Optimal Ocean Acoustic Tomography and Navigation with Moving Sources

by

Max Deffenbaugh

S.M., Oceanographic Engineering, MIT/WHOI Joint Program (1994)
B.S. Electrical Engineering, Princeton University (1991)

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF SCIENCE IN OCEANOGRAPHIC ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION

April 1997

Signature of Author .....
Joint Program in Applied Ocean Science and Engineering
Massachusetts Institute of Technology/Woods Hole Oceanographic Institution
April 11, 1997

Certified by .....
Prof. Henrik Schmidt
Professor of Ocean Engineering, MIT
Thesis Supervisor

Certified by .....
Dr. James Lynch
Associate Scientist, WHOI
WHOI Advisor

Accepted by .....
Prof. Henrik Schmidt
Chairman, Joint Committee for Applied Ocean Science and Engineering
Massachusetts Institute of Technology/Woods Hole Oceanographic Institution

Optimal Ocean Acoustic Tomography and Navigation with Moving Sources

by

Max Deffenbaugh

Submitted to the Massachusetts Institute of Technology/
Woods Hole Oceanographic Institution
Joint Program in Applied Ocean Science and Engineering
on April 13, 1997, in partial fulfillment of the
requirements for the degree of
DOCTOR OF SCIENCE IN OCEANOGRAPHIC ENGINEERING

Abstract

Ocean Acoustic Tomography inverts for the two- or three- dimensional sound speed structure in a volume of water by measuring acoustic travel times along ray paths traversing the volume The sensitivity of the acoustic travel times to particular modes of sound speed variation is highly dependent on the source and receiver positions. Autonomous underwater vehicles provide mobile instrument platforms at relatively low cost. Tomography sources mounted on AUVs can be adaptively repositioned to better image emerging sound speed features. The goal of optimal moving source tomography is to make optimal use of mobile controllable tomography sources in gaining information about the environment. The component technologies for optimal moving source tomography are position estimation, sound speed parameterization and estimation, ray path identification, and vehicle path optimization. This thesis makes contributions in each of these areas.

Thesis Supervisor: Prof. Henrik Schmidt
Professor of Ocean Engineering, MIT

WHOI Advisor: Dr. James Lynch
Associate Scientist, WHOI

Acknowledgments

I would like to thank Henrik Schmidt for serving as thesis advisor and Art Baggeroer, Jim Lynch, Jim Preisig, and Alan Willsky for serving as the thesis committee. I would also like to thank Jim Bellingham for serving as chairman of the thesis defense, and Tom Curtin of the ONR Arctic Programs Office for support of this work.

Contents

1 Introduction **12**

2 Position Estimation **17**

2.1 Positioning Systems 18

2.1.1 Spherical Positioning 19

2.1.2 Hyperbolic Positioning 22

2.1.3 Positioning with Time Synchronization 28

2.2 Position Estimators 34

2.2.1 Vehicle Sitting Still 37

2.2.2 Vehicle Moving 39

2.3 Optimal Navigation 41

2.3.1 Path Parameterization 42

2.3.2 Optimization Algorithm 44

2.3.3 Examples of Optimal Navigation 48

2.4 Conclusion 51

3 Sound Speed Estimation **53**

3.1 The Forward Problem 55
3.1.1 Ray Sampling Functions 59
3.2 The Inverse Problem 62
3.3 Weight Estimators 64
3.3.1 Suboptimal estimators 65
3.3.2 Optimal Estimator 67
3.4 Basis Functions 71
3.4.1 Empirical Orthogonal Functions 71
3.4.2 Optimal Orthogonal Functions 73
3.5 Examples 75
3.5.1 Temperate Example 75
3.5.2 Arctic Example 79
3.6 Using Estimated Profile Covariance Matrices 83
3.6.1 The Covariance Matrix Estimate 83
3.6.2 Weighting the Noise Covariance Diagonal 84
3.6.3 Wrapping Profile Covariance into Noise 85
3.6.4 Comparison of Methods 85
3.7 Optimal Moving Source Tomography 86
3.7.1 Tomographic Resolution 91
3.7.2 Optimal Tomography 94
3.8 Conclusion 98

4 Arrival Matching **99**

4.1 Test Scenario 99
4.2 Simple Ordering 103
4.3 Validation Windows 105
4.4 Independent Matching 107
4.5 Correlated matching algorithm 112
4.6 Advantage of the Correlated Matcher 119
4.7 Future Direction 123
4.8 Conclusion 124

5 Test Experiments **125**

5.1 Multipath Navigation 126
5.1.1 Simulation of Noise Effects 129
5.1.2 Simulation of Blocking 130
5.1.3 Demonstration Plot 132
5.2 Moving Source Tomography 134
5.2.1 Estimator Options 134
5.2.2 Experiment Description 136

6 Conclusion **145**

List of Figures

1-1 Block diagram of moving source tomography system. 14

2-1 Spherical positioning. 20

2-2 Cramér-Rao bound in meters for spherical position estimate. 23

2-3 Hyperbolic positioning. 25

2-4 Cramér-Rao bound in meters for hyperbolic position estimate. 27

2-5 Transition from spherical to hyperbolic performance as shown by position Cramér-Rao bounds in meters. 32

2-6 Cramér-Rao bound for time synchronization estimate in milliseconds. 33

2-7 Vehicle location. 39

2-8 Position estimate accuracy. 40

2-9 Vehicle path. 42

2-10 Position estimate accuracy comparison. 43

2-11 Feasible perturbation region. 45

2-12 Feasible region with bounding rectangle. 47

2-13 Optimal navigation, 31 transmissions. 50

2-14 Optimal navigation, 51 transmissions. 50

2-15 Optimal navigation, 51 transmissions, different starting point. 51

3-1 Profiles for the three ray sampling functions. 59

3-2 A sampling function and ray for profile 1–constant sound speed. . . . 60

3-3 A sampling function and ray for profile 2–the Arctic profile. 61

3-4 A sampling function and ray for profile 3–the temperate profile. 62

3-5 Sound speed profile and eigenrays for the temperate example. 77

3-6 Sampling functions and profile variations for the temperate example. . 78

3-7 The EOF and OOF for the temperate example. 78

3-8 Sound speed profile and eigenrays for the Arctic example. 81

3-9 Sampling functions and profile variations for the Arctic example. . . . 81

3-10 The EOF and OOF for the Arctic example. 82

3-11 Comparison of MSE achieved as a function of σ, temperate case. . . . 87

3-12 Comparison of MSE achieved as a function of σ, Arctic case. 87

3-13 Rays with fixed source. 88

3-14 Rays with moving source. 89

3-15 Rays with fixed source. 90

3-16 Rays with moving source. 90

3-17 Sound speed profile and basis function. 92

3-18 Sound speed estimate variance. 93

3-19 Sound speed profile and basis functions. 95

3-20 Sound speed estimate variance for stationary, moving, and optimal moving source tomography (first region of interest). 96

3-21 Sound speed estimate variance for stationary, moving, and optimal moving source tomography (second region of interest). 97

4-1 Predicted profile and profile variation. 101

4-2 True profiles and eigenrays for predicted profile. 101

4-3 Predicted arrival times and measured arrival times for parameter values of 1 to 20. 102

4-4 Ordered matching. 104

4-5 Error for ordered matching. 105

4-6 Validation window matcher 106

4-7 Estimate error using validation window matcher 107

4-8 Independent matching. 112

4-9 Error for independent matching. 113

4-10 Correlated matching. 115

4-11 Error comparison. 116

4-12 Number of possible matches. 117

4-13 Width of tree. 119

4-14 Region where tomography is possible using standard arrival matching methods. 121

4-15 Region where tomography is possible using correlated matching method. 122

5-1 Block diagram of the multipath positioning system 127

5-2 Positioning demonstration setup (top view) 128

5-3 Ray blocking by shallow region 131

5-4 Error magnitude (in meters) for both systems, and arrival times of eigenrays from the three beacons 133

5-5 Mean sound speed profile and profile variations 137

5-6 Optimal transmission depths 139

5-7 Fraction of errors lying within one theoretical standard deviation (should be 0.683) 139

5-8 Mean sound speed profile and profile variations 141

5-9 Optimal transmission depths 142

5-10 Parameter estimate errors and theoretical standard deviation 142

5-11 Fraction of errors lying within one theoretical standard deviation (should be 0.683) 143

Chapter 1

Introduction

Ocean Acoustic Tomography estimates the two- or three-dimensional sound speed structure in a volume of water by measuring acoustic travel times along ray paths traversing the volume [37, 36]. In a typical application, acoustic sources and receiver arrays are placed around the volume of interest. The acoustic travel times along the multiple eigenrays connecting each source and receiver are predicted based on an initial estimate of parameters describing the sound speed structure. The actual acoustic travel times are then measured for each eigenray. The differences between measured and predicted travel times for the various eigenrays can be linearly related to parameters describing variations in the sound speed structure, assuming that these variations are not too large. Based on this linear relationship, an inversion can be performed to determine the sound speed parameters. This technique has been successfully applied to estimate sound speed fields in a number of experiments [47].

The sensitivity of the acoustic travel times to a particular sound speed variation is highly dependent on the source and receiver locations. By using mobile sources sus-

pended from ships [8] or mounted on autonomous underwater vehicles [14] it is possible to obtain measurements from many more locations and to determine the sound speed field to a much higher level of accuracy than would be possible with fixed sources. This thesis examines the use of mobile sources in ocean acoustic tomography. It lays a foundation of improved and in some cases optimal algorithms for acoustic position measurement, for sound speed parameterization and inversion, and for eigenray identification. It introduces source path planning for optimal measurement accuracy, and simulates the performance of all algorithms in realistic ocean environments.

Throughout this thesis, it is assumed that sound propagation is described well by a ray model, and ray models will be used exclusively. Acoustic tomography has also been performed using characteristics of mode propagation, but these methods are outside the scope of the present work.

The overall structure of a moving source tomography system is shown in Fig. 1-1. This system takes as its input the received acoustic signal and generates as its output source location [13] and parameters describing sound speed variability [10]. The functions of sound speed estimation and position estimation are included in the same system because the problems are tightly coupled. Accurate acoustic positioning depends on accurate sound propagation models, and accurate acoustic tomography requires that the source and receiver positions be known. Each block in Fig. 1-1 is described in greater detail below, and the contributions of the thesis to each block are outlined.

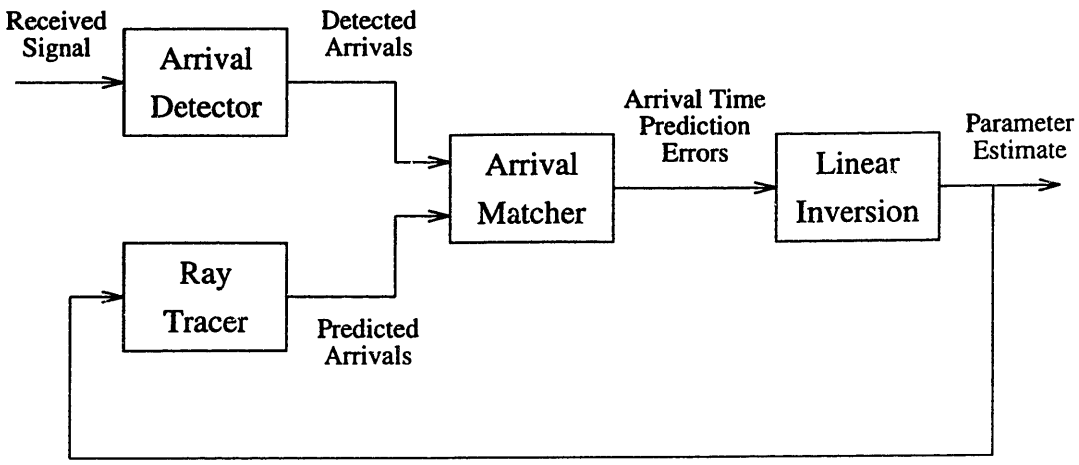

Figure 1-1: Block diagram of moving source tomography system.

Ray Tracing Model The *ray tracing model* block predicts the travel times along the rays connecting each source and receiver based on the current estimate of sound speed parameters and source location. The construction of ray tracing models is well-documented in the literature [26].

Arrival Detector The *arrival detector* block breaks down the received acoustic signal into its component ray arrivals and determines the time of each arrival. The decomposition of a received signal into time delayed, amplitude scaled replicas of the transmitted signal is well documented in the signal processing literature [20, 4, 45]. In cases where the measured travel times are calculated from actual or simulated received signals, the travel time estimators are kept simple and standard in order to illustrate the robustness of the new arrival identification and inversion algorithms.

Arrival Matcher As position or sound speed parameters change, the measured arrivals will tend to either shift linearly in time or disappear. Because of these dual aspects of arrival behavior, the inversion process is divided into two stages. The

first stage is the *arrival matcher* block which attempts to match the measured arrival times with the predicted arrival times to which they are linearly related, allowing for the possibility that some arrivals may not have matches. When a measured arrival is identified with a predicted arrival, a prediction error can be calculated as the difference between the measured arrival time and the predicted arrival time. Chapter 4 introduces the arrival matching problem. The commonly used methods are discussed as well as their shortcomings. Two new arrival matching techniques are presented. The first of these is sub-optimal but is fast and handles arrival fading well. The second is a powerful algorithm which takes into account correlations between the time shifts of different arrivals. The performance of the new algorithms is compared with the standard methods.

Linear Inversion The arrival time prediction errors are used by the *linear inversion* block to update the estimate of sound speed parameters and source location. Chapter 2 considers the estimation of source position and the coupling between position errors and time synchronization errors. A mathematical relationship between spherical and hyperbolic positioning is derived and illustrated graphically, and the performance of a system which estimates both source position and transmit time synchronization is considered. It is observed that the accuracy of a vehicle position estimate depends not only on the present location of the vehicle but also on the path followed in reaching the present location. This leads to formulation and solution of the optimal navigation problem of determining the path to a destination which will result in the least error in the position estimate upon reaching the destination.

Chapter 3 considers the estimation of sound speed profile parameters. The way in which rays sample the sound speed field is formalized by defining ray sampling functions such that the projection of a sound speed profile variation onto a ray sampling function is the travel time change which that profile variation causes in the given ray. These sampling functions prove instrumental in deriving the optimal estimator for sound speed variability and the optimal parameterization for a sound speed field. The ability of a tomography measurement to resolve a given sound speed variation is highly dependent on the source and receiver positions. Based on this dependence, optimal paths are found for moving acoustic sources to focus measurement accuracy at specific features of interest in the environment.

Example Missions Chapters 2 through 4 describe the various component technologies which comprise optimal moving source tomography. Chapter 5 integrates the algorithms of the previous chapters into a single system and tests the system performance in several simulated missions. The ultimate test of the system is of course its performance in the field, and the example missions are in fact a sequence of recommended field tests to demonstrate the algorithms developed in this thesis.

Chapter 2

Position Estimation

This chapter considers the estimation of position from acoustic travel time measurements. Acoustic positioning systems have traditionally fallen into two categories, spherical and hyperbolic [35, 43, 6]. Spherical positioning systems determine position by measuring acoustic travel times from beacons at known locations. To make this travel time measurement, the receiver must know exactly when each beacon transmitted. Hyperbolic positioning systems determine position by measuring *differences* in travel time between signals from the beacons. The hyperbolic receiver does not need to know when the beacons transmitted, only that they all transmitted at the same time or with known delays relative to each other. The spherical system, which must know transmit time exactly, and the hyperbolic system, which does not know transmit time at all can be seen as the two endpoints of a continuum of systems parameterized by the accuracy with which the transmit time is known. This continuum is shown to exist for an arbitrary number of beacons in an arbitrary dimensional space. It is illustrated for the case of three beacons in a two dimensional space.

Hyperbolic positioning produces much larger position estimate errors than does spherical positioning near the baseline extensions. When the beacons transmit at regular time intervals, however, a hyperbolic system can approach spherical performance levels by estimating transmit time in order to synchronize itself with the master beacon clock. Examples of position estimation with clock synchronization are presented for both stationary and moving vehicles. An approximate condition is given under which estimating time synchronization can improve position estimate accuracy.

The accuracy of position and clock synchronization estimates depends on the location of the vehicle with respect to the beacons. Position and particularly clock synchronization cannot change infinitely quickly, and so past measurements as well as the present measurement contain information about the present position and clock synchronization. Putting these two facts together, it is observed that for a moving vehicle, the present position estimate depends not only on where the vehicle is, but also on the path which it followed in getting there. Optimal navigation seeks to select the path between a fixed origin and destination which minimizes the position estimate error achieved at the destination. In this chapter, the optimal navigation problem is developed and solved using the method of simulated annealing, and optimal paths are shown for several examples.

2.1 Positioning Systems

Most underwater acoustic positioning systems are either spherical or hyperbolic in design. These two systems are analyzed, and a third system is introduced–one which

operates in a hyperbolic beacon network but estimates transmit time in addition to position in an attempt to come closer to spherical performance levels.

2.1.1 Spherical Positioning

Spherical positioning systems determine position by measuring acoustic travel times from beacons at known locations. In practice this is often accomplished by transponding with the beacons so that a two-way travel time is measured and then divided by two. The travel times are converted to ranges, and each range measurement defines a sphere, centered at the beacon, on which the receiver must lie. Range measurements from several beacons provide several spheres, the intersection of which is the position of the receiver. In Figure 2-1, the beacons are shown by black circles, and the receiver is shown by a black square. The range from each beacon to the receiver defines a circle around that beacon on which the source must lie. The point where all three circles intersect is the position of the receiver [35, 43].

Positioning is considered in an M dimensional space (in general $M = 2$ or 3), where position is determined by range measurements from N beacons. The range from the nth beacon to the source at $\mathbf{x} = [x_1, x_2, \ldots, x_M]^T$ can be found by the Pythagorean theorem as,

$$r_n(\mathbf{x}) = \sqrt{\sum_{m=1}^{M} (x_m - x_{Bnm})^2}. \tag{2.1}$$

The acoustic travel time measured from the nth beacon is then,

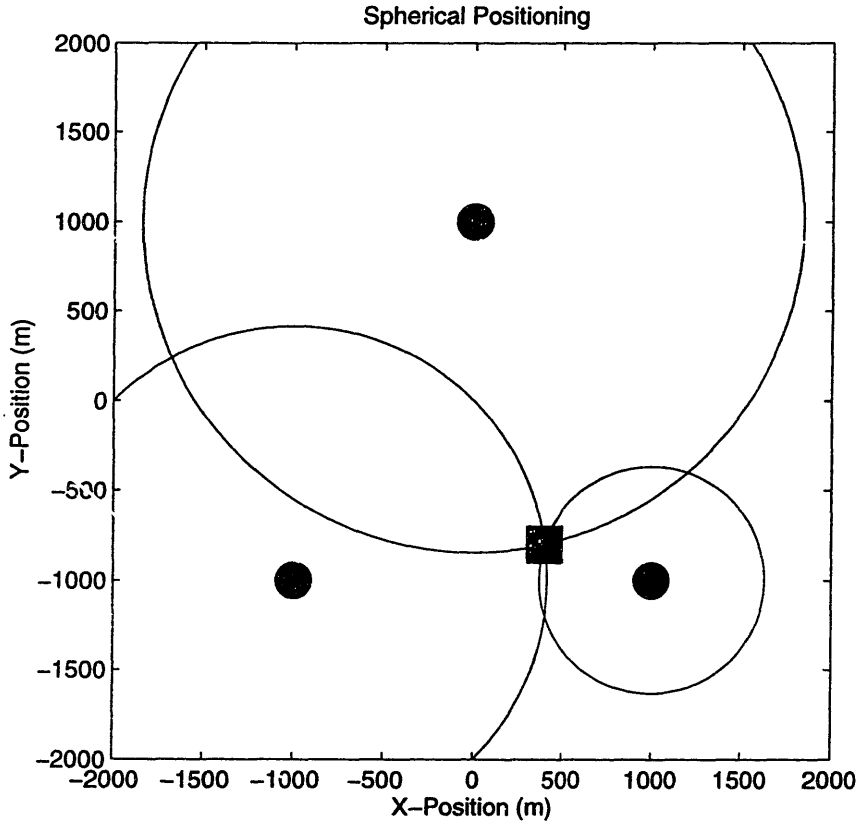

Figure 2-1: Spherical positioning.

$$t_n(\mathbf{x}) = \frac{r_n(\mathbf{x})}{c} + v_n \tag{2.2}$$

where c is the speed of sound (assumed throughout this chapter to be independent of position), and v_n is measurement noise which is assumed to be zero-mean Gaussian.

The vehicle will attempt to determine its position $\mathbf{x}$ from the vector consisting of travel time measurements from each beacon $\mathbf{t}$. The Cramér-Rao bound gives a lower bound on the covariance of an unbiased position estimate formed from this measurement. To calculate this bound, it is necessary to linearize the non-linear equation (2.2) about the position at which the bound is to be calculated, $\bar{\mathbf{x}}$. Define,

$$\tilde{\mathbf{t}} = \mathbf{t}(\mathbf{x}) - E[\mathbf{t}(\bar{\mathbf{x}})] \tag{2.3}$$

$$\tilde{\mathbf{x}} = \mathbf{x} - \bar{\mathbf{x}} \tag{2.4}$$

This linearized equation is,

$$\tilde{\mathbf{t}} = \mathbf{C}\tilde{\mathbf{x}} + \mathbf{v} \tag{2.5}$$

The matrix of partial derivatives $\mathbf{C}$ can be calculated,

$$\mathbf{C}_{nm}(\bar{\mathbf{x}}) = \left.\frac{\partial E[t_n(\mathbf{x})]}{\partial x_m}\right|_{\mathbf{x}=\bar{\mathbf{x}}}$$

$$= \frac{1}{c}\frac{\bar{x}_m - x_{Bnm}}{r_n(\bar{\mathbf{x}})}. \tag{2.6}$$

The Cramér-Rao bound for the position estimate is [39],

$$CRB_S = \left[\mathbf{C}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{C}\right]^{-1}, \tag{2.7}$$

where $\mathbf{R}$ is the measurement noise covariance matrix, $\mathbf{R} = E[\mathbf{v}\mathbf{v}^T]$. The standard deviation of the position estimate given by this bound is shown as a function of position in Figure 2-2. For this figure, the sound speed is 1500 m/s, the measurement errors are independent identically distributed Gaussian random variables with zero mean and standard deviations of 1 ms, and the beacons are placed at the locations indicated by the black circles.

Notice that the position resolution is better toward the center of the beacon network, but is still fairly good in far away corners. Along a baseline extension, two rows of the measurement matrix C become identical, but in this example of three beacons in a two-dimensional space, the position estimate is still uniquely determined along the baseline extensions. Geometrically speaking, two circles of position are tangent along the baseline extensions, but there is a third circle which provides the second position constraint.

2.1.2 Hyperbolic Positioning

Hyperbolic positioning systems determine position by measuring *differences* in travel times between signals from the various beacons, which are assumed to be synchronized

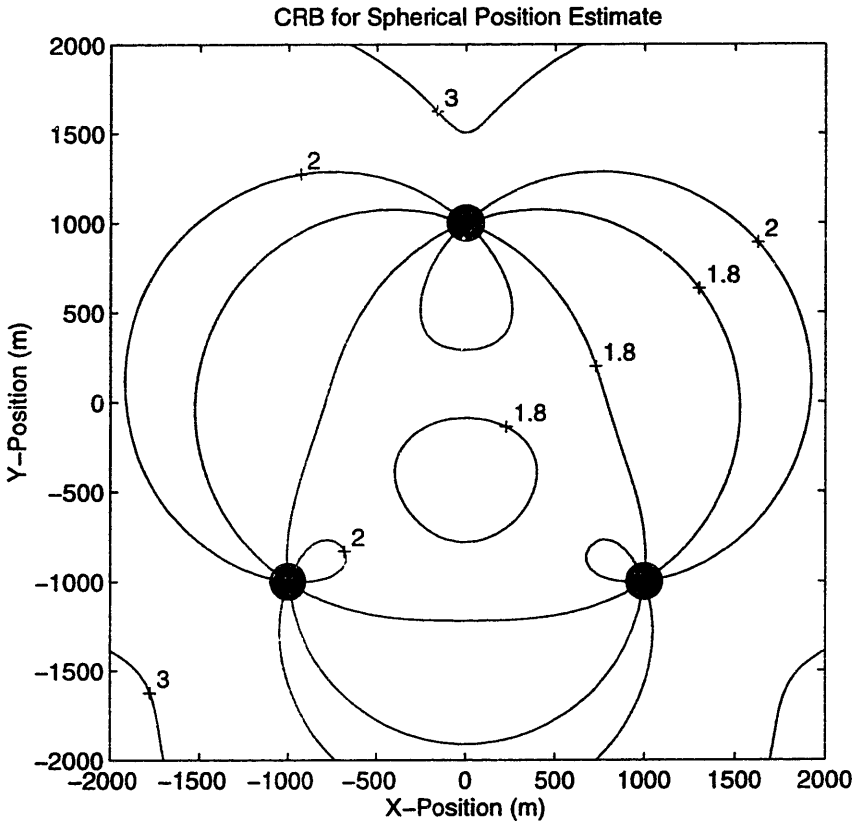

Figure 2-2: Cramér-Rao bound in meters for spherical position estimate.

with eachother but not with the receiver. This synchronization may be accomplished by providing a common time base (such as GPS time) to all beacons or by having slave beacons trigger of the acoustic transmission of a single master beacon. The differences in travel times are converted to differences in ranges to the beacons, and each range difference defines a hyperboloid (3-dimensional hyperbola) on which the receiver must lie. Several of these range differences give several hyperboloids, the intersection of which is taken to be the position of the receiver. In Figure 2-3, the beacons are shown by black circles, and the receiver is shown by a black square. The left hyperbola is determined by the difference in range from the receiver to the top beacon and the receiver to the left bottom beacon. The right hyperbola is determined by the difference in range from the receiver to the top beacon and the receiver to the right bottom beacon. The intersection of these two hyperbola is the position of the receiver. The advantage of the hyperbolic system is that it does not use the time at which the beacons transmit in its calculations, so that the receiver does not need to be synchronized with the beacons. The travel time differences can be found as long as the beacons transmit at the same time, or with known delays relative to each other. The disadvantage of the hyperbolic system is that it provides poorer position resolution than the spherical system, particularly at the fringes of the beacon network [35, 43].

The hyperbolic positioning system measures the differences in travel times to the various beacons. The linearized equation (2.5) can be modified to create the relationship between position and travel time difference by multiplying by the matrix

M

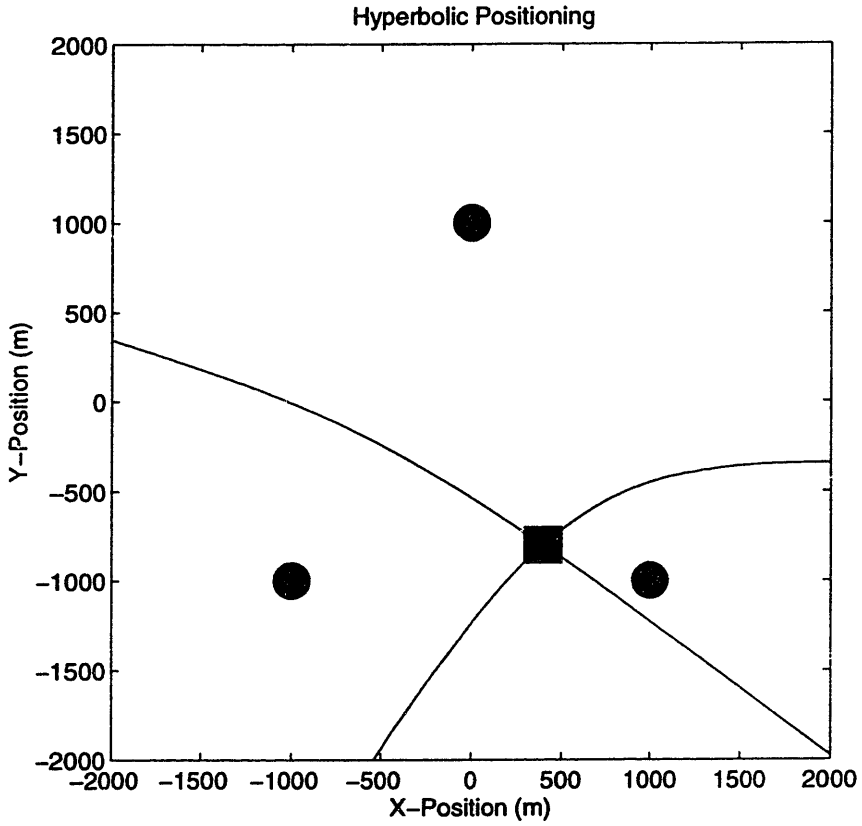

Figure 2-3: Hyperbolic positioning.

$$\mathbf{M} = \left[\begin{array}{c|ccc} 1 & -1 & & 0 \\ \vdots & & \ddots & \\ 1 & 0 & & -1 \end{array}\right]. \tag{2.8}$$

The linearized hyperbolic positioning equation is then,

$$\mathbf{M}\tilde{\mathbf{t}} = \mathbf{MC}\tilde{\mathbf{x}} + \mathbf{Mv}. \tag{2.9}$$

By comparison with the spherical system (2.5) and bound (2.7), the the Cramér-Rao bound for the hyperbolic system is found to be,

$$CRB_H = \left[\mathbf{C^T M^T}\left(\mathbf{MRM^T}\right)^{-1}\mathbf{MC}\right]^{-1} \tag{2.10}$$

The standard deviation of the position estimate given by this bound is shown as a function of position in Figure 2-4. As in Figure 2-2, the sound speed is 1500 m/s, the measurement errors are independent identically distributed Gaussian random variables with zero mean and standard deviations of 1 ms, and the beacons are placed at the locations indicated by the black circles.

Notice the greatly deteriorated position resolution outside the center of the beacon network. The resolution is particularly poor along the baseline extensions. Along a baseline extension, two rows of the hyperbolic measurement matrix $\mathbf{MC}$ become identical. In the example of three-beacons in two-dimensional space, the hyperbolic measurement matrix is square, and when two rows become identical along the baseline extensions, the position estimate becomes underdetermined. Geometrically speaking,

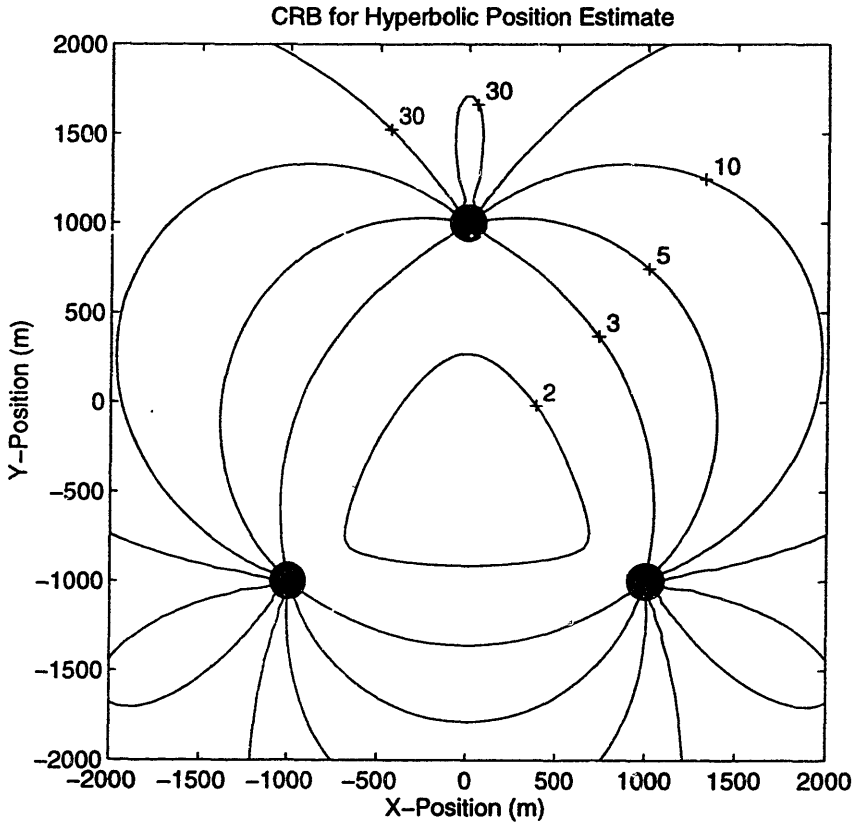

Figure 2-4: Cramér-Rao bound in meters for hyperbolic position estimate.

two hyperbolas of position are tangent along a baseline extension, so with only two hyperbolas available, position is unconstrained in the tangent direction.

2.1.3 Positioning with Time Synchronization

In the spherical positioning system, the transmit time of the beacons must be known exactly so that travel times can be determined. In the hyperbolic positioning system, the transmit time falls out of the equations when the travel time differences are taken. The spherical system, which requires zero variance in the transmit time estimate, and the hyperbolic system, which allows infinite variance in the transmit time estimate, represent the endpoints of a continuum of intermediate systems in which the transmit time is known with a finite, but non-zero, variance [10]. In this section, it is shown that the spherical and hyperbolic bounds are indeed the endpoints of a whole continuum of bounds for systems in which the transmit time variance ranges between zero and infinity.

We will now consider a system in which the beacons all transmit at a predetermined time. The vehicle knows the scheduled transmit time, but its internal clock may be running fast or slow with respect to the beacons. Because of this clock synchronization error, the vehicle will miscalculate all the travel times by a constant amount, which will be called τ. The measured travel time now consists of the true travel time plus a time synchronization error plus measurement noise, and the linearized navigation equation (2.5) can be extended to reflect this.

$$\tilde{\mathbf{t}} = \mathbf{C}\tilde{\mathbf{x}} + \tau\mathbf{1} + \mathbf{v}, \tag{2.11}$$

where $\mathbf{1}$ is the $N \times 1$ vector with all elements equal to one. If τ is treated as a random variable with zero mean and variance σ_τ^2, this time synchronization error can be grouped into the error covariance matrix which becomes $\sigma_\tau^2 \mathbf{1}\mathbf{1}^T + \mathbf{R}$.

Using this new measurement covariance in (2.7), the Cramér-Rao bound on position for a system having a time synchronization error with standard deviation σ_τ is

$$CRB_T = \left[\mathbf{C}^\mathbf{T}\left(\sigma_\tau^2 \mathbf{1}\mathbf{1}^T + \mathbf{R}\right)^{-1}\mathbf{C}\right]^{-1}. \tag{2.12}$$

Next, it is shown that the limit of this equation as $\sigma_\tau \to 0$ is the Cramér-Rao bound for spherical positioning, and the limit as $\sigma_\tau \to \infty$ is the Cramér-Rao bound for hyperbolic positioning.

The Spherical Positioning limit, $\sigma_\tau^2 \to 0$ In the limit $\sigma_\tau^2 \to 0$, performance approaches that of the spherical positioning system in (2.7).

$$\begin{aligned}
\lim_{\sigma_\tau^2 \to 0} CRB_T &= \lim_{\sigma_\tau^2 \to 0} [\mathbf{C}^\mathbf{T}\left(\sigma_\tau^2 \mathbf{1}\mathbf{1}^T + \mathbf{R}\right)^{-1}\mathbf{C}]^{-1} \\
&= \left[\mathbf{C}^\mathbf{T}\mathbf{R}^{-1}\mathbf{C}\right]^{-1} \\
&= CRB_S. \qquad (2.13)
\end{aligned}$$

The Hyperbolic Positioning limit, $\sigma_\tau^2 \to \infty$ Verifying the hyperbolic limiting case is somewhat more complicated. To evaluate the limit, a matrix $\mathbf{N}$ is introduced

$$\mathbf{N} = \left[\begin{array}{c|c} 1 \;\; 1 \;\; \cdots \;\; 1 \\ \hline \mathbf{M} \end{array}\right] = \left[\begin{array}{c} \mathbf{1}^T \\ \hline \mathbf{M} \end{array}\right], \tag{2.14}$$

where the lower submatrix $\mathbf{M}$ is the one defined in (2.8). Note here that the first row of $\mathbf{N}$ is orthogonal to all of its other rows, and that $\mathbf{N}$ is symmetric and invertible.

$$\begin{aligned}
\lim_{\sigma_\tau^2 \to \infty} CRB_T &= \lim_{\sigma_\tau^2 \to \infty} [\mathbf{C^T} \left(\sigma_\tau^2 \mathbf{1}\mathbf{1}^T + \mathbf{R}\right)^{-1} \mathbf{C}]^{-1} \\
&= \lim_{\sigma_\tau^2 \to \infty} \left[\mathbf{C^T N^T} (\mathbf{N}(\sigma_\tau^2 \mathbf{1}\mathbf{1}^T + \mathbf{R})\mathbf{N^T})^{-1} \mathbf{NC}\right]^{-1} \\
&= \left[\mathbf{C^T N^T} \left(\lim_{\sigma_\tau^2 \to \infty} [\mathbf{N}(\sigma_\tau^2 \mathbf{1}\mathbf{1}^T + \mathbf{R})\mathbf{N^T}]^{-1}\right) \mathbf{NC}\right]^{-1} \\
&= \left[\mathbf{C^T N^T} \lim_{\sigma_\tau^2 \to \infty} \left[\begin{array}{c|c} N^2\sigma_\tau^2 + \mathbf{1}^T\mathbf{R}\mathbf{1} & \mathbf{1}^T\mathbf{R}\mathbf{M^T} \\ \hline \mathbf{MR1} & \mathbf{MRM^T} \end{array}\right]^{-1} \mathbf{NC}\right]^{-1}
\end{aligned} \tag{2.15}$$

Using the relation that for a partitioned matrix $\mathbf{A}$, where

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{A_1} & \mathbf{A_2} \\ \hline \mathbf{A_3} & \mathbf{A_4} \end{array}\right] \tag{2.16}$$

the inverse can be written [3],

$$\mathbf{A}^{-1} = \left[\begin{array}{c|c} (\mathbf{A}_1 - \mathbf{A}_2\mathbf{A}_4^{-1}\mathbf{A}_3)^{-1} & -\mathbf{A}_1^{-1}\mathbf{A}_2(\mathbf{A}_4 - \mathbf{A}_3\mathbf{A}_1^{-1}\mathbf{A}_2)^{-1} \\ \hline -\mathbf{A}_4^{-1}\mathbf{A}_3(\mathbf{A}_1 - \mathbf{A}_2\mathbf{A}_4^{-1}\mathbf{A}_3)^{-1} & (\mathbf{A}_4 - \mathbf{A}_3\mathbf{A}_1^{-1}\mathbf{A}_2)^{-1} \end{array}\right], \quad (2.17)$$

the calculation continues from (2.15) after taking the limit of the inverse,

$$\begin{aligned} \lim_{\sigma_T^2 \to \infty} CRB_T &= \left[\mathbf{C}^{\mathbf{T}}\mathbf{N}^{\mathbf{T}} \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \left[\mathbf{MRM}^{\mathbf{T}}\right]^{-1} \end{array}\right] \mathbf{NC}\right]^{-1} \\ &= \left[\mathbf{C}^{\mathbf{T}}\mathbf{M}^{\mathbf{T}}\left(\mathbf{MRM}^{\mathbf{T}}\right)^{-1}\mathbf{MC}\right]^{-1} \\ &= CRB_H. \end{aligned} \quad (2.18)$$

This bound is the same as the one derived for the hyperbolic positioning system in (2.10), which demonstrates that the performance approaches that of the hyperbolic positioning system in the limit where the uncertainty in transmit time becomes infinite.

Figure 2-5 shows the spherical and hyperbolic endpoints of the performance bounds, as well as some intermediate values. It is clear from Figure 2-5 that even a rough estimate of the time synchronization can greatly improve the positioning accuracy, particularly in the regions of the baseline extensions.

In Figure 2-6, the Cramér-Rao bound for the time synchronization estimate is shown. Contour labels are in milliseconds. The ability to estimate time synchronization is greatest in the center of the positioning array, and poorest along the baseline

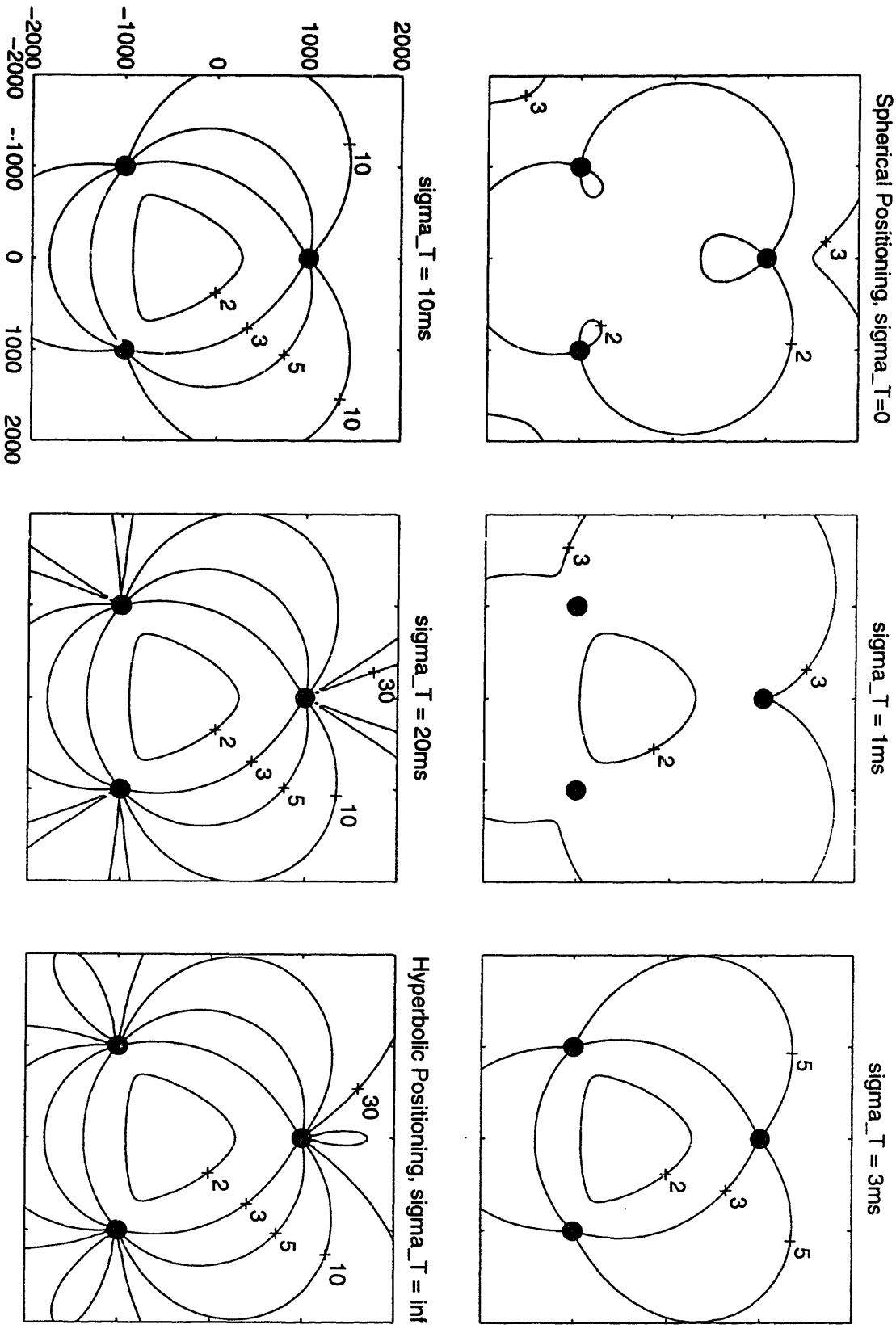

Figure 2-5: Transition from spherical to hyperbolic performance as shown by position Cramér-Rao bounds in meters.

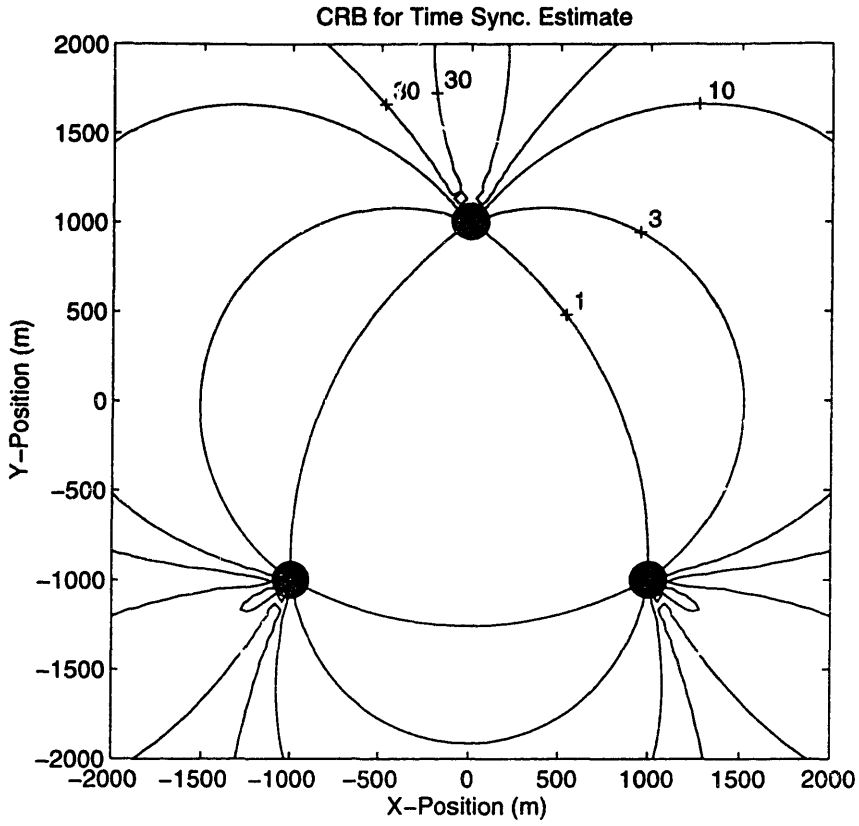

Figure 2-6: Cramér-Rao bound for time synchronization estimate in milliseconds.

extensions.

Position errors and time synchronization errors are coupled. In a hyperbolic positioning system, each point along a hyperbola of position corresponds to a particular time synchronization. In general, when a time synchronization error occurs, the position estimate will be pulled in different directions by the different hyperbolas of position. Along a baseline extension, however, two hyperbolas of position become tangent, and if there are a total of only two hyperbolas of position, time synchronization errors and vehicle position errors along the tangent direction become indistinguishable.

2.2 Position Estimators

The fact that hyperbolic and spherical positioning systems represent the two endpoints of a continuum of possible performance levels has practical consequences. If the beacons in a hyperbolic navigation system repeat their transmissions at regular time intervals, it is possible to produce a running estimate of the transmit time. The accuracy of this synchronization determines where the system lies in the performance continuum between hyperbolic and spherical positioning. In many underwater applications, this synchronization actually affords significant improvement over hyperbolic performance bounds [11]. A simulation is conducted to demonstrate this, and the conditions are discussed under which estimating transmit time provides a substantial improvement in position accuracy.

The Cramér-Rao bounds describe the amount of information available about one set of unconstrained variables from a single noisy observation of another set of variables. If multiple observations are possible, and if the covariance matrix of the variables to be estimated is known, improved estimates can be obtained. The Kalman filter provides a framework for incorporating both the multiple observations over time and the covariance matrix for the variables of interest [22].

In the examples which follow, vehicle position and time synchronization, which we will call the vehicle *state*, will be estimated over the course of simulated vehicle missions. During a mission, the beacons will transmit K times, so the estimator will go through K iterations. There are three steps per iteration, a *prediction* step in which the new vehicle state at the time of a transmission is predicted, a *linearization* step

where a linear approximation of the relationship between arrival time change and state change is calculated centered about the current state estimate, and a *correction* step in which the measured arrival times are used to correct the state prediction. These three steps are described in greater detail below.

Prediction As the vehicle moves between acoustic transmissions, models of vehicle dynamics and clock drift attempt to predict the change in state. In this thesis, we will not touch on models of vehicle dynamics or models of clock drift, but will simply assume that when such models have given their best prediction, an additional error has accumulated in the state described by the zero-mean Gaussian vector $\mathbf{w}(k)$, which has covariance matrix $\mathbf{Q}$. The true state estimate *error* will be the vector $\mathbf{z}$,

$$\mathbf{z} = \begin{bmatrix} \tilde{\mathbf{x}} \\ \\ \tau \end{bmatrix}. \tag{2.19}$$

Let $\mathbf{z}(k|k)$ be the error at the kth iteration based on travel time measurements up to the kth transmission, and $\mathbf{z}(k+1|k)$ will be the state estimate error at the $k+1$th iteration based on measurements up to the kth transmission, *i.e.* the error after the prediction phase but prior to the measurement phase of the $k+1$ iteration.

$$\mathbf{z}(k+1|k) = \mathbf{z}(k|k) + \mathbf{w}(k) \tag{2.20}$$

The covariance matrix of $\mathbf{z}$ will be $\mathbf{P}$, which evolves as,

$$\mathbf{P}(k+1|k) = \mathbf{P}(k|k) + \mathbf{Q} \tag{2.21}$$

Once the state is predicted, the arrival times based on the new state prediction are calculated.

Linearization In the linearization step, the equations relating travel time changes to state changes are relinearized around the new state estimate which was formed in the prediction step. The differences $\mathbf{y}(k)$ between the measured and predicted arrival times can be related to the state estimate errors,

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{z}(k|k) + \mathbf{n}(k) \tag{2.22}$$

where this equation has been linearized about the position prediction $\bar{\mathbf{x}}(k|k-1)$ from the predicted vehicle state vector, so that

$$\mathbf{H}(k) = \left[\mathbf{C}(\bar{\mathbf{x}}(k|k-1)) \,\middle|\, 1 \right]. \tag{2.23}$$

Correction In the correction step, the vehicle will use the measured arrival times to correct its state estimate. The correction added to the predicted state is,

$$\hat{\mathbf{z}}(k|k) = \mathbf{P}(k|k-1)\mathbf{H}^{\mathbf{T}}(k)\left(\mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^{\mathbf{T}}(k) + \mathbf{R}\right)^{-1}\mathbf{y}(k). \tag{2.24}$$

The covariance of the new state estimate acquired by adding this correction to the old predicted state is,

$$\mathbf{P}(k|k) = \mathbf{P}(k|k-1) - \mathbf{P}(k|k-1)\mathbf{H}^{\mathbf{T}}(k)\left(\mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^{\mathbf{T}}(k) + \mathbf{R}\right)^{-1}\mathbf{H}(k)\mathbf{P}(k|k-1) \tag{2.25}$$

This process of prediction, linearization, and correction is carried out for each of the K beacon transmissions which compose the vehicle mission.

Implementation Issues

The time synchronization errors measured in seconds will tend to be small compared to the position errors measured in meters. To avoid the computational problems which this can create, the actual software uses units of milliseconds to express time errors, though the equations in this thesis are setup to use standard MKS units.

There is a complexity for a *moving* vehicle in that the three beacons are not heard by the vehicle at the same location, because the vehicle moves some distance in the time interval between when the first and last beacons are heard. Incorporating this complexity into the calculation adds little insight, so for purposes of this thesis, it will be assumed that the travel times have already been back-corrected so that the travel times from all three beacons correspond to the vehicle being at the same location.

2.2.1 Vehicle Sitting Still

In the first example, the beacons are arranged as shown by the grey spots in Fig. 2-7. The beacons transmit 21 times during this mission, while the vehicle attempts to hold station at $(x, y) = (1000, -1500)$, the point indicated by the circle labeled "vehicle" on

Fig. 2-7. The vehicle will attempt to estimate both position and time synchronization, and the accuracy of its position estimate will be graphed over the course of the 21 receptions. The position estimate accuracy will be compared with the hyperbolic and spherical bounds for three different sizes of clock drift errors. The system errors are chosen to be realistic for a vehicle receiving transmissions every 50 seconds and using a crystal oscillator as its time base. The system errors are as follows:

- Initial position variance (m^2): 100^2 (in each axis)
- Initial time sync. variance (s^2): 1^2
- Travel time measurement variance (s^2): 0.001^2
- Increase in position variance between transmissions (m^2): 20^2 (in each axis)
- Increase in time sync. variance between transmissions (s^2): $(10^{-4})^2, (10^{-3})^2, (10^{-2})^2$

The position errors for spherical and hyperbolic navigation are shown in Fig. 2-8 in solid lines (labeled to the right of the graph). These lines are lower and upper bounds respectively on the position error of the system which attempts to estimate time synchronization. The lower the drift which the receiver clock experiences, the closer to the spherical performance limit the position estimates come. Roughly speaking, the time synchronization estimate is only useful if the clock drift per second converted to an equivalent range drift per second is smaller than range drift of the receiver navigation sensors.

$$c\Delta\tau < \Delta\mathbf{x}, \tag{2.26}$$

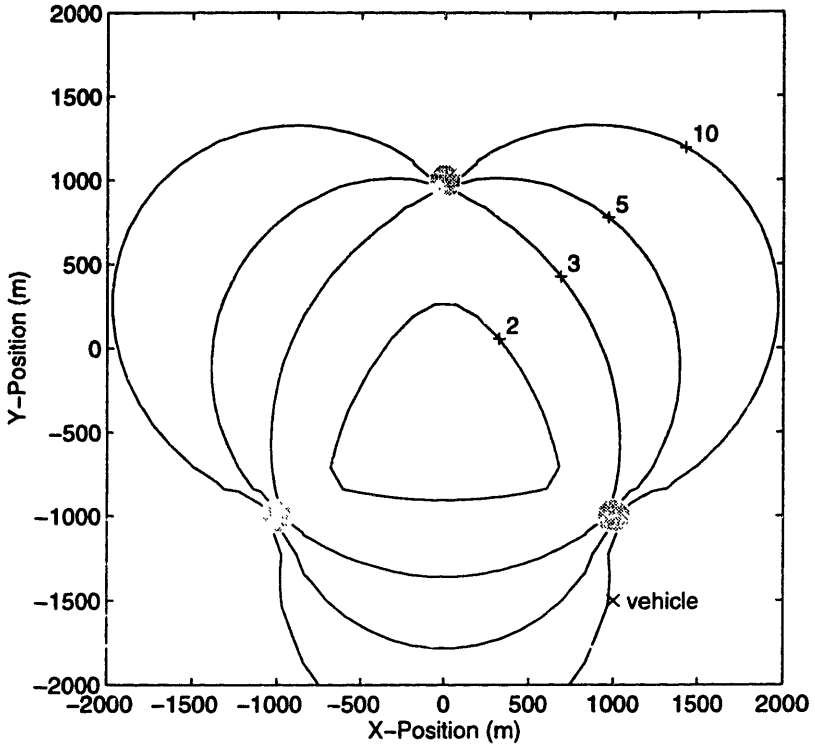

Figure 2-7: Vehicle location.

where c is the speed of sound. When this condition fails, time synchronization information is lost too quickly to be useful compared to position information. A consequence of this is that while this technique is helpful in improving resolution for acoustic positioning, where $c = 1500m/s$, it provides little improvement for radio positioning systems like LORAN, where $c = 3 \times 10^8 m/s$, without a very stable time base.

2.2.2 Vehicle Moving

In this second example, the beacon arrangement is the same, but the vehicle will be moving. The initial transmission will be heard by the vehicle at $(x, y) = (-1000, -1500)$. Between transmissions the vehicle will move 100m in the x-direction, so that on the

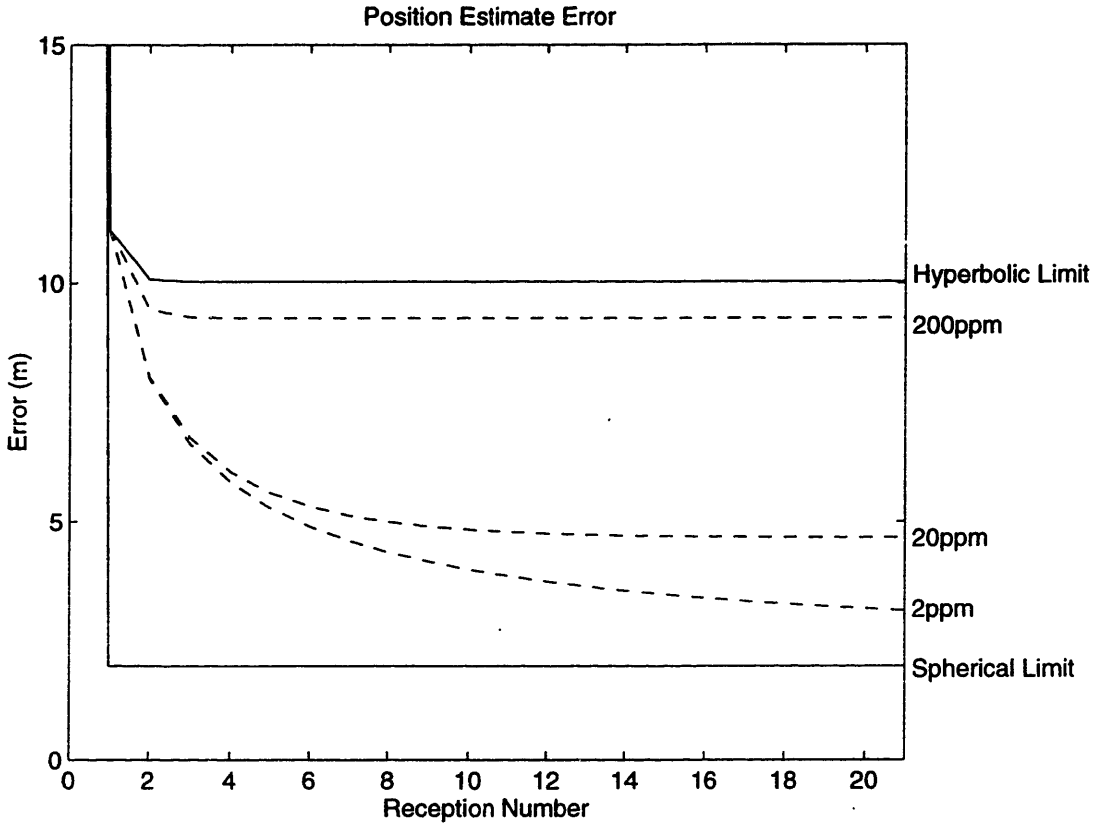

Figure 2-8: Position estimate accuracy.

last transmission it ends its mission at $(x, y) = (1000, -1500)$. This vehicle path is shown in Fig. 2-9, with the dots indicating the points at which transmissions are heard. The accuracy of the position estimate will be graphed over the course of the 21 receptions. The system errors are as in the previous example and are reasonable for a dead reckoning vehicle traveling at 2m/s.

The results of this second mission are shown on Fig. 2-10. The old bounds for the stationary source are shown in light dotted lines for comparison. Note that the ultimate hyperbolic and spherical performance bounds are nearly identical to those achieved in the previous example by a vehicle simply waiting at the destination point. However, the time estimating system ends up with better performance because of the trip it took. The time synchronizing system was able to get improved time information near the middle of its path. Although the benefit of the accurate positions was lost due to higher drift rates in position sensors, the benefit of the accurate time synchronization remained.

2.3 Optimal Navigation

The vehicle has the ability to move, and, in general, some flexibility in the path it selects from its origin to its destination. In this section, a set of feasible vehicle paths is parameterized, and the path is selected from this set which results in *the least position estimate error at the end of the journey.* The selection of such a path is known as the optimal navigation problem.

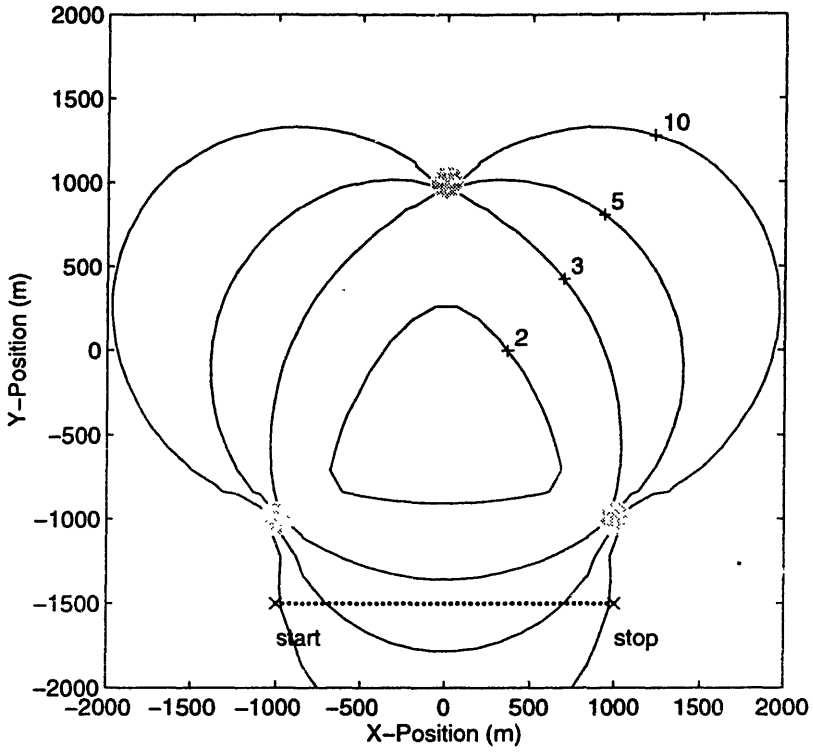

Figure 2-9: Vehicle path.

2.3.1 Path Parameterization

The set of feasible vehicle paths must be parameterized in some way so that the set can be searched to find the optimal path. Path parameterization begins with identifying the features of the path which are significant to the solution of the problem. Information is only obtained along the path when the vehicle receives the beacon transmissions, and thus the only significant points on the vehicle path are the points at which reception occurs. The path is therefore represented as a set of connected reception points. It is assumed that the source is transmitting at evenly spaced time intervals, so the maximum vehicle speed imposes a constraint on the maximum separation between consecutive reception points. A mission time duration specifies the total number of reception points in the path, and the starting point and final destination point are

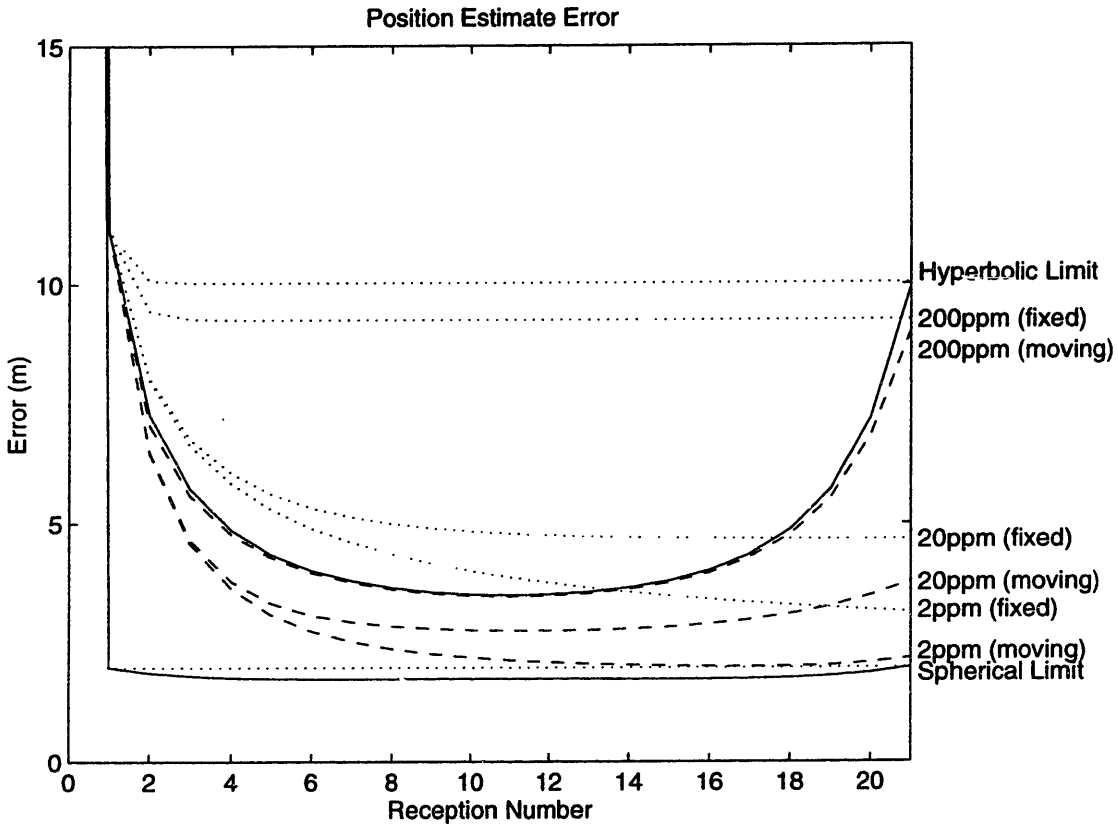

Figure 2-10: Position estimate accuracy comparison.

fixed. The path constraints are listed below:

- The time between vehicle transmissions is a constant Δt.
- The path will have a fixed number of reception points K, which is equivalent to constraining the total mission time to be $(K-1)\Delta t$
- The first transmission point will be at a fixed origin, and the last transmission point will be at a fixed destination.
- The vehicle has a maximum speed v, so the maximum separation between transmission points is $v\Delta t$.

Thus the feasible path consists of a set of K transmission points with the first and last of these points fixed and the maximum distance between consecutive points constrained.

2.3.2 Optimization Algorithm

A simulated annealing algorithm was used to find the optimal path [28, 34]. An initial feasible path is chosen as the working path, and the objective is evaluated for this path. In this case, the initial working path is taken to be the straight path between the origin and the destination, and the objective is the final position error at the destination point.

Next, one of the points in the working path is then perturbed to form a new path. The point of the perturbation is chosen randomly, with all points being equally likely (excluding the start point and end point, which are fixed). The new location of the

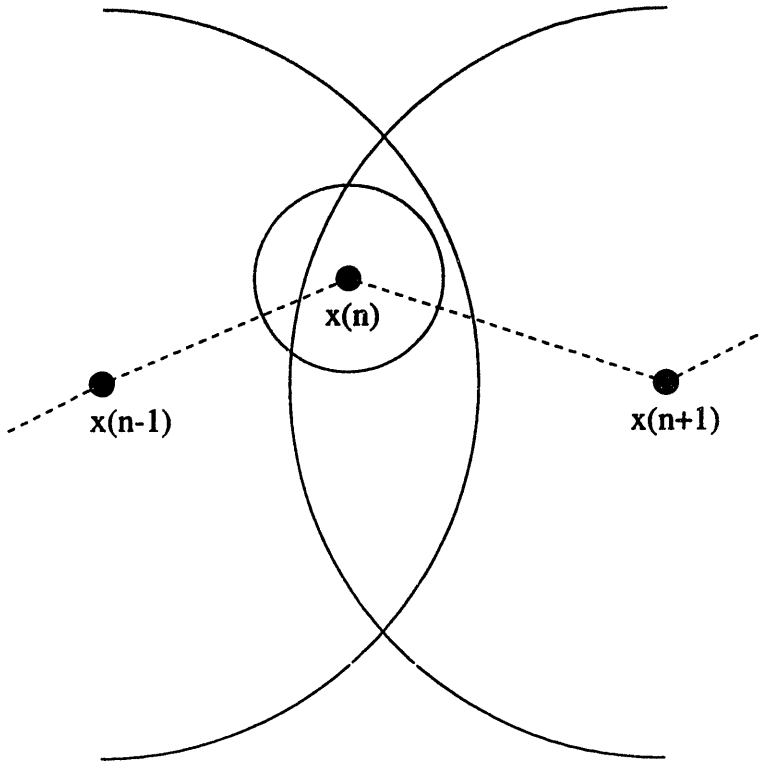

Figure 2-11: Feasible perturbation region.

perturbed point is constrained. It must be no further than $v\Delta t$ from the previous and following points in the path. It is also desirable to decrease the perturbation size as the annealing process continues. To accomplish this, the perturbed point is further constrained on the kth iteration to lie no further than a distance $\rho(k)$ from its original location, where $\rho(k) = \rho_0\alpha^k$ for some α.

The combination of these constraints means that the perturbed point lies in the intersection of three circles: one circle centered at the previous point and having radius $v\Delta t$, one circle at the following point and having the same radius, and one circle centered at the current point location and having radius $\rho(k)$. These constraints are shown in Fig. 2-11.

The desired probability distribution for selecting the perturbed point is one uni-

formly distributed over the feasible region and zero elsewhere. This is difficult to create numerically, so instead, a rectangular region is found which contains the feasible region, and points are guessed with a uniform distribution over this rectangular region until one is found which lies in the feasible region. The rectangular bounding region is found by taking an axis connecting the points preceeding and following the current point. Let d be the length of this axis. The zero of this axis is taken to be the preceding point with positive numbers in the direction of the following point. The coordinate of all feasible points along this axis must lie in the interval, $[d - v\Delta t, v\Delta t]$. A second axis is defined perpendicular to the first, and having its zero at its intersection with the first axis. The coordinate of all feasible points along this second axis must lie in the interval, $\left[-\sqrt{r_{max}^2 - d^2}, \sqrt{r_{max}^2 - d^2}\right]$. The maximum perturbation constraint consists of a circle of radius $\rho(k)$ around the current point. This circle is approximated by the square with sides $2\rho(k)$ which contains it. The intersection of the rectangle from the maximum separation constraint and the square from the maximum perturbation constraint defines a rectangular region which bounds the feasible region. This combined constraint is the shaded region in Fig. 2-12. Perturbed points are guessed with uniform distribution in this rectangular region until one is found which lies in the feasible region.

Once the perturbed path is created, its objective function is evaluated. If the perturbed path achieves a lower (better) objective function value, then the perturbed path becomes the new working path. If the perturbed path achieves a higher (worse) objective function value, then it is only chosen as the new working path with a probability of

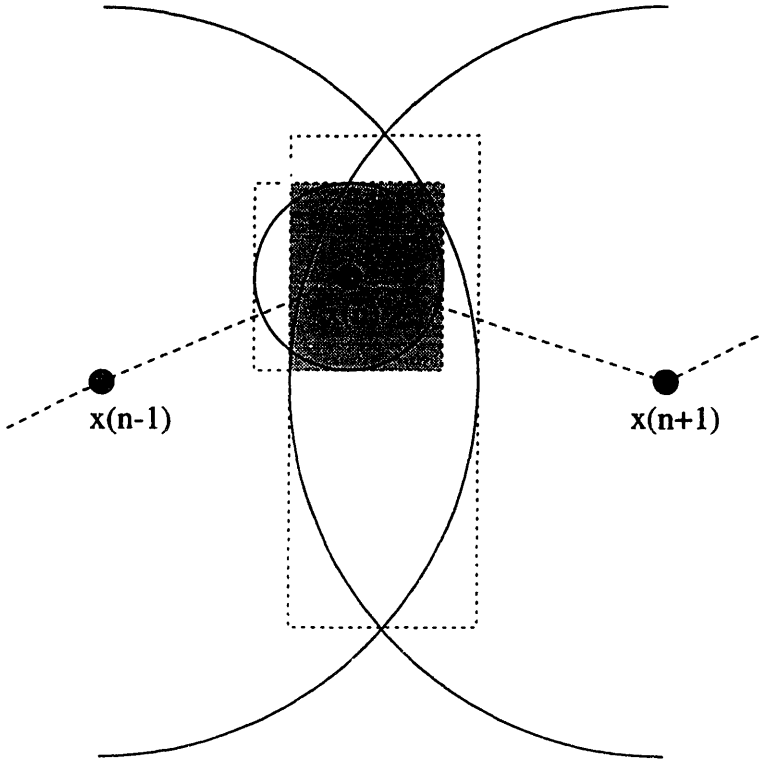

Figure 2-12: Feasible region with bounding rectangle.

$$exp(-\frac{\Phi_{new} - \Phi_{old}}{T(k)}) \tag{2.27}$$

where Φ_{new} and Φ_{old} are the new and old values of the objective function. and $T(k)$ is a temperature term at the kth iteration, $T(k) = T_0\alpha^k$, where α will be the same value as was used in the definition of the maximum perturbation $\rho(k)$. Note that the probability of accepting the path with the higher (worse) objective depends on two factors. First, it depends on how much worse the perturbed path is. It is more likely to choose a path which is a little bit worse than one which is much worse. Second, it depends on the iteration number k. As the algorithm proceeds, it becomes less likely to accept a worse path. This technique of selecting at times worse paths is intended to allow the minimization algorithm to escape from local minima [34].

The algorithm terminates after a fixed number of iterations, and is repeated 5 times to insure that the same optimal path is achieved each time. If the solutions from each run are not the same, then the number of iterations for each run is increased until a consistant answer is obtained.

2.3.3 Examples of Optimal Navigation

The simulated annealing algorithm is now applied to find optimal navigation paths, for the conditions below.

- Maximum vehicle speed: 2 m/s
- Time Between Transmissions: 50 s
- Number of Transmissions K: 31
- Initial position variance (m^2): 100^2 (in each axis)
- Initial time sync. variance (s^2): 1^2
- Travel time measurement variance (s^2): 0.001^2
- Increase in position variance between transmissions (m^2): 20^2 (in each axis)
- Increase in time sync. variance between transmissions (s^2): $(10^{-4})^2$
- Number of Iterations N: 10000
- Initial search radius ρ_0: $500m$
- Initial temperature T_0: 1

- Decay Coefficient α: $0.01^{(1/N)}$

There are two kinds of variables in the vehicle state vector for the optimal navigation case, position and time synchronization. The position estimate error grows rather quickly with time, so that previous position estimates contain far less information about present position than does the present measurement. Time synchronizations, however, drift slowly, so that an accurate time synchronization obtained at some past time continues to be useful well into the future. The usefulness of accurate time synchronization can be seen by comparing the hyperbolic and spherical Cramér-Rao bounds described earlier.

The resulting optimal path is shown in Fig. 2-13 superimposed on the bounds for time synchronization. The vehicle diverts toward the center of the array to acquire better time synchronization. This time synchronization results in much improved resolution in the region near the baseline extension, where hyperbolic positioning would be quite poor.

If the path length is extended to 51 transmissions, the diversion toward the center of the array becomes more clear, as seen in Fig. 2-14. The vehicle in fact moves at its full speed to the center of the array, then sits there, and then moves at full speed to the destination point.

Finally, in Fig. 2-15 the start point of the path is changed. Again, the vehicle travels to the center of the array and sits there before proceeding to its destination.

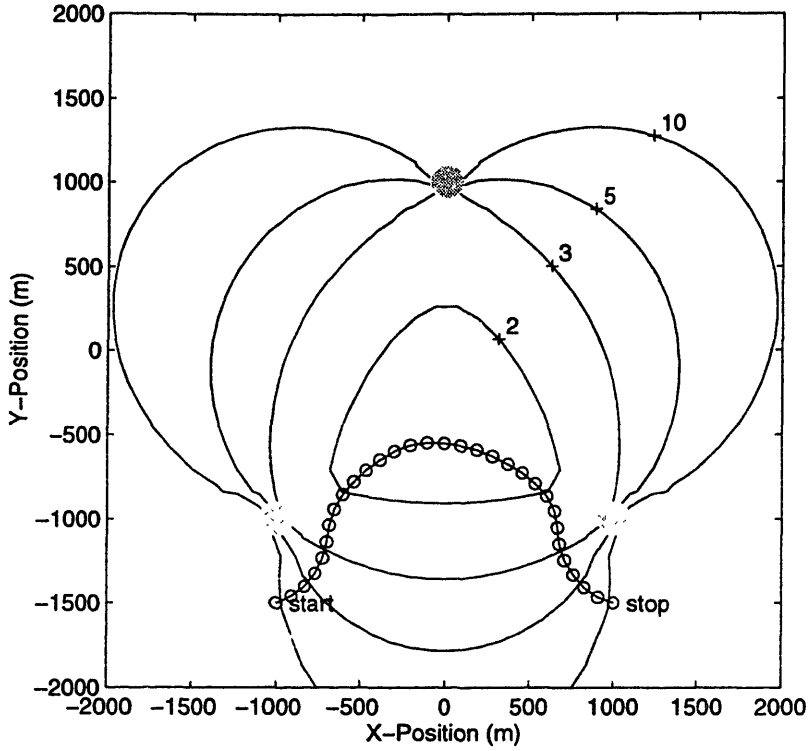

Figure 2-13: Optimal navigation, 31 transmissions.

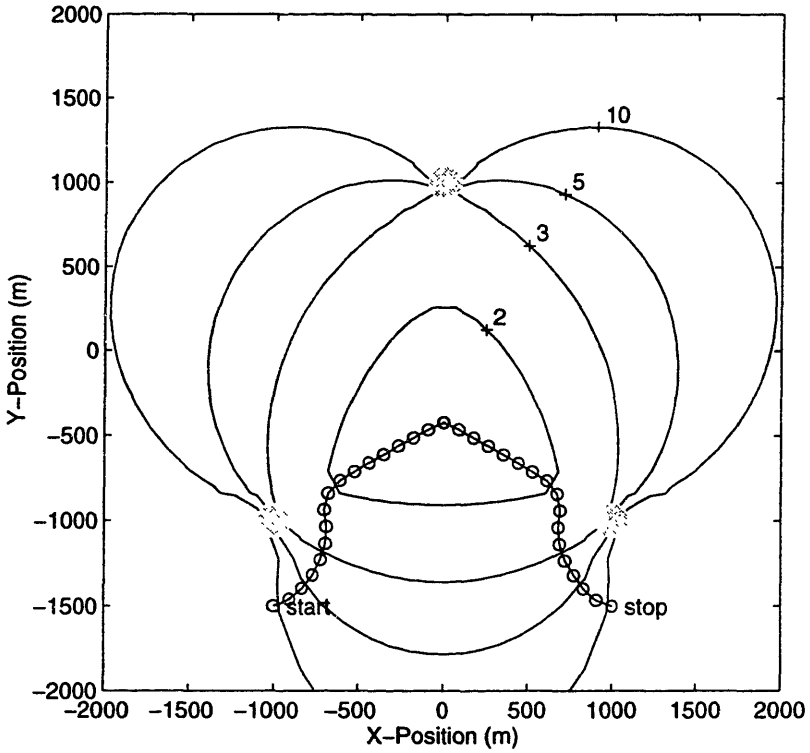

Figure 2-14: Optimal navigation, 51 transmissions.

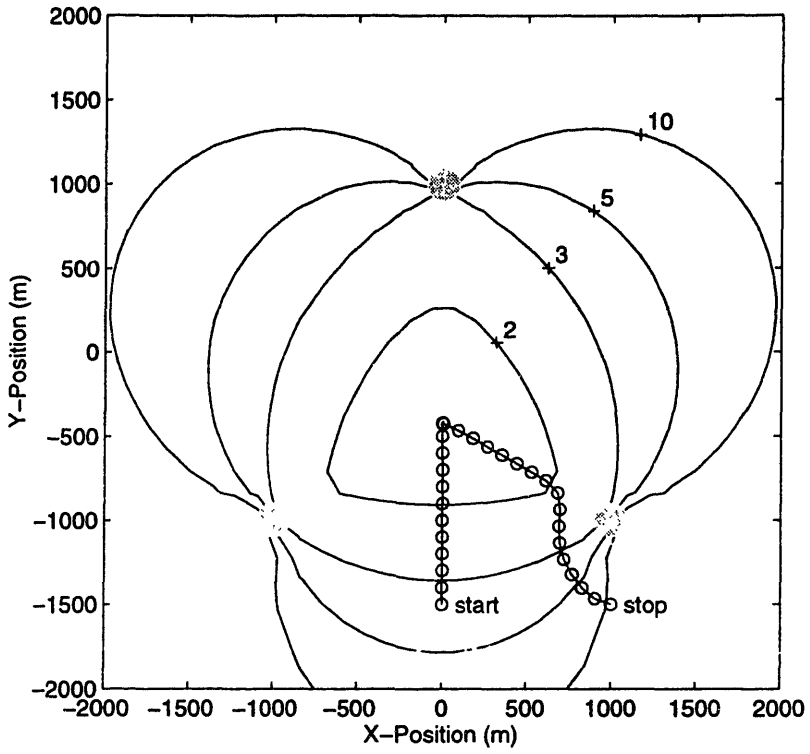

Figure 2-15: Optimal navigation, 51 transmissions, different starting point.

2.4 Conclusion

Spherical and hyperbolic positioning systems represent end points of a continuum of possible systems. This continuum is parameterized by the variance of the transmit time synchronization error. By estimating the time synchronization in a hyperbolic navigation system which transmits at regular intervals, it is possible to improve the position resolution, possibly up to the re- -lution which a spherical system would provide. To achieve substantially improved resolution, however, the drift rate of the clocks multiplied by the sound speed must be small compared to the position drift of the inertial position sensors. In acoustic systems, clock drift multiplied by sound speed is generally much smaller than the drift in position sensors, so time synchronizing position estimation works well. Another consequence of the slow drift of sensors is that

the position estimate at a destination point depends not only on local measurement geometry, but also on the path followed in reaching the destination. The problem of optimal navigation is to find the vehicle path which provides the least error in the final position estimate at the destination. This problem was solved for several examples, and it was found that the optimal path was one which diverted toward the center of the positioning array where the most accurate time synchronization estimates could be obtained.

Chapter 3

Sound Speed Estimation

Ocean acoustic tomography seeks to estimate variations in the sound speed profile within a volume by measuring travel time changes in rays traversing the volume. The analysis of this estimation problem begins with quantifying the travel time effect of an arbitrary sound speed perturbation on an eigenray in section 3.1. Toward this end, ray sampling functions are derived such that the projection of a sound speed variation onto the ray sampling function is the travel time change produced in the ray by the sound speed variation. By discretizing the sound speed variation and the ray sampling function, the forward problem of determining travel time variation from sound speed variation can be written as a linear matrix equation.

The inverse problem of determining sound speed variation from travel time perturbations is considered in section 3.2. The first observation which is made about the inverse problem is that it is very underdetermined. There are far more dimensions of possible sound speed profile variation than there are eigenrays through a typical environment. For this reason, the sound speed profile variation is approximated by a

weighted sum of basis function, and the profile estimate is formed simply by estimating the weights for each basis function.

The fact that tomography works with a reduced order model has implications for the design of the optimal weight estimator–implications which have been largely overlooked in the acoustics literature. The commonly-used weight estimators are suboptimal because they do not account for sound speed profile variations orthogonal to the basis set, some of which may have a travel time effect disproportionate to their size because of the sensitivity characteristics of the tomography measurement. In section 3.3, the optimal weight estimator is derived.

The basis functions used to represent sound speed profile variability are usually chosen using the method of empirical orthogonal functions. However, the empirical orthogonal functions are a suboptimal basis because they are chosen without taking into account how accurately their weights can be measured by the tomography experiment. In section 3.4, an optimal orthogonal function expansion for sound speed profile variability is derived.

In section 3.5, the sound speed profile estimate accuracy using the optimal estimator and the optimal orthogonal function basis is compared with the estimate accuracy using conventional methods. The difference in performance is demonstrated for typical temperate and arctic environments.

The optimal estimator and optimal basis functions were derived under the assumption that the profile covariance matrix was known. In practice, the profile covariance matrix must be estimated from a limited set of historical profile measurements. In section 3.6, the effect of using estimated covariance matrices is considered.

Finally, in section 3.7, the advantage in coverage obtained by mounting tomography sources on moving platforms is demonstrated, and the question of how to optimally utilize moving sources to focus tomographic accuracy at features of interest is analyzed.

3.1 The Forward Problem

In ocean acoustic tomography, the variation $\mathbf{x}$ in a mean sound speed profile within a volume is estimated using a measurement $\mathbf{y}$ of the resulting variation in ray travel times through that volume. A consequence of Fermat's principle is that the size of the travel time variations can be linearly related to the size of the sound speed variations, assuming that the measured ray arrivals in the received acoustic signal can be correctly matched with predicted eigenray paths. Although the sound speed profile change is a continuous function, it is sufficient to treat it as a large vector of sound speed samples taken at closely spaced depths. This allows the linear relationship between $\mathbf{x}$ and $\mathbf{y}$ to be written as,

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{n}. \tag{3.1}$$

Unfortunately, there is noise $\mathbf{n}$ associated with making the measurement $\mathbf{y}$. In the tomography problem, the noise vector $\mathbf{n}$ includes both true measurement noise as well as any non-conformance of the true travel time perturbations to the linear model.

The ray sampling function is defined such that the projection of a sound speed profile perturbation onto the ray sampling function is the travel time change which that profile perturbation causes in the ray. In terms of (3.1), the ray sampling functions

are the rows of the measurement matrix **C**, and they can be calculated as follows. The travel time t_i along the ith ray which follows path S_i through an environment with sound speed profile $c(z)$ is [26],

$$t_i = \int_{S_i} \frac{1}{c(z)} ds \tag{3.2}$$

Add a small sound speed perturbation $x(z)$ to the sound speed profile, and write a Taylor expansion for the perturbed profile, retaining only the first two terms.

$$\frac{1}{c(z) + x(z)} \approx \frac{1}{c(z)} - \frac{x(z)}{c^2(z)}. \tag{3.3}$$

Now consider the effect of this profile perturbation on ray travel time. According to Fermat's principle, the time integral will be independent of a small perturbation in the integral path, so the time may be calculated *as if the path had been unchanged* by the sound speed perturbation. The perturbation in travel time for the ith ray will be $\mathbf{y_i}$.

$$t_i + \mathbf{y_i} = \int_{S_i} \frac{1}{c(z)} ds - \int_{S_i} \frac{x(z)}{c^2(z)} ds \tag{3.4}$$

The relation between the variations is then,

$$\mathbf{y_i} = -\int_{S_i} \frac{x(z)}{c^2(z)} ds \tag{3.5}$$

Changing this path integral to an integral over depth between the starting depth z_{bottom} and the ending depth z_{top} of a ray segment, and adding a term $k_i(z)$ which is

the number of times the ith ray passes the depth z,

$$\mathbf{y_i} = -\int_{z_{bottom}}^{z_{top}} \frac{x(z)k_i(z)}{c^2(z)|sin(\theta(z))|}dz \tag{3.6}$$

where θ is the angle with respect to horizontal of the ray. Using Snell's law, and defining c_t as the sound speed for which the ray would become horizontal, *i.e.* the sound speed at a refractive turning point of the ray,

$$\mathbf{y_i} = -\int_{z_{bottom}}^{z_{top}} \frac{x(z)k_i(z)}{c^2(z)\sqrt{1-\frac{c^2(z)}{c_t^2}}}dz \tag{3.7}$$

For purposes of this thesis, this integral equation must be written in a finite difference form so that the continuous integral above can be represented by the discrete equation,

$$\mathbf{y} = \mathbf{Cx} \tag{3.8}$$

The integral could be put into finite difference form by letting the ith row of $\mathbf{C}$ which corresponds to the ith eigenray be,

$$\mathbf{C_{in}} = -\frac{k_i(n\Delta z)}{c^2(n\Delta z)\sqrt{1-\frac{c^2(n\Delta z)}{c_t^2}}}\Delta z \tag{3.9}$$

However, this function has a singularity at the turning depth of the ray, so the discrete approximation becomes poor near the turning depth. Since the singularity is integrable, and the ultimate goal is to approximate the integral in (3.7), a better solution

is to let $\mathbf{C_{in}}$ be the integral in (3.7) over a depth interval $(n-1)\Delta z$ to $n\Delta z$,

$$\mathbf{C_{in}} = -\int_{(n-1)\Delta z}^{n\Delta z} \frac{k_i(z)}{c^2(z)\sqrt{1-\frac{c^2(z)}{c_t^2}}} dz. \tag{3.10}$$

For sufficiently fine sampling of the profile (small Δz), we may assume that the sound speed profile changes linearly between our sample points, so that $c(z) = c((n-1)\Delta z) + \beta(z-(n-1)\Delta z)$, where $\beta = (c(n\Delta z) - c((n-1)\Delta z))/(\Delta z)$. The next step will be changing the variable of integration from z to c. The variable $k(z)$ has thus far prevented the integrand from having a non-zero value outside of the depth range of the ray. When $k(z)$ becomes discretized to $k(n\Delta z)$, care will have to be taken with the limits of integration so that the integral does not extend outside the depth range of the ray.

$$\mathbf{C_{in}} = -\frac{k_i(n\Delta z)}{\beta} \int_{min(ct,c((n-1)\Delta z))}^{min(ct,c(n\Delta z))} \frac{1}{c^2\sqrt{1-\frac{c^2}{c_t^2}}} dc. \tag{3.11}$$

Using the integral [23],

$$\int \frac{dx}{x^2\sqrt{a^2-x^2}} = -\frac{\sqrt{a^2-x^2}}{a^2 x}. \tag{3.12}$$

the expression for $\mathbf{C_{in}}$ becomes

$$\mathbf{C_{in}} = \frac{k_i(n\Delta z)}{\beta} \left[\frac{\sqrt{1-\frac{c^2}{c_t^2}}}{c} \right]_{min(ct,c((n-1)\Delta z))}^{min(ct,c(n\Delta z))}. \tag{3.13}$$

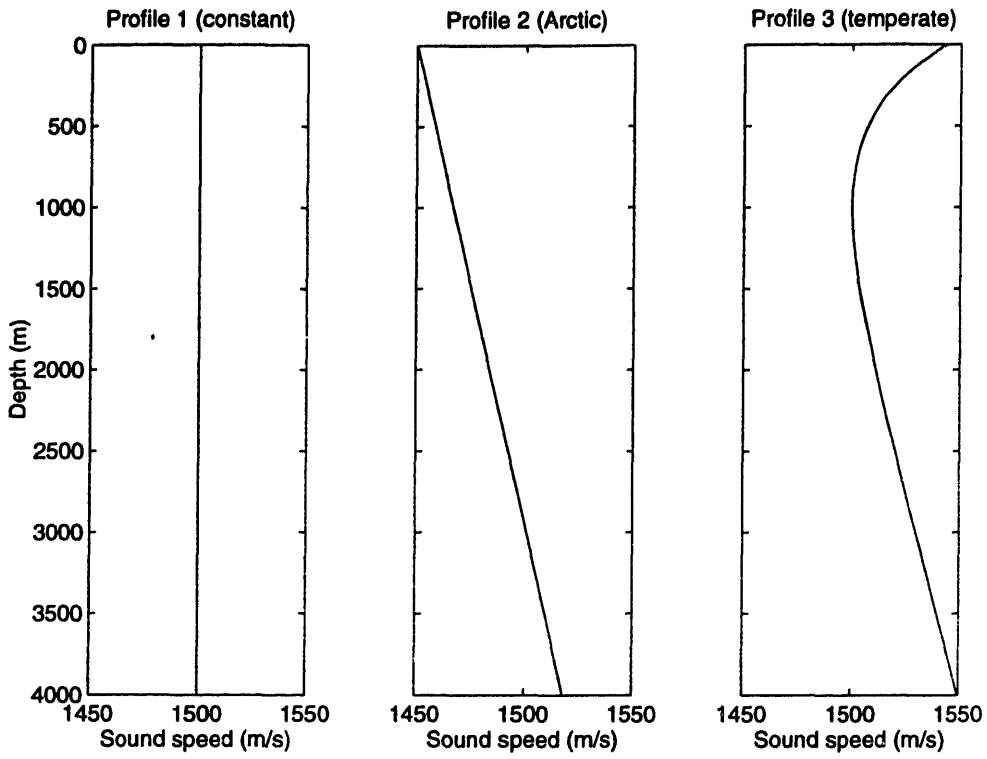

Figure 3-1: Profiles for the three ray sampling functions.

3.1.1 Ray Sampling Functions

Ray sampling functions are shown in Figs. 3-2, 3-3, and 3-4 for rays in environments with the three sound speed profiles shown in Fig. 3-1.

The first example deals with Profile 1 in Fig. 3-1. In this case the sound speed is constant everywhere in the environment. The ray which will be analyzed is the one shown in the right section of Fig. 3-2. This ray leaves a source at 1000m depth and is received 100km away at 3000m depth. The ray does not pass through water below 3000m, and so the ray sampling function is equal to zero below this depth. The ray passes twice through water above 1000m depth, once on the way up and once on the way down, so the ray sampling function above 1000m is twice as large as it is between 1000m and 3000m.

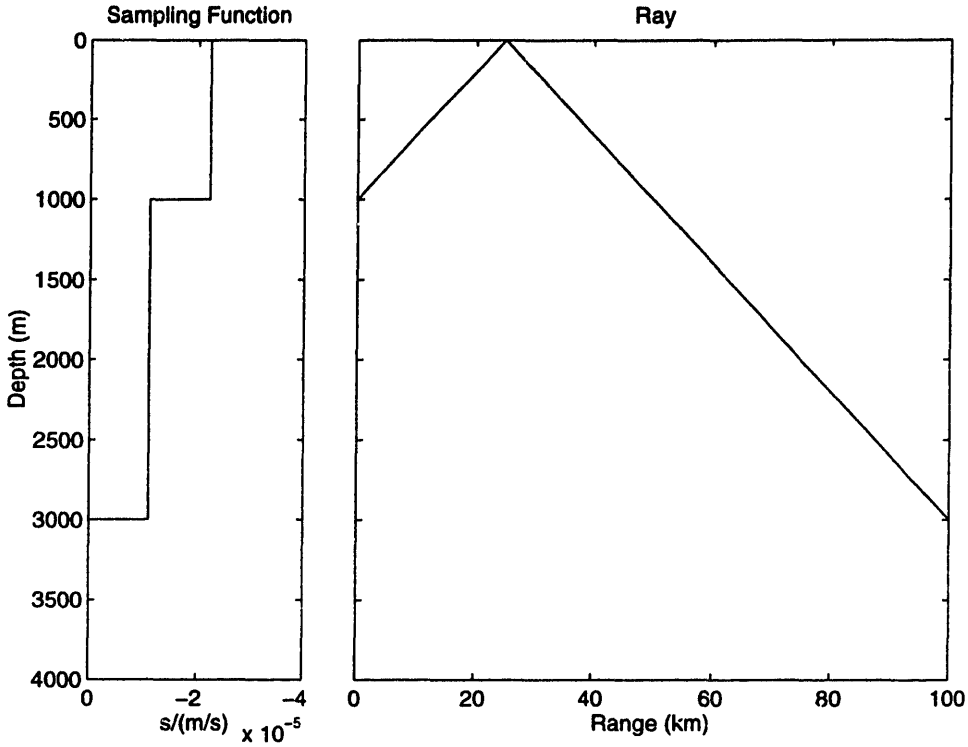

Figure 3-2: A sampling function and ray for profile 1–constant sound speed.

The second example deals with Profile 2 in Fig. 3-1. In this case the sound speed is a linearly increasing function of depth, a simple Arctic profile. The ray which will be analyzed is the one shown in the right section of Fig. 3-3. This ray leaves a source at 1000m depth, is turned by refraction, and is received 30km away at 1000m depth. Rays are most sensitive to sound speed variations at their turning depth, and this is reflected by the sampling function for this ray which becomes large at the turning depth of the ray. The sampling function is actually unbounded but integrable at the turning depth, however this discontinuity is removed while maintaining the same integrated travel time effect using the discretization proposed in (3.10). Note that the ray sampling function is again equal to zero for depths which the ray does not pass through.

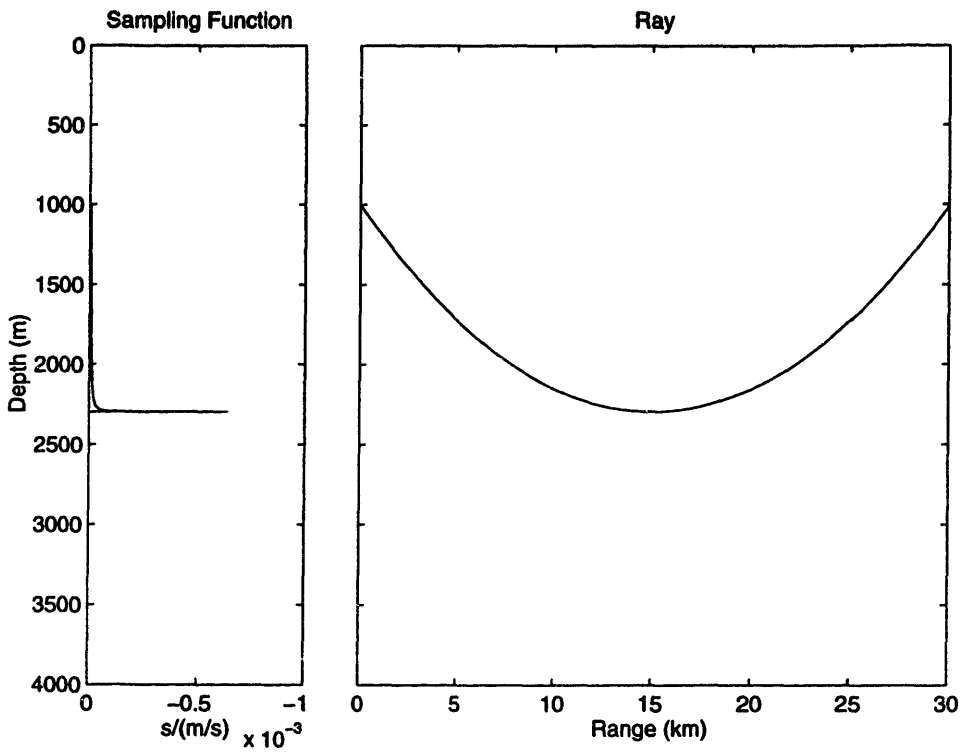

Figure 3-3: A sampling function and ray for profile 2–the Arctic profile.

The third example deals with Profile 3 in Fig. 3-1. In this case the sound speed has a minimum at 1000m depth, a simple temperate profile. The ray which will be analyzed is the one shown in the right section of Fig. 3-4. This ray leaves a source at 1000m depth, is turned by refraction several times both above and below the source depth, and is received 100km away at 1000m depth. Rays are most sensitive to sound speed variations at their turning depth, and this is reflected by the sampling function for this ray which becomes large at both the upper and lower turning depths of the ray. Note that the ray sampling function is 50% larger for depths just above 1000m than for depths just below 1000m. This is because the ray passes six times through depths above 1000m and only four times through depths below 1000m.

It should be noted that point measurements of sound speed can also be represented

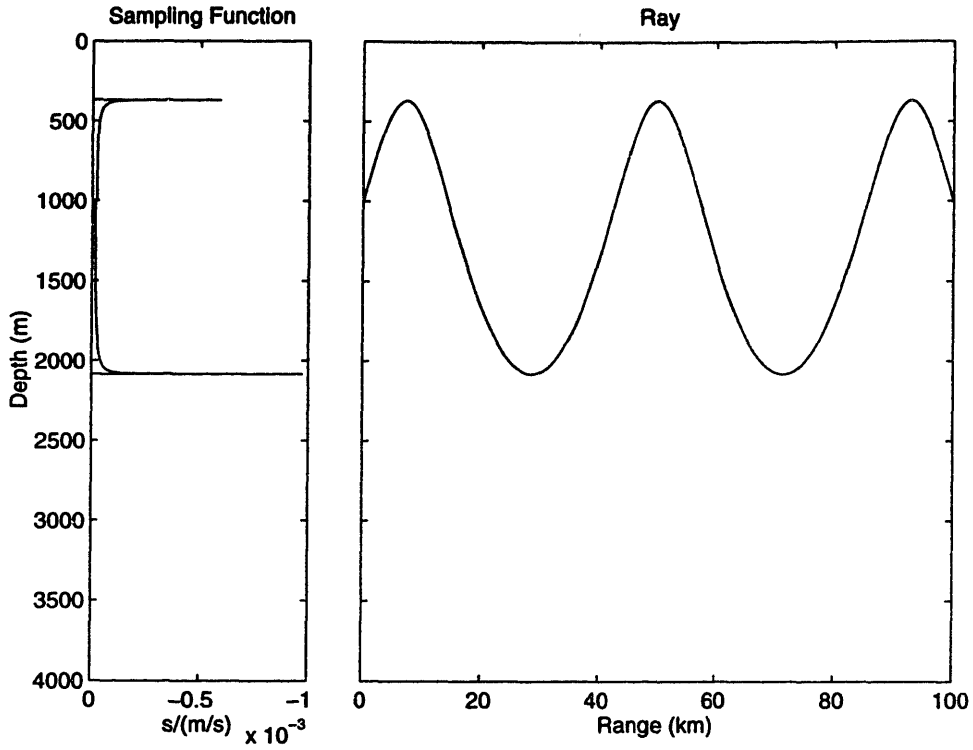

Figure 3-4: A sampling function and ray for profile 3–the temperate profile.

as sampling functions. The sampling function of a point measurement is simply a delta function at the depth of the measurement.

3.2 The Inverse Problem

The goal of tomography is to produce sound speed profile estimates within a volume based on a limited number of travel time measurements along eigenrays passing through the volume. As such, the tomography problem is grossly underdetermined [36]. However, most sound speed profile variability can be represented by a relatively small number of modes of variation about the local mean profile [32]. If the sound speed profile variation is approximated as a weighted sum of a small number of basis functions corresponding to the dominant modes of variation, then a reasonable profile

estimate can be obtained simply by estimating the weights for the basis functions.

The variation $\mathbf{x}$ in a mean sound speed profile within a volume is estimated using a measurement $\mathbf{y}$ of the resulting variation in ray travel times through that volume. In the linear model of (3.1),

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{n}, \tag{3.14}$$

it will be assumed that the sound speed perturbation vector $\mathbf{x}$ and the noise vector $\mathbf{n}$ are zero-mean Gaussian with covariance matrices $\mathbf{P} = \mathbf{E}\left[\mathbf{x}\mathbf{x}^{\mathbf{T}}\right]$ and $\mathbf{R} = \mathbf{E}\left[\mathbf{n}\mathbf{n}^{\mathbf{T}}\right]$.

The vector $\mathbf{x}$ has far more elements than the measurement $\mathbf{y}$ which contains only as many elements as there are eigenrays through the environment. As a linear equation the tomography problem is very underdetermined, and, were it not for knowledge of the covariance matrix $\mathbf{P}$, little could be said about $\mathbf{x}$ from the measurement $\mathbf{y}$. Knowledge about the sound speed profile covariance $\mathbf{P}$ is usually incorporated into the tomography problem by approximating $\mathbf{x}$ as a weighted sum of a small number of orthonormal basis vectors corresponding to the most important modes of variation in $\mathbf{x}$. If these basis vectors compose the columns of $\mathbf{\Phi}$ and the appropriate weight for each basis vector is contained in $\mathbf{a}$, then the approximation is made,

$$\mathbf{x} \simeq \mathbf{\Phi}\mathbf{a} \tag{3.15}$$

where $\mathbf{a} = \mathbf{\Phi}^{\mathbf{T}}\mathbf{x}$. The weight vector $\mathbf{a}$ is much smaller in size than the original unknown vector $\mathbf{x}$, and the inverse problem can be solved.

One step in designing a tomography experiment is to pick a set of basis functions

$\mathbf{\Phi}$ and specify a method for determining an estimate $\hat{\mathbf{a}}$ of the weights for those basis functions based on the measurement $\mathbf{y}$. For purposes of this thesis, the optimal choice of $\hat{\mathbf{a}}$ and $\mathbf{\Phi}$ will be defined as the one which minimizes the expected mean squared error e in the resulting profile estimate,

$$e = \mathbf{E}\left[(\mathbf{x} - \mathbf{\Phi}\hat{\mathbf{a}})^{\mathbf{T}}(\mathbf{x} - \mathbf{\Phi}\hat{\mathbf{a}})\right]. \tag{3.16}$$

This scalar error can be equivalently written as the trace of the covariance matrix,

$$e = \mathbf{tr}\ \mathbf{E}\left[(\mathbf{x} - \mathbf{\Phi}\hat{\mathbf{a}})(\mathbf{x} - \mathbf{\Phi}\hat{\mathbf{a}})^{\mathbf{T}}\right]. \tag{3.17}$$

Section 3.3 derives and analyzes the optimal weight estimator, and section 3.4 derives and analyzes the optimal set of basis functions for representing profile variability.

3.3 Weight Estimators

For now, let us set aside the question of how to select the basis vectors which compose $\mathbf{\Phi}$ and examine the choice of $\hat{\mathbf{a}}$. While novel if less direct techniques have been applied to ray-based inversions, such as neural networks [40] and simulated annealing [7], the most common estimators for sound speed profile weights are ones which express the weight estimate as a matrix function of the travel time variations [36].

Because the system is linear and the variables are Gaussian, the optimal estimator $\hat{\mathbf{a}}$ in the mean squared error sense will also be the linear least squares estimator [15],

so with no loss of generality, it may be assumed that

$$\hat{\mathbf{a}} = \mathbf{K}\mathbf{y}. \tag{3.18}$$

The problem of choosing the optimal estimator is reduced to one of selecting a gain matrix $\mathbf{K}$.

3.3.1 Suboptimal estimators

After selecting a basis set, it is common practice to rewrite (3.1) and (3.16) using the reduced order model for $\mathbf{x}$ in (3.15). (3.1) then becomes

$$\mathbf{y}_r = \mathbf{C}_r\mathbf{a} + \mathbf{n} \tag{3.19}$$

where $\mathbf{C}_r = \mathbf{C}\boldsymbol{\Phi}$. Using (3.15) and the fact $\boldsymbol{\Phi}$ was defined to have orthonormal columns so that $\boldsymbol{\Phi}^{\mathbf{T}}\boldsymbol{\Phi} = \mathbf{I}$, (3.16) becomes

$$e = \mathbf{E}\left[(\mathbf{a} - \hat{\mathbf{a}})^{\mathbf{T}}(\mathbf{a} - \hat{\mathbf{a}})\right], \tag{3.20}$$

or equivalently,

$$e = \mathbf{tr}\ \mathbf{E}\left[(\mathbf{a} - \hat{\mathbf{a}})(\mathbf{a} - \hat{\mathbf{a}})^{\mathbf{T}}\right]. \tag{3.21}$$

Basing the weight estimate on the reduced order measurement model of (3.19), $\hat{\mathbf{a}} = \mathbf{K}\mathbf{y}_r$, the error objective e in (3.21) can be rewritten,

$$e = \mathbf{tr}\left[\mathbf{P}_r - \mathbf{K}\mathbf{C}_r\mathbf{P}_r - \mathbf{P}_r\mathbf{C}_r^{\mathbf{T}}\mathbf{K}^{\mathbf{T}} + \mathbf{K}(\mathbf{C}_r\mathbf{P}_r\mathbf{C}_r^{\mathbf{T}} + \mathbf{R})\mathbf{K}^{\mathbf{T}}\right], \tag{3.22}$$

where $\mathbf{P}_r = \mathbf{E}\left[\mathbf{a}\mathbf{a}^{\mathbf{T}}\right] = \mathbf{\Phi}^{\mathbf{T}}\mathbf{P}\mathbf{\Phi}$. The optimal gain matrix $\mathbf{K}$ is found by setting the derivative of this error objective with respect to the matrix $\mathbf{K}$ to zero. The following symbolic matrix derivatives are helpful in this calculation [18, 42, 19, 44].

$$\frac{d}{d\mathbf{A}}\mathbf{tr}\ [\mathbf{A}\mathbf{B}] = \mathbf{B}^{\mathbf{T}} \tag{3.23}$$

$$\frac{d}{d\mathbf{A}}\mathbf{tr}\ \left[\mathbf{B}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\right] = \mathbf{B}^{\mathbf{T}} \tag{3.24}$$

$$\frac{d}{d\mathbf{A}}\mathbf{tr}\ \left[\mathbf{A}\mathbf{B}\mathbf{A}^{\mathbf{T}}\right] = \mathbf{A}\mathbf{B} + \mathbf{A}\mathbf{B}^{\mathbf{T}} \tag{3.25}$$

Taking the derivative of (3.22) with respect to $\mathbf{K}$ and setting it to zero yields,

$$2\mathbf{P}_r\mathbf{C}_r^{\mathbf{T}} - 2\mathbf{K}(\mathbf{C}_r\mathbf{P}_r\mathbf{C}_r^{\mathbf{T}} + \mathbf{R}) = 0. \tag{3.26}$$

Solving for $\mathbf{K}$ produces the gain matrix for the reduced order estimator,

$$\mathbf{K} = \mathbf{P}_r\mathbf{C}_r^{\mathbf{T}}\left(\mathbf{C}_r\mathbf{P}_r\mathbf{C}_r^{\mathbf{T}} + \mathbf{R}\right)^{-1}. \tag{3.27}$$

Since $\mathbf{C_r} = \mathbf{C\Phi}$ and $\mathbf{P_r} = \mathbf{\Phi^T P \Phi}$, this equation can also be written in expanded form,

$$\mathbf{K} = \mathbf{\Phi^T P \Phi \Phi^T C^T} \left(\mathbf{C \Phi \Phi^T P \Phi \Phi^T C^T} + \mathbf{R} \right)^{-1}. \tag{3.28}$$

3.3.2 Optimal Estimator

If the full measurement equation is retained instead of making the reduced order approximation, a different answer is found [12]. The estimate $\hat{\mathbf{a}}$ will be chosen to minimize the original error objective in (3.17),

$$e = \mathbf{tr}\ \mathbf{E} \left[(\mathbf{x} - \mathbf{\Phi \hat{a}})(\mathbf{x} - \mathbf{\Phi \hat{a}})^{\mathbf{T}} \right], \tag{3.29}$$

and the weight estimator will be based on the full order measurement model of (3.1), so $\hat{\mathbf{a}} = \mathbf{Ky}$. In this case, the error objective which $\mathbf{K}$ must minimize is

$$e = \mathbf{tr}\ \left[\mathbf{P} - \mathbf{\Phi K C P} - \mathbf{P C^T K^T \Phi^T} + \mathbf{\Phi K (C P C^T} + \mathbf{R) K^T \Phi^T} \right]. \tag{3.30}$$

The order of multiplications within a trace can be rearranged without changing the value of the trace, that is, for $\mathbf{A}$ an $n \times m$ matrix and $\mathbf{B}$ and $m \times n$ matrix, $\mathbf{tr\ AB} = \mathbf{tr\ BA}$. Rearranging (3.30), and recalling that the columns of $\mathbf{\Phi}$ are orthonormal so that $\mathbf{\Phi^T \Phi} = \mathbf{I}$, (3.30) becomes,

$$e = \mathbf{tr}\ \left[\mathbf{P} - \mathbf{K C P \Phi} - \mathbf{\Phi^T P C^T K^T} + \mathbf{K (C P C^T} + \mathbf{R) K^T} \right]. \tag{3.31}$$

Taking the derivative of this trace with respect to the matrix $\mathbf{K}$ with the help of the identities in (3.23), (3.24), and (3.25), and setting the derivative to zero yields

$$2\mathbf{\Phi}^{\mathbf{T}}\mathbf{P}\mathbf{C}^{\mathbf{T}} - 2\mathbf{K}(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathbf{T}} + \mathbf{R}) = 0. \tag{3.32}$$

Solving for the optimal gain matrix $\mathbf{K}$ produces,

$$\mathbf{K} = \mathbf{\Phi}^{\mathbf{T}}\mathbf{P}\mathbf{C}^{\mathbf{T}}\left(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathbf{T}} + \mathbf{R}\right)^{-1}. \tag{3.33}$$

Note that this is the Kalman gain for the full order system projected onto the basis $\mathbf{\Phi}$, so the estimate of the weights is the same as if the estimate of the full vector $\mathbf{x}$ were formed and then projected onto the selected basis.

Simple Comparison A simple example illustrates the advantage of the optimal estimator in (3.33) over the reduced order estimator in (3.27). Assume that the noise covariance matrix $\mathbf{R} = \mathbf{I}$, and

$$\mathbf{C} = \begin{bmatrix} 1 & 9 \end{bmatrix}, \qquad \mathbf{P} = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix}. \tag{3.34}$$

The dominant mode of variation in $\mathbf{x}$ is in the direction $\begin{bmatrix} 1 & 0 \end{bmatrix}^{\mathbf{T}}$, since this is the eigenvector corresponding to the larger eigenvalue of $\mathbf{P}$. However the matrix $\mathbf{C}$ makes the measurement vector $\mathbf{y}$ most sensitive to variations in $\mathbf{x}$ along the direction of the eigenvector corresponding to the smaller eigenvalue of $\mathbf{P}$, $\begin{bmatrix} 0 & 1 \end{bmatrix}^{\mathbf{T}}$.

The basis vector Φ will be chosen to represent the dominant mode of variation in $\mathbf{x}$

$$\Phi = \begin{bmatrix} 1 & 0 \end{bmatrix}^{\mathbf{T}}. \tag{3.35}$$

The two-element vector $\mathbf{x}$ will be approximated by a single weight applied to this basis function. The mean squared error for the resulting estimate of $\mathbf{x}$ is given in the table below when the weight is estimated using the suboptimal estimator and the optimal estimator.

Suboptimal	Optimal
67.5 (actual)	9.1
0.9 (predicted)	

There are two error terms listed under the suboptimal estimator. The error labeled "actual" is in fact the mean squared error achieved by the suboptimal estimator. Before the estimators are applied, the mean squared error associated with $\mathbf{x}$ is $\mathbf{tr}\ \mathbf{P} = 10$. The error after the application of the suboptimal estimator is 67.5, far worse than before. The job of an estimator–any estimator–is to account for all of the measurement vector. It does this by attributing some of the measurement to a weight change and the rest to noise. In the reduced order model used by the suboptimal estimator, neither the weight change nor the noise model accommodate well the large measurement changes produced by the unmodeled direction of variation in $\mathbf{x}$. Yet all measurement change

must be accounted for, so the estimator is forced to introduce large weight shifts to explain the measurement change within the context of the reduced order model.

Not only does the suboptimal estimator perform poorly, but within the context of the reduced order model, there is no indication of the performance problem. The error which would be predicted if the reduced order model were accurate is a very satisfying 0.9. The reduced order model takes no account of the unmodeled direction of variation in $\mathbf{x}$ which, though smaller than the modeled variation, has the greater effect on the measurement $\mathbf{y}$. The reduced order model not only results in an estimator with poor performance but also produces a false prediction of estimate accuracy.

The optimal estimator produces a realistic evaluation of its own performance, because it is based on a full order error model. The initial uncertainty in $\mathbf{x}$ was 10. Using the measurement, the optimal estimator is only able to reduce this to 9.1. This is a result of the fundamental ambiguity in the measurement, which the suboptimal estimator does not account for, since it uses a reduced order model where the ambiguity is not reflected. The optimal estimator makes the best use possible of the limited information afforded by the measurement.

In this example, a simple system was created specifically to reveal the weakness of the suboptimal estimator. The simulations in Section 3.5 will demonstrate that these problems are not only a property of carefully crafted examples, but in fact are present in realistic tomography experiments.

3.4 Basis Functions

Having determined the optimal estimator $\hat{\mathbf{a}}$ for the weights of arbitrary basis vectors $\mathbf{\Phi}$, the question of choosing the optimal basis vectors is now addressed. The earliest tomography experiments used constant velocity layers to parameterize sound speed [37]. Presently, the method of Empirical Orthogonal Functions (EOF) is commonly used in tomography to select a set of basis vectors to represent variations in the sound speed profile. This method of EOFs is known in other fields as principal component analysis or the Karhunen-Loéve expansion, and is well described in [27]. It seems to have been first developed by [29]. Early oceanographic applications were representing currents [30] and temperature fields [9], and since that time it has been widely used to represent sound velocity data. In the method of EOFs, historical profiles are used to estimate a profile covariance matrix, and the eigenvectors corresponding to the largest eigenvalues of the covariance matrix are taken for the basis vectors. First, the method of empirical orthogonal functions is reviewed. Then, a method for generating the optimal set of basis functions is derived. Finally, the functions generated by the two methods are tested in the simple example of the previous section, and their performance compared.

3.4.1 Empirical Orthogonal Functions

In selecting the optimal basis function, the same error objective from (3.17) will be used, except that now the free variable is the basis set $\mathbf{\Phi}$ instead of the estimator gain matrix $\mathbf{K}$. *Empirical orthogonal functions* (EOFs) are the set of basis functions which

minimize the error objective under the assumption that the weights are determined exactly, that is, $\hat{\mathbf{a}} = \mathbf{a} = \mathbf{\Phi}^{\mathbf{T}}\mathbf{x}$. The error objective is,

$$e = \mathbf{tr}\ \mathbf{E}\left[(\mathbf{x} - \mathbf{\Phi}\hat{\mathbf{a}})(\mathbf{x} - \mathbf{\Phi}\hat{\mathbf{a}})^{\mathbf{T}}\right]. \tag{3.36}$$

If it is assumed that the coefficient weights are determined correctly, then the objective function becomes,

$$e = \mathbf{tr}\ \left[\mathbf{P} - \mathbf{\Phi}\mathbf{\Phi}^{\mathbf{T}}\mathbf{P} - \mathbf{P}\mathbf{\Phi}\mathbf{\Phi}^{\mathbf{T}} + \mathbf{\Phi}\mathbf{\Phi}^{\mathbf{T}}\mathbf{P}\mathbf{\Phi}\mathbf{\Phi}^{\mathbf{T}}\right]. \tag{3.37}$$

Using the fact that within a trace the order of multiplication can be rearranged and the fact that $\mathbf{\Phi}^{\mathbf{T}}\mathbf{\Phi} = \mathbf{I}$, the error is rewritten,

$$e = \mathbf{tr}\ \left[\mathbf{P} - \mathbf{\Phi}^{\mathbf{T}}\mathbf{P}\mathbf{\Phi}\right]. \tag{3.38}$$

This function is clearly minimized if the basis vectors which form the columns of $\mathbf{\Phi}$ are the eigenvectors corresponding to the largest eigenvalues of the covariance matrix $\mathbf{P}$. These columns of $\mathbf{\Phi}$ are the EOFs. The first use of EOFs to represent sound speed profile variability seems to have been for the purpose of compressing large archives of historic profiles [32]. In this application, the EOFs are optimal, because the weights can be calculated directly as $\mathbf{a} = \mathbf{\Phi}^{\mathbf{T}}\mathbf{x}$.

3.4.2 Optimal Orthogonal Functions

Acoustic tomography differs from the data compression application of EOFs in that the weights are estimated from travel time variations rather than being calculated exactly. The travel time variations are not equally sensitive to all modes of profile variation, and as a result, all weights cannot be estimated with the same accuracy. The process of creating EOFs does not take into account these measurement resolution issues. This section presents a method for calculating *optimal orthogonal functions* (OOFs) which are the basis that provides the smallest mean squared error in the final profile estimate by taking into account not only the sizes of the different profile variations but also the ability of the tomographic experiment to measure the variations [12].

The error objective which will be minimized is, as before, (3.17),

$$e = \mathbf{tr}\ \mathbf{E}\left[(\mathbf{x} - \mathbf{\Phi}\hat{\mathbf{a}})(\mathbf{x} - \mathbf{\Phi}\hat{\mathbf{a}})^{\mathbf{T}}\right]. \tag{3.39}$$

In acoustic tomography, the weights must be estimated from the data, so $\hat{\mathbf{a}} = \mathbf{Ky}$. Using (3.1), this error objective can then be rewritten,

$$e = \mathrm{tr}\ \left[\mathbf{P} - \mathbf{\Phi KCP} - \mathbf{PC^T K^T \Phi^T} + \mathbf{\Phi K(CPC^T} + \mathbf{R)K^T \Phi^T}\right]. \tag{3.40}$$

Using the optimal $\mathbf{K}$ from (3.33), and rearranging the multiplication order within the trace, the error objective becomes,

$$e = \mathbf{tr}\ \left[\mathbf{P} - \mathbf{\Phi^T PC^T}\left(\mathbf{CPC^T} + \mathbf{R}\right)^{-1}\mathbf{CP\Phi}\right]. \tag{3.41}$$

By comparison with (3.38) it is clear that the function is minimized when the columns of $\mathbf{\Phi}$ are the eigenvectors corresponding to the largest eigenvalues of the matrix $\mathbf{PC^T}\left(\mathbf{CPC^T}+\mathbf{R}\right)^{-1}\mathbf{CP}$. These eigenvectors which comprise the columns of $\mathbf{\Phi}$ are the OOFs. They minimize an error function which contains information about both the size of the profile variations, $\mathbf{P}$, and the sensitivity of the tomography measurement to these variations, $\mathbf{C}$.

Simple Comparison The two techniques for creating basis sets are applied to the simple example problem where $\mathbf{R}=\mathbf{I}$, and

$$\mathbf{C}=\begin{bmatrix}1 & 9\end{bmatrix}, \qquad \mathbf{P}=\begin{bmatrix}9 & 0\\ 0 & 1\end{bmatrix} \tag{3.42}$$

The resulting EOF and the resulting OOF are given in the table below, along with the mean squared errors for both when used in conjunction with the optimal estimator.

Using EOFs	Using OOFs
$\mathbf{\Phi}=\begin{bmatrix}1\\ 0\end{bmatrix}$	$\mathbf{\Phi}=\begin{bmatrix}0.71\\ 0.71\end{bmatrix}$
$e=9.1$	$e=8.2$

The EOF is simply chosen to be the largest mode of variation in $\mathbf{x}$. The OOF includes a component of the larger mode as well as a component of the smaller mode in

$\mathbf{x}$. Because of the choice of $\mathbf{C}$, the smaller mode has a greater effect on the measurement than the larger mode, and the balance between mode size and measurement sensitivity in this example is such that the OOF happens to contain equal components of each mode.

The OOF improves the estimate by spanning the subspace where the greatest *reduction in error* is possible using the tomographic measurement rather than simply spanning the subspace where the greatest *a priori uncertainty* exists. Any two orthonormal bases which span the same subspace will produce the same estimate error. A consequence of this is that the improvement in mean squared error from using the OOF instead of the EOF will never be greater than the amount of variance in $\mathbf{x}$ unmodeled by the EOF. In this simple example, 10% of the variance in $\mathbf{x}$ is not modeled by the EOF, and the reduction in mean squared error is about 9% of the total variance in $\mathbf{x}$.

3.5 Examples

Two examples are given to demonstrate the effect of the estimator and basis choice on inversion accuracy in environments with canonical temperate and arctic sound speed profiles.

3.5.1 Temperate Example

In the first example, a 4000m deep temperate ocean with a Munk sound speed profile [38] is considered. The source and receiver are both at 1000m depth and are separated

by 100km. The sound speed profile and the eigenrays connecting source and receiver are shown in Fig. (3-5). Added to this mean profile are two possible sound speed profile perturbations which are shown in the right half of Fig. (3-6). One corresponds to a surface warming and the other to axial warming adjusted so that the two variations are orthonormal. If a profile covariance matrix were formed for this ocean, these would be its only two eigenvectors with non-zero eigenvalues. In this example, the eigenvalue or variance of the weight for the surface warming eigenvector will be 400, and for the axial warming 100. The basis functions are orthonormal, so the total variance in the profile is 500 meters-squared per second-squared, with 80% of the variance of the profile in the direction of the surface warming variation, and only 20% in the direction of the axial warming. The noise covariance matrix is $\mathbf{R} = (0.01)^2\mathbf{I}$ seconds-squared.

On the left side of Fig. (3-6) are the ray sampling functions shown on the same axes for all four eigenrays. These sampling functions are the rows of the matrix $\mathbf{C}$, so the inner product of each sampling function with a profile variation is the travel time change which that profile variation will cause in a particular eigenray [10]. These sampling functions are derived from the ray travel time equations and described in greater detail in the Appendix. It is important to note here that most of the area of the sampling functions is in the region where the axial warming is large and the surface warming variation is small. This means that the ray travel times will be more sensitive to the small axial warming variation than to the large surface warming change. This condition of being more sensitive to the small mode and less sensitive to the large mode is similar to that demonstrated in the simple example of the previous section.

In this example, a single basis function will be chosen to approximate the two

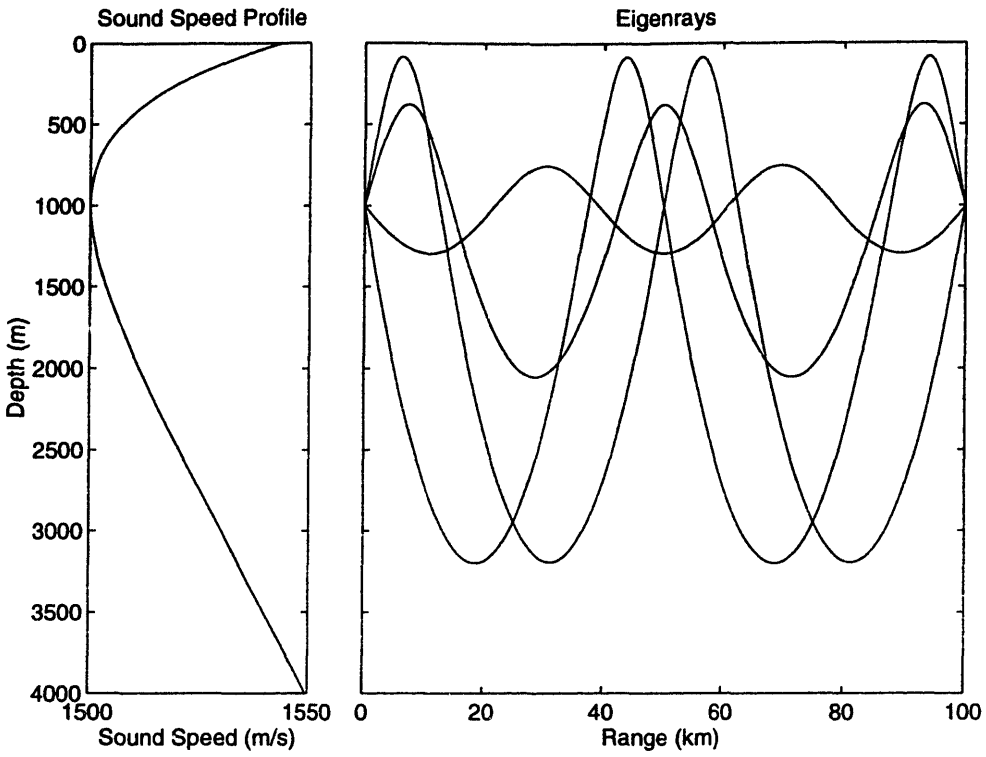

Figure 3-5: Sound speed profile and eigenrays for the temperate example.

independent modes of variation in the profile. When the method of EOFs is used to select this basis function, the mode with the larger eigenvalue, the surface warming mode, is chosen as shown on the left side of Fig. (3-7). The OOF method, however, selects a basis function which contains a component of both the larger mode and the more measurable mode. This OOF is shown on the right side of Fig. (3-7).

The table below shows the error in the profile estimate when the suboptimal estimator is used with the EOF and when the optimal estimator is used with both the EOF and then with the OOF.

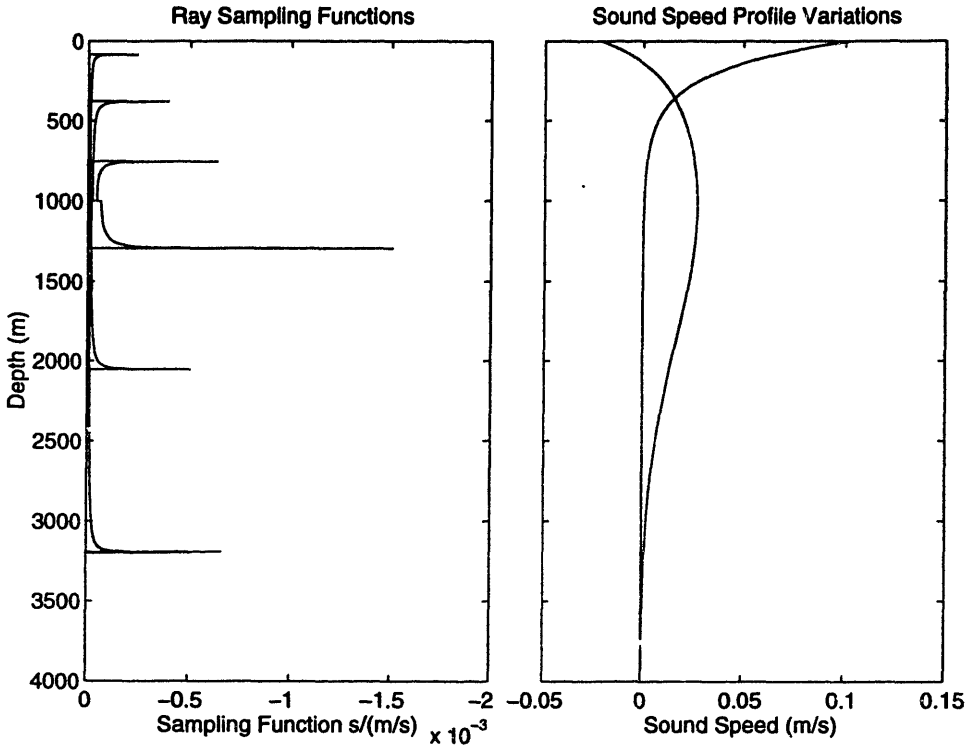

Figure 3-6: Sampling functions and profile variations for the temperate example.

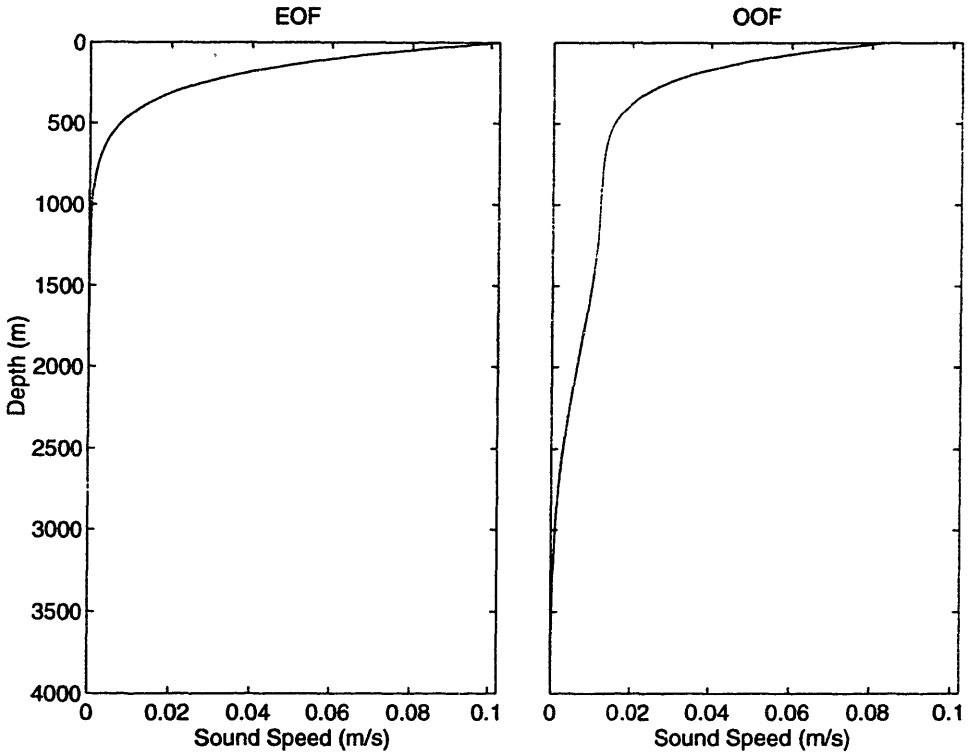

Figure 3-7: The EOF and OOF for the temperate example.

Total Variance of Sound Speed Profile Estimate $\left(\frac{m^2}{s^2}\right)$		
Suboptimal Estimator	Optimal Estimator	
EOF	EOF	OOF
447 (actual)		
	383	369
255 (predicted)		

As in the simple example, the suboptimal estimator predicts an error much smaller than its true error because it takes no account of the unmodeled mode of sound speed profile variation. The optimal estimator makes a significant improvement in estimate accuracy because it takes into account the effect of the unmodeled mode. Using the OOF with the optimal estimator makes a slight additional improvement. The EOF already represents 80% of the profile variation, so the maximum improvement possible with a different basis is 20% of the total variance. The OOF achieves a small improvement of about 3% of the initial total sound speed profile variance.

3.5.2 Arctic Example

In the second example, a 4000m deep Arctic ocean with a linear profile is considered. As before, the source and receiver are both at 1000m depth and are separated by 100km. The sound speed profile and the eigenrays connecting source and receiver are shown in Fig. (3-8). The same two orthonormal sound speed profile perturbations

are considered as shown in the right half of Fig. (3-9). As before, the variance of the weight for the surface warming eigenvector will be 400 and for the deep warming 100, so that the total profile variance is 500 meters-squared per second-squared, with 80% of the variance of the profile in the direction of the surface warming vector, and only 20% in the direction of the deep warming. As before, the noise covariance matrix is $\mathbf{R} = (0.01)^2\mathbf{I}$ seconds-squared. On the left side of Fig. (3-9) are the ray sampling functions shown on the same axes for all the eigenrays. It is important to note here that most of the area of the sampling functions is in the region where the deep warming is large and the surface warming variation is small. This means that the ray travel times will be more sensitive to deep warming than surface warming.

A single basis function will be chosen to approximate the two independent modes of variation in the profile. When the method of EOFs is used to select this basis function, the mode with the larger eigenvalue, the surface warming mode, is chosen as shown on the left side of Fig. (3-10). The optimal orthogonal function method, however, selects a basis function which contains a component of both the larger mode and more measurable mode. The OOF is shown on the right side of Fig. (3-10).

The table below shows the error in the profile estimate when the suboptimal estimator is used with the EOF and when the optimal estimator is used with the EOF and then with the OOF.

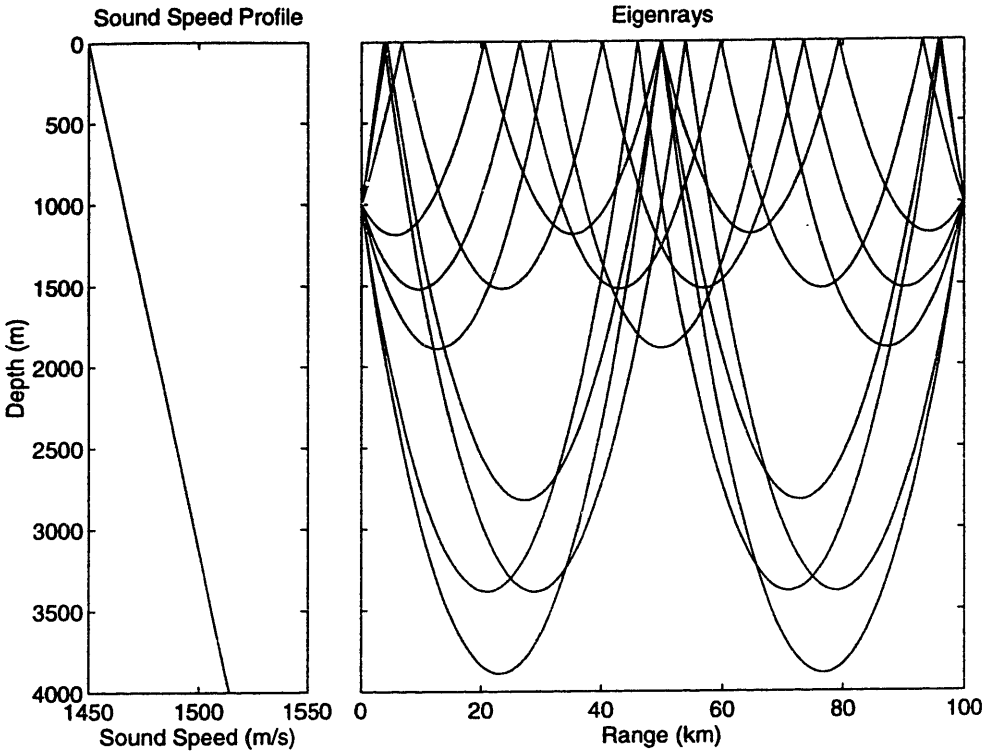

Figure 3-8: Sound speed profile and eigenrays for the Arctic example.

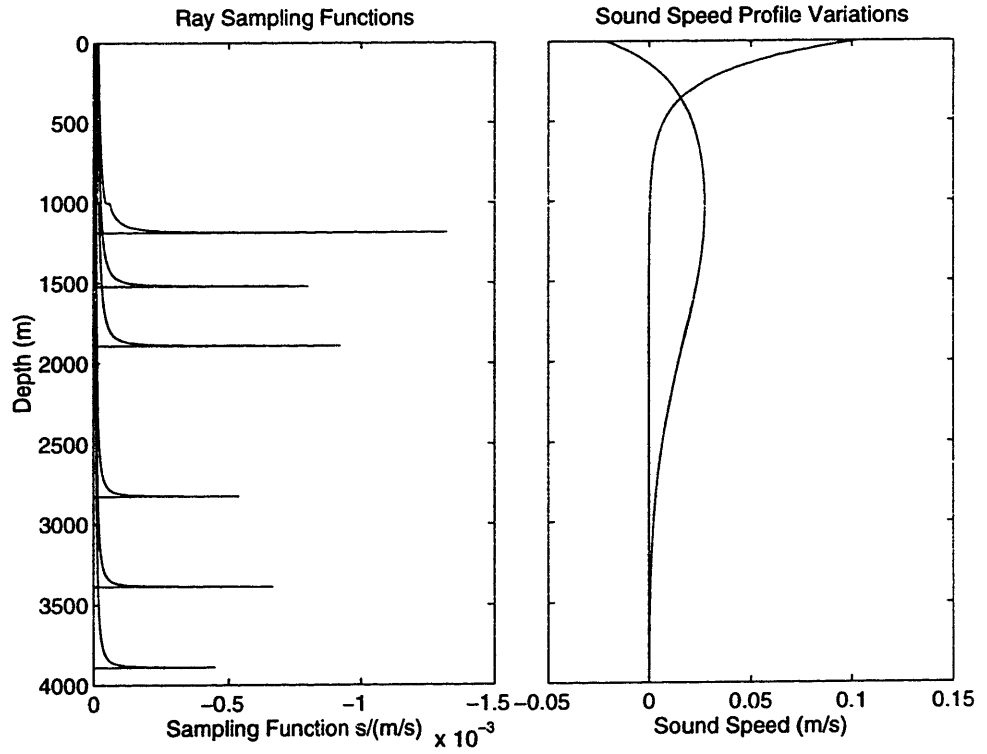

Figure 3-9: Sampling functions and profile variations for the Arctic example.

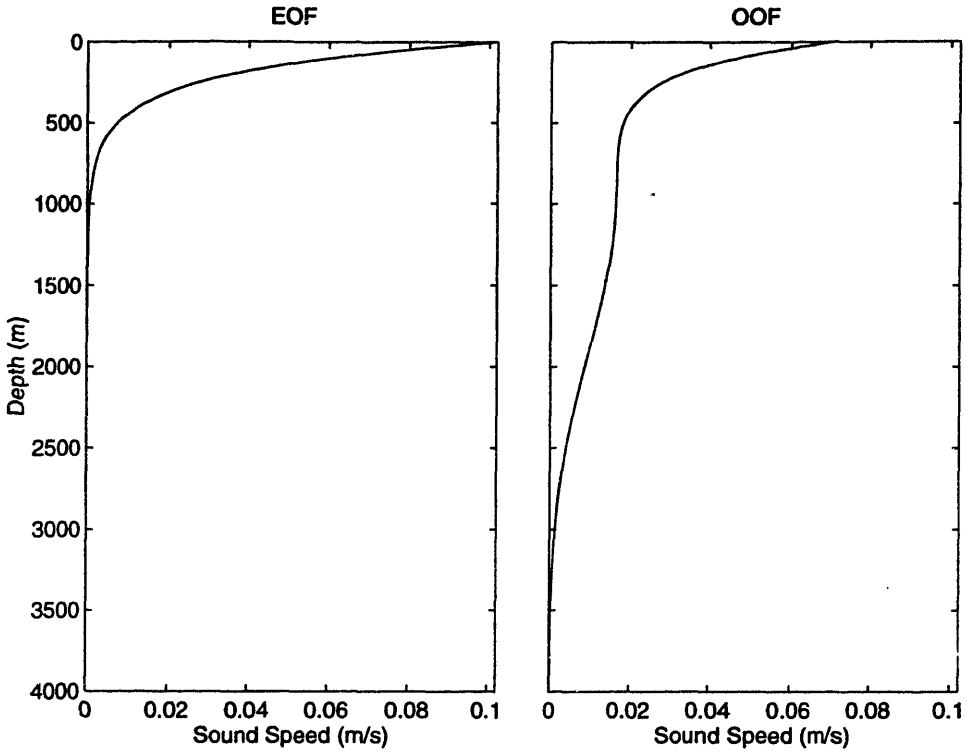

Figure 3-10: The EOF and OOF for the Arctic example.

Total Variance of Sound Speed Profile Estimate $\left(\frac{m^2}{s^2}\right)$

Suboptimal Estimator	Optimal Estimator	
EOF	EOF	OOF
653 (actual)	386	332
132 (predicted)		

As in the simple example, the suboptimal estimator predicts an error much smaller than its true error because it takes no account of the unmodeled mode of sound speed profile variation. The optimal estimator makes a significant improvement in estimate

accuracy because it takes into account the effect of the unmodeled mode. Using the OOF with the optimal estimator makes a slight additional improvement. The EOF already represents 80% of the profile variation. so the maximum improvement possible with a different basis is 20% of the total variance. The OOF achieves an improvement of about 11% of the initial total sound speed profile variance.

3.6 Using Estimated Profile Covariance Matrices

Up to this point, it has been assumed that the profile covariance matrix $\mathbf{P}$ is known. In practice, this covariance matrix is estimated from a finite set of historical profile measurements combined with whatever physical constraints are appropriate to the environment. There are usually much fewer profile measurements than there are points in each measurement, so the estimate of $\mathbf{P}$, $\hat{\mathbf{P}}$ is far from being full rank. This means that the estimate $\hat{\mathbf{P}}$ is in fact a reduced order model of the true covariance $\mathbf{P}$, and as such renders the inversion subject to the some of the same frailties as intentional model order reduction. This section considers what can be done in practical terms to improve the profile estimate, recognizing that the profile covariance estimate is imperfect.

3.6.1 The Covariance Matrix Estimate

In most tomography experiments, a profile mean and covariance matrix are estimated from an ensemble of actual profile measurements, $\{\xi_1, \xi_2, \ldots, \xi_N\}$. From these

measurements, a mean is estimated as,

$$\bar{\xi} = \frac{1}{N}\sum_{n=1}^{N}\xi_n, \tag{3.43}$$

and a covariance is estimated as,

$$\hat{\mathbf{P}} = \frac{1}{N-1}\sum_{n=1}^{N}(\xi_n - \bar{\xi})(\xi_n - \bar{\xi})^{\mathbf{T}}. \tag{3.44}$$

In general there many fewer sample profiles than the dimension of $\hat{\mathbf{P}}$, so beyond a few dominant eigenvalues and their associated eigenvectors the estimate is poor.

3.6.2 Weighting the Noise Covariance Diagonal

The common method for dealing with this problem is to make the approximation,

$$\mathbf{P} = \mathbf{\Phi}\mathbf{D}\mathbf{\Phi}^{\mathbf{T}} \tag{3.45}$$

where $\mathbf{D}$ has along its diagonal the largest eigenvalues of $\hat{\mathbf{P}}$ from (3.44), and $\mathbf{\Phi}$ has as its columns the associated eigenvectors. To deal with the fact that certain directions of variation in $\mathbf{x}$ have been ignored in this approximation, the diagonal of the noise covariance matrix $\mathbf{R}$ is increased to accommodate not only measurement noise, but also the measurement effect of unmodeled directions of variation in $\mathbf{x}$.

$$\mathbf{R}_{\text{eff}} = \mathbf{R} + \sigma\mathbf{I} \tag{3.46}$$

3.6.3 Wrapping Profile Covariance into Noise

The problem with this blind increase of the diagonal of the measurement error covariance is that it does not reflect the true directions of measurement variation caused by profile changes. A better method is to write $\mathbf{x}$ as the sum of its reduced order approximation and an approximation error term, $\tilde{\mathbf{x}}$

$$\mathbf{x} = \mathbf{\Phi a} + \tilde{\mathbf{x}}. \tag{3.47}$$

Using this new representation in (3.1),

$$\mathbf{y} = \mathbf{C\Phi a} + \mathbf{C}\tilde{\mathbf{x}} + \mathbf{n}. \tag{3.48}$$

Grouping the $\mathbf{C}\tilde{\mathbf{x}} + \mathbf{n}$ together as "noise", and assuming that the unmodeled profile variations have covariance $\mathbf{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\mathbf{T}}\right] = \sigma\mathbf{I}$, the new noise covariance matrix becomes.

$$\mathbf{R_{new}} = \mathbf{R} + \sigma\mathbf{C}\mathbf{C}^{\mathbf{T}}. \tag{3.49}$$

3.6.4 Comparison of Methods

In Fig. (3-11), the mean squared error resulting from both methods in the temperate profile case is shown as a function of the weight σ. The dashed line shows the effect of the method in (3.46), and the solid line shows the effect of the method in (3.49). The significant feature is the depth of the minimum. The second method, which takes into account the direction of measurement variation caused by unmodeled profile changes,

produces a lower minimum.

In Fig. (3-12), the mean squared error resulting from both methods in the Arctic profile case is shown as a function of the weight σ. The dashed line shows the effect of the method in (3.46), and the solid line shows the effect of the method in (3.49). The significant feature is the depth of the minimum. The second method, which takes into account the direction of measurement variation caused by unmodeled profile changes, produces a minimum which is slightly lower, 394 versus 396.

Note that the limit in both cases for large σ is 500, the *a priori* error in the profile. When σ is large, the noise has been made to dominate the measurement, and the measurement contains essentially no information. When the added noise is small, the performance approaches that of the suboptimal estimator. Note that $\sigma = 0$ does NOT mean the measurement is taken to be noise free. Rather it means that there is no additional weight added to the existing noise covariance matrix $\mathbf{R}$.

3.7 Optimal Moving Source Tomography

The more ray paths through the environment which are available, the more accurate the inversion will be. Using mobile acoustic sources, it is possible to obtain a greater diversity of ray paths and, in general, more information about the environment than with fixed sources. Fig. 3-13 shows a top view of a typical tomography problem. Four moorings, represented by black circles, have been put in place. Each mooring has a source and a receiver array. The goal of the experiment is to localize a front, represented by the wavy line, which is in within the mooring configuration. The

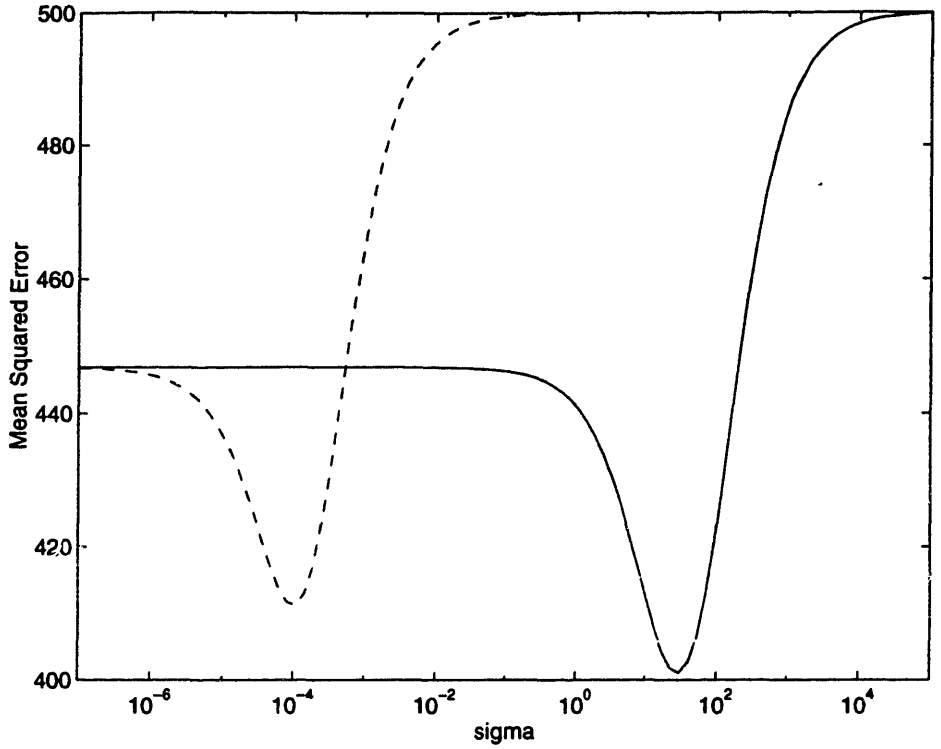

Figure 3-11: Comparison of MSE achieved as a function of σ, temperate case.

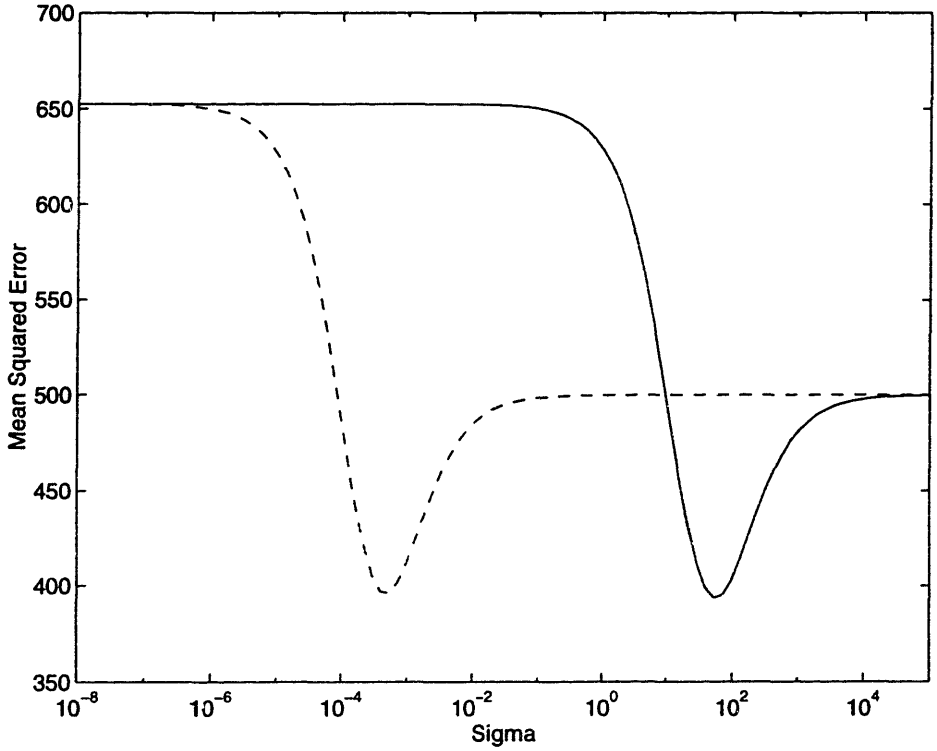

Figure 3-12: Comparison of MSE achieved as a function of σ, Arctic case.

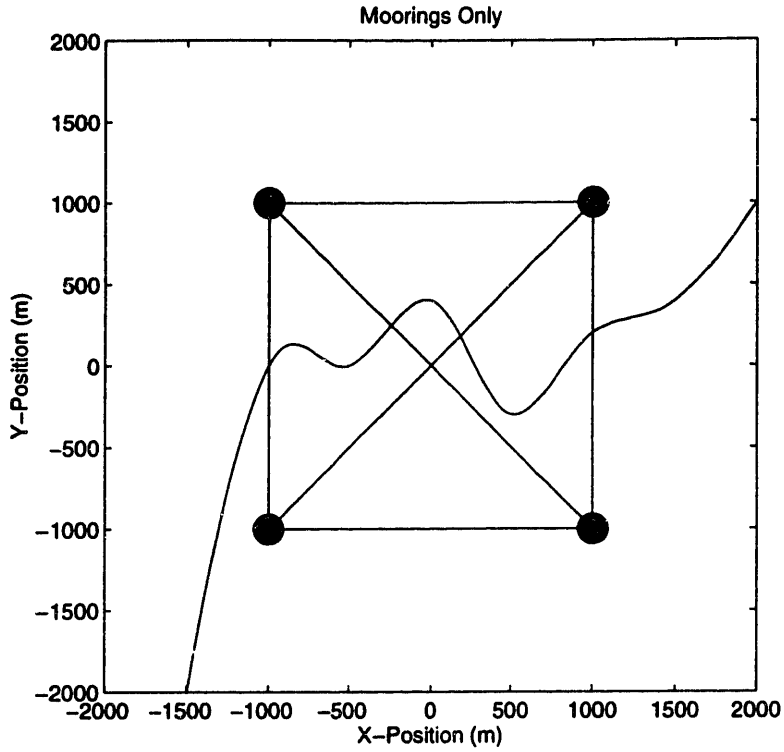

Figure 3-13: Rays with fixed source.

acoustic paths available for this localization are the straight lines shown. In Fig. 3-14, and AUV carrying a source moves around the outside of the array while transmitting. The circles represent the transmission points. With the AUV, there are many more acoustic paths which interact with the front, and localization of the front will be much more accurate. The usefulness of moving horizontally to obtain better sampling has been recognized in the acoustics literature [8].

While the source is moving, the environment is changing, and in large scale tomographic measurements, it may not be possible to move a source quickly enough to get an effectively contemporaneous image of the environment. Even if temporal-spatial aliasing is a problem for horizontally moving sources, it may still be possible to benefit from moving source technology by moving the source vertically.

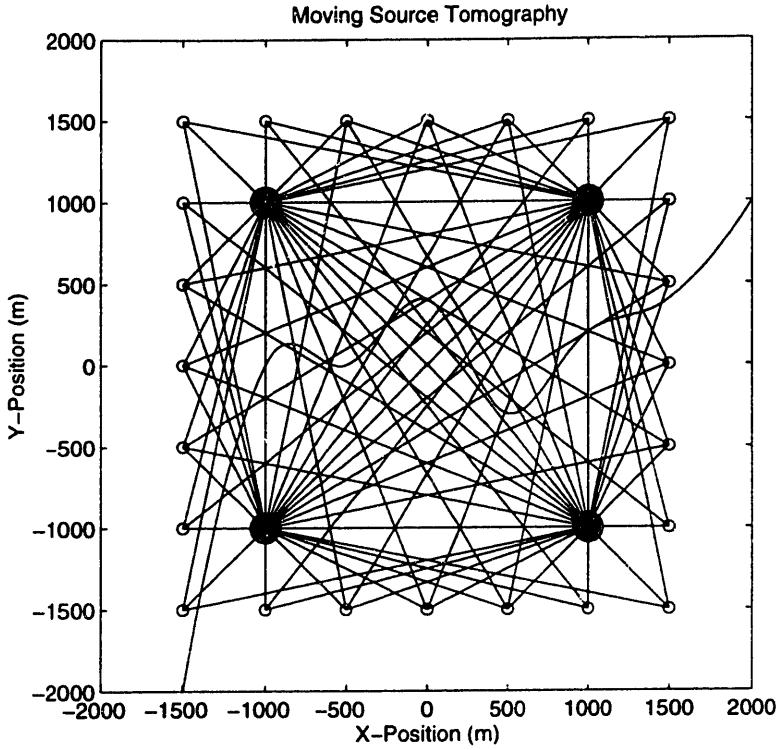

Figure 3-14: Rays with moving source.

Fig. 3-15 shows the ray paths through a slice of temperate ocean for a fixed source at 1000m depth and seven receivers throughout the water column at 100km range. If the number of receivers is increased, additional rays will fill in the spaces between the existing rays, however the shallow shadow zone between 15 and 35km range will still remain as will the deep shadow zone between 50 and 65km range. If the source moves, however, it is able to project sound into the shadow zones, and also provide a much larger number of rays to aid in the inversion, as shown in Fig. 3-16.

This section takes moving source tomography one step further and asks the question of where a moving source should go to obtain the most information about the environment, or about a specific feature within the environment.

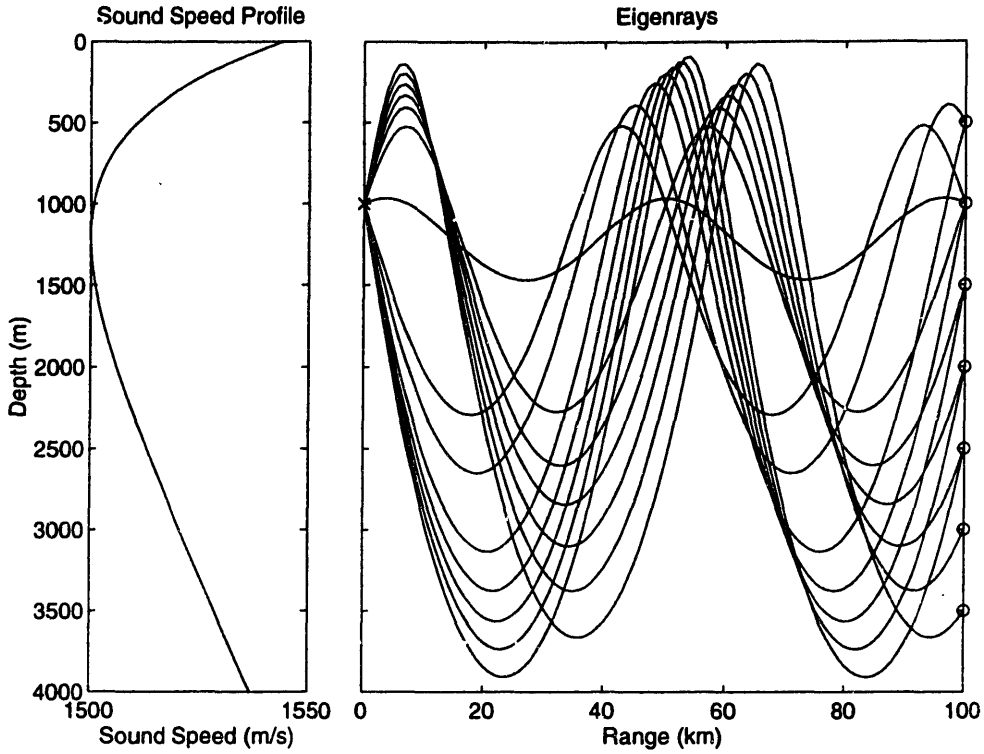

Figure 3-15: Rays with fixed source.

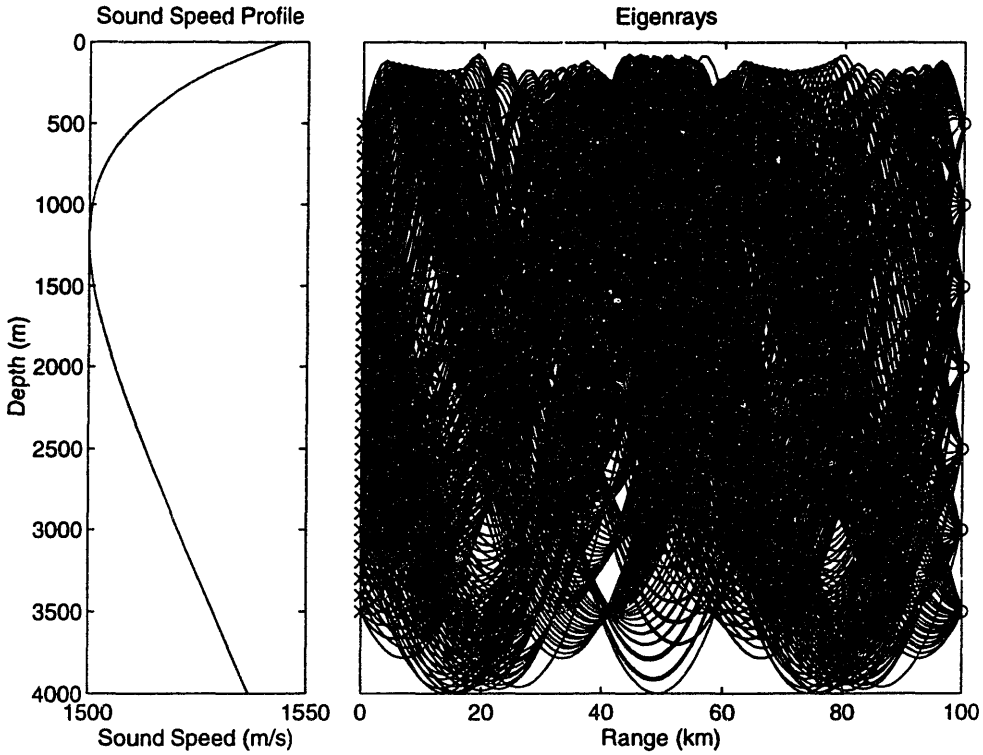

Figure 3-16: Rays with moving source.

3.7.1 Tomographic Resolution

Moving the source provides more constraints on the inversion by providing a greater diversity of ray paths. It also can be used to focus measurement resolution at a particular region in the environment. This principle is shown in the example below. Consider an environment of 200m deep water and 4km range. The mean sound speed profile and a single basis function representing all profile variability are shown in Fig. 3-17. The 4km range from source to receiver is divided into ten 400m segments, with the weight for the basis function of equal variance and uncorrelated from segment to segment. The two pictures in Fig. 3-18 show the variance of the profile estimate throughout the environment for two different source locations. Lighter is larger errors. In every range section, the variance has the shape of the one basis function, largest at 100m, and smallest at the surface and bottom, but the size of the weight error changes depending on how the rays sample that range. There are three eigenrays which sample the environment. In the top figure, in the first range division (0-400m), the rays are near the surface where the profile variation is small, and as a result, the errors are still fairly large. In the second range division (400-800m), the rays are deeper, and pass through depths where the profile variation is larger, so the weight error is smaller. In the third range division (800-1200m), the rays pass through the depths where the profile variation is largest, and so the weight error is even smaller. As the rays head into deeper water for range segments four, five, and six (1200-2400m), they are further and further away from the large part of the profile variation, and so the weight errors begin increasing again. Oddly, the error is quite small in range bin seven (2400-

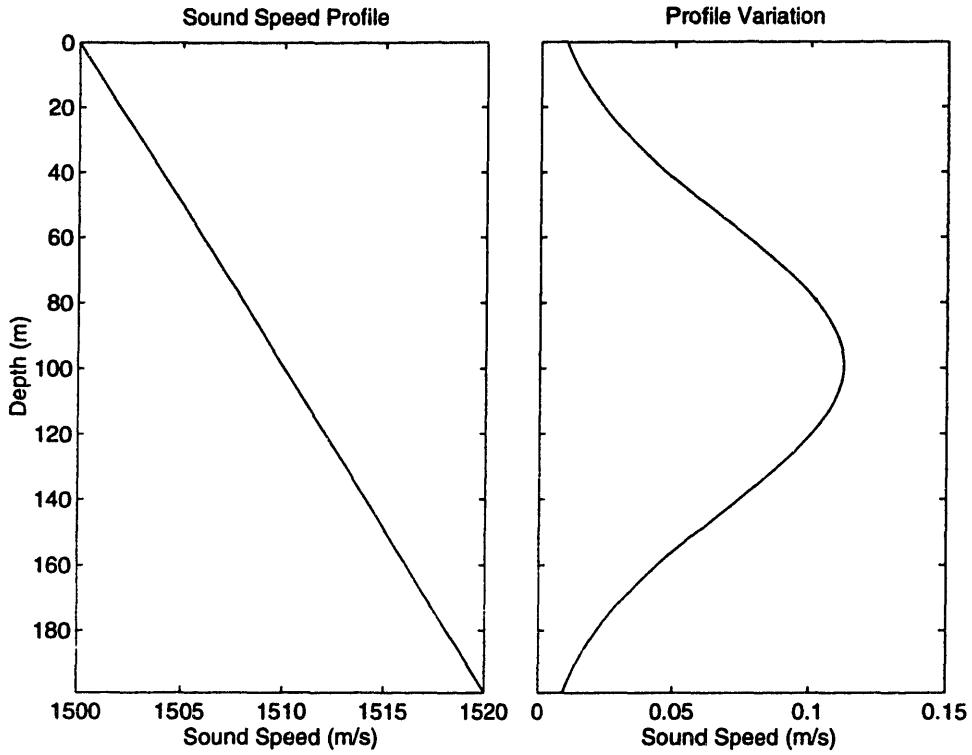

Figure 3-17: Sound speed profile and basis function.

2800m), even though two of the three rays are below the depth of significant profile variation. The reason for the excellent accuracy in segment seven is that in the earlier segments, the travel times of all three rays are influenced approximately the same amount by the variation, while in segment seven, two of the rays are unaffected by the profile variation, but one of them is very sensitive to the variation. Thus a variation in range segment seven produces a travel time shift in the rays which is nearly orthogonal to the variation caused by profile changes in all the other range segments.

When the source is moved, the location of these regions of high accuracy changes, as shown in the bottom picture of Fig. 3-18. Here range segments six (2000-2400m) and eight (2800-3200m) have good resolution. This is again due to the characteristics of the ray sampling. Refer to the ray which reflects off the surface and bottom once the

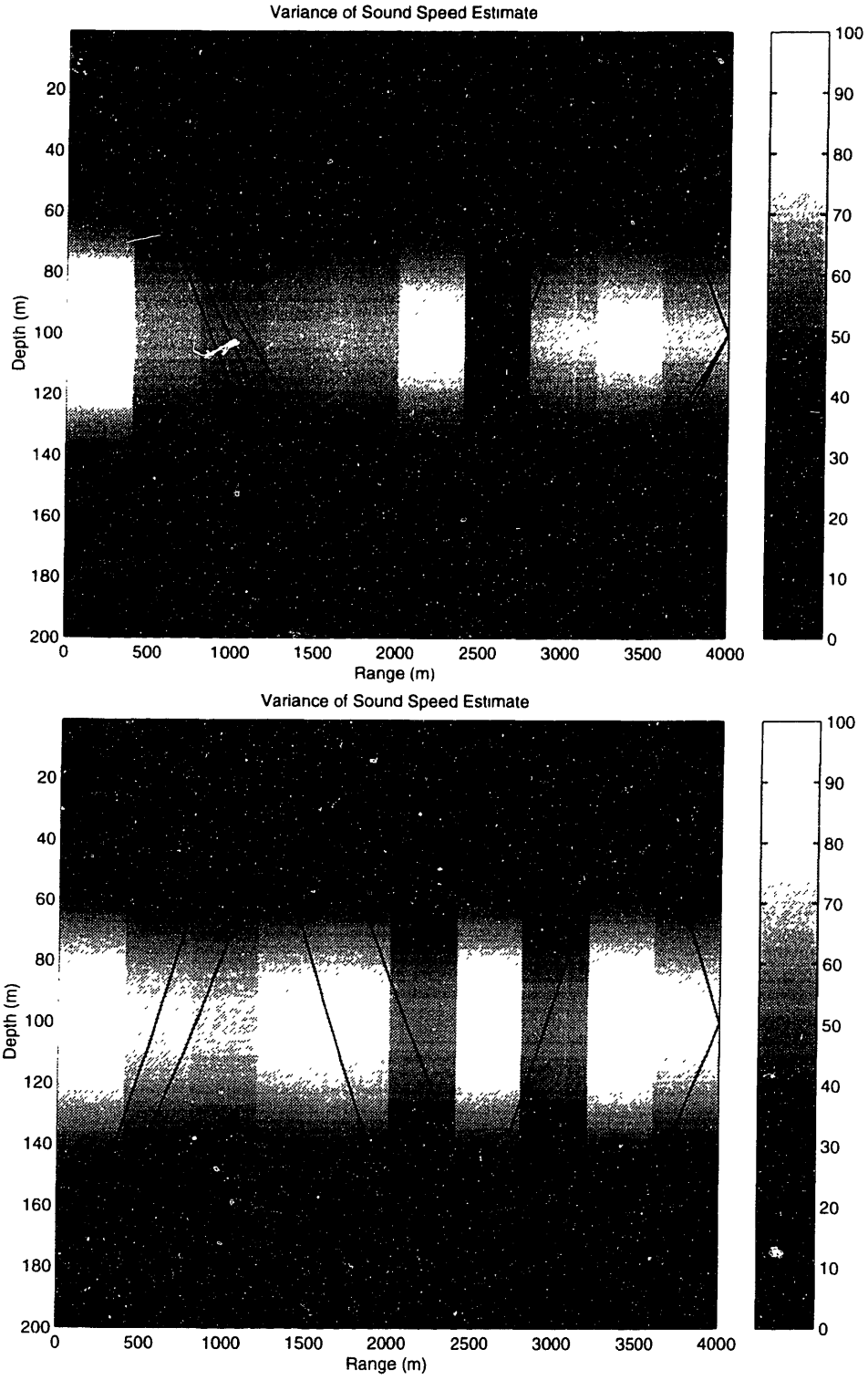

Figure 3-18: Sound speed estimate variance.

SB-ray, and the ray which reflects of the surface twice and bottom once the SBS-ray. In range bin six, the SB-ray is very sensitive to the profile variation, while the SBS-ray is insensitive to it because of their depths. In range bin eight, the sensitivities are reversed. In the other bins the rays both have somewhat similar sensitivities to the profile variation.

3.7.2 Optimal Tomography

The accuracy of the tomographic measurement depends on how the various rays sample the region of interest in the environment, which depends on the source depth. In this example, a set of four transmission depths is found which give the maximum accuracy in a prescribed region of interest. In this example, a tomography problem is considered in a slice of 200m deep water at 4km range. The environment is divided horizontally into 8 range slices of 500m each. The sound speed profile variation within each range slice is represented by a sum of five basis functions, which allow representation of piecewise linear profiles. The mean sound speed profile and the basis functions are shown in Fig. 3-19. There is a single receiver in this environment at a depth of 50m. The source is able to move vertically, and the optimization problem we will consider is one of choosing the depths at which the source will transmit. The source is allowed to transmit four times, and the transmission depths will be chosen to minimize the integrated variance over the region enclosed by a box in the figures which follow. In the top picture of Fig. 3-20 the boxed region of interest is between 100 and 150m depth and between 500 and 1500m range. The shading of the plot indicates

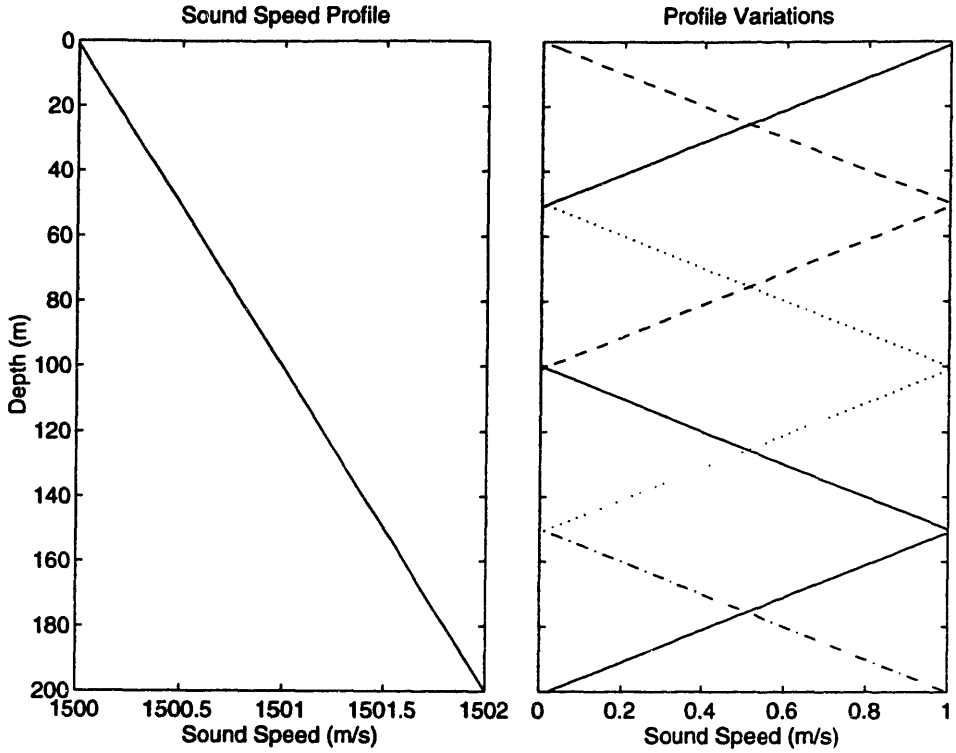

Figure 3-19: Sound speed profile and basis functions.

the variance of the sound speed estimate as a function of position. The lower picture of Fig. 3-20 shows the estimate variance for each of the 40 weights in this example (5 weights for each range segment). The dotted line is the variance for a stationary source transmitting 4 times at 100m depth. The dashed line is the variance for a moving source transmitting at 40, 80, 120, and 160m. The solid line is the variance for a moving source transmitting at the optimal depths. The circles on the solid line indicate the four parameters which influence the variance of the sound speed estimate in the focusing region. In Fig. 3-21, the plots of Fig. 3-20 are repeated, except that the boxed region of interest is between 0 and 50m depth and between 1500 and 2500m range, and new optimal transmission depths for the new region of interest have been determined.

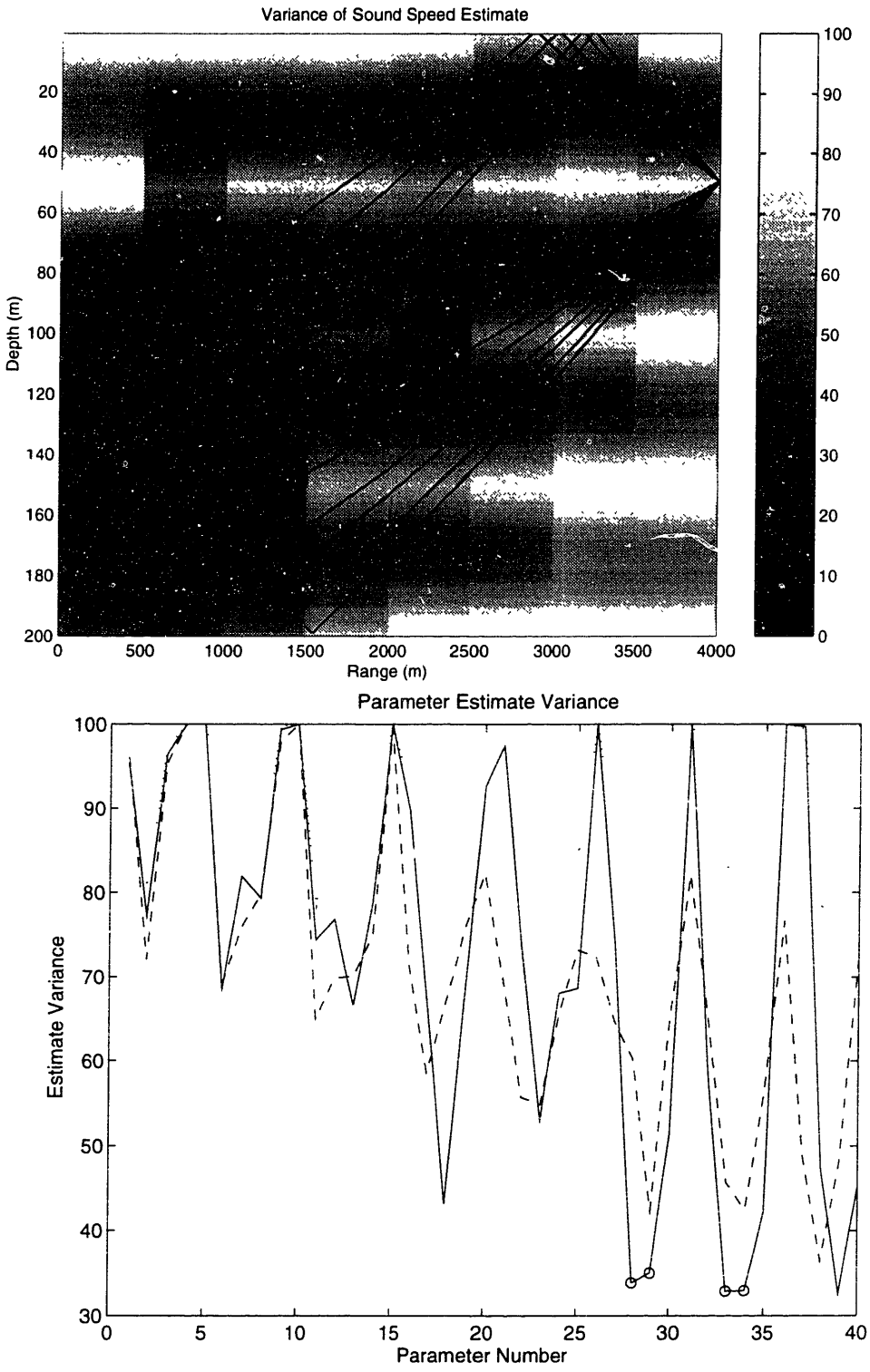

Figure 3-20: Sound speed estimate variance for stationary, moving, and optimal moving source tomography (first region of interest).

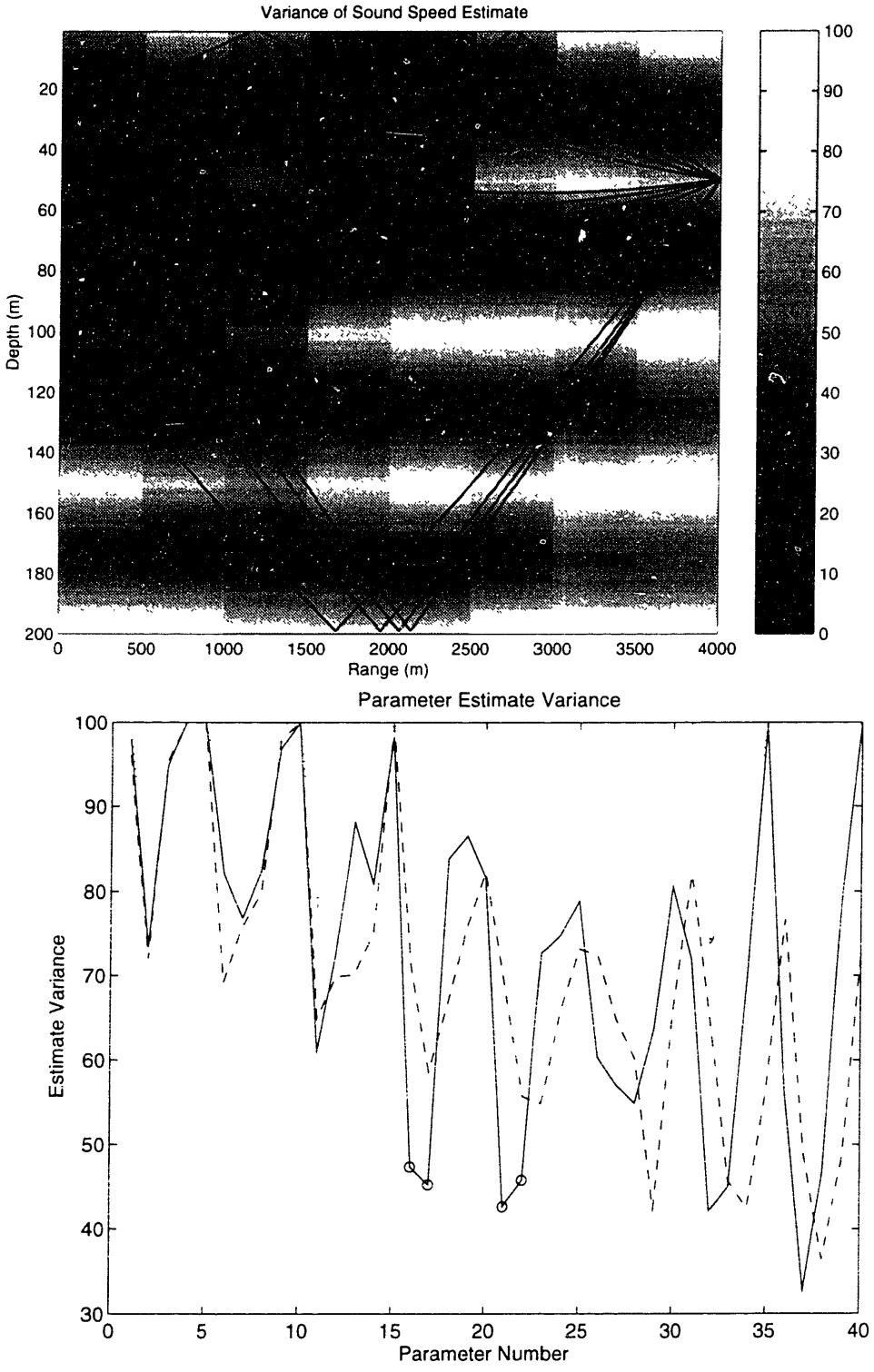

Figure 3-21: Sound speed estimate variance for stationary, moving, and optimal moving source tomography (second region of interest).

3.8 Conclusion

The travel time measurement in ocean acoustic tomography is not equally sensitive to all modes of sound speed profile variability. To be optimal, a profile parameterization or inversion must take into account both the expected size of the profile variations and the resolution with which each variation can be measured. An optimal parameterization and inversion were derived which take both of these factors into consideration, and the accuracy enhancement which these techniques offer was demonstrated in tomography examples for typical temperate and arctic environments. In addition to optimizing the parameterization and the estimator, it is possible to optimize source locations for the best resolution in a region of interest. This optimization problem was also demonstrated here.

Chapter 4

Arrival Matching

Before a tomographic inversion can be performed, the measured ray arrivals in the received signal must be matched with predicted arrivals to generate travel time prediction errors. If the predicted arrivals are identified with the wrong measured arrivals, errors will result in the inversion. This chapter examines the problem of arrival matching. A test environment is described in Section 4.1, and four different matching algorithms are described and evaluated in Sections 4.2, 4.3, 4.4, and 4.5. Finally, the advantages of the new correlated matching algorithm are analyzed in Section 4.6.

4.1 Test Scenario

If the predicted acoustic environment exactly matches the true acoustic environment, then the predicted arrivals will occur at the same times as the measured arrivals, and all of the methods described will identify the arrivals perfectly. What differentiates the methods is their ability to correctly identify arrivals when the true environment is

different from the predicted environment. To evaluate the methods presented in this chapter, a test scenario is created. The acoustic environment is described by a sound speed profile. The predicted sound speed profile is a constant 1500m/s, as shown on the left in Fig. 4-1. The true sound speed profile $c(z)$ is the predicted profile $c_0(z)$ with some amount of a profile variation $\phi(z)$ added to it,

$$c(z) = c_0(z) + \alpha\phi(z) \tag{4.1}$$

The parameter α determines how much of the profile variation is included in the true profile. The profile variation used in this trial is shown on the right of Fig. 4-1. Values of α ranging from 1 to 20 were used to generate the family of 20 profiles shown on the left of Fig. 4-2. The eigenrays connecting the source at 70m depth with the receiver at 120m depth for the predicted profile are shown on the right of Fig. 4-2.

Figure 4-3 shows the predicted arrival times for the predicted profile numbered from one to seven along the x-axis. The x's above the axis show the measured arrival times corresponding to values of α ranging from 1 to 20. Note that the travel times tend to change linearly with the parameter value. The matching algorithm will attempt to determine which of these linear arrival trends each measured arrivals is part of and match the measured arrival with the predicted arrival at the bottom of the linear trend. Note that while the graphs show the measured arrival times for all parameter values, the matching algorithm will only be given measured arrival times for a single parameter value at a time, and it will not know what that parameter value is.

In the sections which follow, each matching algorithm will be applied to this test

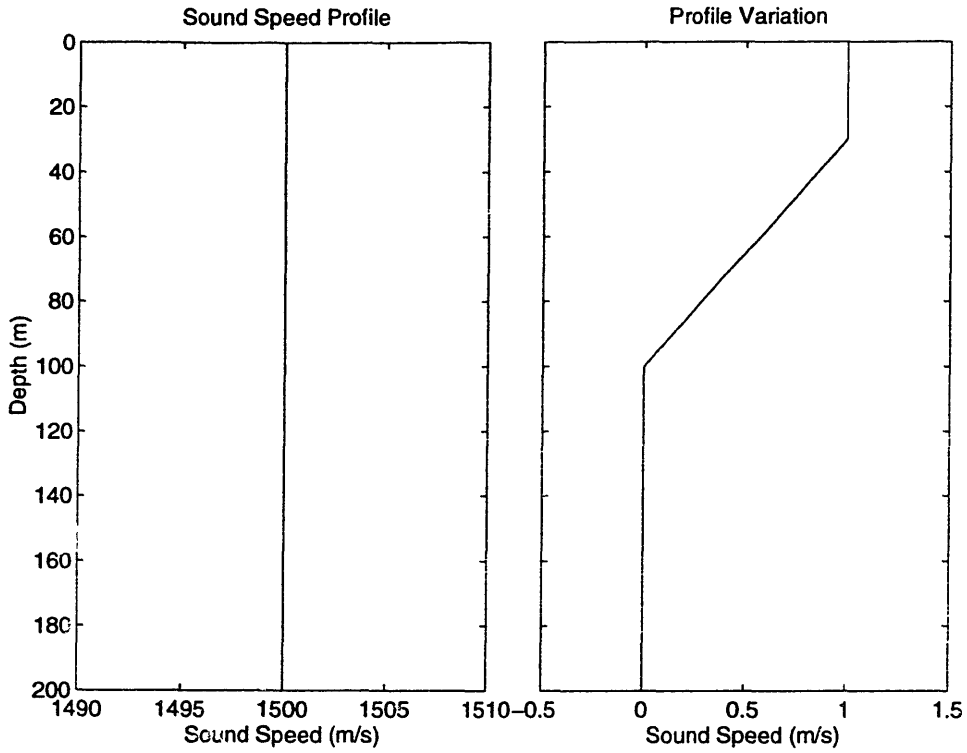

Figure 4-1: Predicted profile and profile variation.

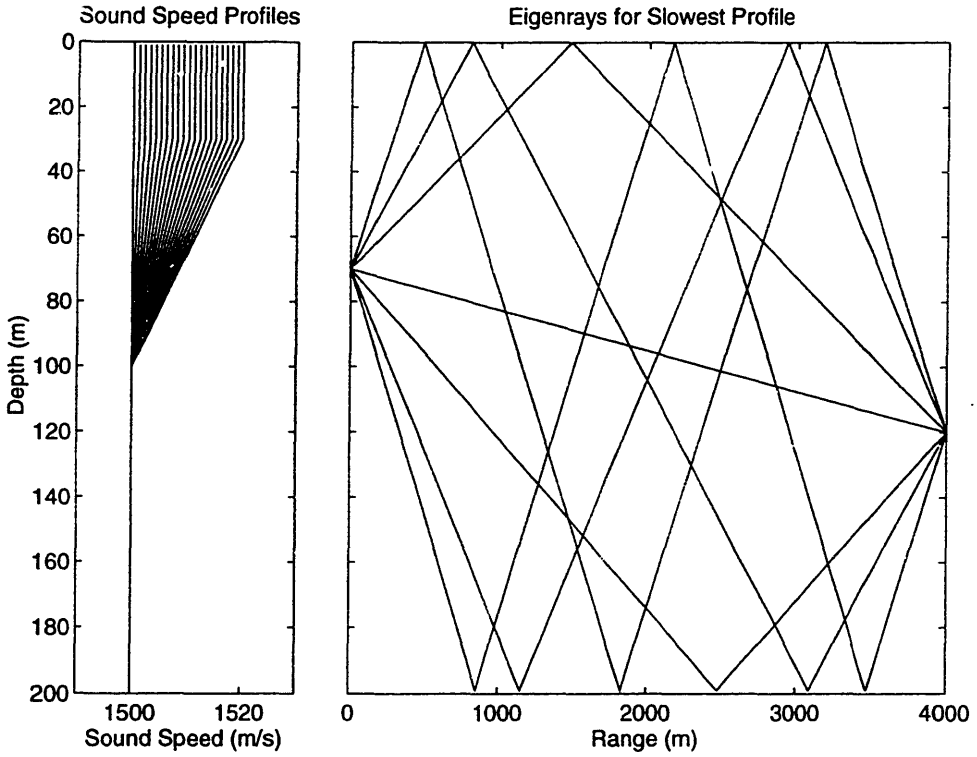

Figure 4-2: True profiles and eigenrays for predicted profile.

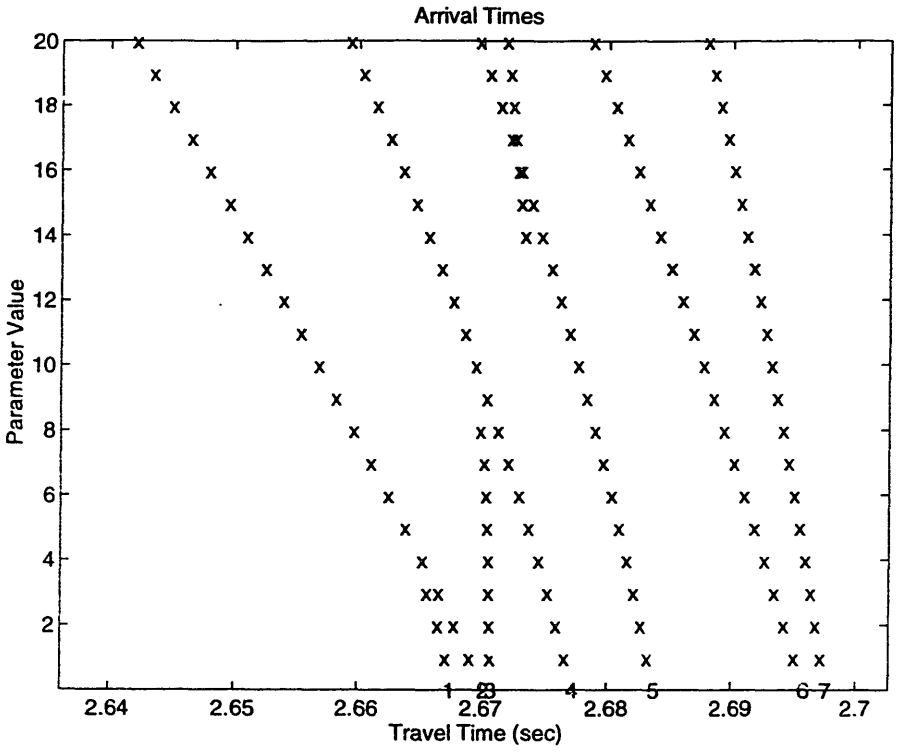

Figure 4-3: Predicted arrival times and measured arrival times for parameter values of 1 to 20.

data set. Once a matching is selected, a Gauss-Markov estimate of the parameter value will be made, based on the linearized model of the relationship between travel time prediction error $\mathbf{t}$ and parameter α,

$$\mathbf{t} = \mathbf{C}\alpha + \mathbf{n}. \tag{4.2}$$

The noise vector $\mathbf{n}$ is zero-mean Gaussian with covariance matrix $\mathbf{R} = \sigma_n^2\mathbf{I}$, where $\sigma_n = 0.001$ seconds and α has variance σ_α^2 equal in each case to the true value of α squared.

For each value of α, the selected match will be shown by replacing the x's for the measured arrivals in Fig. 4-3 with the numbers of the predicted arrivals to which

each measured arrival was matched. A correct matching would then show all measured arrivals along each linear trend matched to the predicted arrival at the starting point of that trend. Once the matching for a parameter value is established and the resulting arrival time differences are calculated, the parameter value is estimated using the Gauss-Markov estimator,

$$\alpha = \sigma_\alpha^2 \mathbf{C}^\mathbf{T}(\mathbf{C}\sigma_\alpha^2\mathbf{C}^\mathbf{T} + \sigma_n^2)^{-1}\mathbf{t}. \tag{4.3}$$

The parameter estimate error will then be shown as a function of parameter value for all matching methods.

4.2 Simple Ordering

The simplest matching algorithm matches arrivals according to their order of reception. The earliest measured arrival is matched to the earliest predicted arrival. The next measured arrival is matched to the next predicted arrival, and so forth until all the arrivals of interest have been matched.

The results of applying this matching algorithm to the test data set are shown in Fig. 4-4. When the parameter value is less than 4, the matching works well. When the parameter value reaches 4, one of the ray paths disappears, and as result, the paths are mismatched. Note that the appearance and disappearance of ray paths is a common phenomenon. A path also disappears at parameter value 9 and one appears at parameter value 14. The matchings shown in Fig. 4-4 are used as the basis for estimating the parameter value, and the resulting parameter estimate error is shown

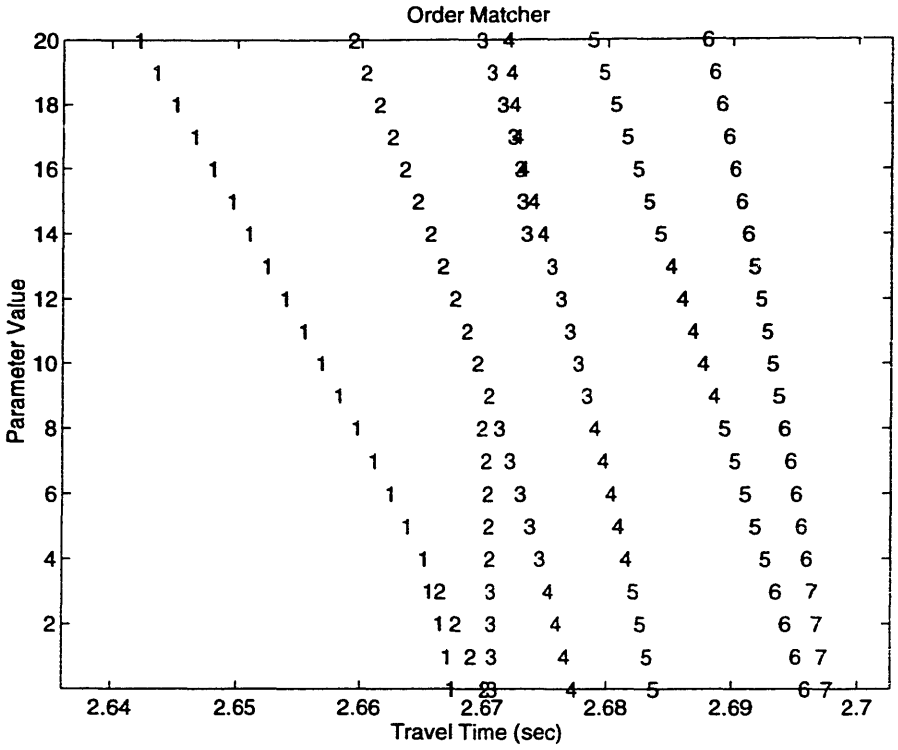

Figure 4-4: Ordered matching.

as a function of the true parameter value in Fig. 4-5. Note that the estimate errors are small for parameter values of 1,2, and 3, since all the paths were identified correctly in these cases. A large jump in the error occurs at parameter value 4, since this is the first value for which paths are incorrectly identified. Other discontinuities in the error occur at parameter value nine where a second ray path disappears and at 14 where a new path appears.

The order matching method will produce the correct matching as long as arrivals do not appear, disappear, or change in order. Although arrivals may be consistent in long-range deep-ocean tomography, there is often significant fading in shallow water or in the shallow Arctic sound channel. A missing or appearing arrival will cause all subsequent arrivals to be incorrectly identified, so fading environments call for a more

Figure 4-5: Error for ordered matching.

robust arrival identification algorithm.

4.3 Validation Windows

A method which is more robust to fading than the simple ordering is one using validation windows. A time window is drawn around each predicted arrival, and if one and only one measured arrival falls within this window, it is matched to the predicted arrival.

In this example, the window for each arrival is centered around the predicted arrival time, and the width of the window is the smaller of the distance to the previous arrival and the distance to the subsequent arrival. The application of this matching algorithm to the test data set is shown in Fig. 4-6. The validation windows used are

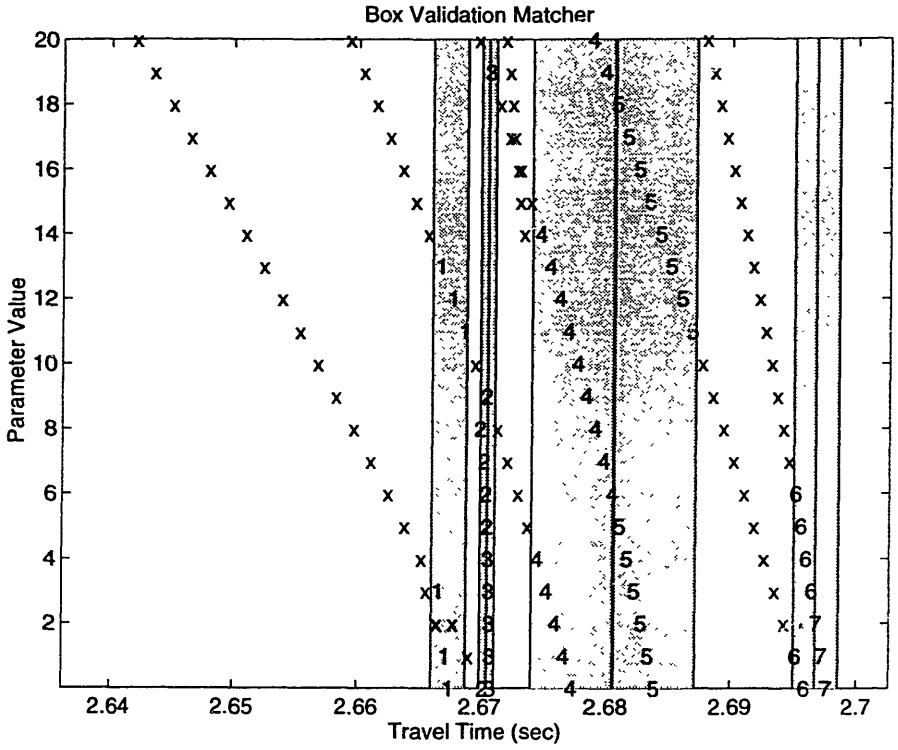

Figure 4-6: Validation window matcher

shown as shaded regions on the graph. Note that when one and only one measured arrival falls within a window, it is matched to the predicted arrival within that window. The method fails when the shifts in arrival time are greater than the widths of the windows. At a parameter value as small as 1 this method has failed to identify the second arrival. At parameter value 3, it mistakes arrival 7 for arrival 6, and significant inversion errors result, as shown in Fig. 4-7.

The validation window method is most useful when arrival fading occurs before the arrival time shifts become significant compared to the separation between arrivals. The validation window method is often used for deep-ocean tomography where the arrival time separations can be quite large, but the method is problematic in shallow water where the arrival time separations are smaller.

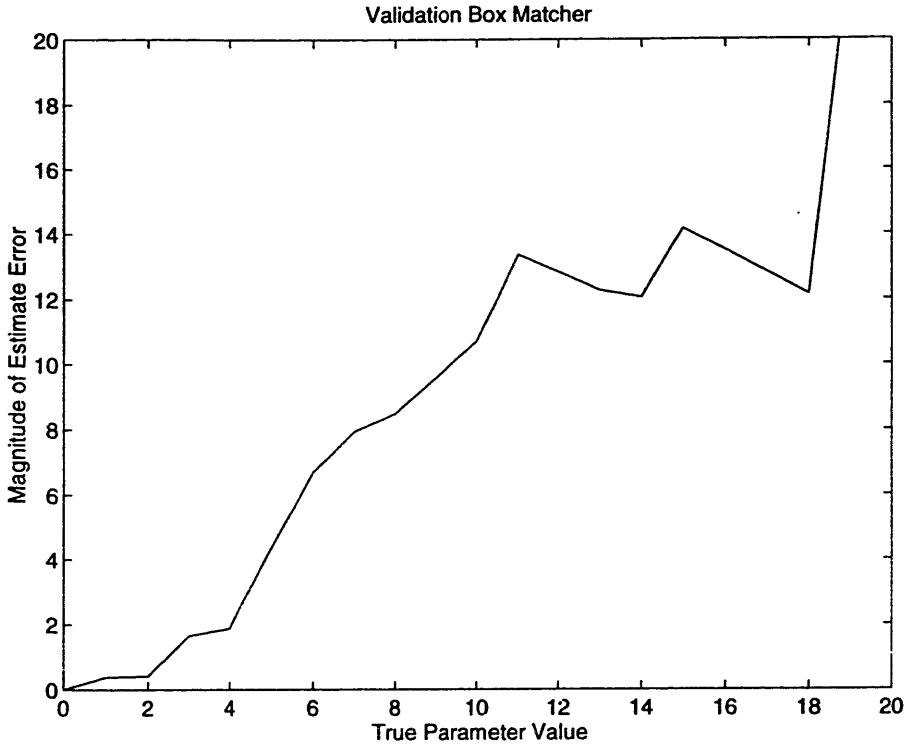

Figure 4-7: Estimate error using validation window matcher

4.4 Independent Matching

The linear model relating travel time variations to parameter variations provides information about the behavior of arrivals which can be useful in solving the matching problem. This section describes a matching algorithm which uses the linear model to determine the travel time variance for each measured arrival and based on these variances finds the most likely match between measured and predicted arrivals. This algorithm also allows for the possibility that some arrivals may not have matches. The formulation of this algorithm allows the matching problem to be posed as an assignment problem and solved using standard linear programming techniques [10].

Let the predicted travel times be the elements of a vector $\mathbf{t_p}$, and the corresponding measured travel times be elements of a vector $\mathbf{t_m}$. The travel time prediction error t

is,

$$\mathbf{t} = \mathbf{t_m} - \mathbf{t_p}. \tag{4.4}$$

The travel time prediction errors can be related to parameter changes according to the linear model,

$$\mathbf{t} = \mathbf{C}\mathbf{x} + \mathbf{n}. \tag{4.5}$$

where the parameter vector $\mathbf{x}$ and the measurement noise vector $\mathbf{n}$ are independent and Gaussian with covariance matrices $\mathbf{P}$ and $\mathbf{R}$ respectively. The probability density for the ith element of the vector of matched measured arrivals is,

$$p(\mathbf{t_{m_i}}) = \frac{1}{\sqrt{(2\pi)\sigma_i^2}} \exp\left(-\frac{(\mathbf{t_{m_i}} - \mathbf{t_{p_i}})^2}{\sigma_i^2}\right), \tag{4.6}$$

where $\sigma_i^2 = (\mathbf{CPC^T} + \mathbf{R})_{ii}$. The travel time measurements are correlated since they all depend on the same parameter vector $\mathbf{x}$. However, for purposes of this algorithm, it will be assumed that they are *independent* so that the joint probability density for the whole vector of matched measured arrivals becomes,

$$p(\mathbf{t_m}) = \prod_{i=1}^{P} p(\mathbf{t_{m_i}}) \tag{4.7}$$

The goal of the independent matcher is to form the vector $\mathbf{t_m}$ by ordering a subset of the measured arrivals in such a way as to maximize the above likelihood function (4.7). It is possible that some of the predicted arrivals may not have suitable matches

in the set of measured arrivals. In this case, the probability for the missing element of $\mathbf{t_m}$ is replaced by a constant penalty value in (4.7). This allows the matcher to leave some predicted arrivals unmatched, but assigns a penalty for doing so.

Taking the natural logarithm of (4.7), removing some constant terms, and multiplying by -1, the maximum likelihood problem can be rewritten as an equivalent minimization problem with objective function,

$$f_{obj}(\mathbf{t_m}) = \sum_{i=1}^{P} \frac{(\mathbf{t_{m_i}} - \mathbf{t_{p_i}})^2}{\sigma_i^2}. \tag{4.8}$$

If an arrival is left unmatched, then the corresponding term in this sum is replaced by a constant α.

This minimization problem can be posed as an *assignment problem* and solved with standard linear programming techniques. A cost matrix for the assignment problem is defined where the columns of the matrix correspond to the M measured arrivals and the rows correspond to the P predicted arrivals. The elements of the matrix are then,

$$\Phi_{ij} = \text{cost of matching the } i\text{th prediction to the } j\text{th measurement} \tag{4.9}$$

$$\Phi_{ij} = \frac{(\mathbf{T_j} - \mathbf{t_{p_i}})^2}{\sigma_i^2}, \tag{4.10}$$

where $\mathbf{T_j}$ is the jth measured arrival (before a subset of these measurements are ordered in the vector $\mathbf{t_m}$).

The possibility that some arrivals may not have matches is handled by the assignment of "dummy" rows and columns [25]. The fixed penalty α is assessed for leaving an arrival unmatched. To incorporate this, $M-1$ dummy rows are added to the P prediction rows so that as many as $M-1$ of the M measured arrivals may remain unmatched if necessary, but at least one measured arrival will be matched to a predicted arrival. The cost penalty for these unmatched measurements is set to the value α. Similarly, $P-1$ dummy columns are added to the M measurement columns, so that as many as $P-1$ of the P predicted arrivals may remain unmatched if necessary, but at least one predicted arrival will be matched to a measured arrival. The cost penalty for these unmatched predictions is also α. A small constant ϵ is subtracted from the penalty for matching the jth dummy column with the jth dummy row. This ensures that the algorithm will not waste time seeking a "best match" between the dummy rows and columns. The cost matrix constructed in this way is shown below. In this example $P < M$,

	m_1	m_2	$\cdots$	m_M	u_1	u_2	$\cdots$	u_{P-1}
p_1	Φ_{11}	Φ_{12}	$\cdots$	Φ_{1M}	α	α	$\cdots$	α
p_2	Φ_{21}	Φ_{22}	$\cdots$	Φ_{2M}	α	α	$\cdots$	α
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
p_P	Φ_{P1}	Φ_{P2}	$\cdots$	Φ_{PM}	α	α	$\cdots$	α
u_1	α	α	$\cdots$	α	$\alpha-\epsilon$	α	$\cdots$	α
u_2	α	α	$\cdots$	α	α	$\alpha-\epsilon$		α
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\ddots$	
u_{P-1}	α	α	$\cdots$	α	α	α		$\alpha-\epsilon$
u_P	α	α	$\cdots$	α	α	α	$\cdots$	α
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
u_{M-1}	α	α	$\cdots$	α	α	α	$\cdots$	α

(4.11)

In this thesis, Mack's method was used to solve this assignment problem [5].

The results of this independent matcher are shown in Fig. 4-8. The linear model informs the matcher about which arrivals will experience large time shifts as the parameter value changes, like arrival 2, and which will experience small time shifts, like arrival 3. This information enables good matchings to be made until the parameter value reaches 6. This technique retains only the variances of the arrival times from the model, not the covariances between arrival times. As a result, for parameter value 6, it sees nothing inconsistent with attributing an increases in travel time to arrivals 2 and 4 and at the same time attributing decreases in travel time to arrivals 6 and 7. The estimate error plot in Fig. 4-9 shows the large jump in error at parameter value

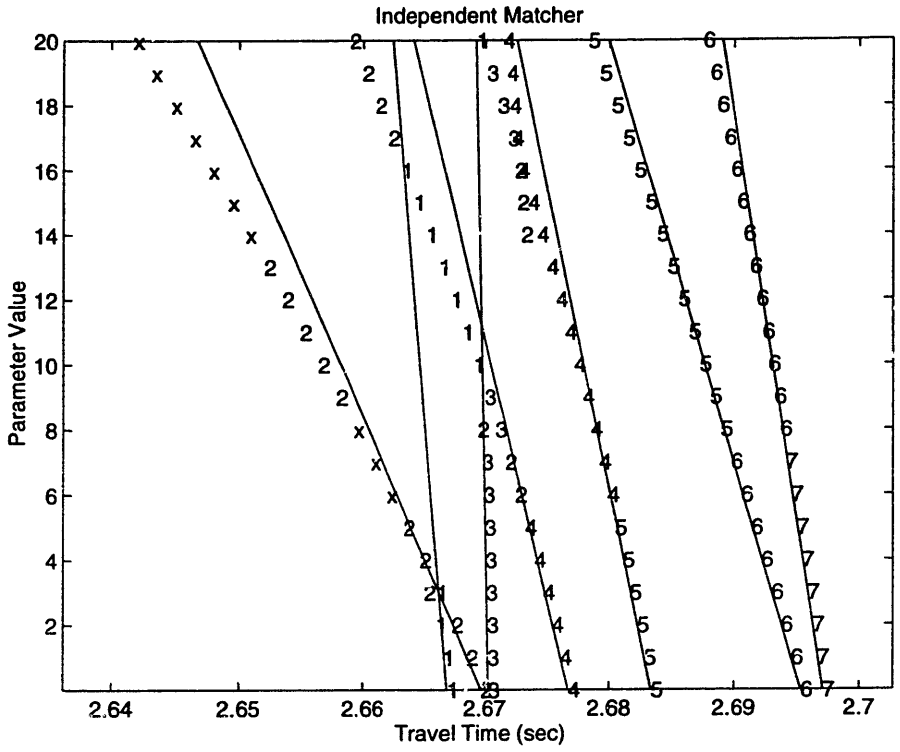

Figure 4-8: Independent matching.

6 where the first significant arrival identification errors are made.

In summary, the independent matching algorithm uses information about which arrivals move significantly and which will tend to remain fixed with parameter value changes to offer some improvement in performance over conventional methods.

4.5 Correlated matching algorithm

The correlated matching algorithm proposed in this section takes into account correlations between the time shifts in the various arrivals, fully utilizing the information provided by the model in (4.5).

Based on the statistics of the linear model (4.5) and using (4.4), a joint probability density function can be written for the matched measured arrival vector $\mathbf{t_m}$.

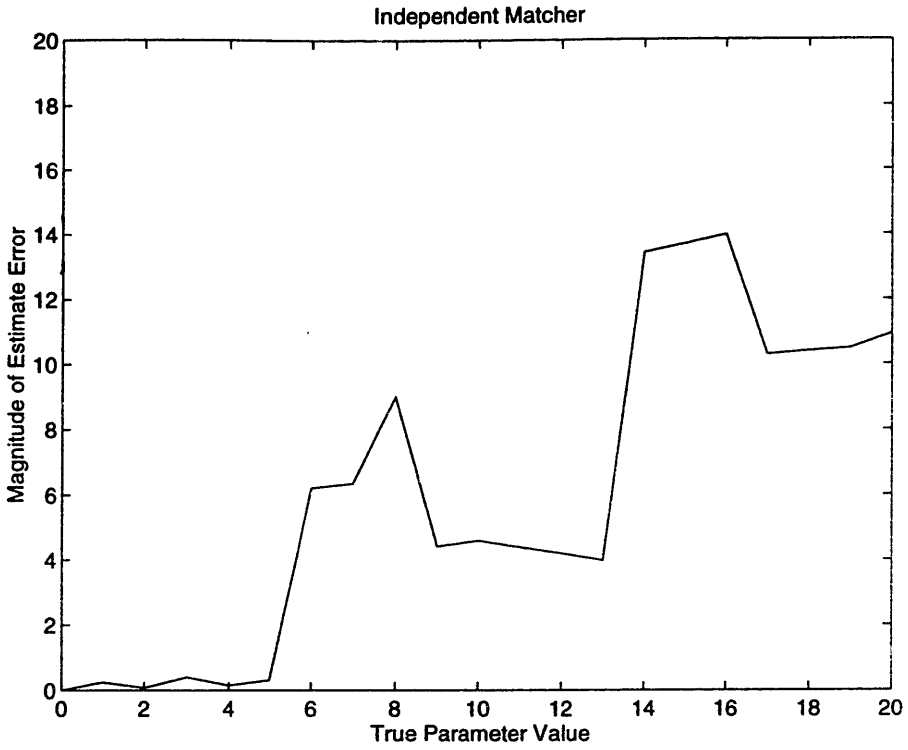

Figure 4-9: Error for independent matching.

$$p(\mathbf{t_m}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{t_m} - \mathbf{t_p})^{\mathbf{T}} \mathbf{\Sigma}^{-1} (\mathbf{t_m} - \mathbf{t_p})\right), \tag{4.12}$$

where $\mathbf{\Sigma} = \mathbf{CPC^T} + \mathbf{R}$. The goal of the independent matcher is to form the vector $\mathbf{t_m}$ by ordering a subset of the measured arrivals in such a way as to maximize the above likelihood function (4.12).

It is possible that some of the predicted arrivals in $\mathbf{t_p}$ may not have suitable matches in the set of measured arrivals. In this case, the empty spaces in $\mathbf{t_m}$ are filled in with the *expected value* for these measured arrival times, given the information about the parameter vector contained in the arrival times which were matched. Specifically, let $\mathbf{t'_p}$ and $\mathbf{t'_m}$ be the matched prediction and measurements respectively, and let $\mathbf{t''_p}$ be the unmatched predictions. If the prediction vector is,

$$\mathbf{t_p} = \left[\frac{\mathbf{t'_p}}{\mathbf{t''_p}}\right] \tag{4.13}$$

and the measurement matrix is also partitioned as

$$\mathbf{C} = \left[\frac{\mathbf{C'}}{\mathbf{C''}}\right] \tag{4.14}$$

then the completed matched measurement vector $\mathbf{t_m}$ becomes,

$$\mathbf{t_m} = \left[\frac{\mathbf{t'_m}}{\mathbf{t''_p} + \mathbf{C''PC'^T}\left(\mathbf{C'PC'^T} + \mathbf{R'}\right)^{-1}(\mathbf{t'_m} - \mathbf{t'_p})}\right] \tag{4.15}$$

Where $\mathbf{R'}$ the measurement noise covariance matrix for $\mathbf{t'_m}$.

In addition to filling in expected values for the missing arrivals, (4.12) is also multiplied by a constant penalty factor for each unmatched arrival. This provides a disincentive to leaving arrivals unmatched. Taking the natural logarithm of (4.12), and removing terms independent of $\mathbf{t_m}$, the problem of finding the maximum likelihood matching for $\mathbf{t_m}$ becomes one of minimizing the objective function below,

$$f_{obj}(\mathbf{t_m}) = (\mathbf{t_m^T} - \mathbf{t_p^T})\mathbf{\Sigma^{-1}}(\mathbf{t_m^T} - \mathbf{t_p^T}) + u(\tau_\mathbf{m})\alpha \tag{4.16}$$

where $u(\tau_\mathbf{m})$ is the number of arrivals left unmatched in $\mathbf{t_m}$, and α is the penalty for an unmatched arrival.

The correlated matcher which minimizes (4.16) provides correct matches for much larger changes in parameter value by fully utilizing the information contained in the

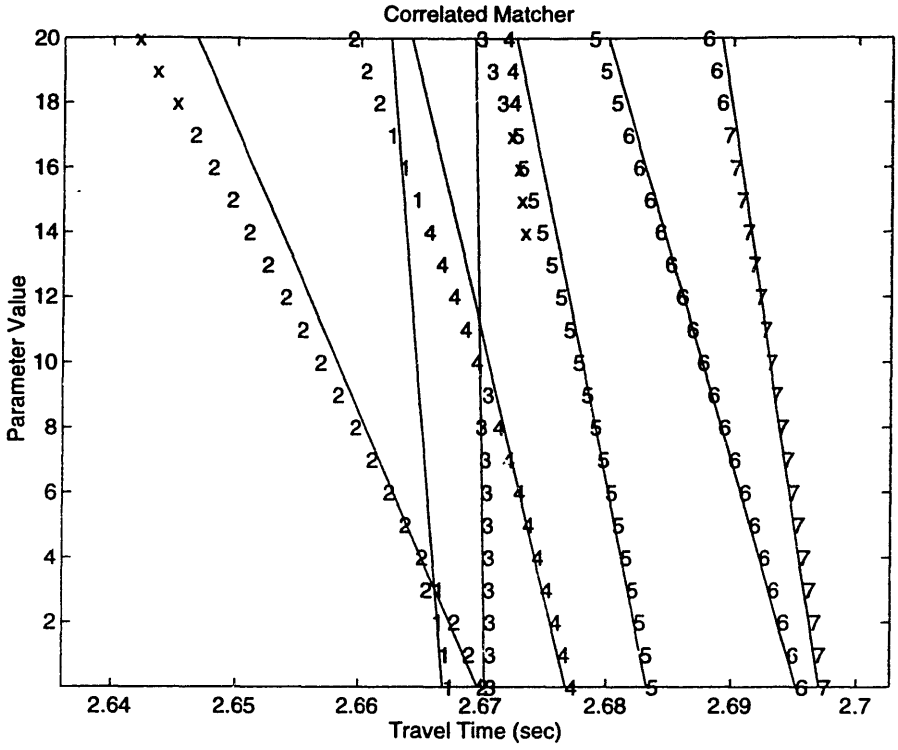

Figure 4-10: Correlated matching.

linear model about correlations in arrival time changes. It starts to fail at parameter value 15 where non-linearities in the travel time shift become significant compared to the time separation between arrivals. The first major failure occurs at parameter value 18 when the non-linearities have become quite large. This major failure at parameter value 18 is reflected on the estimate error graph in Fig. 4-11, where the solid line is the error for the correlated matching algorithm, and the other lines were the error performances for the other algorithms. Note that error is slowly increasing up to this point as a result of non-linearities in the travel time shifts which are not accounted for by the linear model.

Implementation Issues For a small number of arrivals and a small number of beacons, it is practical to evaluate the objective function in (4.16) for all possible

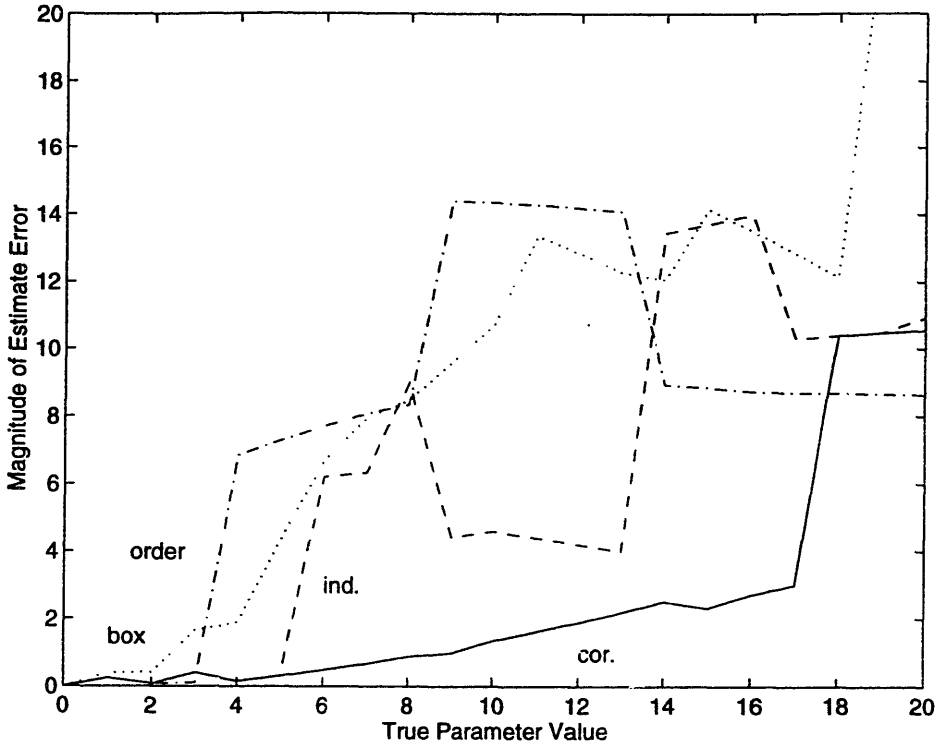

Figure 4-11: Error comparison.

choices of $\mathbf{t_m}$. For more than a few arrivals and beacons, however, the calculation would become cumbersome. Specifically, for N predicted arrival times and M measured arrival times, the number of possible match vectors is:

$$\sum_{n=0}^{\min(N,M)} \frac{M!}{(M-n)!} \begin{pmatrix} N \\ N-n \end{pmatrix} \tag{4.17}$$

The summation is over the number of predictions which will be matched out of the total of N predictions. The maximum number of matches possible is the lower of N and M. There are n predictions which must be matched from the M possible measurements giving the $\frac{M!}{(M-n)!}$ term. This leaves $N-n$ unmatched predictions,

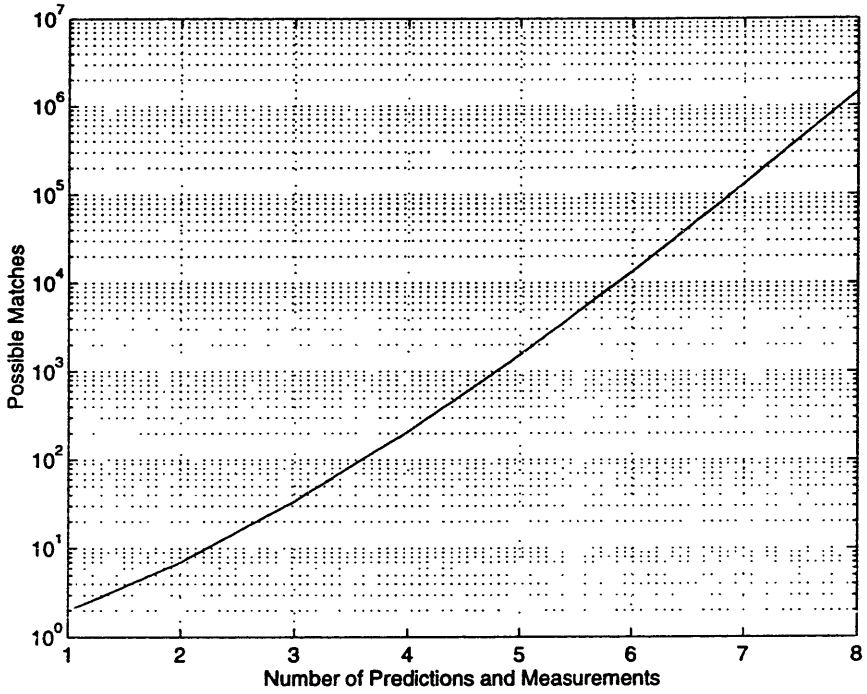

Figure 4-12: Number of possible matches.

which may be chosen in $\begin{pmatrix} N \\ N-n \end{pmatrix}$ ways. In the example problem which has been considered, there were seven predicted arrivals and typically seven measured arrivals. In this situation, there would be 130,992 possible combinations which would have to be evaluated. Fig. 4-12 shows the scaling of the number of possible matches with the number of predicted and measured arrivals. It is assumed in the graph that the number of predicted and measured arrivals is the same.

In an experiment with multiple receivers in the water, the arrival times at any one receiver contain information about the whole environment and therefore about the arrival times at all other receivers. This means that the objective functions for all the receivers must be evaluated together. As a result, the total number of matches which

must be tried is the product of the numbers of matches for each of the individual receivers. It is clear that for systems beyond a very limited size, this can become cumbersome.

Most of the possible matches, however, are quite obviously wrong. For example, the first arrival in the predicted multipath sequence is unlikely to be matched with the last arrival in the measured sequence, and the value of the objective function for such a match reflects this. As a result, the set of possible matches can be quickly pruned using a branch-and-bound algorithm [31]. In such an approach the match vector $\mathbf{t_m}$ is constructed one element at a time. As each element is matched, upper and lower bounds can be found for the objective function value for the *best* match among all possible matches for the remaining undecided elements. An upper bound is the value obtained by assuming that all remaining undecided arrivals find measurements equal to their expected values. A lower bound is the value obtained by assuming that all remaining undecided arrivals are unmatched. Using these bounds, a branch-and-bound algorithm is able to quickly search out the optimal choice of $\mathbf{t_m}$. Fig. 4-13 shows the number of active branches in the search tree as a function of depth through the tree for the matchings generated in Fig. 4-10. At small depths, the tree grows, though not nearly as fast as it would if there were no pruning. At the greater depths near the end of the tree, the bounds on each branch tighten, enabling more pruning, and the tree width actually begins shrinking. Overall, the total number of trial matches which must be evaluated is far smaller than it would be for an exhaustive search.

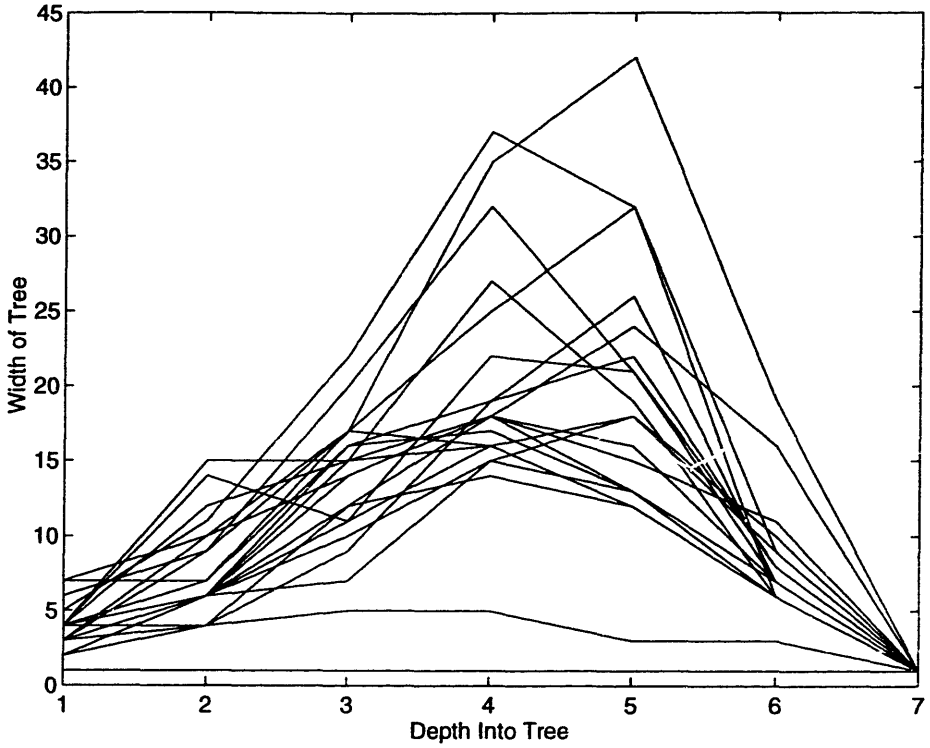

Figure 4-13: Width of tree.

4.6 Advantage of the Correlated Matcher

The systems described in this thesis seek to invert a linearized forward model of the form

$$\mathbf{t} = \mathbf{C}\mathbf{x} + \mathbf{n}, \tag{4.18}$$

where $\mathbf{t}$ is the difference between the predicted and measured travel times, and $\mathbf{x}$ is the corresponding difference between the estimated and true parameters, where these parameters may be source position and time synchronization as described in chapter 2 or weights for an orthogonal function expansion of the sound speed profile as described in chapter 3. Associated with the measurement is a noise vector $\mathbf{n}$, which

includes both true measurement noise as well as the travel time effect of unmodelled sound speed features and non-linearities.

For the inversion to be successful, two conditions are necessary. First, it must be possible to accurately identify the measured arrivals with certain predicted eigenrays, since an incorrect ray path identification will lead to large estimation errors. Second, the travel time effect of the parameter change $\mathbf{Cx}$ must be larger than the noise $\mathbf{n}$, otherwise the measurement will contain little information about $\mathbf{x}$. These two conditions are shown graphically in Fig. 4-14. The horizontal axis of this figure is the standard deviation of the noise $\sigma_{\mathbf{n}}$ divided by the time separation between adjacent arrivals T, and the vertical axis of this figure is the standard deviation of the travel time changes due to parameter variation $\sigma_{\mathbf{Cx}}$ divided by the time separation between adjacent arrivals T. The vertically lined region is where the total standard deviation of travel time variation, including both parameter and noise effects, is greater than the separation between adjacent arrivals T. This is the region where identification tends to become unreliable by conventional methods. The horizontally lined region is where to travel time effect of parameter changes is smaller than the travel time effect of noise and therefore the inversion is poor. Thus with conventional matching algorithms, the system is limited to operation in the unshaded region labeled *conventional tomography.*

While the difficulties of ray path identification for scenarios of larger travel time shifts has been noted in the literature [33], little effort has been devoted to improved identification algorithms. The disinterest in the problem seems to be due to the fact that conventional methods are often adequate for the early arrivals in the deep ocean where most ray-based tomography has occured. In shallow water, however,

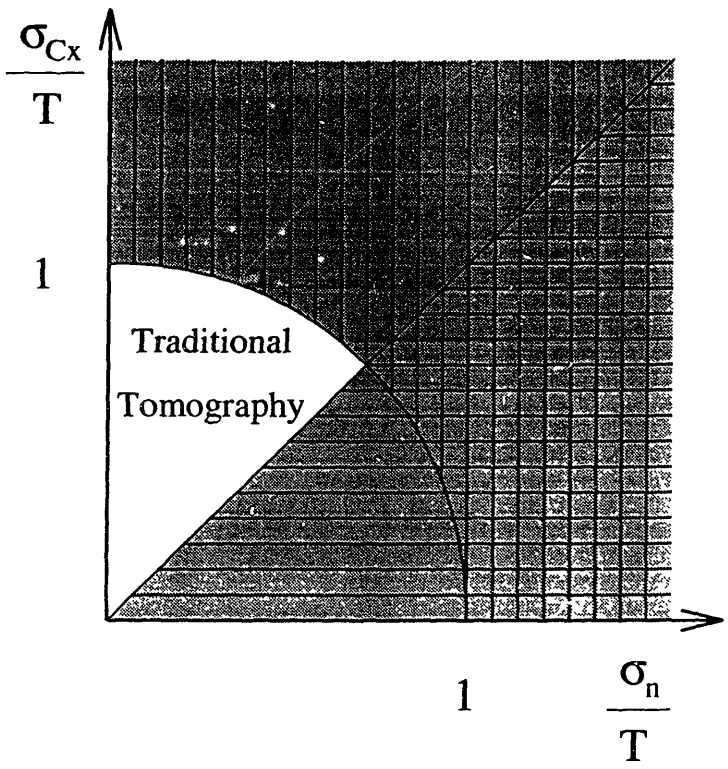

Figure 4-14: Region where tomography is possible using standard arrival matching methods.

ray path identification is more challenging. An algorithm has been presented which fully utilizes the linear model for travel time shifts to allow arrival identification in cases where the parameter-induced travel time shifts may be larger than the time spacing between arrivals. By accounting for the predicted linear shifts in arrivals, this correlated matching algorithm is able to identify arrivals correctly as long as the measurement noise (or non-linearities, which are treated by the model as measurement noise) are not larger than the arrival separations, even for large parameter induced travel time shifts. In Fig. 4-15, the vertically lined region is where arrival identification fails using the new correlated matching algorithm. The horizontally lined region is where the inversion is poor due to the travel time effect of noise being larger than the travel time effect of the parameters of interest. With the new correlated matching

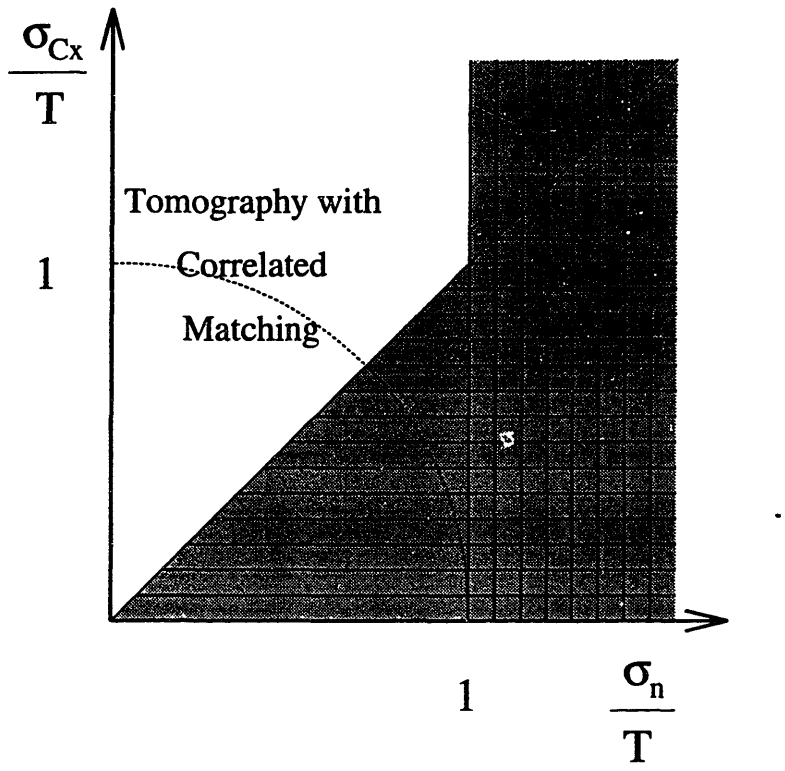

Figure 4-15: Region where tomography is possible using correlated matching method.

algorithm, the system can operate in the unshaded region of Fig. 4-15 which is much larger than the region of operation using conventional matching techniques shown in Fig. 4-14.

This expanded region of arrival identifiability has several useful applications. It makes tomographic inversions possible in environments which are changing quickly. In slowly changing environments, it makes it possible to do all arrival matching from a single initial prediction, instead of having to track arrival shifts. Unlike a tracking system, this system would contain no state, and therefore have no trouble recovering from false matchings. Finally, if an exhaustive search is to be done of a parameter space which is large enough to present significant non-linearity, it allows the parameter space to be carved up into fewer linear search regions than would be possible with

conventional matching techniques, since the size of the search regions is limited by the size of the non-linearity rather than a maximum acceptable size for linear shifts.

4.7 Future Direction

The demonstrations in this section have considered the problem of matching a single set of measured arrival times, or perhaps a set of measured arrival times enhanced by averaging over a period short enough that the parameter values remain constant. Tomography experiments, however, will run over a period of time where parameter values change significantly. Thus the problem of arrival matching becomes a problem of arrival tracking, which has some interesting solution techniques [1].

The most powerful matching technique presented in this section was the correlated matcher which used an objective function that fully exploited the linear model. The problem with the correlated matcher is that even using a branch and bound algorithm it is still rather slow. It may be advantageous, then, to consider some suboptimal matching search strategies that are faster than the branch and bound search. One possible approach to matching would be to begin with only the most certain of the matches, and then use the information obtained from the certain matches to achieve better accuracy in handling the less certain matches [41]. Some help may also be found in the image processing literature, where the matching problem appears in other contexts [17].

4.8 Conclusion

To perform a tomographic or position inversion, it is first necessary to identify the multipath arrivals in the received signal with physical paths through the environment, so that the sensitivity of each arrival to the various sound speed parameters can be determined. In deep ocean tomography, the time spacing between ray arrivals is typically large compared to the parameter induced changes in arrival times, so ray path identification is not difficult. In shallow water, however, ray path identification can be more challenging. An algorithm has been presented which allows tomography using rays where the parameter induced arrival time shifts may be larger than the time spacing between arrivals. The algorithm is also robust to the unexpected appearance and disappearance of subsets of the measured and predicted arrivals.

Chapter 5

Test Experiments

The examples in this chapter will tie together the various contributions of the thesis, and at the same time provide examples in simulation of recommended field trials. In section 5.1, the correlated matching algorithm will be used to enhance navigation accuracy where ambient noise and unknown bottom topography cause unexpected disappearance and appearance of arrivals. The simulation is designed to reflect what can be achieved with *existing* navigation hardware, so the arrival time measurements are made by a simulated wide-band narrow-band detector of the sort commonly used for acoustic navigation receivers. In section 5.2, vertical moving source tomography is demonstrated in a simulation which includes measurement noise, and accounts for errors introduced by the matching algorithms.

5.1 Multipath Navigation

Acoustic positioning systems typically use the first acoustic arrival from each beacon to determine ranges to known beacon locations, and subsequent multipath arrivals are blanked out. As long as the first arrival is reliably present, these systems are dependable, but in many realistic positioning scenarios, the expected first arrival may be blocked by underwater obstacles or masked by noise. If a subsequent multipath arrival is mistaken for the missing first arrival, a position error will result [46].

One attempt at positioning in a fading multipath environment deployed extra redundant beacons and selected for each position estimate only those beacons whose travel times produced a mutually consistent position estimate. This made the system robust to the loss of first arrivals from a few of the beacons. [16]. When it is possible to predict where additional arrivals in the multipath structure will appear, the multipath arrivals can be used to produce a positioning system which is robust to fading without having to add redundant beacons. This thesis presents a positioning system which uses the full multipath structure of the received signal to make the system robust to the fading of individual arrivals. Localization based on multipath delays has been demonstrated by many authors [21, 24]. A unique feature of the system presented here is the ability to deal with missing arrivals and with travel time prediction errors which are larger than the arrival separation.

The system determines its position by a two-step process. First, the detected arrivals in the multipath structure are identified with physical ray paths through the environment using the new correlated matching algorithm which is robust to the dis-

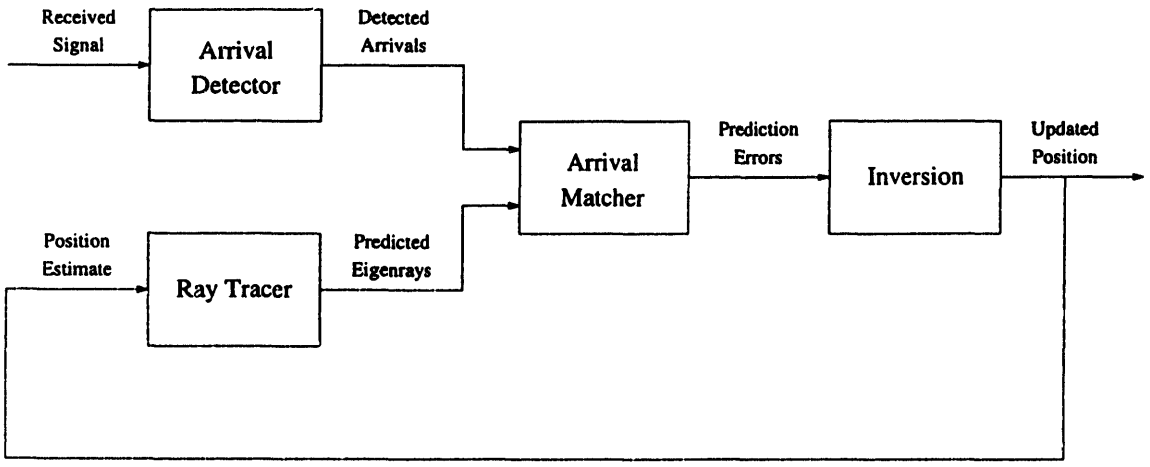

Figure 5-1: Block diagram of the multipath positioning system

appearance of a subset of the predicted and measured arrivals. Then, the differences in arrival times between the measured arrivals and the predicted ray arrivals are used in a linear inversion to produce a position estimate. The multipath positioning system was developed for use in the shallow under-ice sound channel in the Arctic where the expected first arrival may fade in and out due to small changes in the sound speed profile [13, 2]. The operation of the system is simulated in a typical coastal environment where arrivals become unexpectedly absent due to blocking by unknown bottom topography, and where a high ambient noise level often produces missed arrivals and false detections [14].

The structure of the multipath utilization algorithm is shown in Figure 5-1. The received acoustic signal is broken down into a set of ray arrival times by the arrival detector. At the same time, a ray tracing model predicts which eigenrays it expects to see based on its estimate of its current position. The arrival matcher tries to associate each of the predicted eigenrays with one of the detected arrivals, while allowing for the possibility that there may be some blocked arrivals or false detections. Once the

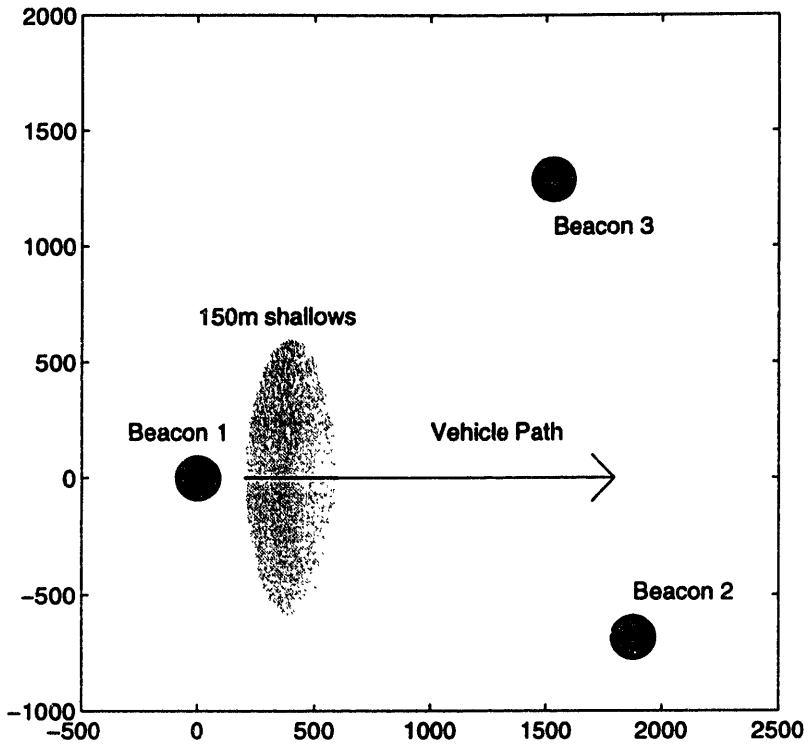

Figure 5-2: Positioning demonstration setup (top view)

detected arrivals have been associated with eigenrays, arrival time prediction errors can be calculated as the differences between the detected arrival times and the arrival times predicted for the associated eigenrays. These arrival time prediction errors are used in an inversion to improve the position estimate [10].

Two important causes of positioning errors in conventional positioning system are missed arrivals or false arrivals caused by noise, and blocking of rays by bottom topography. A simulation was conducted comparing the performance of a conventional positioning system with the multipath system described here under such conditions.

5.1.1 Simulation of Noise Effects

The simulation results presented here assume a receiver which uses a wide-band / narrow-band detector of the sort which is used in most acoustic releases and transponders. In this system, the received signal is filtered by a wide band filter, and then hard limited. The output of the hard limiter has constant power. It is followed by a narrow band filter tuned to the beacon frequency. If most of the constant power in the limiter output is concentrated at the beacon frequency (as when the beacon signal is present), then a large signal comes out of the narrow band filter. If the constant power in the limiter output is evenly distributed over frequency (as when noise only is present), then a small signal comes out of the narrow band filter. A fixed threshold detector on the output of the narrow band filter is triggered by a certain signal-to-noise ratio at the limiter input, regardless of the absolute signal and noise levels. This wide-band / narrow-band detector eliminates the need for careful gain readjustments when the system is moved to environments with different signal and noise levels and it is easy to implement in hardware so it is widely used for transponders and acoustic releases. This common circuit is chosen for this simulation to show that the new multipath navigation algorithm can be added with only a software modification to many existing navigation systems. The acoustical specifications for the simulated system are:

- Source Level: 190dB re 1 μPa
- Ambient Noise Level: 130dB re 1 μPa (Vehicle noise)
- Ping Frequency: 10kHz

- Ping Duration: 3ms
- Wide Band Filter: 8-16kHz, 8-Pole Butterworth
- Narrow Band Filter: 9.7-10.3kHz (3dB Bandwidth), 2-Pole

The relatively high ambient noise specification represents the noise environment for a positioning system mounted on an AUV. The largest sources of noise in this case are motors and gears (and sometimes noise from inductors in the switching DC to DC converters!)

5.1.2 Simulation of Blocking

Underwater obstacles can lead to the unexpected disappearance of one or more multipath arrivals from the blocked source. This effect is introduced by placing a shallow region in the simulated environment. The underwater obstacle (a shallow region) and the beacons and receiver are arranged as described below.

- Water Depth (Normally): 200m
- Water Depth (Shallow Region): 150m
- Shallow Region Width: 400m (along the acoustic path to the vehicle)
- Beacon Depths (all 3 beacons): 175m
- Receiver Depth: 50m
- Sound Speed: 1500m/s

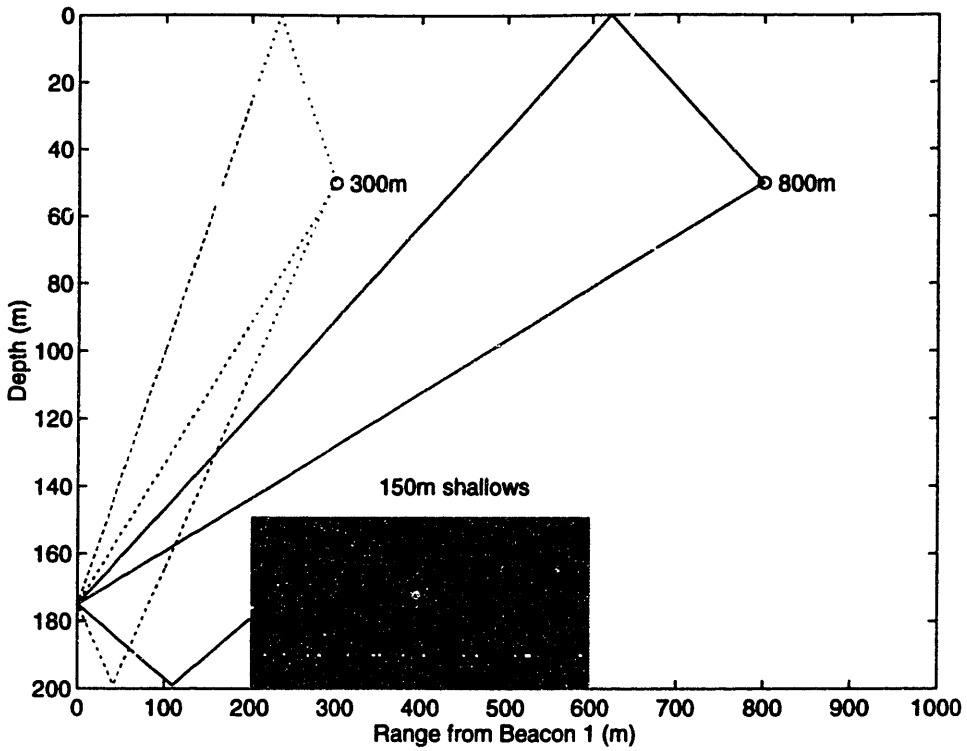

Figure 5-3: Ray blocking by shallow region

The horizontal arrangement of the beacons and the shallow region is shown in Figure 5-2.

The effect of the shallow region is to block some of the ray paths from the beacon which is behind it. The geometry of the positioning network is such that it doesn't effect the other two beacons. This blocking is shown in the ray trace in Figure 5-3. The eigenrays are shown for two source locations, 300m range (dotted lines) and 800m range (solid lines). Note that the bottom-reflected path is blocked at the longer range. As the vehicle continued to even longer ranges, eventually the direct path would also be blocked, leaving only the surface reflection.

5.1.3 Demonstration Plot

The vehicle follows the path shown on Figure 5-2 traveling along a strait line from a position 200m from the first beacon to a position 1800m from the first beacon. In the left frame of Figure 5-4, the dashed line shows the magnitude of the position error at 50 meter intervals for a conventional positioning system, and the solid line shows the error for the new multipath positioning system. The arrival times from each of the three beacons are shown to the right of the error to aid in understanding the cause of the errors. These times are shifted so that the direct path would come in at $t = 0$. The dots represent times when arrivals are expected (without knowing about the shallow region), and the circles represent times when arrivals were actually detected. At all locations and all times, the receiver is subject to noise-induced false detections. If a false detection precedes the first arrival, the conventional system will mistake it for the first arrival, resulting in a range error. Noise may also mask a true arrival. In this case, the conventional system mistakes a subsequent arrival in the multipath structure for the first arrival, resulting in a range error. In addition to these noise induced errors, there is loss of arrivals due to blocking by the shallow region. At ranges of 500m and greater, the receiver is shielded from the bottom reflected arrival (second arrival) from Beacon 1. At ranges of 1000m and greater, the receiver is also shielded from the direct path arrival from Beacon 1 so the conventional system consistently makes position errors. Since the multipath positioning system uses multiple arrivals from each beacon, it is immune to the disappearance of arrivals which causes such large errors for the conventional system.

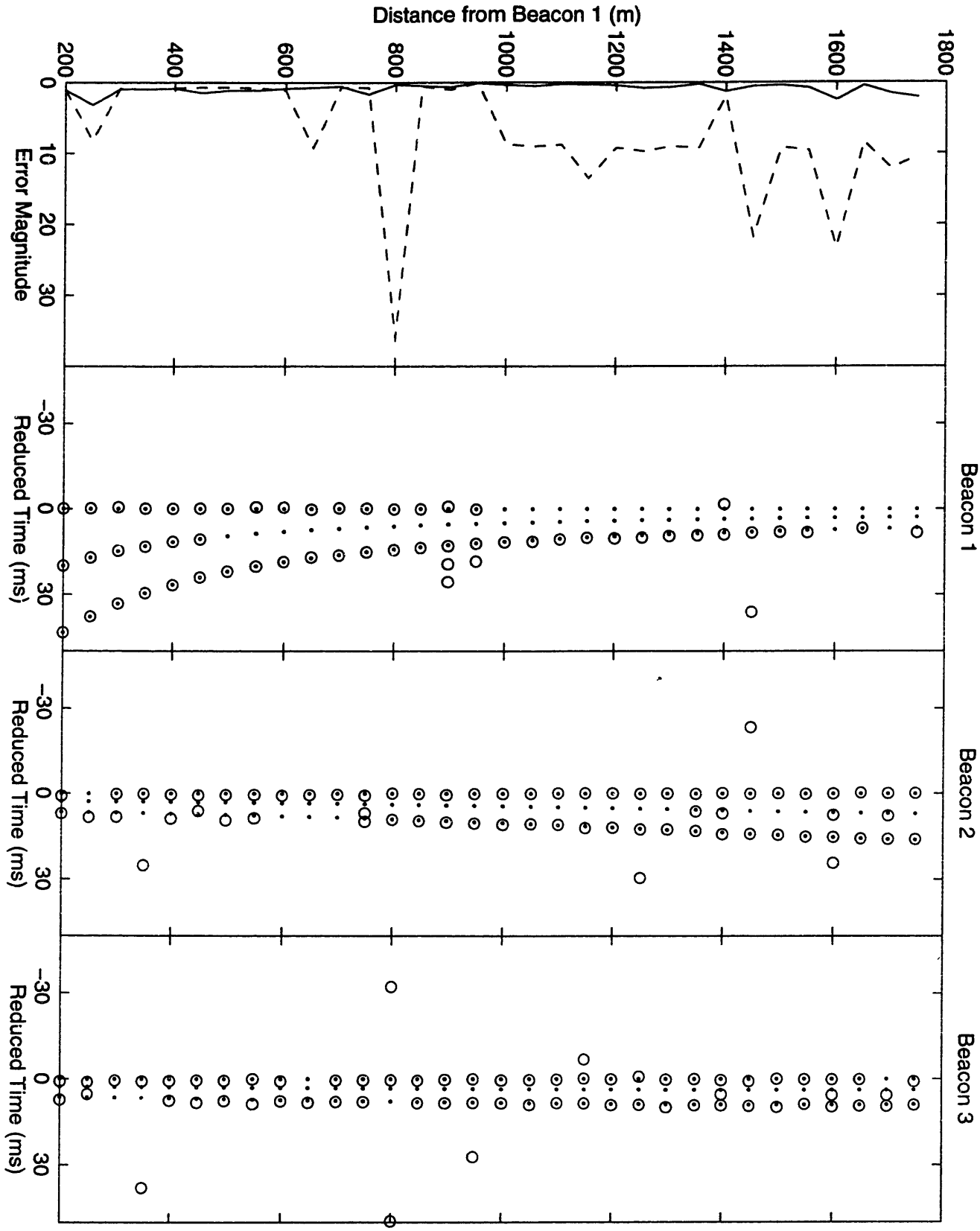

Figure 5-4: Error magnitude (in meters) for both systems, and arrival times of eigenrays from the three beacons

In summary, when the position of multipath arrivals can be predicted, the information provided by the multipath can be used to create a positioning system which is robust to the disappearance or unexpected appearance of a subset of the arrivals. The speed of the arrival matching algorithm and the fact that the system can utilize existing conventional receiver electronics, make feasible the upgrading of many existing positioning systems with only a software change.

5.2 Moving Source Tomography

It was demonstrated in Chapter 3 that moving the acoustic source can focus tomographic resolution at environmental features of interest, assuming that ray paths are identified correctly. An important question then is whether, with the possibility of ray path identification errors, a system can still achieve predicted performance levels. In this example, optimal source locations are found and the tomography problem is solved in a shallow water environment. Repeated trials allow characterization of the true system performance taking into account the effect of arrival matching errors.

5.2.1 Estimator Options

In the examples which follow, sound speed parameters, contained in a *state vector*, will be estimated over the course of a simulated mission. During a mission, the source will transmit K times, and is free to move vertically between transmissions.

The state vector could be estimated recursively over the course of K iterations, with one iteration per source transmission. There would be three steps per iteration, a

prediction step in which the new vehicle state at the time of a transmission is predicted, a *linearization* step where a linear approximation of the relationship between arrival time change and state change is calculated centered about the current state estimate, and a *correction* step in which the measured arrival times are used to correct the state prediction.

The recursive estimator provides the same final parameter estimate as would be obtained if all of the transmissions were taken together in a single inversion *assuming the arrival matchings are the same.* However, the matchings obtained in conjunction with a recursive estimator *are often not the same* as the matchings which would be obtained by taking all the arrivals together. At each iteration where a correct matching is made, information is obtained about the true parameter values. This means the predicted arrival times will be closer to the measured arrival times for the next iteration, and the matching will be more accurate. On the other hand, when an incorrect matching is made, a poor parameter estimate will be obtained, which will make the next matching even less reliable, though the estimated parameter covariance matrix will give no indication of this increasing parameter estimate error. Thus the recursive system will tend either to converge on a good environment estimate and good matchings or diverge resulting in continuing poor matchings.

The recursive estimation process allows the matcher to utilize the correlations between arrival times from one transmission to the next by passing on an updated state estimate. It does not utilize the full arrival time covariance matrix, however. For example, it does not utilize the correlations between the travel time shifts in the last transmission and the first transmission to aid in matching for the first transmission.

For this reason, recursive estimators will generally perform worse than a one-time estimator using the correlated matching algorithm.

The disadvantage of a one-time inversion using the correlated matching algorithm is the computational burden presented by the correlated matcher, and real-time implementations may want to consider obtaining an initial environment estimate using several transmissions simultaneously, with the number of transmissions incorporated in each inversion decreasing as the environmental estimate improves. For purposes of this simulation, however, the computational burden of the correlated matcher is not large, and so one-time inversions will be used.

5.2.2 Experiment Description

An acoustic source is attached to a cable crawler which is able to move vertically on a mooring cable. It is assumed in this example that the source position and transmit time are known. The receiver consists of a single hydrophone at a depth of 50m. The water depth is 200m in this simulation, and the source and receiver are separated by a 4000m range.

Example 1: Two Parameters / Six Rays

In the first example, the water between source and receiver is horizontally uniform, and the sound speed is described as a mean sound speed profile shown in the left half of Fig. 5-5, with the two variations represented by the basis functions in the right half of Fig. 5-5. The weights for these two variations are considered to be independent.

In the example which follows the source will transmit twice, with the optimal depths

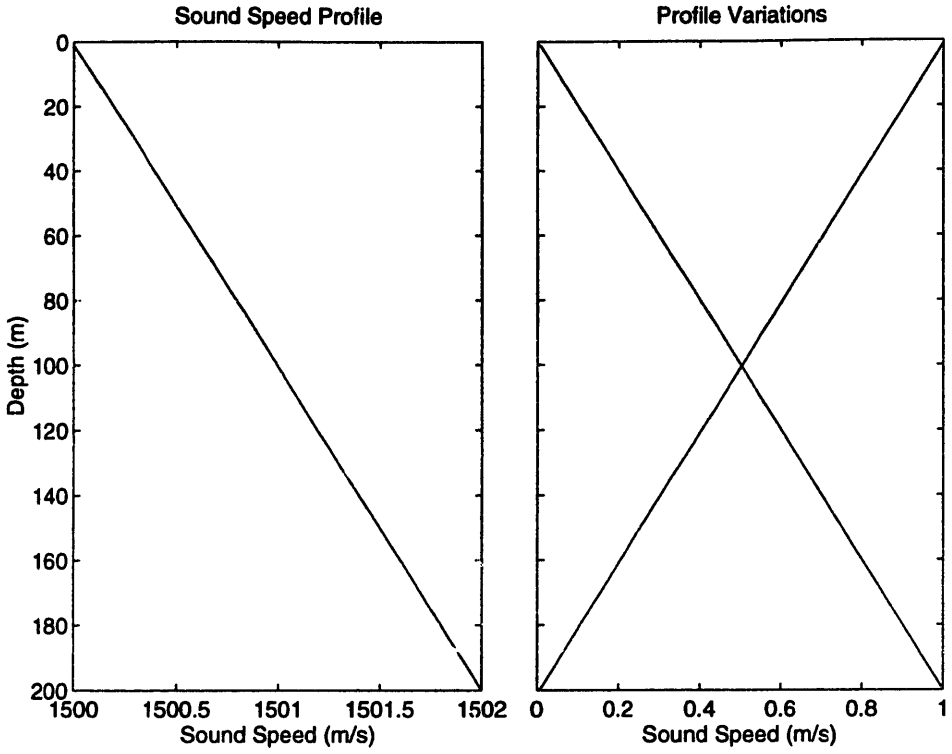

Figure 5-5: Mean sound speed profile and profile variations

for transmission selected as in Chapter 3 to minimize the total variance throughout the environment, assuming a weight variance of one. The true weight values will be selected as independent identically distributed Gaussian random variables with variance σ_z^2. Based on the true parameter vector, the true ray travel times from each source location to the receiver will be determined. These true travel times will then be corrupted by adding measurement noise which is independent identically distributed Gaussian random variables with variance σ_n^2. The matching algorithms described in Chapter 4 will be employed to match measured and predicted arrivals, and an inversion will be performed based on the results of each matching. The parameter estimate error will be recorded for each inversion, and the estimate variance for each parameter will be determined experimentally by averaging the results of 1000 trials. These variance

estimates will be compared with the theoretical variance which is calculated based on the assumption of perfect matching, and the differences are analyzed for various values of σ_z^2 with $\sigma_n = 0.001s$.

The optimal depth selection is shown in Fig. 5-6 to minimize the total weight error assuming a weight variance of one. Parameter estimates are formed for 1000 trials. One test of how well the true system performance compares with the theoretical performance bounds is to determine what fraction of the parameter estimate errors falls within the theoretical one standard deviation limit. If the theoretical bound is correct, then this ratio should be 0.683. The ratio is shown in Fig. 5-7 for various values of σ_z. The solid line indicates the ratios for the correlated matcher. The dashed line indicates the ratios for the uncorrelated matcher. The dash-dot line indicates the ratios for the validation window matcher, and the dotted line indicates the ratios for the order matcher.

The correlated and uncorrelated matchers have equal ratios when the parameter variances are small, since the observed travel time shifts are uncorrelated being due mostly to noise. The measured ratios approach their theoretical values since all predicted ray paths are present, non-linearities are small, and the probability of incorrect identifications is also small.

As the parameter variances increase, three effects cause the ratio to decrease. First, some of the predicted arrivals may not have matches anymore among the measured arrivals due to fading. If the remaining matches are correctly matched, the estimate variance will still be larger than its theoretical value, since the information provided by the faded path has been lost. This results in a small decrease in the ratio. Second,

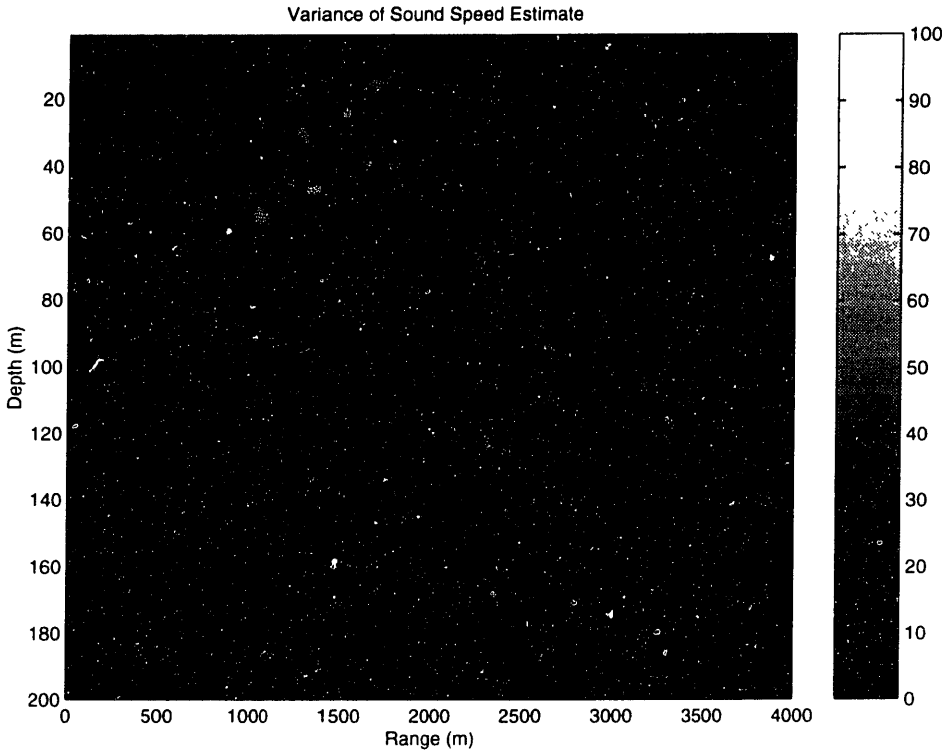

Figure 5-6: Optimal transmission depths

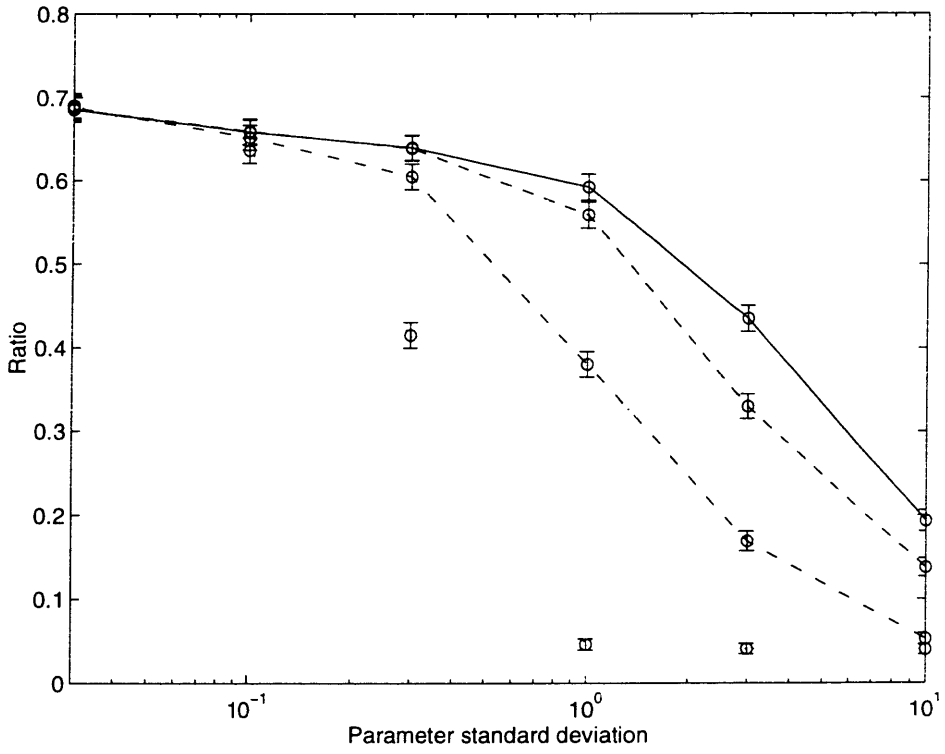

Figure 5-7: Fraction of errors lying within one theoretical standard deviation (should be 0.683)

some of the arrivals will begin to be incorrectly matched. An incorrect matching can cause a large inversion error, lowering the ratio. Third, when the parameter changes become large, non-linearities become important. These non-linearities have the effect of measurement noise, though they are not accounted for in the measurement noise covariance matrix. Thus, they make the effective measurement noise larger than the modeled measurement noise, and cause a decrease in the ratio.

Note that the correlated matcher remains closer to its theoretical bound than any of the other matchers, due to its lower likelihood of making incorrect matches. It is, however, still subject to reductions in the ratio due to faded arrivals and non-linearity.

Example 2: Eight Parameters / Nine Rays

In this second example, The water between source and receiver is divided into 4 range segments, and within each segment, the sound speed is described as a mean sound speed profile shown in the left half of Fig. 5-8, with variations represented by the basis functions in the right half of Fig. 5-8. The weights for these variations are considered to be independent.

The source will transmit three times. The optimal depth selection is shown in Fig. 5-9 to minimize the error in the estimate of parameter 6, the weight of the mid-column variation in the second range division from the left for $\sigma_z^2 = 10$. The parameter estimate errors for the 1000 trials are shown for each parameter as pluses on Fig. 5-10. The horizontal axis is the parameter number. The even parameter numbers correspond to the mid-water variation and the odd parameter numbers correspond to the surface variation, with parameter numbers 1 and 2 corresponding to the range division nearest

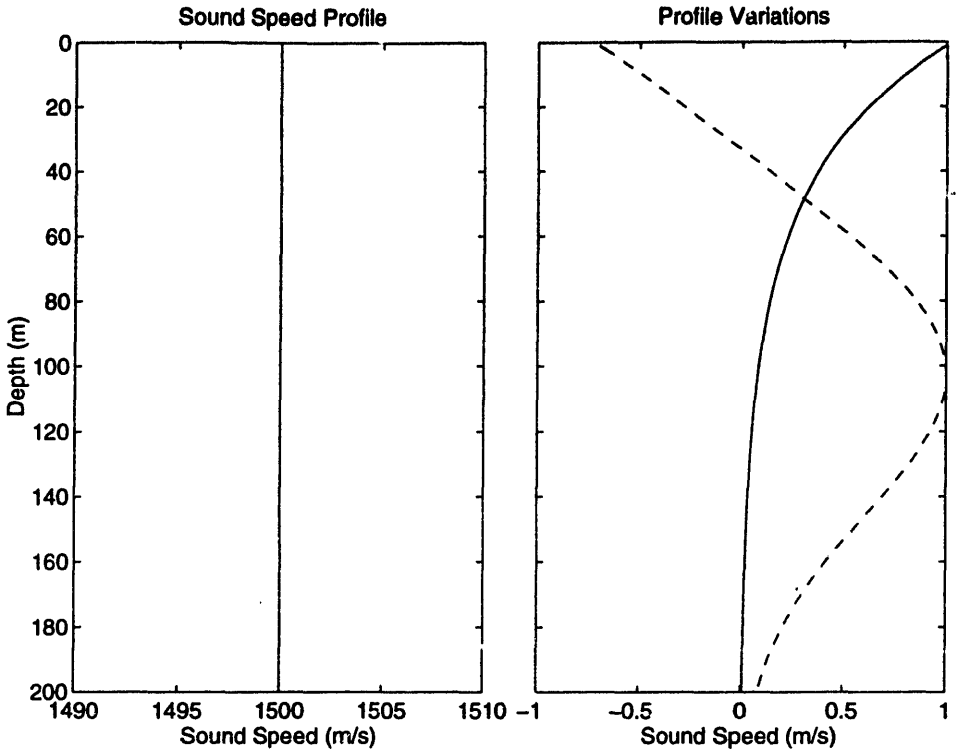

Figure 5-8: Mean sound speed profile and profile variations

the receiver and increasing toward the source. The theoretical standard deviations for a perfect matching are shown as lines. These errors were based on $\sigma_z = 1$, and $\sigma_n = 0.0001s$.

The fraction of the parameter estimates which have errors within one standard deviation is shown in Fig. 5-11 for various values of σ_z. The solid line indicates the ratios for the correlated matcher. The dashed line indicates the ratios for the uncorrelated matcher. The dash-dot line indicates the ratios for the validation window matcher.

In the first example there were a total of six rays sampling the environment and only two parameters to estimate. In this example, there are nine rays and eight parameters, so the inverse here is only slightly overdetermined. Since there are so

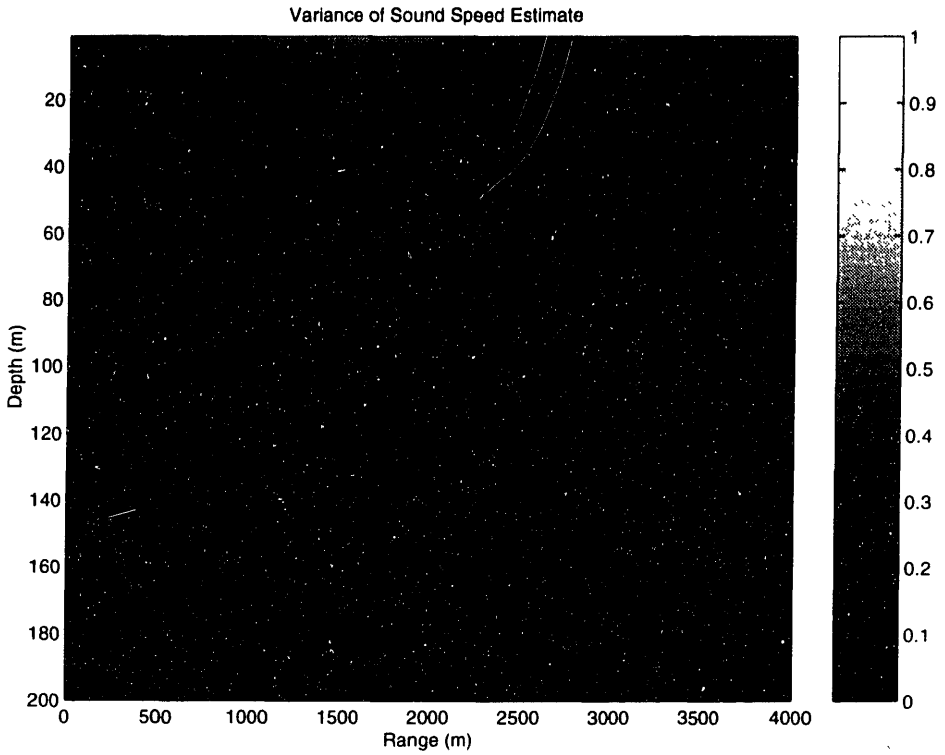

Figure 5-9: Optimal transmission depths

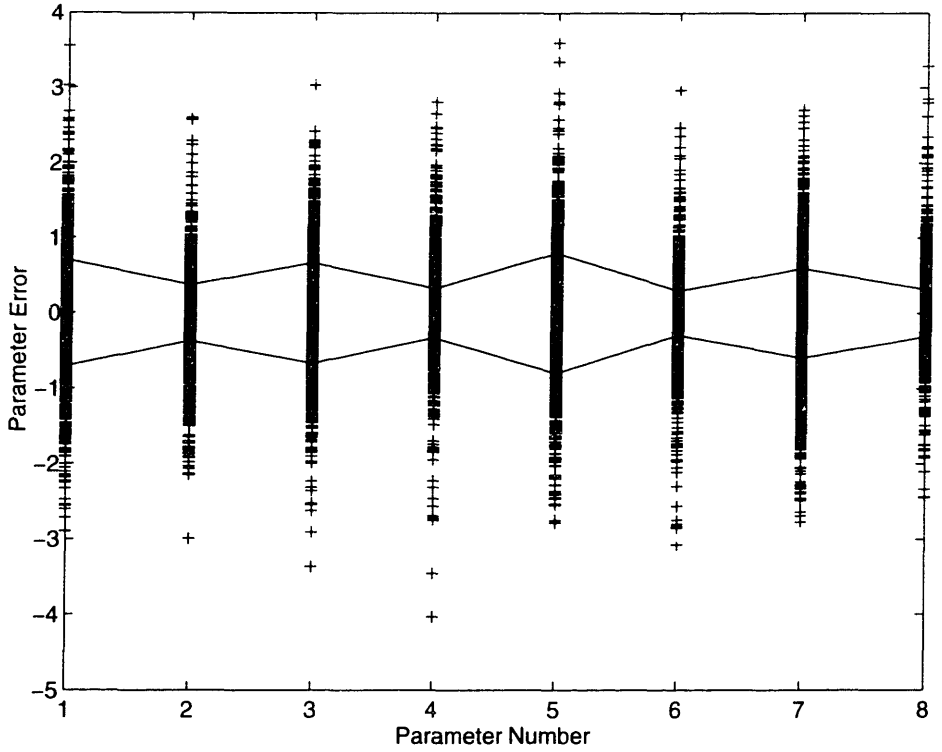

Figure 5-10: Parameter estimate errors and theoretical standard deviation

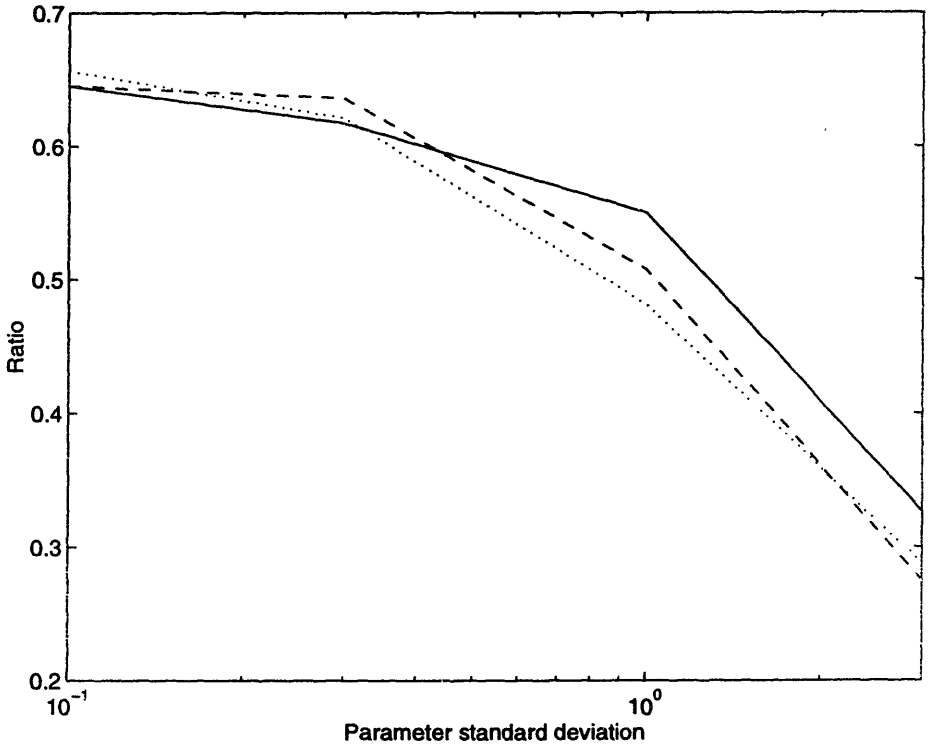

Figure 5-11: Fraction of errors lying within one theoretical standard deviation (should be 0.683)

many more parameters relative to the number of ray paths, the correlated matcher does not offer as large a benefit in performance as it did for the very over-determined case.

Chapter 6

Conclusion

This thesis contributes techniques and analysis tools for optimal moving source tomography in ray environments, as well as offering insights into the estimation problems which underly moving source tomography.

The first problem in moving source tomography is determining where the sources are located. The standard navigation techniques of spherical and hyperbolic positioning are shown to be two end points of a continuum of possible systems. It is then shown that hyperbolic systems can move along this continuum toward spherical performance limits if the time synchronization between the vehicle clock and the master beacon clock is estimated. A rule of thumb is given for when such time synchronizing systems offer significant position accuracy improvement over hyperbolic systems. Finally, it is observed that navigation accuracy depends on both present and past vehicle positions, and optimal navigation is defined as the problem of determining the vehicle path from an origin to a destination such that the position estimate error is minimized upon reaching the destination.

The second problem in moving source tomography is representation and estimation of sound speed profile variability. The fundamental contribution of the thesis in this area is the derivation of ray sampling functions which describe the sensitivity of ray travel time variations to sound speed profile variations at any depth. These ray sampling functions allow the derivation of an optimal orthogonal function expansion for sound speed profile variability which leads to more accurate tomographic inversions than are possible with the commonly used method of empirical orthogonal functions. The ray sampling functions also allow derivation of a minimum variance reduced order estimator for the sound speed profile, which again offers improved performance over standard methods, particularly in the rejection of errors due to unmodelled profile variations. The ray sampling functions illustrate that tomographic resolution at a given region in the environment is highly dependent on source and receiver locations. This leads to posing the optimal moving source tomography problem of finding the locations where a moving source should transmit in order to minimize the variance of the sound speed estimate in a certain region of interest.

The third problem in moving source tomography is ray path identification. Two new algorithms are presented for ray path identification. The common thread in both algorithms is that they incorporate the linear model for travel time variability that includes the effect of both parameter changes and noise. The better (though slower) of these two algorithms accounts for correlations between travel time shifts to allow accurate arrival identification over much larger ranges of parameter uncertainty than is possible with commonly used techniques. This algorithm enables tomography in a broader range of environments, and also enables moving source tomography where the

travel time changes due to source motion can thwart standard arrival identification techniques.

The contributions of the preceding chapters are brought together in Chapter 5, which presents realistic simulations demonstrating the application of the algorithms. These simulations serve not only to demonstrate the capabilities of the algorithms developed in this thesis, but are also intended as suggestions for simple proof of concept demonstrations.

Moving source tomography is a powerful tool for improving the information return from oceanographic experiments. Constraints of energy and time make it important to utilize moving sources in an optimal way. It is the author's hope that this thesis has added to our understanding of the moving source tomography problem, and that the techniques presented here will find useful application in efficient oceanographic and seismic imaging.

Bibliography

[1] Yaakov Bar-Shalom. *Tracking and Data Association.* Academic Press, Boston, 1988.

[2] James G. Bellingham, Max Deffenbaugh, John J. Leonard, Josko Catipovic, and Henrik Schmidt. Arctic under-ice survey operations. *Proc. 8th Int. Symp. Unmanned Untethered Submersible Technology*, pages 50–59, 1993.

[3] William L. Brogan. *Modern Control Theory, 3rd ed.* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1991.

[4] Alfred M. Bruckstein, Tie-Jun Shan, and Thomas Kailath. The resolution of overlapping echos. *IEEE Trans. Acoust., Speech, and Signal Processing*, ASSP-33(6):1357–1367, 1985.

[5] Brian D. Bunday. *Basic Linear Programming.* Edward Arnold Publishers, Baltimore, 1980.

[6] J. A. Cestone, R. J. Cyr, Gordon Roesler, and E. St. George. Recent developments in acoustic underwater navigation. *J. Navigation*, 30(2):246–280, 1977.

[7] Michael D. Collins and W. A. Kuperman. Focalization: Environmental focusing and source localization. *J. Acoust. Soc. Am.*, 90(3):1410–1422, 1991.

[8] B. Cornuelle, W. Munk, and P. Worcester. Ocean acoustic tomography from ships. *J. Geophys. Res.*, 94(C5):6232–6250, 1989.

[9] Russ E. Davis. Predictability of sea surface temperature and sea level pressure anomalies over the north pacific ocean. *J. Phys. Ocean.*, 6(3):249–266, 1976.

[10] Max Deffenbaugh. A matched field processing approach to long range acoustic navigation. Master's thesis, Massachusetts Institute of Technology, February 1994.

[11] Max Deffenbaugh, James G. Bellingham, and Henrik Schmidt. The relationship between spherical and hyperbolic positioning. *Proc. Oceans '96*, pages 590–5, 1996.

[12] Max Deffenbaugh and Henrik Schmidt. Improved empirical orthogonal functions for ocean acoustic tomography. *J. Acoust. Soc. Am.*, 99(4):2451, 1996.

[13] Max Deffenbaugh, Henrik Schmidt, and James G. Bellingham. Acoustic navigation for arctic under-ice AUV missions. *Proc. Oceans '93*, pages I–204–9, 1993.

[14] Max Deffenbaugh, Henrik Schmidt, and James G. Bellingham. Acoustic positioning in a fading multipath environment. *Proc. Oceans '96*, pages 596–600, 1996.

[15] R. Deutsch. *Estimation Theory.* Prentice-Hall, Englewood Cliffs, NJ, 1965.

[16] G. L. Duckworth. A robust algorithm for tracking of drifting acoustic arrays in the arctic. *Conference Record: Twenty-first Asilomar Conference on Signals, Systems, and Computers*, pages 103–108, 1987.

[17] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis.* John Wiley and Sons, New York, 1973.

[18] Paul S. Dwyer. Some applications of matrix derivatives in multivariate analysis. *J. Am. Stat. Assoc.*, 62:607–625, 1967.

[19] Paul S. Dwyer and M. S. Macphail. Symbolic matrix derivatives. *Annals of Math. Stat.*, 19:517–534, 1948.

[20] J. E. Ehrenberg, T. E. Ewart, and R. D. Moore. Signal-processing techniques for resolving individual pulses in a multipath signal. *J. Acoust. Soc. Am.*, 63(6):1861–1865, 1978.

[21] Benjamin Friedlander. Accuracy of source localization using multipath delays. *IEEE Trans. Aero. and Elect. Sys.*, 24(4):346–359, 1988.

[22] Arthur Gelb. *Applied Optimal Estimation.* The M.I.T. Press, Cambridge, MA, 1989.

[23] I.S. Gradshteyn and I.M. Ryzhik. *Table of Integrals, Series, and Products, 4th Ed.* Academic Press, Inc, San Diego, New York, Berkeley, 1980.

[24] J. C. Hassab. Passive tracking of a moving source by a single observer in shallow water. *J. Sound and Vibration*, 44(1):127–145, 1976.

[25] James P. Ignizio. *Linear Programming in Single and Multiple Objective Systems.* Prentice-Hall, Englewood Cliffs, NJ, 1982.

[26] F.B. Jensen, W.A. Kuperman, M.B. Porter, and H. Schmidt. *Computational Ocean Acoustics.* American Institute of Physics, Woodbury, New York, 1994.

[27] I. T. Jolliffe. *Principal Component Analysis.* Springer-Verlag, Berlin, 1986.

[28] S. Kirkpatrick, Jr. C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[29] D. D. Kosambi. Statistics in function space. *J. Indian Math. Soc.*, 7:76–88, 1943.

[30] P. K. Kundu, J. S. Allen, and L. R. Smith. Modal decomposition of the velocity field near the oregon coast. *J. Phys. Ocean.*, 5:683–704, 1975.

[31] E. L. Lawler and D. E. Woods. Branch and bound methods: A survey. *Operations Research*, 14:699–719, 1966.

[32] L. R. LeBlanc and F. H. Middleton. An underwater acoustic sound velocity data model. *J. Acoust. Soc. Am.*, 67(6):2055–2062, 1980.

[33] J. A. Mercer and J. R. Booker. Long-range propagation of sound through oceanic mesoscale structures. *J. Geophys. Res.*, 88(C1):689–699, 1983.

[34] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 24(3):1087–1092, 1953.

[35] P. H. Milne. *Underwater Acoustic Positioning Systems.* E. and F. N. Spon, Ltd., New York, 1983.

[36] Walter Munk, Peter Worchester, and Carl Wunsch. *Ocean Acoustic Tomography.* Cambridge University Press, New York, 1995.

[37] Walter Munk and Carl Wunsch. Ocean acoustic tomography: a scheme for large scale monitoring. *Deep-Sea Research*, 26A:123–161, 1979.

[38] Walter H. Munk. Sound channel in an exponentially stratified ocean, with application to SOFAR. *J. Acoust. Soc. Am.*, 55(2):220–226, 1974.

[39] N. E. Nahi. *Estimation Theory and Applications.* John Wiley and Sons, Inc., New York, 1969.

[40] Joseph C. Park and Robert M. Kennedy. Remote sensing of ocean sound speed profiles by a perceptron neural network. *IEEE Journal of Ocean Engineering*, 21(6):216–224, 1996.

[41] James C. Preisig. Personal Communication, April 1997.

[42] Gerald S. Rogers. *Matrix Derivatives.* Marcel Dekker, Inc., New York, 1980.

[43] G. J. Sonnenberg. *Radar and Electronic Navigation, 6th ed.* Butterworth and Co., Boston, 1988.

[44] Derrick S. Tracy and Paul S. Dwyer. Multivariate maxima and minima with matrix derivatives. *J. Am. Stat. Assoc.*, 64:1576–1594, 1969.

[45] R. J. Tremblay, D. W. Lytle, and G. C. Carter. A practical approach to the estimation of amplitude and time-delay parameters of a composite signal. *IEEE Journal of Ocean Engineering*, OE-12(1):273–278, 1987.

[46] Henry VanCalcar. What you need to know to improve LBS submersible tracking performance. *Proc. ROV '85*, pages 43–51, 1985.

[47] Peter F. Worcester, Bruce D. Cornuelle, and Robert C. Spindel. A review of ocean acoustic tomography: 1987-1990. *Reviews of Geophysics, Supplement*, pages 557–570, 1991.