want to find the expected mean and covariance of this uncertainty. We begin by again substituting (B.33) into the defining

relation for (B.6) of $\hat{\mathbf{x}}$.

$$\hat{\mathbf{x}} = \mathbf{H}\tilde{\mathbf{b}} + \mathbf{H}\Delta\mathbf{b} \tag{B.38}$$

$$= \tilde{\mathbf{x}} + \Delta\hat{\mathbf{x}}$$

As before I define $\Delta\hat{\mathbf{x}} = \mathbf{H}\Delta\mathbf{b}$. Again if $\Delta\mathbf{b}$ has zero mean then $\Delta\hat{\mathbf{x}}$ will also. Hence, $\hat{\mathbf{x}}$ (and hence $\mathbf{x}_R$) is not biased. Finally, the covariance of the uncertainty in $\mathbf{x}_R$ introduced by $\hat{\mathbf{x}}$ is given simply as

$$\mathbf{C}_{\mathbf{x}} = \mathrm{E}[\Delta\hat{\mathbf{x}}(\Delta\hat{\mathbf{x}})^T]$$

$$= \mathrm{E}[\mathbf{H}\Delta\mathbf{b}(\mathbf{H}\Delta\mathbf{b})^T] \tag{B.39}$$

$$= \mathbf{H}\mathrm{E}[\Delta\mathbf{b}(\Delta\mathbf{b})^T]\mathbf{H}^T$$

$$= \mathbf{H}\mathbf{D}\mathbf{H}^T$$

APPENDIX C
COSO DATA HYPOCENTERS

Coso Local Earthquakes											
event number	latitude			longitude			depth (km)	origin time				
1	36	0	54.37	117	52	4.16	1.9	75	294	2	4	15.92
2	36	0	4.32	117	45	40.15	2.3	75	327	13	47	13.62
3	36	2	42.54	117	49	32.86	1.8	76	116	16	14	33.31
4	36	2	41.33	117	49	28.50	1.8	76	122	20	2	.97
6	36	0	46.58	117	47	24.08	2.5	78	349	21	57	24.73
7	36	2	22.69	117	45	55.84	3.0	77	15	7	38	14.98
8	36	1	23.85	117	48	46.00	.7	77	26	17	43	3.24
9	36	1	59.26	117	45	55.46	1.4	77	30	2	8	14.72
10	36	3	54.83	117	50	.10	2.0	77	77	4	29	31.71
11	36	3	59.03	117	49	58.79	2.2	77	77	4	58	51.57
12	35	59	56.21	117	47	15.90	2.0	77	89	1	46	12.71
1311	36	0	2.71	117	47	10.79	1.9	77	89	7	12	49.94
14	35	59	54.98	117	47	27.31	1.2	77	94	8	4	15.76
15	36	0	4.36	117	47	21.73	1.1	77	98	4	5	10.09
16	36	5	48.76	117	51	37.46	2.7	76	98	8	43	57.60
17	36	2	50.87	117	49	32.63	2.0	76	85	3	31	39.63
32	36	4	29.96	117	49	19.15	3.1	76	51	9	18	28.84
42	36	2	53.26	117	47	46.48	3.8	76	3	5	16	27.00
52	36	0	20.85	117	51	24.25	3.9	76	22	14	30	7.73
62	36	0	52.10	117	50	28.33	5.1	76	156	23	3	17.30
72	36	0	50.01	117	50	16.95	5.1	76	157	12	36	45.20
82	35	59	53.82	117	45	37.77	3.0	76	180	13	39	40.28
92	36	0	48.79	117	50	21.33	4.5	76	349	6	21	23.10
102	36	5	47.30	117	52	18.76	4.8	77	30	5	13	41.90
112	36	3	5.80	117	46	51.42	4.5	77	42	2	29	2.07
122	36	3	33.76	117	48	57.03	3.1	77	64	17	16	22.01
132	36	3	31.82	117	48	53.68	3.4	77	67	1	20	50.12
142	36	3	27.20	117	48	48.34	3.3	77	67	2	53	9.64
152	36	3	33.06	117	48	51.99	3.2	77	67	6	51	23.22
162	36	3	24.24	117	48	47.66	3.4	77	67	7	43	1.12
172	36	3	38.08	117	48	54.34	3.6	77	67	12	26	56.06
182	36	3	23.70	117	48	48.58	4.1	77	69	5	17	6.66
192	36	3	7.71	117	45	32.64	4.4	77	74	22	55	.80
202	36	5	5.14	117	49	11.60	4.4	77	243	11	53	56.29
212	36	2	6.76	117	48	41.44	4.9	77	271	2	5	21.37
222	36	2	10.32	117	48	43.46	4.7	77	261	11	50	17.01
1322	36	2	50.22	117	47	9.83	4.8	75	314	13	9	38.00
23	36	6	21.23	117	52	25.57	10.7	76	108	6	10	32.49
33	36	4	10.03	117	47	47.55	6.6	76	337	13	53	6.38

Coso Local Earthquakes													
event number	latitude			longitude			depth (km)	origin time					
43	36	4	9.73	117	47	43.39	6.9	76	337	14	7	32.25
63	36	3	59.77	117	47	54.55	7.5	76	337	14	23	6.95
73	36	4	8.29	117	47	49.98	6.4	76	337	17	13	50.87
83	36	4	15.18	117	47	39.96	6.4	76	337	17	27	36.67
93	36	4	7.49	117	47	48.45	6.4	76	337	19	44	6.75
103	36	4	13.29	117	47	43.96	6.7	76	337	21	2	8.52
113	36	4	20.88	117	47	33.37	6.1	76	337	21	15	39.48
123	36	0	52.22	117	50	20.25	5.5	76	154	5	32	36.34
133	36	0	49.83	117	50	17.52	5.4	76	154	5	39	2.59
143	36	0	46.32	117	50	27.16	5.9	76	154	5	44	30.41
153	36	0	55.88	117	50	26.33	5.3	76	154	6	18	34.85
163	36	0	58.06	117	50	29.36	5.2	76	154	12	51	3.76
173	36	0	58.16	117	50	24.92	5.2	76	154	13	14	47.53
183	36	0	45.18	117	50	18.25	6.0	76	158	0	9	56.64
193	36	4	11.96	117	47	48.22	6.2	76	338	1	27	21.82
203	36	4	10.41	117	47	44.51	7.2	76	337	14	10	15.62
213	36	4	16.00	117	47	26.25	6.2	77	8	1	0	57.67
223	35	59	50.47	117	48	31.32	7.0	77	35	20	10	28.43
233	36	4	40.70	117	46	25.38	6.0	77	55	12	50	39.11
243	36	1	36.63	117	49	1.39	7.4	77	271	9	51	17.15

VITA
Gary Lee Pavlis

Birth:

July 28, 1953
Wagner, South Dakota

Parents:

William Pavlis
Stella B. C. Pavlis

Education:

1971

Geddes High School
Geddes, South Dakota

1975

South Dakota State University
Brookings, South Dakota
B.S. Engineering Physics

1982

University of Washington
Seattle, Washington
Ph.D. Geophysics