Adaptive Matched Field Processing in an Uncertain Propagation Environment

RLE Technical Report No. 567

James C. Preisig

January 1992

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139-4307

This work was supported in part by the Defense Advanced Research Projects Agency monitored by the Office of Naval Research under Grant N00014-89-J-1489, in part by the Office of Naval Research under Grant N00014-91-J-1628, in part by the National Science Foundation under Grant MIP 87-14969, in part by a National Science Foundation Graduate Fellowship, and in part by a General Electric Foundation Graduate Fellowship in Electrical Engineering.

Adaptive Matched Field Processing in an Uncertain Propagation Environment

by

JAMES CALVIN PREISIG

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
and the
WOODS HOLE OCEANOGRAPHIC INSTITUTION
January 1992

Abstract

Adaptive array processing algorithms have achieved widespread use because they are very effective at rejecting unwanted signals (i.e., controlling sidelobe levels) and in general have very good resolution (i.e., have narrow mainlobes). However, many adaptive high-resolution array processing algorithms suffer a significant degradation in performance in the presence of environmental mismatch. This sensitivity to environmental mismatch is of particular concern in problems such as long-range acoustic array processing in the ocean where the array processor's knowledge of the propagation characteristics of the ocean is imperfect. An Adaptive Minmax Matched Field Processor has been developed which combines adaptive matched field processing and minmax approximation techniques to achieve the effective interference rejection characteristic of adaptive processors while limiting the sensitivity of the processor to environmental mismatch.

The derivation of the algorithm is carried out within the framework of minmax signal processing. The optimal array weights are those which minimize the maximum conditional mean squared estimation error at the output of a linear weight-and-sum beamformer. The error is conditioned on the propagation characteristics of the environment and the maximum is evaluated over the range of environmental conditions in which the processor is expected to operate. The theorems developed using this framework characterize the solutions to the minmax array weight problem, and relate the optimal minmax array weights to the solution to a particular type of Wiener filtering problem. This relationship makes possible the development of an efficient algorithm for calculating the optimal minmax array weights and the associated estimate of the signal power emitted by a source at the array focal point. An important feature of this algorithm is that it is guarenteed to converge to an exact solution for the array weights and estimated signal power in a finite number of iterations.

The Adaptive Minmax Matched Field Processor can also be interpreted as a two-stage Minimum Variance Distortionless Response (MVDR) Matched Field Processor. The first stage of this processor generates an estimate of the replica vector of the signal emitted by a source at the array focal point, and the second stage is a traditional MVDR Matched Field Processor implemented using the estimate of the signal replica vector.

Computer simulations using several environmental models and types of environmental uncertainty have shown that the resolution and interference rejection capability of the Adaptive Minmax Matched Field Processor is close to that of a traditional MVDR Matched Field Processor which has perfect knowledge of the characteristics of the propagation environment and far exceeds that of the Bartlett Matched Field Processor. In addition, the simulations show that the Adaptive Minmax Matched Field Processor is able to maintain it's accuracy, resolution and interference rejection capability when it's knowledge of the environment is only approximate, and is therefore much less sensitive to environmental mismatch than is the traditional MVDR Matched Field Processor.

Thesis Supervisor: Alan V. Oppenheim
Title: Distinquished Professor of Electrical Engineering

Thesis Supervisor: Arthur B. Baggeroer
Title: Ford Professor of Electrical and Ocean Engineering

Acknowledgments

While the completion of a doctorate is looked upon as an individual accomplishment, it is rarely accomplished without the help of a significant cast of supporting characters. Such is the case here, for it has been the friendship, guidance, and support of many people that has helped make the last six years some of the most exciting and stimulating of my life and have help me through the times when things got tough.

Over the past several years, Al Oppenheim has been a source of inspiration and stimulation, has tolerated my often unhealthy addiction to bicycle racing, and deftly alternated between providing the encouragement and critical evaluation which I needed. His guidance and friendship has helped me grow both professionally and personally. Art Baggeroer has similarly taken the time and effort to push me to develop more fully in my work over the past few years. For their concern, caring and effort, I am eternally grateful.

I would also like to thank Udi Weinstein, Bruce Musicus and Rob Freund for the time and friendship which they have freely given to help me both technically and personally in the past years. Henrik Schmidt has also provided much help with his constant feedback on acoustics issues and his continual help with using SAFARI in doing my work.

All my colleagues in the Digital Signal Processing Group, past and present, have been intrumental in making the last three years fun and challenging. Particular among these have been Greg Wornell, who has ridden with me on this roller coaster for the last six years and has been an accomplice in creating an unholy mess in my kitchen on many occasions; and Steve Isabelle, who, as an honorary member of TEAM THESIS 91, tolerated far more repetitions of the music of the Nylons and of Andrew Lloyd Webber than was reasonable. Special thanks also go to John Buck, Andy Singer, and Giovanni Aliberti who have patiently answered so many of my dumb computer questions that they must be convinced (rightly so) that I am a hopeless case when it comes to computer system management. My other friends including Jim and Carol Bowen, John and Dana Richardson, Josko and Liz Catipovic, and all my cohorts in the world of bicycle racing have been constant sources of friendship and excitement and have helped me keep the world in a little better perspective through all of this. And finally, the love and understanding of my family has been the pillar of support which has kept me upright over the years. Without them, none of this would have happened.

Financially, I have relied upon the generousity of others to keep me fed and in school here at MIT and WHOI. For their support, I would like to thank the National Science Foundation, the General Electric Foundation, the Office of Naval Research, the Defense Advanced Research Projects Agency, and the Woods Hole Oceanographic Institution.

To all of you, and to all of the many people who are not specifically listed above but who have been there and given of themselves whenever I needed their help and friendship, I say thanks. You have helped make it a wonderful 6 years.

Contents

List of Figures

Chapter 1

Introduction

The signals received by spatial arrays of sensors are often composed of the sum of signals emitted by sources at different locations. In order to estimate the signal, or the parameters of the signal, emitted by a source at a particular location, the array processor must often separate that signal from the other signals which are received. This separation of signals based upon the location of the source is referred to as spatial filtering. Thus, the spatial filtering of signals received by an array of sensors to generate estimates of the parameters of the signals emitted by sources at locations of interest is an important operation in many array processing applications.

Array processors achieve spatial discrimination through filtering by exploiting the fact that the spatial characteristics of a propagating signal as received at an array of sensors depend upon the location of the source of the signal. However, the spatial characteristics of a propagating signal also depend upon the characteristics of the medium through which the signal is propagating. Therefore, if a processor has inaccurate or incomplete information concerning the characteristics of the propagation environment, it may be unable to determine the spatial characteristics which should be exhibited by a signal emitted by a source at the location of interest. In this case, the processor may have difficulty in accomplishing the spatial filtering necessary to estimate the parameters of the signal of interest. This work proposes an approach to array processing which yields a processor capable of operating with only approximate environmental information while at the same time achieving levels of spatial discrim-

ination which are close to those achieved by adaptive processors having accurate and detailed environmental information.

The remainder of this chapter contains general background information on array processing. Section 1.1 discusses general linear, adaptive, and matched field processing. Section 1.2 describes a parameterization of the spatial characteristics of propagating signals, which is useful for the class of algorithms considered herein. The problems which array processors exhibit when the environmental information is inaccurate, and possible approaches to developing processors which are able to operate effectively with inaccurate or imprecise information are reviewed in Section 1.3. This section also introduces the minmax signal processing approach, which is proposed herein to address the problem of array processing with only approximate environmental information.

The theoretical foundations of the minmax approach, based on the Minmax Characterization Theorm, are developed in Chapter 2. This theorem sets forth the necessary and sufficient conditions which must be met by any solution to a general class of minmax problems. The details of the proposed array processor, referred to as the Adaptive Minmax Matched Field Processor, are presented. A computationally efficient algorithm which is guaranteed to converge to an exact solution of the minmax optimization problem of interest is developed by exploiting the special structure imposed on the solution by the Minmax Characterization Theorem. Finally, an approach to bounding the minmax performance achievable by any processor is proposed.

In Chapter 3, the structure imposed by the Minmax Characterization Theorem is again exploited to relate the Adaptive Minmax Matched Field Processor to Capon's Minimum Variance Distortionless Response (MVDR) Matched Field Processor [9, 11]. The relationship developed leads to a qualitative analysis of the processor. This analysis motivates a small change to the algorithm developed in Chapter 2. A quantitative analysis of the algorithm based on results of numerical simulations is also presented. These numerical results are generated for both deterministic time-invariant and random time-varying propagation environments. The results for the latter case motivate another small change to the algorithm which is also detailed.

Chapter 4 addresses the problem of generating a priori estimates of the spatial/temporal characteristics of a propagating signal as a function of environmental conditions and source location. This chapter does not present original work. Instead, it presents results developed by others [29, 30, 31, 32, 33, 40, 41] on the propagation of signals through random media and outlines how this work can be applied to generating estimates of the spatial/temporal signal characteristics. Finally, the results generated herein are summarized and future work is discussed in Chapter 5.

1.1 Linear, Adaptive, and Matched Field Processing

Many types of array processors either implicitly or explicitly incorporate a linear weight-and-sum beamformer to implement spatial filtering. This filtering allows the processor to discriminate among signals based upon the location of the source of the signals. Given an input $\underline{y}$ (the joint temporal/spatial filtering of the sampled vector time series $\underline{y}[n]$ will not be considered in this introduction), the output of a linear weight-and-sum beamformer is $\hat{x} = \underline{w}^h \underline{y}$ where $\underline{w}$ is the array weight vector, the superscript h denotes complex conjugate transpose (i.e., Hermitian), and $\hat{x}$ is an estimate of the signal emitted by a source at a location of interest.

Linear beamformers enjoy widespread use for several reasons. First, they generally have the lowest computational complexity of the available methods of implementing a spatial filter (given an N element array, the filtering operation is an $\mathcal{O}(N)$ operation and, if required, the calculation of the array weights to minimize a squared error criterion is often an $\mathcal{O}(N^3)$ operation). Second, when the array weights are chosen to minimize a mean-squared estimation error criterion, the solution for the optimal array weights is a convex quadratic minimization problem and is analytically tractable. Third, linear filtering preserves the actual time-series of the signal of interest which is important in many applications. Finally, when the received signal consists of the sum of a signal of interest and interfering signals,the spatial correlation of the interfering signals is different from that of the signal of interest, and the signal of interest is correlated across the aperture of the array, the linear beamformer is effective at

filtering out the interfering signals and generating an estimate of th signal of interest.

Another class of array processors which enjoys widespread use it the adaptive array processor. Adaptive array processors use observations of, or information about, the signal, noise, and propagation environments to adjust the characteristics of the processor to minimize or maximize some performance criterion. The processors are able to efficiently use the degrees of freedom available to the processor to adjust to the environment in which the processor is operating. [6] The most widely-used type of adaptive processors are those incorporating adaptive linear beamformers [7, 8]. Two such examples are Capon's MVDR Processor [9] and the Applebaum Beamformer [10]. These processors use observations of the combined signal and noise environment to adaptively adjust the array weight vector to optimally pass the signal of interest through the filter and while controlling the sidelobes of the filter's spatial response to reject the interfering signals contained in the received signal. In order to distinguish between the signal emitted by a source at the location of interest from all other signals, the processor uses a priori estimates of the spatial characteristics of the signals of interest. These a priori estimates depend upon the manner in which the propagating signals are modeled.

Traditionally, array processors have modeled propagating signals as plane waves following a straight line path from the source to the array of sensors. This corresponds to an implicit model of the propagation medium as being homogeneous and infinite in extent and the source being far from the array. The propagation of acoustic waves through the ocean is not modeled accurately in this manner. Both the time-invariant and the time-varying temperature, salinity, and pressure structures of the ocean are spatially-variant. When coupled with the finite extent (principally the finite depth) of the oceans, these spatially-variant structures cause acoustic signals to propagate in a manner which deviates significantly from that predicted by the plane-wave model.

This deviation has both adverse and advantageous consequences. The adverse consequence is that, if a plane-wave model is used by the processor, the spatial characteristics of the signal of interest may not match those estimated by the processor. In this situation, which is referred to as model mismatch, the processor may treat the

signal of interest as an interfering signal and attempt to reject it. A more detailed discussion of this problem is contained in Section 1.3. The advantageous consequence is that, if the processor uses a fairly accurate environmental and propagation model, it is possible to achieve source localization accuracies which far exceed those which are available in an infinte, homogeneous medium [11].

A class of processors which has been developed to take advantage of this improved accuracy and to eliminate the model mismatch problems caused by the use of the plane-wave model is referred to as the Matched Field Processor. First proposed in [12], these processors use fairly complete environmental and propagation models to make a priori estimates of the spatial structure of received signals as a function of environmental condition and source location. The processors use these spatial structure estimates to operate on the received sound field and generate estimates of signal parameters of interest. The spatial structure of the signal of interest is parameterized by the signal replica vector as defined in the following section.

1.2 The Signal Replica Vector

The signal replica vector is a parameterization of the spatial characteristics of a propagating signal as a function of the location of the source of the signal and the propagation characteristics of the medium. Traditionally, the signal replica vector is defined for a narrowband signal propagating through a time-invariant medium. In this case, the signal replica vector is, to within a complex scaling factor, a replica of the deterministic narrowband signal emitted by a source at the location $\underline{z}$ as received at the array of sensors. Thus, given that a source at the location $\underline{z}$ emits the complex exponential $Ae^{j2\pi ft}$ and the medium is time-invariant, the signal received by the array of sensors can be expressed as

$$\underline{x}(t) = B\, e^{j2\pi ft} \underline{q}(f, \underline{z}, \underline{\phi}),$$

where A is a complex random variable, $\underline{q}(f, \underline{z}, \underline{\phi})$ is the signal replica vector, $\underline{\phi}$ is a parameterization of the characteristics of the propagation environment and the receiving array (e.g., the sound speed profile, the depth of the ocean, the sensor locations, etc.), and $B = cA$ for some complex constant c which depends on the signal attenuation and progagation delay between the source and the sensor array and the manner in which the replica vector in normalized. The signal replica vector is usually normalized so that its magnitude equals one.

For this work, the signal replica vector is defined in a stochastic signal framework as

$$\underline{q}(f, \underline{z}, \underline{\phi}) \triangleq \frac{\mathrm{E}[\underline{X}(f, \underline{z}) X_k^*(f, \underline{z}) \mid \underline{\phi}]}{\mathrm{E}[X_k(f, \underline{z}) X_k^*(f, \underline{z}) \mid \underline{\phi}]}, \tag{1.1}$$

where $X_k(f, \underline{z})$ is the discrete-time Fourier transform at the frequency f of the signal emitted by a source at the location $\underline{z}$ **as received** at the k^{th} array sensor, $\underline{X}(f, \underline{z})$ is the discrete-time Fourier transform of the same signal **as received** at the entire array of sensors, and the k^{th} sensor is the reference sensor of the array. Thus, the signal replica vector is the normalized cross-correlation between the discrete-time Fourier transform of the signal of interest as received at the reference sensor and the same signal as received at the entire array of sensors. It is important to note that the signal replica vector is defined in terms of the propagating signal as received at the array of sensors. This new defintion is used for two reasons. First, it explicitly allows the parameterization of the spatial structure of a signal emitted by stochastic source and which propagates through a random medium. Second, this parameterization incorporates all of the the information concerning spatial structure of the signal of interest which can be exploited by a linear processor which is optimized to minimize a mean-squared error criterion.

When the source at the location $\underline{z}$ emits the complex exponential $Ae^{j2\pi ft}$ and the medium is time-invariant as described previously, the signal replica vector as defined in (1.1) is, to within a complex scaling factor, identical to the traditionally defined replica vector also described previously.

The term in the denominator of (1.1) is a normalization term which yields

$q_k(f, \underline{z}, \underline{\phi}) = 1$. Another normalization convention is proposed in Subsection 3.1.3. A different definition of the replica vector is proposed in Subsection 3.2.3. This definition is similar in concept to (1.1) in that it is based upon the spatial cross-correlation of the signal of interest and, in a time-invariant propagation environment, these definitions are roughly equivalent. However, in a random time-varying medium, they are different, and the definition proposed in Subsection 3.2.3 yields better results.

1.3 Array Processor Performance in Uncertain Propagation Environments

As mentioned earlier, array processors exploit the fact that the spatial characteristics of a signal as received at an array of sensors depend on the location of the source of the signal in order to differentiate among signals emitted by sources at different locations. High resolution processors are able to discriminate among signals whose spatial characteristics, parameterized here by the signal replica vector, differ only slightly. While the ability to discriminate among signals whose replica vectors differ only slightly provides good spatial resolution, it also makes the processor sensitive to changes in the propagation characteristics of the environment. A small change in the characteristics of the propagation medium resulting in a small change in the signal replica vector, may cause the processor to inaccurately estimate the location of the source of the signal.

Adaptive processors, such as Capon's MVDR Processor [9], are particularly sensitive to inaccuarate or imprecise knowledge of the characteristics of the propagation environment (referred to as model mismatch). This sensitivity stems from the fact that, if the processor incorrectly calculates the signal replica vector, then the received signal emitted by a source at the location of interest will not be recognized as such. Consequently, the processor will attempt to reject (i.e., filter out) that signal. The sensitivity of various adaptive and non-adaptive processors to model mismatch has been analyzed extensively [13, 14, 15].

Several approaches to reducing the sensitivity of adaptive processors to model

mismatch have been proposed. The most commonly proposed approach, which is applicable to linear processors, is to add an additional constraint to the array weight optimization problem which places an upper bound on the norm of the array weight vector. A survey of these methods is contained in [8]. The motivation for these approaches is that the sensitivity of a processor to spatially uncorrelated perturbations to the nominal spatial characteristics of a signal is proportional to the norm-squared of the array weight vector. A related approach, referred to as the Generalized Cross-Spectral Method [16], is to add a penalty function proportional to the norm-squared of the array weight vector to the criterion, which is minimized or maximized by the selection of the optimal array weight vector.

Another approach to reducing the sensitivity of linear processors to model mismatch is to constrain the response of the linear weight-and-sum beamformer over a range of environmental conditions. One example is the Multiple Constraints Method [17], which accomplishes this goal by placing equality constraints on the response of the beamformer at a number of locations surrounding the location of interest. Relying on the fact that the signal replica vector is a smooth function of both the source location and the environmental conditions, the equality constraints at different locations surrounding the location of interest also constrain the response of the beamformer at the location of interest for various environmental conditions which are close to the nominal environmental condition. A more direct approach using the multiple constraints approach [18] places equality constraints on the response of the beamformer for a number of environmental conditions which result from small perturbations to a nominal environmental condition. A related approach [19] uses inequality constraints on the response of the processor to insure that for various environmental conditions, the actual response is within some tolerance factor of the desired response.

A third approach to reducing the sensitivity of array processors to model mismatch, which is not limited in applicability to linear beamformers, is the random environment approach. here, the environmental parameters are considered random parameters with known probability distributions. The array weight vector or the estimate of the signal parameters are chosen to mininimize or maximize a criterion which

is averaged over the possible values of the environmental parameters. A Maximum A Posteriori source location estimator utilizing this approach is proposed in [20].

There are several drawbacks to these approaches. First, the approaches using linear equality constraints lose one degree of freedom in the beamformer for each constraint. This level of reduction in degrees of freedom may not be necessary to accomplish the desired goal. Second, the selection of the response levels, the norm bounds, and the tolerance factors is, in general, an ad hoc procedure without clearly defined criteria. Finally, the processors developed using the random environment approach may exhibit poor performance for particular sets of environmental conditions even though their average performance is good.

This work seeks to develop an array processor which exhibits the efficient use of degrees of freedom and the interference rejection capability characteristic of adaptive array processors, and the source localization capability characteristic of the matched field processors while operating with only approximate information about the propagation characteristics of the medium. The minmax signal processing approach is proposed to develop such a processor. The minmax approach requires that an error criterion which is a function of the environmental conditions as well as the processor characterists be defined. Using this criterion as a measure of processor performance, the maximum value of the criterion taken over a user-specified range of environmental parameters is minimized. If this is done in an adaptive manner, the processor should be able to efficiently use its degrees of freedom to improve the performance of the processor for the environmental conditions where the performance is most critical.

The use of the minmax approach to develop a processor which is insensitive to modeling uncertainties has been studied previously ([21] and references therein). However, the signal processing techniques developed therein are not applicable to the problem of achieving spatial discrimination in an uncertain propagation environment, and are not adaptive in the sense described in Subsetion 1.1. Therefore, the Adaptive Minmax Matched Field Processor described in Chapter 2 is proposed to achieve the goal of this work.

Chapter 2

Minmax Array Processing

For the reasons stated in Chapter 1, a minmax approach is used here to develop an adaptive array processor which is robust with respect to uncertainties in the propagation environment. Section 2.1 presents a general minmax framework for signal processing along with a characterization theorem for the solutions to a large class of minmax signal processing problems. Using this theorem as a basis, an algorithm for adaptive minmax matched field processing is developed in Section 2.2. Section 2.3 addresses the implementation of the Adaptive Minmax Matched Field Processor. A new algorithm is developed to solve a particular class of quadratic minmax problems which includes the minmax portion of the Adaptive Minmax Matched Field Processor. This algorithm has the desirable property of being guaranteed to converge to an exact solution in a finite number of iterations. Finally, a new approach to the development of minmax estimation error bounds is proposed in Section 2.4.

2.1 The Minmax Signal Processing Framework

The framework in which the array processing algorithm described in Section 2.2 is developed is minmax signal processing. In general terms the framework addresses the problem of developing a processor whose worst-case performance evaluated over a given class and range of uncertainties is as favorable as possible. Specifically, let $g(\underline{y}, \underline{w})$ be a processor parameterized by the vector $\underline{w}$ which operates on an observed

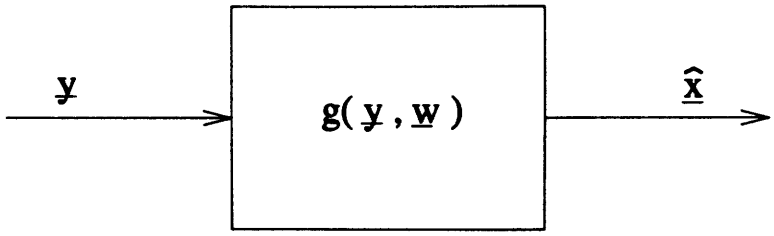


Figure 2-1: The Minmax Signal Processor

signal $\underline{y}$ to generate an estimate of some signal or parameter of interest $\hat{\underline{x}}$ (Figure 2-1). The $\hat{(\cdot)}$ symbol denotes an estimate of the variable over which it is positioned (e.g., $\hat{\underline{x}}$ denotes an estimate of the vector $\underline{x}$.). The set of allowable values for the parameter vector $\underline{w}$ is denoted by W. In the case where $g(\underline{y}, \underline{w})$ is a linear filter with N taps, the vector $\underline{w}$ could contain the filter weights and W could be the space of N-dimensional complex numbers $\mathcal{C}^N$.

The parameters which govern the relationship between the observed signal $\underline{y}$ and the signal or parameter of interest $\underline{x}$ are referred to as the environmental parameters and denoted by the vector $\underline{\phi}$. In the context of array processing problems where $\underline{y}$ is the received signal and $\underline{x}$ is a particular signal of interest, the vector $\underline{\phi}$ could contain the location of the array sensors or the phase, gain, and directional characteristics of those sensors. It could also contain a parameterization of the interfering signals or the characteristics of the propagation medium. The ability of any particular processor as determined by the choice of $\underline{w}$ to estimate $\underline{x}$ depends upon the particular environmental condition under which the processor operates. Thus, a particular value of $\underline{w}$ which yields good processor performance under one environmental condition may yield very poor performance under another environmental condition. A real valued error function $\varepsilon(\underline{w}, \underline{\phi})$ is used as a figure of merit to evaluate the performance of any particular processor operating under any particular environmental condition.

If the processor has perfect knowledge of the environmental conditions (e.g., $\underline{\phi} = \underline{\phi}_o$), then the processor parameters can be chosen to minimize $\varepsilon(\underline{w}, \underline{\phi}_o)$. However, in many situations the processor does not have perfect knowledge of the environmental conditions under which it must operate, but instead knows only that the uncertain environmental parameter $\underline{\phi}$ falls within some range denoted by the set Φ. The pro-

cessor should then be designed to operate over this entire range.

As discussed in Section 1.3, one possible approach to designing the processor to operate over Φ is to treat $\underline{\phi}$ as a random parameter with an assigned pdf (probability distribution function) $p_{\underline{\phi}}$ and then select $\underline{w}$ to minimize the average value of $\varepsilon(\underline{w}, \underline{\phi})$ taken over Φ with respect to $p_{\underline{\phi}}$. That is,

$$\underline{w}_{opt} = \arg \min_{\underline{w} \in W} \int_{\Phi} p_{\underline{\phi}}(\underline{\phi}_o)\ \varepsilon(\underline{w}, \underline{\phi}_o)\ d\underline{\phi}_o.$$

However, this approach requires that a pdf be explicitly assigned to $\underline{\phi}$ and does not necessarily solve the problem of the processor performance being very poor for particular environmental conditions under which it may have to operate.

The minmax signal processing framework makes it possible to avoid these problems when selecting $\underline{w}$ by treating $\underline{\phi}$ as a *nonrandom* parameter. Then, under the assumptions that $\varepsilon(\underline{w}, \underline{\phi})$ is a continuous function of $\underline{\phi}$ for every $\underline{w} \in W$ and Φ is a compact set contained in a metric space, the worst-case performance of the processor over the range of the environmental parameters is defined as

$$\Delta(\underline{w}) \triangleq \max_{\underline{\phi} \in \Phi} \varepsilon(\underline{w}, \underline{\phi}).$$

$\Delta(\underline{w})$ is referred to as the extremal value for the processor parameter vector $\underline{w}$. The optimal minmax processor parameter vector is defined as that which minimizes this extremal value. Mathematically, this is stated as

$$\underline{w}_{opt} \triangleq \arg \min_{\underline{w} \in W} \Delta(\underline{w}) = \arg \min_{\underline{w} \in W} \max_{\underline{\phi} \in \Phi} \varepsilon(\underline{w}, \underline{\phi}).$$

As in any optimization problem, the specification of the necessary conditions which must be met by the optimal solution and the sufficient conditions which guarantee that a solution is optimal are of central importance. The specification of such conditions for minmax optimization problems requires the definition of extremal points, extremal point sets, the convex hull of a set of points, and the gradient operator. An extremal point is any environmental point $\underline{\phi}$ contained in Φ at which the error function $\varepsilon(\underline{w}, \underline{\phi})$

achieves the extremal value $\Delta(\underline{w})$. The extremal point set, denoted by $M(\underline{w})$, is the set of all extremal points. That is,

$$M(\underline{w}) \triangleq \{\underline{\phi} \in \Phi \mid \varepsilon(\underline{w}, \underline{\phi}) = \Delta(\underline{w})\}.$$

Given any set of points A contained in a metric space $\mathcal{S}$, the convex hull of the set A in $\mathcal{S}$, denoted by $\mathcal{H}(A)$, consists of all points $\underline{s} \in \mathcal{S}$ which can be expressed as the convex combination of the points $\underline{a}_i \in A$. That is,

$$\begin{aligned} \mathcal{H}(A) \;=\; & \{\underline{s} \in \mathcal{S} \mid \exists J > 0,\ \underline{a}_i \in A, \text{ and } p_i \subset \mathbb{R}, \text{ for } i = 1, \ldots, J \\ & \text{s.t. } p_i > 0 \ \ i = 1, \ldots, J\,, \ \ \sum_{i=1}^{J} p_i = 1 \ \text{ and } \ \underline{s} = \sum_{i=1}^{J} p_i \underline{a}_i\}. \end{aligned}$$

A final required definition is that of the gradient operator. Let $\varepsilon(\underline{w}, \underline{\phi})$ be any real valued scalar function of the vectors $\underline{w}$ and $\underline{\phi}$. Then the gradient operator of ε with respect to $\underline{w}$ is any vector function of $\underline{w}$ and $\underline{\phi}$, denoted by $\nabla_{\underline{w}} \varepsilon \left(\underline{w}, \underline{\phi}\right)$, which is continuous with respect to $\underline{w}$ and $\underline{\phi}$ and for which the following is true: There exists a real, positive scalar constant k such that for any particular processor parameter vector $(\underline{w}_o)$ and environmental condition $(\underline{\phi}_o)$, the incremental change in ε corresponding to an incremental change in $\underline{w}$ away from $\underline{w}_o$, denoted by $\delta \underline{w}$, is equal to $k < \nabla_{\underline{w}} \varepsilon \left(\underline{w}_o, \underline{\phi}_o\right), \delta \underline{w} >$, where $<\ ,\ >$ denotes the inner product. A formal statement of this definition is contained in Appendix A. The definition of the gradient as a vector of partial derivatives is not used because the error function used later in this chapter is not differentiable with respect to the elements of the complex vector $\underline{w}$.

Given the preceding definitions, the following Minmax Characterization Theorem, which is a generalization of that given in Chapter 6 of [1] for the case of minmax approximation with differentiable functions, states the conditions which characterize the optimal solution to a general class of minmax problems.

Theorem 1 Let Φ be a compact set contained in a metric space denoted by Γ, W be an open set of a Euclidian metric space denoted by E, $\varepsilon : W \times \Phi \rightarrow \mathbb{R}$ be a continuous function on both W and Φ for which, at each $\underline{w} \in W$, a directional derivative with respect to $\underline{w}$ can be defined on Φ,

and $\nabla_{\underline{w}}\varepsilon\left(\underline{w}, \underline{\phi}\right)$ be the gradient of ε with respect to $\underline{w}$. Then a necessary condition for $\underline{w}_o \in W$ to be a solution to the following minmax problem

$$\underline{w}_{opt} = \arg \min_{\underline{w} \in W} \max_{\underline{\phi} \in \Phi} \varepsilon(\underline{w}, \underline{\phi}),$$

is that

$$\underline{0} \in \mathcal{H}\left(\{\nabla_{\underline{w}}\varepsilon\left(\underline{w}_o, \underline{\phi}\right) \mid \underline{\phi} \in M(\underline{w}_o)\}\right).$$

If, in addition, ε is a convex function of $\underline{w}$ and W is a convex set, this condition is a necessary and sufficient condition for $\underline{w}_o \in W$ to be the solution to the stated minmax problem.

A proof of this theorem is contained in Appendix A. This theorem states that a necessary (and sufficient if ε is convex on W and W is itself convex) condition for the optimality of $\underline{w}_o$ is that the origin, denoted by $\underline{0}$, is contained in the convex hull of the set of gradients of ε with respect to $\underline{w}$ evaluated at the extremal points of $\varepsilon(\underline{w}_o, \underline{\phi})$.

The following example may be useful to clarify the definitions and the concepts introduced thus far and to provide an intuitive interpretation of the Minmax Characterization Theorem.

Example: Assume that $\underline{w}$ is a two-dimensional real vector, $W = \mathbb{R}^2$, ϕ is a real scalar variable, Φ is the closed interval between zero and one (i.e. $\Phi = [0, 1]$), and $\varepsilon(\underline{w}, \phi)$ is a real-valued scalar function which is convex with respect to $\underline{w}$ for all ϕ and continuous with respect to ϕ.

For some $\underline{w}_1$, Figure 2-2a shows the error function plotted as a function of ϕ with the extremal value $\Delta(\underline{w}_1)$ and the extremal point set $M(\underline{w}_1) = \{\phi_1, \phi_2, \phi_3\}$ labeled. The gradients of ε with respect to $\underline{w}$ evaluated at $\underline{w}_1$ and each of the three extremal points are shown in Figure 2-2b. The convex hull of this set of vectors is the shaded region. The origin is not in this convex hull and therefore $\underline{w}_1$ is not an optimal solution. This can be seen by noting that, if we can choose a direction vector such as $\underline{d}$ for which $\underline{d}^t \nabla_{\underline{w}}\varepsilon\left(\underline{w}_1, \phi_i\right) < 0$ for $i = 1, 2, 3$, then the initial change in ε evaluated at each of the extremal points will be negative as we move away from $\underline{w}_1$ in the direction of $\underline{d}$. Since the value of ε is simultaneously reduced at each of the extremal points as we move away from $\underline{w}_1$, the extremal value of ε will be reduced as we move away from $\underline{w}_1$ and therefore $\underline{w}_1$ cannot be an optimal solution.

For some $\underline{w}_2$, Figure 2-3a shows the error function plotted as a function of ϕ with the extremal value $\Delta(\underline{w}_2)$ and the extremal point set $M(\underline{w}_2) = \{\phi_4, \phi_5, \phi_6\}$ labeled. The gradients of ε with respect to $\underline{w}$ evaluated at $\underline{w}_2$ and each of the three extremal points are shown in Figure 2-3b with the convex hull denoted as before. In this case, the origin is in the

convex hull and therefore $\underline{w}_2$ is an optimal solution. This can be seen by noting that, for any direction vector such as $\underline{d}$ which we choose, the inner product $\underline{d}^t \nabla_{\underline{w}} \varepsilon(\underline{w}_2, \phi_i)$ will be greater than zero for at least one of the three gradient vectors (in the case shown, $\underline{d}^t \nabla_{\underline{w}} \varepsilon(\underline{w}_2, \phi_6) > 0$). Therefore, as we move away from $\underline{w}_2$ in any direction $\underline{d}$, the initial change in ε evaluated at one or more of the extremal points will be positive. Since the value of ε increases at one or more of the extremal points as we move away from $\underline{w}_2$, the extremal value of ε will be increased as we move away from $\underline{w}_2$. Therefore, $\underline{w}_2$ is a locally optimal solution. However, since ε is a convex function of $\underline{w}$ for all ϕ, $\Delta(\underline{w})$ is also a convex funtion of $\underline{w}$. (See Lemma A.2 in Appendix A). Therefore, $\underline{w}_2$ is a globally optimal solution.

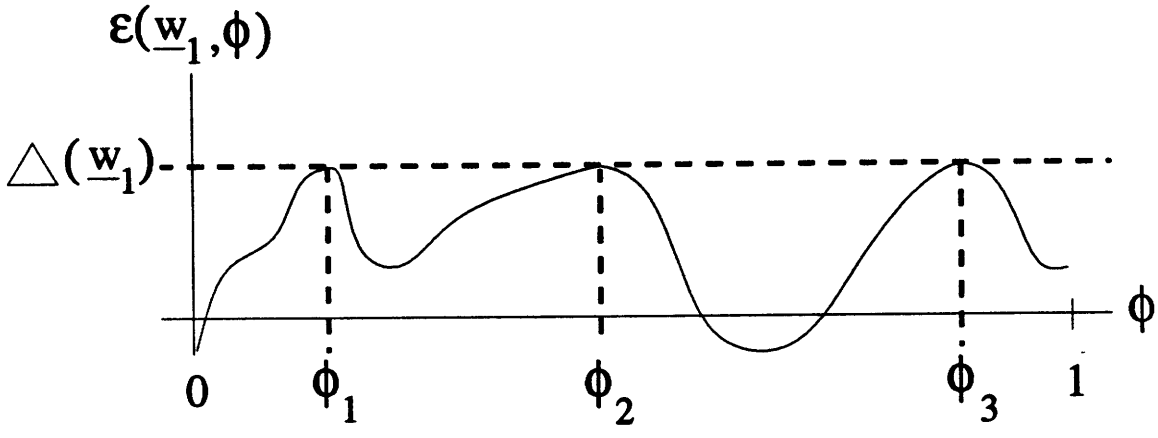
2.2 The Adaptive Minmax Matched Field Processor

2.2.1 Signal Model

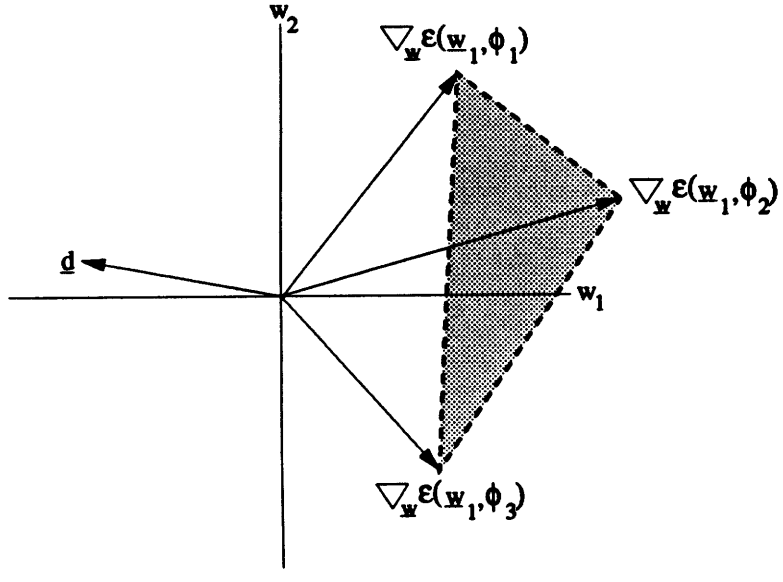
The Adaptive Minmax Matched Field Processor takes as its input the signal received by an array of sensors which has been low-pass filtered to prevent frequency domain aliasing and then sampled. This input is denoted by the vector time series $\underline{y}[m]$. This input signal is assumed to be the sum of propagating background noise generated by spatial spread sources, such as, breaking surface waves, sensor noise which is assumed to be spatially white, and propagating signals generated by spatially-discrete point sources such as marine mammals, ships, etc.. $\underline{x}[m, \underline{z}]$ denotes the time sampled received signal which was emitted by a point source at the spatial location $\underline{z}$. $\underline{n}[m]$ denotes the sum of the sensor noise and the received propagating background noise. It is assumed that $\underline{n}[m]$ and $\underline{x}[m, \underline{z}]$ are uncorrelated zero mean wide-sense stationary random processes for each $\underline{z}$ and that $\underline{x}[m, \underline{z}_1]$ and $\underline{x}[m, \underline{z}_2]$ are uncorrelated for any two source locations $\underline{z}_1 \neq \underline{z}_2$. Thus, $\underline{y}[m]$ is a zero mean random process represented by

$$\underline{y}[m] = \underline{n}[m] + \sum_{\underline{z}} \underline{x}[m, \underline{z}].$$

The modeling of $\underline{x}[m, \underline{z}]$ as a zero mean random process can include a signal emitted by a stochastic source propagating through either a deterministic or a random environment, or a signal emitted by a deterministic source propagating through a

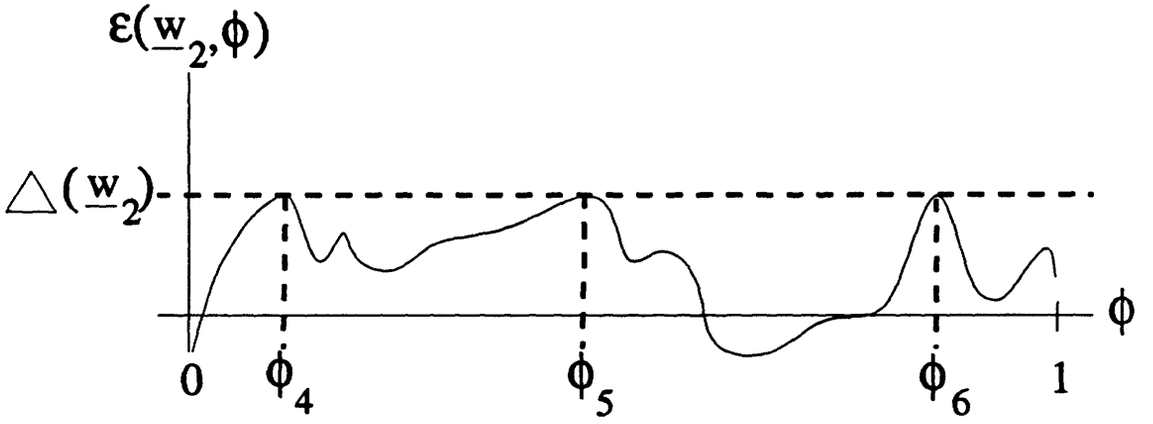


(a) Error Function vs. ϕ

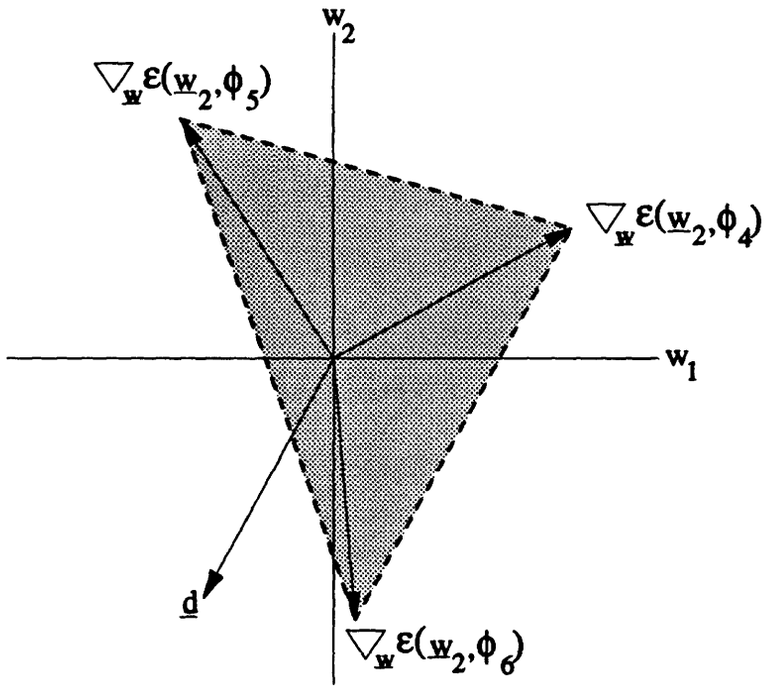


(b) Extremal Point Gradients and Convex Hull

Figure 2-2: Non-Optimal Solution: $\underline{w} = \underline{w}_1$



(a) Error Function vs. ϕ



(b) Extremal Point Gradients and Convex Hull

Figure 2-3: Optimal Solution: $\underline{w} = \underline{w}_2$

random environment.

2.2.2 Processor Structure

The Adaptive Minmax Matched Field Processor was developed to achieve the effective sidelobe control which is characteristic of adaptive processors such as Capon's MVDR Processor [9], and the improvement in spatial resolution provided by matched field processing techniques such as those presented in [11], without exhibiting the extreme sensitivity to mismatch in the estimation of the characteristics of the propagation environment which is exhibited by these algorithms and techniques [13, 17]. The quantity estimated by the processor is the average power in a selected frequency component of the signal emitted by a point source at a location of interest as received at one array sensor (the reference sensor). The location of interest is referred to as the array focal point. The signal emitted by a point source at the array focal point and received at the reference sensor is referred to as the desired signal and denoted by $x_k[m, \underline{z}]$. Here the k^{th} sensor is the reference sensor and $\underline{z}$ is the array focal point. Unlike the case of traditional array processors, the desired signal **is not** the signal as emitted by a source at the array focal point. Instead, the desired signal is the signal emitted by a source at the array focal point **as received** at the reference sensor. The array focal point can be swept through space and the selected frequency can be swept through the frequency spectrum to generate an estimate of the average power in the desired signal as a function of spatial location and temporal frequency. This estimate is denoted by $\hat{\sigma^2}(f, \underline{z})$.

Conceptually, the processor which generates this estimate consists of three modules (Figure 2-4). The first module divides the time-sampled signal received by the array $\underline{y}[m]$ into segments M samples in length which may be overlapping, and computes the vector discrete-time Fourier transform of each segment at the selected frequency

$$\underline{Y}^l(f) = \sum_{m=0}^{M-1} \underline{y}^l[m] e^{-j2\pi f m \Delta_t}.$$

l indicates the segment number, $\underline{y}^l[m]$ is the m^{th} sample of the l^{th} segment, and Δ_t

is the sampling period. Here, f is the frequency expressed in cycles per second which satisfies $| f | \leq \frac{1}{2\Delta_t}$. f is **not** the normalized frequency expressed in cycles per sample which satisfies $| f | \leq \frac{1}{2}$. The linearity of the Fourier transform yields

$$\underline{Y}^l(f) = \sum_{\underline{z}} \underline{X}^l(f, \underline{z}) + \underline{N}^l(f),$$

where the summation is over the locations of the point sources. The transformed segments are known as "snapshots" and $\underline{X}^l(f, \underline{z})$ denotes the snapshot of the l^{th} segment of $\underline{x}[m, \underline{z}]$. $Y_i^l(f)$ denotes the discrete-time Fourier transform of the l^{th} segment of the signal received by the i^{th} array sensor. In effect, the first module is a temporal filter which selects the frequency component of interest in the received signal.

The snapshots of the received signal are the inputs to the second module which is a linear weight-and-sum beamformer. This beamformer computes an estimate of the Fourier transform of the l^{th} segment of the desired signal using

$$\hat{X}_k^l(f, \underline{z}) = \underline{w}^h \underline{Y}^l(f),$$

where $\underline{w}$ is the array weight vector. The beamformer is a spatial filter which attempts to pass only the desired signal (i.e., that which was emitted by a source at the array focal point $\underline{z}$) while rejecting all other signals received by the array sensors. The final module computes an estimate of the average power in the desired signal. The overbar indicates the sample mean taken over all l. That is, if L is the number of segments used in estimating $\sigma^2(f, \underline{z})$, then

$$\hat{\sigma}^2(f, \underline{z}) = \frac{1}{L} \sum_{l=1}^{L} | \hat{X}_k^l(f, \underline{z}) |^2 .$$

While this structure is the same as that used by many array processors such as Capon's MVDR Processor, the unique feature of this processor is the manner in which the array weight vector $\underline{w}$ is calculated. For this processor, the array weight vector is the solution to a minmax optimization problem where the error ε is a measure of the spatial filter's ability to pass the desired signal without distortion while rejecting the

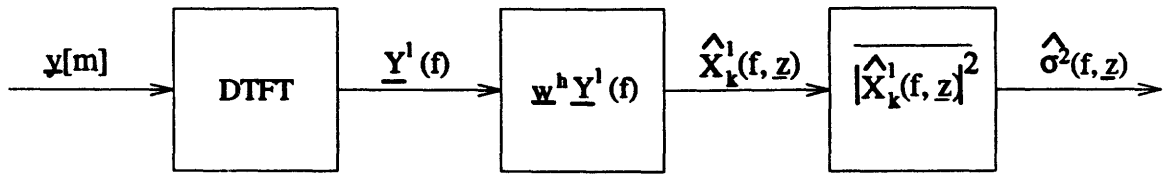


Figure 2-4: Array Processor Structure

interfering signals in a given propagation environment.

2.2.3 The Minmax Array Weight Problem

For any particular array focal point, frequency, array weight vector, and propagation environment, the error function for the Adaptive Minmax Matched Field Processor is the a priori mean-squared error in the estimation of $X_k(f,\underline{z})$ conditioned on the characteristics of the propagation environment. That is

$$\begin{aligned}\varepsilon(f,\underline{z},\underline{w},\underline{\phi}) &= \mathrm{E}[\mid X_k(f,\underline{z}) - \hat{X}_k(f,\underline{z}) \mid^2 \mid \underline{\phi}] \\ &= \mathrm{E}[\mid X_k(f,\underline{z}) - \underline{w}^h\, \underline{Y}(f) \mid^2 \mid \underline{\phi}],\end{aligned} \tag{2.1}$$

where the characteristics of the propagation environment are parameterized by the vector $\underline{\phi}$.

For a given array focal point and frequency, the optimal array weights are defined as

$$\underline{w}_{opt}(f,\underline{z}) = \arg \min_{\underline{w}\in\mathcal{C}^N} \max_{\underline{\phi}\in\Phi} \varepsilon(f,\underline{z},\underline{w},\underline{\phi}), \tag{2.2}$$

where N is the number of array sensors and Φ is the user specified range of the environmental parameters over which the processor must operate.

Under the assumption stated earlier that the desired signal and the interfering signals are uncorrelated, (2.1) can be rewritten as

$$\begin{aligned}\varepsilon(f,\underline{z},\underline{w},\underline{\phi}) &= \mathrm{E}[X_k(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}] - \\ 2\,\mathrm{Real}(\mathrm{E}[\underline{X}(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}]^h\, \underline{w}) &+ \underline{w}^h\, \mathrm{E}[\underline{Y}(f)\underline{Y}(f)^h \mid \underline{\phi}]\, \underline{w},\end{aligned} \tag{2.3}$$

where the superscript $*$ denotes complex conjugate.

The expectation in the last term of (2.3) is the cross-spectral correlation matrix of the received signal conditioned on the environmental parameter $\underline{\phi}$. The cross-spectral correlation matrix is the parameterization used by the processor to characterize the spatial structure of the total signal field, and it is the input to the processor which enables the processor to adapt to reject unwanted signals. Here, the matrix will not be treated as a function of the particular environmental conditions or the characteristics of any particular propagating signal. Instead it wil be treated as a property of the total signal field. Therefore, the conditioning of the expectation in the last term of (2.3) is dropped and the actual ensemble cross-spectral correlation matrix, $\mathrm{S}(f)$, is used. In most cases, this ensemble cross-spectral correlation matrix is unknown to the processor. Therefore, the sample cross-spectral correlation matrix given by $\hat{\mathrm{S}}(f) \triangleq \frac{1}{L}\sum_{l=1}^{L} \underline{Y}^l(f)\underline{Y}^l(f)^h$ will be substituted for $\mathrm{S}(f)$. Nothing in the derivation of the algorithm in the remainder of this chapter depends upon this substitution.

The expectation in the second term of (2.3) can be expressed as

$$\mathrm{E}[X_k(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}]\frac{\mathrm{E}[\underline{X}(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}]}{\mathrm{E}[X_k(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}]}.$$

The quotient is the signal replica vector defined in Section 1.2 as

$$\underline{q}(f,\underline{z},\underline{\phi}) \triangleq \frac{\mathrm{E}[\underline{X}(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}]}{\mathrm{E}[X_k(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}]}.$$

Therefore, the second term can be expressed as

$$2\,\mathrm{E}[X_k(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}]\mathrm{Real}(\underline{q}^h(f,\underline{z},\underline{\phi})\,\underline{w}) \tag{2.4}$$

The signal replica vector in this factorization is the means by which the a priori model of the dependence of the desired signal's spatial characteristics on the environmental conditions is incorporated into the processor.

The expression $\mathrm{E}[X_k(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}]$ appears in the first term of (2.3) and, as a result of the factorization in (2.4), will also appear in second term of (2.3). This

expression is the conditional average power in the desired signal, and will be replaced by the actual average power in the desired signal $\sigma^2(f,\underline{z}) \triangleq \mathrm{E}[X_k(f,\underline{z})X_k^*(f,\underline{z})]$. Given the factorization and the substitutions detailed above, the error criterion can be expressed as

$$\varepsilon(f,\underline{z},\underline{w},\underline{\phi},\sigma^2(f,\underline{z})) = \sigma^2(f,\underline{z}) - 2\sigma^2(f,\underline{z})\,\mathrm{Real}(\underline{q}^h(f,\underline{z},\underline{\phi})\,\underline{w}) + \underline{w}^h\,\hat{\mathrm{S}}(f)\,\underline{w}, \quad (2.5)$$

where the dependence of the error on the average power in the desired signal is explicitly shown.

The optimal array weights minimize the maximum value of this error taken over the operating range of the environmental parameters. Conceptually, they can be considered those of a data-adaptive Wiener filter which is robust with respect to changes in the spatial correlation of the signal to be estimated.

The Adaptive Minmax Processor described in this subsection can also be interpreted as an efficient implementation of a bank of MVDR Matched Field Processors, each using a different assumed value of the environmental parameter vector ($\underline{\phi}$) and therefore of the signal replica vector ($\underline{q}(f,\underline{z},\underline{\phi})$) (Figure 2-5). The range of assumed values of $\underline{\phi}$ is the range of environmental conditions over which the processor is designed to operate. The Adaptive Minmax Processor output is the output of the MVDR Processor with the largest estimated average power. The derivation of this interpretation is detailed in Section 3.1 where the processor bank interpretation is equivalent to the Two-Stage MVDR Matched Field Processor interpretation.

2.2.4 The Minmax Array Processing Algorithm

A problem in calculating the solution to the minmax problem in (2.2) is that the error criterion and therefore the optimal array weights are functions of $\sigma^2(f,\underline{z})$. However, the array processor does not have knowledge of the true value of $\sigma^2(f,\underline{z})$, but estimates it to be the sample average power in the output of the weight-and-sum beamformer.

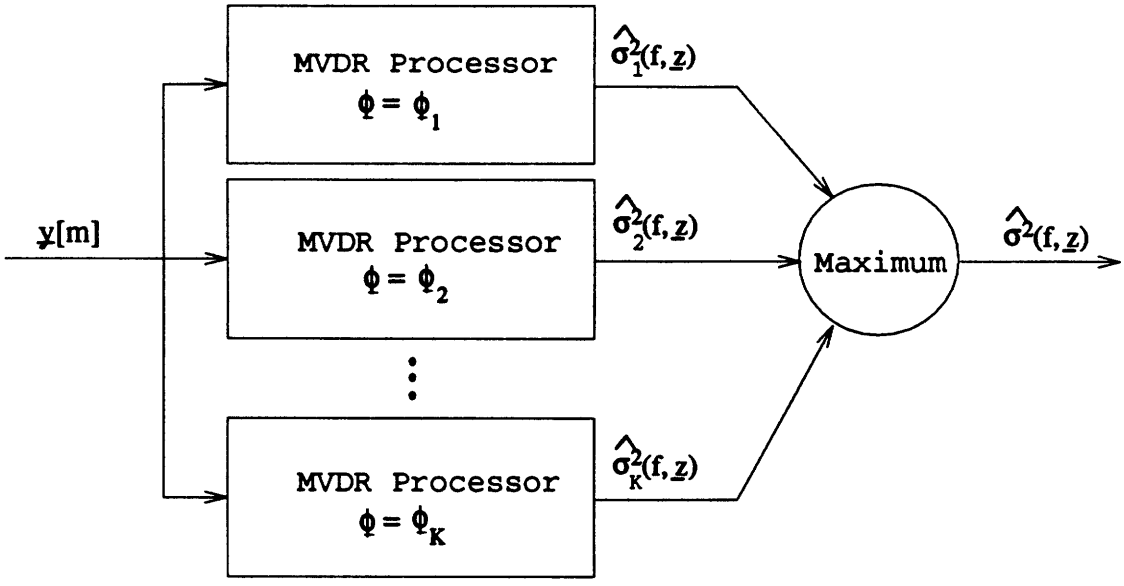


Figure 2-5: The MVDR Processor Bank

That is,

$$\hat{\sigma}^2(f,\underline{z}) = \overline{|\hat{X}_k^l(f,\underline{z})|^2} = \frac{1}{L}\sum_{l=1}^{L}\underline{w}^h\underline{Y}^l(f)\underline{Y}^l(f)^h\underline{w} = \underline{w}^h\hat{\mathrm{S}}(f)\underline{w}.$$

Therefore, the error criterion and optimal array weights depend upon the average power in the frequency component of interest in the desired signal and the estimate of this average power depends upon the array weights used by the beamformer. This interdependence makes it necessary to jointly calculate the optimal array weights and estimate the average power.

This joint calculation and estimation problem is addressed by requiring that the average power in the desired signal used when calculating the optimal array weights weights be equal to the estimated average power in the desired signal resulting from the use of those weights. The joint array weights calculation/power estimation problem can be posed as finding $\underline{w}_{opt}(f,\underline{z},\hat{\sigma}^2(f,\underline{z}))$ and $\hat{\sigma}^2(f,\underline{z})$ so that

$$\underline{w}_{opt}(f,\underline{z},\hat{\sigma}^2(f,\underline{z})) = \arg\min_{\underline{w}\in\mathcal{C}^N}\max_{\underline{\phi}\in\Phi}\epsilon(f,\underline{z},\underline{w},\underline{\phi},\hat{\sigma}^2(f,\underline{z})), \tag{2.6}$$

and

$$\hat{\sigma}^2(f,\underline{z}) = \underline{w}_{opt}^h(f,\underline{z},\hat{\sigma}^2(f,\underline{z}))\ \hat{\mathrm{S}}(f)\ \underline{w}_{opt}(f,\underline{z},\hat{\sigma}^2(f,\underline{z})), \tag{2.7}$$

where the dependence of the optimal array weights on the average power is explicitly shown.

A trivial solution to the problem expressed in (2.6) and (2.7) is $\underline{w}_{opt}(f,\underline{z},0) = \underline{0}$ and $\hat{\sigma}^2(f,\underline{z}) = 0$. The existence of a non-trivial solution and an algorithm for jointly finding the nonzero $\underline{w}_{opt}(f,\underline{z},\hat{\sigma}^2(f,\underline{z}))$ and $\hat{\sigma}^2(f,\underline{z})$ which satisfy (2.6) and (2.7) is based upon the following theorem, a proof of which is given in Appendix B.

Theorem 2 Let σ_o^2 be any real positive number, Φ be a compact set contained in a metric space, $\underline{q}(f,\underline{z},\underline{\phi})$ be a continuous function on Φ, and

$$\underline{w}_{opt}(f,\underline{z},\sigma_o^2) = \arg \min_{\underline{w}\in\mathcal{C}^N} \max_{\underline{\phi}\in\Phi} \varepsilon(f,\underline{z},\underline{w},\underline{\phi},\sigma_o^2).$$

Then for any real non-negative σ^2, the solution to the problem

$$\underline{w}_{opt}(f,\underline{z},\sigma^2) = \arg \min_{\underline{w}\in\mathcal{C}^N} \max_{\underline{\phi}\in\Phi} \varepsilon(f,\underline{z},\underline{w},\underline{\phi},\sigma^2)$$

is given by

$$\underline{w}_{opt}(f,\underline{z},\sigma^2) = (\sigma^2/\sigma_o^2)\ \underline{w}_{opt}(f,\underline{z},\sigma_o^2).$$

Therefore, given $\underline{w}_{opt}(f,\underline{z},\sigma_o^2)$ for any real positive σ_o^2, the solution to (2.6) can be expressed as

$$\underline{w}_{opt}(f,\underline{z},\hat{\sigma}^2(f,\underline{z})) = (\hat{\sigma}^2(f,\underline{z})/\sigma_o^2)\ \underline{w}_{opt}(f,\underline{z},\sigma_o^2). \tag{2.8}$$

Subsituting (2.8) into (2.7) yields

$$\hat{\sigma}^2(f,\underline{z}) = (\hat{\sigma}^2(f,\underline{z})/\sigma_o^2)^2\ (\underline{w}_{opt}^h(f,\underline{z},\sigma_o^2)\ \hat{\mathrm{S}}(f)\ \underline{w}_{opt}(f,\underline{z},\sigma_o^2)). \tag{2.9}$$

Solving (2.9) for $\hat{\sigma}^2(f,\underline{z})$ yields

$$\hat{\sigma}^2(f,\underline{z}) = (\sigma_o^2)^2\ (\underline{w}_{opt}^h(f,\underline{z},\sigma_o^2)\ \hat{\mathrm{S}}(f)\ \underline{w}_{opt}(f,\underline{z},\sigma_o^2))^{-1}. \tag{2.10}$$

The optimal array weights which are consistent with this average power estimate can be calculated using (2.8). Therefore, the following algorithm can be used to solve the

joint optimal array weights calculation and average power estimation problem defined by (2.6) and (2.7).

1. Assign any real, positive value to σ_o^2. Calculate $\underline{w}_{opt}(f, \underline{z}, \sigma_o^2)$ as given by

$$\underline{w}_{opt}(f, \underline{z}, \sigma_o^2) = \arg \min_{\underline{w} \in C^N} \max_{\underline{\phi} \in \Phi} \epsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2).$$

2. $\hat{\sigma}^2(f, \underline{z}) = (\sigma_o^2)^2 \, (\underline{w}_{opt}^h(f, \underline{z}, \sigma_o^2) \, \hat{\mathrm{S}}(f) \, \underline{w}_{opt}(f, \underline{z}, \sigma_o^2))^{-1}$

3. $\underline{w}_{opt}(f, \underline{z}, \hat{\sigma}^2(f, \underline{z})) = (\hat{\sigma}^2(f, \underline{z})/\sigma_o^2) \, \underline{w}_{opt}(f, \underline{z}, \sigma_o^2)$

Step 1 can be implemented using any complex minmax approximation algorithm capable of handling quadratic forms. The development of an efficient algorithm to solve this particular minmax problem is detailed in Section 2.3.

2.3 Solution of the Minmax Problem

Step 1 of the array processing algorithm developed in Section 2.2 requires the solution of a quadratic minmax problem. A major impediment to the implementation of minmax signal processing algorithms has been their relatively high computational complexity. The minmax signal processing solutions which have gained widespread use are those for which either analytic solutions are available (e.g., the Dolph-Chebyshev window [22]) or those for which computationally efficient algorithms have been developed (e.g., real linear-phase minmax filter design using the Parks-McClellan algorithm [23]). By exploiting the special structure of the quadratic minmax problem contained in Step 1 of the array processing algorithm, the algorithm developed in this section to solve the minmax problem is relatively efficient computationally and is guaranteed to converge in a finite number of iterations.

2.3.1 Characterization of $\underline{w}_{opt}(f, \underline{z}, \sigma_o^2)$

From Step 1 of the algorithm in Section 2.2, the minmax problem which must be solved is

$$\underline{w}_{opt}(f, \underline{z}, \sigma_o^2) = \arg \min_{\underline{w} \in \mathcal{C}^N} \max_{\underline{\phi} \in \Phi} \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2),$$

where $\varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2)$ is the conditional mean-squared estimation error and can be expressed as

$$\varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2) = \sigma_o^2 \; - \; 2\sigma_o^2 \text{Real}(\underline{q}^h(f, \underline{z}, \underline{\phi}) \; \underline{w}) + \underline{w}^h \; \hat{\text{S}}(f) \; \underline{w}. \tag{2.11}$$

The following characterization theorem for the minmax array weight problem states the necessary and sufficient conditions satisfied by $\underline{w}_{opt}(f, \underline{z}, \sigma_o^2)$. A proof of this theorem is contained in Appendix B.

Theorem 3 Let Φ be a compact set contained in a metric space and $\underline{q}(f, \underline{z}, \underline{\phi})$ be a continuous function on Φ. Then a sufficient condition for $\underline{w}_o$ to be a solution to the following minmax problem

$$\underline{w}_{opt}(f, \underline{z}, \sigma_o^2) = \arg \min_{\underline{w} \in \mathcal{C}^N} \max_{\underline{\phi} \in \Phi} \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2),$$

is that

$$\exists J \; > \; 0 \, , \tag{2.12}$$

and

$$\exists \tilde{M}(\underline{w}_o) = \{\underline{\phi}_1, \ldots, \underline{\phi}_J\} \subseteq M(\underline{w}_o), \tag{2.13}$$

such that

$$\underline{0} \in \mathcal{H}\left(\{(\hat{\text{S}}(f)\underline{w}_o \; - \; \sigma_o^2 \underline{q}(f, \underline{z}, \underline{\phi})) \; | \; \underline{\phi} \in \tilde{M}(\underline{w}_o)\}\right). \tag{2.14}$$

A necessary condition for $\underline{w}_o$ to be a solution to the following minmax problem

$$\underline{w}_{opt}(f, \underline{z}, \sigma_o^2) = \arg \min_{\underline{w} \in \mathcal{C}^N} \max_{\underline{\phi} \in \Phi} \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2),$$

is that

$$\exists J \in \{1, \ldots, 2N + 1\} \tag{2.15}$$

for which (2.13) and (2.14) are satisfied.

(2.14) is equivalent to

$$\underline{0} = \sum_{i=1}^{J} p_i \, (\hat{\mathrm{S}}(f)\underline{w}_o \; - \; \sigma_o^2 \underline{q}(f, \underline{z}, \underline{\phi}_i)), \tag{2.16}$$

where $p_i \geq 0$ and $\sum_{i=1}^{J} p_i \; = \; 1$. Algebraic manipulation of (2.16) yields the following expression for $\underline{w}_o$.

$$\underline{w}_o \; = \; \sigma_o^2 \, \hat{\mathrm{S}}(f)^{-1} \sum_{i=1}^{J} p_i \, \underline{q}(f, \underline{z}, \underline{\phi}_i)$$

Therefore, the following corollary to Theorem 3 states an equivalent set of necessary and sufficient conditions satisfied by $\underline{w}_{opt}(f, \underline{z}, \sigma_o^2)$.

Corollary 1 Let Φ be a compact set contained in a metric space and $\underline{q}(f, \underline{z}, \underline{\phi})$ be a continuous function on Φ. Then a sufficient condition for $\underline{w}_o \in \mathcal{C}^N$ to be a solution to the following minmax problem

$$\underline{w}_{opt}(f, \underline{z}, \sigma_o^2) = \arg \min_{\underline{w} \in \mathcal{C}^N} \max_{\underline{\phi} \in \Phi} \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2),$$

is that

$$\exists J \; > \; 0, \tag{2.17}$$

$$\exists \tilde{M}(\underline{w}_o) = \{\underline{\phi}_1, \ldots, \underline{\phi}_J\} \subseteq M(\underline{w}_o), \tag{2.18}$$

and

$$\exists p_1, \ldots, p_J \in \mathbb{R} \, , \;\; p_1, \ldots, p_J \geq 0 \, , \;\; \sum_{i=1}^{J} p_i \; = \; 1 \, , \tag{2.19}$$

such that

$$\underline{w}_o \; = \; \sigma_o^2 \, \hat{\mathrm{S}}(f)^{-1} \sum_{i=1}^{J} p_i \, \underline{q}(f, \underline{z}, \underline{\phi}_i). \tag{2.20}$$

A necessary condition for $\underline{w}_o \in \mathcal{C}^N$ to be a solution to the following minmax problem

$$\underline{w}_{opt}(f, \underline{z}, \sigma_o^2) = \arg \min_{\underline{w} \in \mathcal{C}^N} \max_{\underline{\phi} \in \Phi} \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2),$$

is that

$$\exists J \in \{1, \ldots, 2N + 1\} \tag{2.21}$$

for which (2.18) through (2.20) are satisfied.

Therefore, if the appropriate set of extremal points and convex weights can be determined, the optimal array weight vector can be calculated directly. The minmax problem can thus be reformulated as jointly finding the $J, \underline{\phi}_1, \ldots, \underline{\phi}_J, p_1, \ldots, p_J$, and

$\underline{w}_o \in \mathcal{C}^N$ which satisfy (2.17) through (2.20). The key to finding the appropriate set of extremal points, convex weights, and array weight vector lies in reformulating the minmax estimation problem as a Wiener filtering problem with the uncertain environmental parameter treated as a random parameter.

2.3.2 The Least Favorable PMF Random Parameter Framework

From Section 2.1, in the minmax signal processing framework the uncertain environmental parameter is treated as a nonrandom parameter. However, an efficient method for calculating the optimal minmax array weights can be developed by treating the uncertain environmental parameter as a random parameter with a particular probability function and then solving for the minimum mean-squared error array weights (i.e., Wiener filter weights). As a computational necessity and to ensure that $\underline{q}(f, \underline{z}, \underline{\phi})$ is a continuous function on Φ, the range of the environmental parameter will be sampled (i.e., $\Phi = \{\underline{\phi}_1, \ldots, \underline{\phi}_K\}$), and the minmax problem will be solved on this discrete set of environmental conditions. The issues associated with the effect of this sampling are treated in Subsection 3.1.4. Therefore, the probability function assigned to the environmental parameters will take the form of a pmf (probability mass function) rather than the form of a pdf. The pmf will be denoted by $\underline{p} \in \mathbb{R}^K$ and is defined by

$$p_i \triangleq \text{Probability}[\underline{\phi} = \underline{\phi}_i].$$

Since $\underline{p}$ is a pmf, it must satisfy

$$p_i \geq 0 \text{ and } \sum_{i=1}^{K} p_i = 1 .$$

These are the same conditions which must be satisfied by the convex weights used to calculate the points in the convex hull of a set of points and therefore by the weights which are used to calculate the optimal minmax weight vector in (2.20). This fact will be used to relate the Wiener filter weight vector to the optimal minmax weight vector.

For any pmf and array weight vector, the mean-squared estimation error is

$$\varepsilon(f,\underline{z},\underline{w},\underline{p},\sigma_o^2) \triangleq \mathrm{E}[\mid X_k^l(f,\underline{z}) - \hat{X}_k^l(f,\underline{z}) \mid^2] = \sum_{i=1}^{K} p_i\, \varepsilon(f,\underline{z},\underline{w},\underline{\phi}_i,\sigma_o^2). \qquad (2.22)$$

Substituting (2.5) into (2.22) and carrying out the algebraic manipulation yields

$$\varepsilon(f,\underline{z},\underline{w},\underline{p},\sigma_o^2) = \sigma_o^2 - 2\,\sigma_o^2\,\mathrm{Real}((\sum_{i=1}^{K} p_i\, \underline{q}(f,\underline{z},\underline{\phi}_i))^h\underline{w}) + \underline{w}^h\hat{\mathrm{S}}(f)\underline{w}\,. \qquad (2.23)$$

Define the Wiener filter weight vector to be

$$\underline{w}_{mmse}(f,\underline{z},\sigma_o^2,\underline{p}) \triangleq \arg\min_{\underline{w}\in\mathcal{C}^N} \varepsilon(f,\underline{z},\underline{w},\underline{p},\sigma_o^2).$$

Then, unconstrained complex quadratic minimization methods yield

$$\underline{w}_{mmse}(f,\underline{z},\sigma_o^2,\underline{p}) = \sigma_o^2\,\hat{\mathrm{S}}(f)^{-1} \sum_{i=1}^{K} p_i\, \underline{q}(f,\underline{z},\underline{\phi}_i). \qquad (2.24)$$

(2.20) and (2.24) differ only in the respect that in (2.20) the summation is over J extremal points contained in $M(\underline{w}_o)$ while in (2.24) the summation is over all environmental conditions contained in Φ. Therefore, if a pmf $\underline{p}$ can be found such that p_i is greater than zero only if $\underline{\phi}_i \in M(\underline{w}_{mmse}(f,\underline{z},\sigma_o^2,\underline{p}))$, then the summation in (2.24) will effectively be over only the extremal points contained in $M(\underline{w}_{mmse}(f,\underline{z},\sigma_o^2,\underline{p}))$. In this case, the sufficient conditions in Corollary 1 will be satisfied by K, $p_1,\ldots,p_K$, $\underline{\phi}_1,\ldots,\underline{\phi}_K$, and $\underline{w}_{mmse}(f,\underline{z},\sigma_o^2,\underline{p})$.

The key to finding the correct pmf can be discerned by observing the behavior of the mean-squared error function $\varepsilon(f,\underline{z},\underline{w}_{mmse}(f,\underline{z},\sigma_o^2,\underline{p}),\underline{p},\sigma_o^2)$ and the conditional mean-squared error function $\varepsilon(f,\underline{z},\underline{w}_{mmse}(f,\underline{z},\sigma_o^2,\underline{p}),\underline{\phi},\sigma_o^2)$ as $\underline{p}$ is allowed to vary. $\varepsilon(f,\underline{z},\underline{w}_{mmse}(f,\underline{z},\sigma_o^2,\underline{p}),\underline{p},\sigma_o^2)$, which will be abbreviated as $\varepsilon(f,\underline{z},\underline{w}_{mmse},\underline{p},\sigma_o^2)$, is the minimum mean-squared estimation error achievable by any array weights given the pmf $\underline{p}$. That is,

$$\varepsilon(f,\underline{z},\underline{w}_{mmse},\underline{p},\sigma_o^2) = \min_{\underline{w}\in\mathcal{C}^N} \varepsilon(f,\underline{z},\underline{w},\underline{p},\sigma_o^2)\,.$$

$\varepsilon(f, \underline{z}, \underline{w}_{mmse}(f, \underline{z}, \sigma_o^2, \underline{p}), \underline{\phi}, \sigma_o^2)$, which will be abbreviated as $\varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}), \underline{\phi}, \sigma_o^2)$, is the conditional mean-squared estimation error achieved by the Wiener array weights for the pmf $\underline{p}$ conditioned on the environmental variable $\underline{\phi}$.

Define $\underline{w}_{opt}(f, \underline{z}, \underline{\phi}_i, \sigma_o^2)$ as

$$\underline{w}_{opt}(f, \underline{z}, \underline{\phi}_i, \sigma_o^2) \triangleq \arg \min_{\underline{w} \in \mathcal{C}^N} \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}_i, \sigma_o^2) = \sigma_o^2 \hat{\mathrm{S}}(f)^{-1} \underline{q}(f, \underline{z}, \underline{\phi}_i).$$

Then, $\underline{w}_{mmse}(f, \underline{z}, \sigma_o^2, \underline{p})$ can be expressed as

$$\underline{w}_{mmse}(f, \underline{z}, \sigma_o^2, \underline{p}) = \sum_{i=1}^{K} p_i \, \underline{w}_{opt}(f, \underline{z}, \underline{\phi}_i, \sigma_o^2) \, .$$

Therefore, as p_m is increased incrementally and p_n is decreased incrementally for some m and n ($\underline{p}$ is a pmf and therefore p_n must be decreased for some n when p_m is increased for some m), $\underline{w}_{mmse}(f, \underline{z}, \sigma_o^2, \underline{p})$ should become more like $\underline{w}_{opt}(f, \underline{z}, \underline{\phi}_m, \sigma_o^2)$ and less like $\underline{w}_{opt}(f, \underline{z}, \underline{\phi}_n, \sigma_o^2)$. Therefore, $\varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}), \underline{\phi}_m, \sigma_o^2)$ should decrease and $\varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}), \underline{\phi}_n, \sigma_o^2)$ should increase. Furthermore, since

$$\varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}, \sigma_o^2) = \sum_{i=1}^{K} p_i \, \varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}), \underline{\phi}_i, \sigma_o^2) \, , \tag{2.25}$$

$\varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}, \sigma_o^2)$ should increase if $\varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}), \underline{\phi}_m, \sigma_o^2) > \varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}), \underline{\phi}_n, \sigma_o^2)$. With this intuition in mind, consider the following example.

Example: Assume that ϕ is a discrete-valued real scalar variable and that for some pmf $\underline{p}_o$, $\varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}_o), \phi, \sigma_o^2)$ is as shown in Figure 2-6. Then, if we select any extremal point $\phi_m \in M(\underline{w}_{mmse}(\underline{p}_o))$ and any nonextremal point $\phi_n \notin M(\underline{w}_{mmse}(\underline{p}_o))$ for which $p_{o_n} > 0$, we can increase p_{o_m} and decrease p_{o_n} incrementally. The effect will be that the mean-squared estimation error $\varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}_o, \sigma_o^2)$ and the conditional mean-squared estimation error conditioned on $\phi = \phi_n$ ($\varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}_o), \phi_n, \sigma_o^2)$) will increase, and the conditional mean-squared estimation error conditioned on $\phi = \phi_n$ ($\varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}_o), \phi_n, \sigma_o^2)$) will decrease. Furthermore, since $\varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}_o), \phi, \sigma_o^2)$ will be decreased at the extremal point, the extremal value $\Delta(\underline{w}_{mmse}(\underline{p}_o))$ will be decreased. Conceptually, this process can be repeated until $p_n = 0$ for all nonextremal points $\phi_n \notin M(\underline{w}_{mmse}(\underline{p}_o))$. When this is the case, the sufficient conditions in Corollary 1 will be satisfied and $\underline{w}_{mmse}(\underline{p}_o)$ will be the optimal minmax

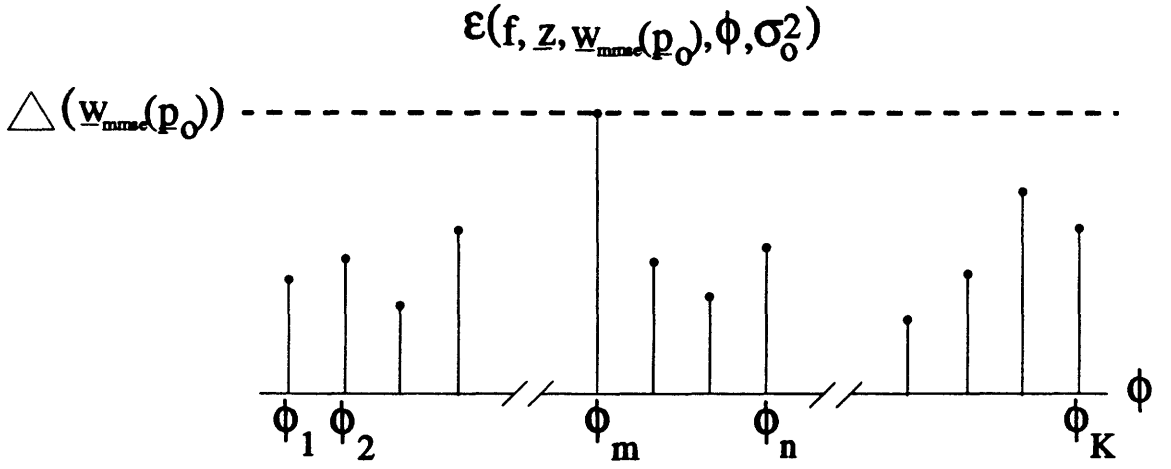


Figure 2-6: Conditional Mean-Squared Estimation Error

array weights. Any further change to $\underline{p}_o$ involving an extremal and a nonextremal point will require lowering the probability corresponding to the extremal point and raising the probability corresponding to the nonextremal point. This would result in lowering the mean-squared estimation error $\varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}_o, \sigma_o^2)$. Thus, the desired $\underline{p}_o$ should be that which maximizes $\varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}_o, \sigma_o^2)$.

The intuitive result illustrated in the preceding example is formalized in the following theorem, a proof of which is contained in Appendix B.

Theorem 4 Let P be the set of all possible pmfs which may be assigned to Φ and define the least favorable pmf as

$$\underline{p}_{lf} \stackrel{\Delta}{=} \arg\max_{\underline{p}\in P} \min_{\underline{w}\in C^N} \varepsilon(f, \underline{z}, \underline{w}, \underline{p}, \sigma_o^2) = \arg\max_{\underline{p}\in P} \varepsilon(f, \underline{z}, \underline{w}_{mmse}(f, \underline{z}, \sigma_o^2, \underline{p}), \underline{p}, \sigma_o^2) . \tag{2.26}$$

Then

$$\underline{w}_{opt}(f, \underline{z}, \sigma_o^2) \stackrel{\Delta}{=} \arg\min_{\underline{w}\in C^N} \max_{\underline{\phi}\in\Phi} \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2) = \underline{w}_{mmse}(f, \underline{z}, \sigma_o^2, \underline{p}_{lf}) .$$

This theorem states that the least favorable pmf is the pmf for which the minimum mean-squared estimation error is maximized and the Wiener filter weight vector for the least favorable pmf is also the optimal minmax array weight vector.

2.3.3 Solving for the Least Favorable PMF

By combining (2.23) and (2.24) and carrying out algebraic manipulation, the minimum mean-squared estimation error can be expressed as

$$\varepsilon(f,\underline{z},\underline{w},\underline{p},\sigma_o^2) \;=\; \sigma_o^2\,(1-\sigma_o^2\,(\sum_{i=1}^{K} p_i\,\underline{q}(f,\underline{z},\underline{\phi}_i))^h\,\hat{\mathrm{S}}(f)^{-1}\,(\sum_{i=1}^{K} p_i\,\underline{q}(f,\underline{z},\underline{\phi}_i)))\,. \quad (2.27)$$

Defining the matrix $Q(f,\underline{z})$ as $Q(f,\underline{z}) \triangleq [\underline{q}(f,\underline{z},\underline{\phi}_1),\ldots,\underline{q}(f,\underline{z},\underline{\phi}_K)]$, (2.27) can be rewritten as

$$\varepsilon(f,\underline{z},\underline{w},\underline{p},\sigma_o^2) \;=\; \sigma_o^2\,(1-\sigma_o^2\,\underline{p}^t Q(f,\underline{z})^h\hat{\mathrm{S}}(f)^{-1}Q(f,\underline{z})\underline{p})\,.$$

Finding the $\underline{p}$ to maximize this quantity is equivalent to finding the $\underline{p}$ to minimize the matrix quadratic product in the second term. Therefore, (2.26) can be rewritten as

$$\underline{p}_{lf} \triangleq \arg\max_{\underline{p}\in P}\min_{\underline{w}\in C^N}\varepsilon(f,\underline{z},\underline{w},\underline{p},\sigma_o^2) \;=\; \arg\min_{\underline{p}\in P}\underline{p}^t Q(f,\underline{z})^h\hat{\mathrm{S}}(f)^{-1}Q(f,\underline{z})\underline{p}\,. \quad (2.28)$$

Since $Q(f,\underline{z})^h\hat{\mathrm{S}}(f)^{-1}Q(f,\underline{z})$ is a Hermitian matrix and $\underline{p}$ is a real vector,

$$\underline{p}^t Q(f,\underline{z})^h\hat{\mathrm{S}}(f)^{-1}Q(f,\underline{z})\underline{p} \;=\; \underline{p}^t\,\mathrm{Real}(Q(f,\underline{z})^h\hat{\mathrm{S}}(f)^{-1}Q(f,\underline{z}))\,\underline{p}\,.$$

Defining the real matrix $T(f,\underline{z}) \triangleq \mathrm{Real}(Q(f,\underline{z})^h\hat{\mathrm{S}}(f)^{-1}Q(f,\underline{z}))$, explicitly defining the set P, and letting the vector $\underline{e} = [1,\ldots,1]^t$, the optimization problem in (2.28) can be expressed as the following real constrained quadratic minimization problem.

$$\underline{p}_{lf} \;=\; \arg\min_{\substack{\underline{p}\geq 0\\ \underline{e}^t\underline{p}=1}}\;\underline{p}^t\,T(f,\underline{z})\,\underline{p} \quad (2.29)$$

A solution to (2.29) is guaranteed to exist because $\underline{p}^t\,T(f,\underline{z})\,\underline{p}$ is a continuous function of $\underline{p}$ and the set P is a compact set. Since $\hat{\mathrm{S}}(f)$ is a positive definite Hermitian matrix, $Q(f,\underline{z})^h\hat{\mathrm{S}}(f)^{-1}Q(f,\underline{z})$ is a positive semi-definite Hermitian matrix.

Therefore, $T(f,\underline{z})$ is a positive semi-definite symmetric matrix and $\underline{p}^t\, T(f,\underline{z})\, \underline{p}$ is a convex function of $\underline{p}$. There are a number of algorithms available for solving linearly constrained convex quadratic minimization problems such as (2.29). An efficient algorithm, based on complementary pivot theory, was proposed by Lemke in [24] and is described in a more readable form in Chapter 11 of [25]. The basic intuition behind the use of the complementary pivot theory to solve a quadratic problem is that the necessary and sufficient conditions, known as the Kuhn-Tucker conditions for $\underline{p}_o$ to be a solution to the problem in (2.29), are largely a set of linear equations. The Kuhn-Tucker conditions can be written as [25]

$$\underline{p}_o \geq 0,\ \ \underline{e}^t\underline{p}_o = 1,\ \ \exists u \in \mathbb{R},\ \text{ and } \exists \underline{v} \geq 0 \text{ s.t.} \tag{2.30}$$

$$\underline{v}^t\underline{p}_o \ = \ 0 \text{ and} \tag{2.31}$$

$$\underline{v} \ - \ T(f,\underline{z})\,\underline{p}_o \ + u\,\underline{e} \ = \ 0\,. \tag{2.32}$$

Given that $\underline{p}_o \geq 0$ and $\underline{v} \geq 0$, (2.31) requires that if $p_{o_i} > 0$ then $v_i = 0$ and if $v_i > 0$ then $p_{o_i} = 0$. This condition is known as the complementary slackness condition; p_{o_i} and v_i are known complements of each other and together they are known as a complementary pair. The primary fact is that, with the exception of (2.31), the Kuhn-Tucker conditions are a set of linear equations (equalities and inequalities) and any solution to this set which also satisfies the complementary slackness condition will be a solution to (2.29).

If $\exists i \in \{1,\ldots,K\}$ s.t. $T(f,\underline{z})_{ii} \leq T(f,\underline{z})_{ji}\ \forall j \neq i$, then letting $p_i = 1,\ \ p_j = 0\ \forall j \neq i,\ \ v_i = 0,\ \ v_j = T(f,\underline{z})_{ji} - T(f,\underline{z})_{ii}\ \forall j \neq i$, and $u = T(f,\underline{z})_{ii}$ will satisfy the Kuhn-Tucker conditions. If such a solution is not apparent, the complementary pivot algorithm can be used to find a solution.

(2.30) through (2.32) are not in a form which allows direct solution using the complementary pivot algorithm. The algorithm presented in [24] and [25] does not allow the inclusion of the unrestricted variable u and the vector $\underline{e}$ which are associated with the linear equality constraint that $\underline{e}^t\underline{p}_o \ = \ 1$. However, with some algebraic

manipulation, the Kuhn-Tucker conditions can be rewritten as

$$\tilde{p} \geq 0, \text{ and } \exists \tilde{v} \geq 0 \text{ s.t.} \tag{2.33}$$

$$\tilde{v}^t \tilde{p} = 0 \text{ and} \tag{2.34}$$

$$\tilde{v} - M(f,\underline{z})\,\tilde{p} = \underline{r}(f,\underline{z}), \tag{2.35}$$

where $\tilde{p}^t = [v_1, p_{o_2}, p_{o_3}, \ldots, p_{o_K}]$ and $\tilde{v}^t = [p_{o_1}, v_2, v_3, \ldots, v_K]$. Using the following notation for $T(f,\underline{z})$

$$T(f,\underline{z}) = \begin{bmatrix} T_{11}(f,\underline{z}) & \underline{t}^t(f,\underline{z}) \\ \tilde{t}(f,\underline{z}) & \tilde{T}(f,\underline{z}) \end{bmatrix},$$

$\underline{r}(f,\underline{z})$ is given by

$$\underline{r} = \begin{bmatrix} 1 \\ \tilde{t}(f,\underline{z}) - T_{11}(f,\underline{z})\underline{e} \end{bmatrix};$$

and $M(f,\underline{z})$ is the following matrix

$$M(f,\underline{z}) = \begin{bmatrix} 0 & -\underline{e}^t \\ \underline{e} & \tilde{M}(f,\underline{z}) \end{bmatrix},$$

and $\tilde{M}(f,\underline{z}) = \tilde{T}(f,\underline{z}) + T_{11}(f,\underline{z})\,\underline{e}\,\underline{e}^t - \underline{e}\,\underline{t}^t(f,\underline{z}) - \tilde{t}(f,\underline{z})\,\underline{e}^t$. Any $\tilde{p}$ and $\tilde{v}$ which satisfy (2.33) through (2.35) will yield a $\underline{v}$ and $\underline{p}_o$ for which (2.30) through (2.32) will be satisfied for some u; and any $\underline{v}$ and $\underline{p}_o$ for which a u exists, satisfying (2.30) through (2.32), will yield a $\tilde{p}$ and $\tilde{v}$ which satisfy (2.33) through (2.35) . Therefore, a solution to (2.30) through (2.32) can be found by solving (2.33) through (2.35); and (2.33) through (2.35) are in a form for which a solution can be found directly using the complementary pivot algorithm.

The complementary pivot algorithm used to find a solution to (2.33) through (2.35) is referred to as Scheme I in [24]. The algorithm finds a solution by introducing a slack variable z_o and a vector with positive entries $\underline{d}$ and then conceptually solving the problem

$$\min z_o \text{ s.t.} \tag{2.36}$$

$$z_0,\ \tilde{\underline{v}},\ \tilde{\underline{p}} \geq 0\,, \tag{2.37}$$

$$\tilde{\underline{v}}^t \tilde{\underline{p}} = 0\,, \text{ and} \tag{2.38}$$

$$\tilde{\underline{v}} - M(f,\underline{z})\,\tilde{\underline{p}} - \underline{d}\,z_o = \underline{r}(f,\underline{z}). \tag{2.39}$$

Since a solution to (2.33) through (2.35) is guaranteed to exist, the solution to (2.36) through (2.39) will be $z_o = 0$ with a $\tilde{\underline{v}}$ and $\tilde{\underline{p}}$ which satisfy (2.33) through (2.35).

The details of the complementary pivot algorithm can be examined in [24] or [25]. In concept, the complementary pivot algorithm is very similar to the simplex method [25] for solving linear programming problems. The algorithm is solving for $2K+1$ real nonnegative variables and at the solution to (2.36) through (2.39), $z = 0$. (2.39) is a set of K linear equations and is satisfied at each iteration of the algorithm. By construction, at each iteration of the algorithm, at least $K+1$ of the unknown variables equal zero. Using the terminology from the simplex method, the term "non-basic variables" will refer to $K+1$ of the variables which equal zero. The term "basic variables" will refer to the other K variables. Collectively, the basic variables are referred to as the basis. It is not required that every basic variable be non-zero. It is only required that every non-basic variable equals zero.

At each iteration of the algorithm, a new basis is selected in a manner which guarantees that (2.37), (2.38), and (2.37) are satisfied. When the variable z leaves the basis, which guarantees that $z = 0$, the algorithm terminates. The fundamental difference between the simplex method and the complementary pivot algorithm is that in the complementary pivot algorithm, only one variable from each complementary pair can be in the basis at any given time. This is enforced to guarantee that (2.38) will be satisfied.

The critical property of the algorithm for the purpose this work is its convergence property. The essential points of the convergence proof for the complementary pivot algorithm [26] are that, at each iteration the basis changes, no basis can be visited by the algorithm more than once, and there are a finite number of possible bases. These three points lead to the conclusion that the algorithm must terminate in a finite number of iterations.

The following two claims, for which proofs are contained in Appendix B, are necessary in order to analyze the convergence of the algorithm in a more rigorous manner.

Claim 1 $M(f, \underline{z})$ is a co-positive-plus matrix.

Claim 2 The set of equations (2.33) through (2.35) are consistent.

Given Claims 1 and 2, Theorem 11.1.8 on page 446 of [25] states that, if each almost complementary basic feasible solution to (2.37) through (2.39) is nondegenerate, then the complementary pivot algorithm will terminate in a finite number of iterations with a solution to (2.33) through (2.35).

The outstanding condition on which this finite convergence property of the algorithm depends is that each almost complementary basic feasible solution to (2.37) through (2.39) is nondegenerate. In [26] this is referred to as the system being nondegenerate. Lemma 4 on page 616 of [26] states that almost every vector $\underline{d}$ will yield a nondegenerate system. Therefore, in practice, system degeneracy is usually not a problem. However, should the degeneracy of the system be a concern, a modification of the complementary pivot algorithm can be used (Section 7 of [26]). Given Claims 1 and 2, Theorem 2 on page 618 of [26] states that the modified algorithm will converge in a finite number of iterations with a solution to (2.33) through (2.35). A clear explanation of the modified algorithm is given on pages 80 and 81 of [27]. The form in which the modified algorithm is presented in [27] is slightly different from the form used in [26], but the two algorithms are identical.

Conceptually, the modification to the complementary pivot algorithm is that the system of linear equalities (2.39) is modified to

$$\tilde{v} - M(f, \underline{z})\, \tilde{p} - \underline{d}\, z_o = \underline{r}(f, \underline{z}) + I\, \underline{\delta}\,, \tag{2.40}$$

where I is the identity matrix, $\underline{\delta} = [\delta, \delta^2, \ldots, \delta^K]^t$, and δ is an arbitrarily small positive constant which remains unspecified throughout the execution of the algorithm. This small perturbation to (2.39) can be shown to always create a nondegenerate system (2.37), (2.38), and (2.40). In practice, the modification requires simply that the

scalar comparisons of the updated ratios $\frac{\overline{r}_i(f,\underline{z})}{d_i}$ in Step 1 of the original complementary pivot algorithm (page 440 of [25]) be replaced by lexicographic comparisons of the updated vector ratios $\frac{\overline{\underline{r}(f,\underline{z})\, I_i}}{d_i}$, where $\overline{\underline{r}(f,\underline{z})\, I_i}$ denotes the i^{th} row of the updated righthand side matrix $[\underline{r}(f,\underline{z})\ I]$.

2.3.4 Least Favorable PMF form of the Array Processing Algorithm

Given the least favorable pmf, the optimal minmax array weights can be calculated by applying Theorem 4 and (2.24). The power estimate can be calculated using (2.10). These can be combined to yield the following three-step algorithm for implementing the Adaptive Minmax Matched Field Processor which does not require an a priori assumption about the average signal power.

1. Use the (modified) complementary pivot algorithm to calculate

$$\underline{p}_{lf} \;=\; \arg \min_{\substack{\underline{p} \geq \underline{0} \\ \underline{e}^t \underline{p} = 1}} \; \underline{p}^t \, T(f,\underline{z}) \, \underline{p} \, .$$

2. $\hat{\sigma}^2(f,\underline{z}) = (\underline{p}_{lf}^t \, T(f,\underline{z}) \, \underline{p}_{lf})^{-1}$

3. $\underline{w}_{opt}(f,\underline{z},\hat{\sigma}^2(f,\underline{z})) = \hat{\sigma}^2(f,\underline{z}) \, \hat{\mathrm{S}}(f)^{-1} \, Q(f,\underline{z}) \, \underline{p}_{lf}$

2.4 Minmax Estimation Error Bounds

In Section 2.1, a general framework for minmax signal processing was described and a theorem characterizing optimal minmax estimators was developed. However, this characterization required that the estimator be parameterized by a vector $\underline{w}$ and that the problem be reduced to finding the optimal parameter vector $\underline{w}$. This parameterization requires that the "optimal" estimator lie within a particular class of estimators (e.g., linear estimators). While the resulting estimator is the best from within the chosen class of estimators, it is usually not possible to make a definitive statement about

how the estimator's performance compares to that of arbitrary unspecified classes of estimators. Therefore, it is useful to develop lower bounds on the performance (as measured by the error functions ε and Δ) of **any** estimator. While not explicitely calculating any bounds, this section proposes an approach to the development of such performance bounds for the case where ε is the conditional mean-squared estimation error. In the development of this approach, the form of the optimal minmax estimator and an achievable lower bound are derived.

The problem for which the bound is proposed is the estimation of a scalar parameter $\theta \in \mathbb{R}$ based upon observations $x \in X$ where X is the observation space. Given any estimator $g : X \rightarrow \mathbb{R}$, the mean-squared estimation error of θ is assumed to depend on some environmental parameters $\phi \in \Phi$. In the literature, these parameters are also referred to as nuisance parameters. The bound developed will be a lower bound on the extremal value

$$\Delta(g) \triangleq \max_{\phi \in \Phi} \varepsilon(g, \phi)$$

for the case where

$$\varepsilon(g, \phi) = \mathrm{E}[(\theta - g(x))^2 \mid \phi], \tag{2.41}$$

the conditional expectation is taken over all θ and x; and g is any function mapping X into $\mathbb{R}$.

The approach to bounding $\Delta(g)$ is developed using the least favorable pmf framework where the environmental parameter ϕ is considered to be a random parameter. This is the same framework used in Section 2.3 to develop an efficient algorithm for solving the minmax problem. As in this earlier development, the set of possible values for the environmental paramters Φ is sampled to yield a finite discrete set and the vector $\underline{p}$ is the pmf for ϕ on this set. In the earlier development, the sampling was needed for computational reasons as well as to ensure that the error measure and its gradient were both continuous on Φ. However, here the sampling is needed only for computational reasons.

The following theorem, which is needed to prove Theorem 6, provides some insight

into the proposed approach to bounding $\Delta(g)$. A proof of this theorem is contained in Appendix A.

Theorem 5 Let $\varepsilon(g,\phi)$ be any performance measure for the estimator g given the environmental condition ϕ. Let $\Phi = \{\phi_1,\ldots,\phi_K\}$ and let $\underline{p} \in P = \{\underline{p} \in \mathbb{R}^K \mid \underline{p} \geq 0 \text{ and } \underline{e}^t\,\underline{p} = 1\}$ be any pmf assigned to ϕ on Φ. Let $\varepsilon(\underline{p},g) = \sum_{i=1}^K p_i \varepsilon(g,\phi_i)$ and let $\beta(\underline{p})$ be any global lower bound on $\varepsilon(\underline{p},g)$. That is,

$$\beta(\underline{p}) \leq \min_{g:X\to\mathbb{R}} \varepsilon(\underline{p},g).$$

Then, $\beta(\underline{p})$ is also a lower bound on $\Delta(g)$. That is,

$$\Delta_o \triangleq \min_{g:X\to\mathbb{R}} \Delta(g) \geq \beta(\underline{p}).$$

The implication of Theorem 5 is that, if $\varepsilon(g,\phi)$ is the conditional mean-squared estimation error as given in (2.41), then $\varepsilon(\underline{p},g)$ is the mean-squared estimation error and the traditional methods of bounding $\varepsilon(\underline{p},g)$ (e.g., the Cramer-Rao bound, the Weiss-Weinstein bound, etc.) can be applied to bounding $\Delta(g)$. Since the inequality $\beta(\underline{p}) \leq \Delta_o$ holds for any pmf, $\underline{p}$ can be chosen to maximize the bound being used $\beta(\underline{p})$. However, this may not be useful unless the derived bounds are reasonably tight.

The following theorem, a proof of which is contained in Appendix A, specifies the optimal minmax estimator and an achievable bound on $\Delta(g)$, and sheds some light on the tightness which can be expected of the bounds developed using the approach in the preceding paragraph.

Theorem 6 Let $\Phi = \{\phi_1,\ldots,\phi_K\}$, $P = \{\underline{p} \in \mathbb{R}^K \mid \underline{p} \geq 0 \text{ and } \underline{e}^t\,\underline{p} = 1\}$, and $g_o : X \times P \to \mathbb{R}$ be given by

$$g_o(x,\underline{p}) = \frac{\sum_{i=1}^K p_i\, p_{x|\phi}(x \mid \phi_i)\mathrm{E}[\theta \mid x,\phi_i]}{\sum_{i=1}^K p_i\, p_{x|\phi}(x \mid \phi_i)}, \tag{2.42}$$

where $p_{x|\phi}(x \mid \phi_i)$ is the conditional pdf or pmf of the observation x given that the environmental parameter $\phi = \phi_i$. Let $\varepsilon(g,\phi) = \mathrm{E}[(\theta - g(x))^2 \mid \phi]$.

Let the least favorable pmf $\underline{p}_{lf} \in P$ be defined as

$$\underline{p}_{lf} \triangleq \arg\max_{\underline{p}\in P} \sum_{i=1}^K p_i\, \varepsilon(g_o(x,\underline{p}),\phi_i).$$

Then

$$g_o(x,\underline{p}_{lf}) = \arg\min_{g:X\to\mathbb{R}} \max_{\phi\in\Phi} \varepsilon(g,\phi),$$

and the quantity

$$\sum_{i=1}^{K} p_{lf_i}\, \varepsilon(g_o(x, \underline{p}_{lf}), \phi_i)$$

is an achievable lower bound on

$$\Delta(g) \;=\; \max_{\phi \in \Phi} \varepsilon(g, \phi).$$

If ϕ is considered a random parameter with the pmf $\underline{p}$, then (2.42) can be rewritten as

$$g_o(x, \underline{p}) \;=\; \mathrm{E}[\theta \mid x],$$

and $g_o(x, \underline{p})$ is the minimum mean-squared error estimator of θ given x. $\varepsilon(\underline{p}, g)$ is the mean-squared estimation error achieved by $g(x)$ given the pmf $\underline{p}$; and $\varepsilon(\underline{p}, g_o(x, \underline{p}))$ is the minimum achievable mean-squared estimation error given the pmf $\underline{p}$. Therefore, $\underline{p}_{lf}$ is the pmf for which the minimum achievable mean-squared estimation error is maximized, hence the term least favorable pmf.

In most situations, it will be impractical to compute the least favorable pmf $\underline{p}_{lf}$, the optimal minmax estimator $g_o(x, \underline{p}_{lf})$, or the associated achievable minmax estimation error bound $\Delta_o = \sum_{i=1}^{K} p_{lf_i}\, \mathrm{E}[(\theta - g_o(x, \underline{p}_{lf}))^2 \mid \phi_i]$. However, Δ_o is an achievable minmax bound and equals the achievable mean-squared estimation error bound given the least favorable pmf. Therefore, it is reasonable to expect that the bound development approach outlined earlier (i.e., choosing the pmf to maximize a mean-squared estimation error bound such as the global Cramer-Rao bound) will yield minmax bounds which are at least as tight as the mean-squared estimation error bound. The further investigation of this approach is left for future work.

Chapter 3

Analysis and Interpretation of the Adaptive Minmax Matched Field Processor

The analysis of the algorithm developed in Chapter 2 is complicated by the lack of illustrative analytical solutions to the minmax problem. The analysis of the characteristics of the Adaptive Minmax Matched Field Processor presented in Section 3.1 is based upon the interpretation of the processor as a Two-Stage MVDR Matched Field Processor and is qualitative rather than quantitative in nature. This analysis motivates a modification to the definition of the signal replica vector which is presented in Subsection 3.1.3. A quantitative analysis of the algorithm based upon numerical simulations is presented in Section 3.2. The cases which are considered involve propagation in a deterministic ideal waveguide, a deterministic horizontally-stratified ocean, and a random ideal waveguide. The results for random ideal waveguide are presented in Subsection 3.2.3. They motivate a further modification of the definition of the signal replica vector and show clearly the shortcomings of the MVDR, Bartlett, and Adaptive Minmax Processors when signal coherence over the array aperture is reduced. Finally, Section 3.3 contains a brief analysis of the computational complexity of the modified complementary pivot algorithm used to solve the minmax problem in Step 1 of the Adaptive Minmax Matched Field Processor.

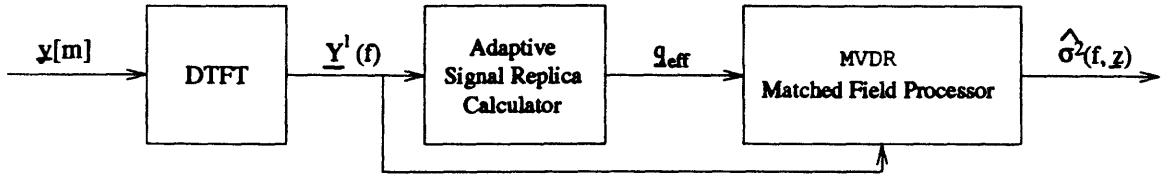


Figure 3-1: The Two-Stage MVDR Matched Field Processor

3.1 MVDR Interpretation of the Adaptive Minmax Matched Field Processor

The Adaptive Minmax Matched Field Processor developed in Chapter 2 can be interpreted as the combination of an algorithm which calculates an effective replica vector, which will be denoted by $\underline{q}_{eff}$, and a MVDR Matched Field Processor [9, 11] which uses $\underline{q}_{eff}$ as the replica vector of the desired signal (Figure 3-1). For several reasons, this interpretation is useful. First, it relates the minmax array processor to an array processor whose properties are well-understood. Second, it makes possible a qualitative analysis of the properties of the minmax array processor. Finally, the interpretation motivates a modification of the replica vector normalization convention used in the minmax array processor, which improves the performance of the processor. Subsection 3.1.1 details the new interpretation of the minmax array processor. Subsection 3.1.2 uses this interpretation to analyze some properties of the minmax array processor while Subsection 3.1.3 motivates a modification to the minmax array processor as developed in Chapter 2 and details the modified algorithm. Finally, Subsection 3.1.4 uses this interpretation to formulate guidelines for the required sampling of Φ in creating the discrete set of replica vectors over which the minmax optimization is carried, and discusses some factors which effect the selection of the uncertainty range of environmental parameters over which the processor must operate.

The structure of the MVDR Processor [9] is identical to the structure the Adaptive Minmax Processor as shown in Figure 2-4. The array weights of the matched field

implementation of the MVDR Processor [11] in a deterministic medium are given by

$$\underline{w}_{opt} = \arg \min_{\underline{w} \in C^N} \underline{w}^h \hat{S}(f) \underline{w} \qquad (3.1)$$
$$\text{such that } \underline{q}^h(f, \underline{z}, \underline{\phi}_o)\underline{w} = 1,$$

where the replica vector is defined in the traditional manner as described in Section 1.2 and $\underline{\phi}_o$ is the parameterization of the assumed environmental conditions. The solution to (3.1) is given by

$$\underline{w}_{opt} = \frac{\hat{S}(f)^{-1} \underline{q}(f, \underline{z}, \underline{\phi}_o)}{\underline{q}^h(f, \underline{z}, \underline{\phi}_o) \hat{S}(f)^{-1} \underline{q}(f, \underline{z}, \underline{\phi}_o)}, \qquad (3.2)$$

and the resulting estimate of the average power in the signal emitted by the source at the array focal point is

$$\hat{\sigma}^2(f, \underline{z}) = (\underline{q}^h(f, \underline{z}, \underline{\phi}_o) \hat{S}(f)^{-1} \underline{q}(f, \underline{z}, \underline{\phi}_o))^{-1}. \qquad (3.3)$$

The relationship between the Adaptive Minmax Matched Field Processor and the MVDR Matched Field Processor which is developed in this section is built upon the similarity between the **form** of the solutions in (3.2) and (3.3) and the solutions for the weights and estimated average power in Steps 2 and 3 of the algorithm detailed in Subsection 3.1.1 which implements the Adaptive Minmax Processor. In comparing these solutions, it is understood that two processors use different definitions of the signal replica vector, and that under some signal and environmental conditions and vector normalization conventions, the two definitions are equivalent.

3.1.1 The Two-Stage MVDR Matched Field Processor

The interpretation of the Adaptive Minmax Matched Field Processor as the Two-Stage MVDR Matched Field Processor shown in Figure 3-1 is motivated by noting that the algorithm detailed in Subsection 2.3.4 can be rewritten as

1. Use the (modified) complementary pivot algorithm to calculate

$$\underline{p}_{lf} = \arg \min_{\substack{\underline{p} \geq \underline{0} \\ \underline{e}^t \underline{p} = 1}} \underline{p}^t \, Q^h(f, \underline{z}) \hat{\mathrm{S}}(f)^{-1} Q(f, \underline{z}) \, \underline{p} \, .$$

2. $\hat{\sigma}^2(f, \underline{z}) = (\underline{p}_{lf}^t \, Q^h(f, \underline{z}) \hat{\mathrm{S}}(f)^{-1} Q(f, \underline{z}) \, \underline{p}_{lf})^{-1}$

3. $\underline{w}_{opt}(f, \underline{z}, \hat{\sigma}^2(f, \underline{z})) = \hat{\sigma}^2(f, \underline{z}) \, \hat{\mathrm{S}}(f)^{-1} \, Q(f, \underline{z}) \, \underline{p}_{lf}$

The set $\{Q(f, \underline{z})\underline{p} \mid \underline{p} \geq \underline{0}$ and $\underline{e}^t \underline{p} = 1\}$ is the convex hull of the set of column vectors in Q(f, z) (i.e. $\{\underline{q}(f, \underline{z}, \underline{\phi}_1), \ldots, \underline{q}(f, \underline{z}, \underline{\phi}_K)\}$). Defining $\tilde{Q}(f, \underline{z})$ to be this set of replica vectors (column vectors of $Q(f, \underline{z})$), the following algorithm is equivalent to the Adaptive Minmax Matched Field Processor.

1. Use the (modified) complementary pivot algorithm to calculate

$$\underline{q}_{eff} = \arg \min_{\underline{q} \in \mathcal{H}\left(\tilde{Q}(f, \underline{z})\right)} \underline{q}^h \, \hat{\mathrm{S}}(f)^{-1} \underline{q} \, .$$

2. $\hat{\sigma}^2(f, \underline{z}) = (\underline{q}_{eff}^h \, \hat{\mathrm{S}}(f)^{-1} \underline{q}_{eff})^{-1}$

3. $\underline{w}_{opt}(f, \underline{z}, \hat{\sigma}^2(f, \underline{z})) = \frac{\hat{\mathrm{S}}(f)^{-1} \, \underline{q}_{eff}}{\underline{q}_{eff}^h \, \hat{\mathrm{S}}(f)^{-1} \underline{q}_{eff}}$

Steps 2 and 3 of this algorithm are the MVDR Matched Field Processor given the replica vector $\underline{q}_{eff}$.

3.1.2 Analysis of the Two Stage MVDR Matched Field Processor

From Steps 1 and 2 of the Two-Stage MVDR Matched Field Processor, $\underline{q}_{eff}$ is the vector contained in $\mathcal{H}\left(\tilde{Q}(f, \underline{z})\right)$ which maximizes the power passed through the resulting MVDR Matched Field Processor. A principal characteristic of the processor is brought to light by this interpretation.

The processor treats any energy in any of the rank one subspaces spanned by the replica vectors in $\mathcal{H}\left(\tilde{Q}(f, \underline{z})\right)$ as energy of the desired signal and attempts to pass

as much of that energy as possible through the MVDR Matched Field Processor. The processor does this by first hypothesizing a rank one signal subspace. It chooses the basis for this subspace ($\underline{q}_{eff}$) from within $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$ to maximize the energy in this subspace as measured by the MVDR Matched Field Processor. This hypothesis of a rank one signal subspace spanned by some replica vector in $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$ causes performance problems in a random propagation medium as demonstrated by the results in Subsection3.2.3).

This characteristic of the processor is seen explicitly when Step 1 of the algorithm is written as

$$\underline{q}_{eff} = \arg \min_{\underline{q} \in \mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)} \sum_{i=1}^{N} \lambda_i^{-1} \mid \underline{v}_i^h \underline{q} \mid^2,$$

where (λ_i, $\underline{v}_i$) are the i^{th} eigenvalue and eigenvector, respectively, of $\hat{\mathrm{S}}(f)$. Assume that all vectors in $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$ have approximately the same norm (the effect of vectors with widely different norms is analyzed in Subsection 3.1.3). Since $\hat{\mathrm{S}}(f)^{-1}$ is Hermitian, its eigenvectors will comprise an orthonormal set which spans $\mathcal{C}^N$. Therefore, $\sum_{i=1}^{N} \mid \underline{v}_i^h \underline{q} \mid^2 = \mid \underline{q} \mid^2$. Since all $\underline{q} \in \mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$ have approximately the same norm, $\sum_{i=1}^{N} \lambda_i^{-1} \mid \underline{v}_i^h \underline{q} \mid^2$ can be minimized only by adjusting the relative magnitudes of the projections of $\underline{q}$ on each of the eigenvectors. Within the constraints imposed by $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$, the solution $\underline{q}_{eff}$ will have a minimal projection on the eigenvectors whose corresponding eigenvalues are small (i.e., λ_i^{-1} is large) and a larger projection on the eigenvectors whose corresponding eigenvalues are large (i.e., λ_i^{-1} is small). Conceptually, this corresponds to the processor maximizing the projection of $\underline{q}_{eff}$ on the portion of the subspace of the received signal (desired signal plus noise) with the largest average power.

The characteristics of the processor and the tradeoffs which it makes in calculating $\underline{q}_{eff}$ can be seen more clearly by considering the following special case. Assume that the desired signal is a deterministic narrowband signal and that the medium is deterministic. Then the subspace of the desired signal will have rank one and the

cross-spectral correlation matrix of the received signal will be

$$\mathrm{S(f)} = \sigma_n^2 \mathrm{S}_n(f) + \sigma_s^2 \underline{q}_{act} \underline{q}_{act}^h.$$

Here, $S_n(f)$ is the normalized cross-spectral correlation matrix of the noise, $\underline{q}_{act}$ is the actual replica vector of the desired signal, and σ_n^2 and σ_s^2 are the average power of the noise and the desired signal, respectively. For this example, it is assumed that $\hat{\mathrm{S}}(f) = \mathrm{S}(f)$.

Several concepts and quantities are useful in analyzing the processor for this special case [13]. The first of these is the array gain denoted by G. Considering the array processor to be a spatial filter, the array gain is a function of the filter weights. It is defined as the ratio of the signal to noise ratio at the output of the filter to the signal to noise ratio at the input of the filter where the signal to noise ratio is defined as $S/N \triangleq \frac{\sigma_s^2}{\sigma_n^2}$. For the special case under consideration, the array gain is given by

$$G(\underline{w}) = \frac{|\, \underline{w}^h \underline{q}_{act} \,|^2}{\underline{w}^h \mathrm{S}_n(f) \underline{w}}. \tag{3.4}$$

In Section 3.2, the rate of decay of the array gain as a function of changes in $\underline{q}_{act}$ will be used as a measure of the robustness of a processor. It can be shown that the linear filter weights which maximize the array gain are given by $\underline{w}_{max} = \mathrm{S}_n(f)^{-1} \underline{q}_{act}$. Substituting these weights into (3.4) yields $G_{max} = \underline{q}_{act}^h \, \mathrm{S}_n(f)^{-1} \, \underline{q}_{act}$ and the maximum achievable signal to noise ratio at the output of a linear filter as $(S/N)_{max} = \frac{\sigma_s^2}{\sigma_n^2} \, \underline{q}_{act}^h \, \mathrm{S}_n(f)^{-1} \, \underline{q}_{act}$.

A useful measure of the similarity of two vectors is based on the generalized angle between the vectors. For any positive-definite Hermitian matrix C, the inner product between the vectors $\underline{a}, \underline{b} \in C^N$ can be defined as $\underline{a}^h \, C \, \underline{b}$. Given this inner product, the cosine-squared of the generalized angle between $\underline{a}$ and $\underline{b}$ is defined as

$$\cos^2(\underline{a}, \underline{b}; C) \triangleq \frac{|\, \underline{a}^h \, C \, \underline{b} \,|^2}{(\underline{a}^h \, C \, \underline{a}) \, (\underline{b}^h \, C \, \underline{b})}.$$

The cosine-squared will always be greater than or equal to zero and less than or equal

to one. It will equal one if $\underline{b}$ is a scalar multiple of $\underline{a}$ and will equal zero if $\underline{a}$ and $\underline{b}$ are orthogonal with respect to the inner product defined earlier. Treating the inner product and the cosine-squared as measures of the similarity of $\underline{a}$ and $\underline{b}$, the inclusion of the matrix C places greatest weight on the similarity or dissimilarity of $\underline{a}$ and $\underline{b}$ in the subspace spanned by the eigenvectors of C with the largest eigenvalues, and places little emphasis on the similarity or dissimilarity in the subspace spanned by the eivenvectors with small eigenvalues. These measures and concepts provide the tools needed to analyze some aspects of the performance of the processor.

Using the identity

$$(A + \underline{b}\,\underline{b}^h)^{-1} = A^{-1} - A^{-1}\underline{b}\,\underline{b}^h A^{-1}\,(1 + \underline{b}^h A^{-1}\underline{b})^{-1},$$

the adaptive replica vector calculation problem can be expressed as

$$\underline{q}_{eff} = \arg \min_{\underline{q} \in \mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)} (\underline{q}^h\, \mathrm{S}_n(f)^{-1}\, \underline{q}) \left(\frac{(S/N)_{max} + 1}{(S/N)_{max}} - \cos^2(\underline{q}_{act}, \underline{q}; \mathrm{S}_n(f)^{-1}) \right). \tag{3.5}$$

From this expression, it can be seen that, when the maximum achievable signal to noise ratio is low, the processor will use most of the available degrees of freedom to minimize $\underline{q}_{eff}^h\, \mathrm{S}_n(f)^{-1}\, \underline{q}_{eff}$ (i.e., select $\underline{q}_{eff}$ to lie in the noisy subspace) and will place little emphasis on maximizing $\cos^2(\underline{q}_{act}, \underline{q}_{eff}; \mathrm{S}_n(f)^{-1})$ (i.e., select $\underline{q}_{eff}$ to match $\underline{q}_{act}$ as closely as possible in the relatively noise-free subspace. A low maximum achievable signal to noise ratio could be the result of either a low input signal to noise ratio or a low G_{max} (i.e., the actual replica rector $\underline{q}_{act}$ lying in a noise subspace with a large average power). If the latter is the cause, then the expected difference between $\underline{q}_{eff}$ and $\underline{q}_{act}$ which results from $\underline{q}_{eff}$ lying in the noisy subspace will be small because $\underline{q}_{act}$ lies in the noisy subspace and the processor may be able to choose $\underline{q}_{eff}$ to lie in the noisy subspace and be close to $\underline{q}_{act}$. However, if the former is the cause and $\underline{q}_{act}$ does not lie in the noisy subspace, the expected difference between $\underline{q}_{eff}$ and $\underline{q}_{act}$ will be larger. This increased mismatch is the price which is paid for a significant amount of noise power existing in a rank one subspace spanned by a replica vector

in $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$. However, the results developed by Cox [13] indicate that for cases such as this, where the input signal to noise ratio is low, the array gain will not decrease significantly as the mismatch between the actual and effective replica vectors increases. The large amount of noise power which will be passed through the filter in this case is consistent with the processor selecting $\underline{q}_{eff}$ to pass as much power as possible through the resulting MVDR linear filter.

If the maximum achievable signal to noise ratio is high, the processor will place an increased emphasis using the available degrees of freedom to maximize $\cos^2(\underline{q}_{act},\underline{q}_{eff};\mathrm{S}_n(f)^{-1})$. Thus, as the maximum achievable signal to noise ratio increases, the processor will reduce the mismatch between $\underline{q}_{act}$ and $\underline{q}_{eff}$. The tendency of the processor to place an increased emphasis on increasing $\cos^2(\underline{q}_{act},\underline{q}_{eff};\mathrm{S}_n(f)^{-1})$ as the maximum achievable signal to noise ratio is increased is verified numerically in Section 3.2.

The effect of the adaptive replica vector calculator on the array gain of the resulting processor can be seen by rewriting (3.4) as [13]

$$G(\underline{q}_{eff}) = \frac{G_{max}\,\cos^2(\underline{q}_{act},\underline{q}_{eff};\mathrm{S}_n(f)^{-1})}{1+\frac{\sigma_A^2}{\sigma_n^2}G_{max}(2+\frac{\sigma_A^2}{\sigma_n^2}G_{max})\,(1-\cos^2(\underline{q}_{act},\underline{q}_{eff};\mathrm{S}_n(f)^{-1}))}.$$

Here, the array gain is expressed as a function of the $\underline{q}_{eff}$ because the array weights are a function of $\underline{q}_{eff}$. Consider the effect on the array gain if $\frac{\sigma_A^2}{\sigma_n^2}$ is increased , $\underline{q}_{eff}$ does not change, and $\cos^2(\underline{q}_{act},\underline{q}_{eff};\mathrm{S}_n(f)^{-1}) \neq 1.0$. In this case, $G(\underline{q}_{eff})$ will decrease. However, from the preceding analysis, the processor adjusts to an increase in $\frac{\sigma_A^2}{\sigma_n^2}$ (and the corresponding increase in $(S/N)_{max}$) by adjusting $\underline{q}_{eff}$ to increase $\cos^2(\underline{q}_{act},\underline{q}_{eff};\mathrm{S}_n(f)^{-1})$. This adjustment reduces the mismatch between $\underline{q}_{eff}$ and $\underline{q}_{eff}$ and therefore reduces the loss in array gain.

The preceding example provides a qualitative analysis of the Adaptive Minmax Matched Field Processor using the interpretation of the processor as a Two-Stage MVDR Matched Field Processor. The assumption throughout this analysis has been that all of the replica vectors contained in $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$ have approximately the same norm. Should this assumption be violated, the characteristics of the processor may differ greatly from what would be expected based upon this analysis. The effect of

replica vectors with widely different norms on the performance of the processor and a modification to the algorithm which mitigates this effect are covered in the following subsection.

3.1.3 Replica Norm Considerations and a Normalization Modification

In defining the signal replica vector in Section 1.2, the normalization convention was that the k^{th} term equals one where k is the number of the reference sensor. This convention was adopted to facilitate the factorization of the cross-correlation function $\mathrm{E}[\underline{X}(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}]$ into $\sigma^2(f,\underline{z})\underline{q}(f,\underline{z},\underline{\phi})$ where $\sigma^2(f,\underline{z})$ is the average power in the desired signal as received at the reference sensor. When using large aperture arrays in the ocean environment, it is possible for different environmental conditions to yield different ratios of the average power in the desired signal at the reference sensor to the norm of the cross-correlation vector $\mathrm{E}[\underline{X}(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}]$. For example, if the receiving array is a large vertical array and a deterministic Normal Mode propagation model is used (see Section 4.3 for a discussion of the normal mode propagation model.), then different environmental conditions will result in different mode shapes and there will be different distributions of signal energy as a function of depth. This will result in the ratio of the average power at the reference sensor to the norm of the cross-correlation vector varying with different environmental conditions. Therefore, the normalized cross-correlation vectors (replica vectors) for different environmental conditions may have different norms.

To understand the effect that replica vectors with different norms may have on the adaptive replica vector calculation problem, note that (3.5) can be expressed as

$$\underline{q}_{eff} = \arg \min_{\underline{q}\in\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)} \mid \underline{q} \mid^2 \left[\frac{\underline{q}^h \, \mathrm{S}_n(f)^{-1} \, \underline{q}}{\mid \underline{q} \mid^2} \left(\frac{(S/N)_{max}+1}{(S/N)_{max}} - \cos^2(\underline{q}_{act}, \underline{q}; \mathrm{S}_n(f)^{-1})\right)\right]. \tag{3.6}$$

The factor in the brackets [ ] is independent of the norm of $\underline{q}$, depending upon only the direction of $\underline{q}$. The processor will jointly adjust the norm and the direction

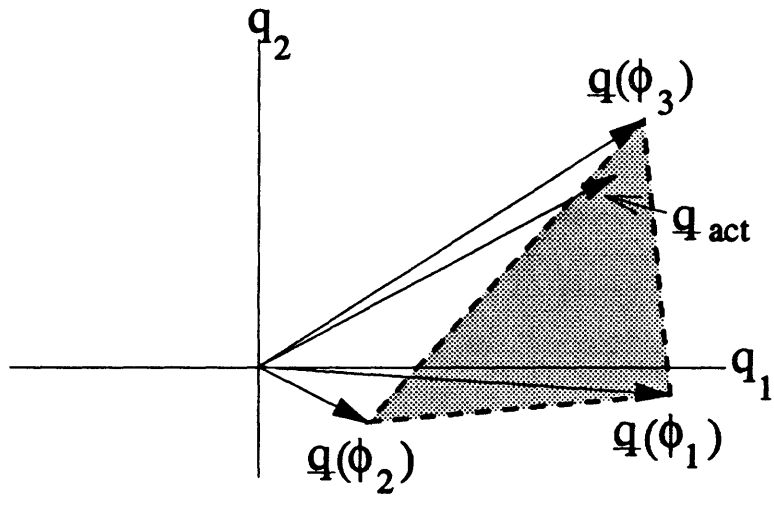


Figure 3-2: The Convex Hull of Replicas with Different Norms

of $\underline{q}$ within the constraints imposed by $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$ to balance the minimization of the norm-squared of $\underline{q}$ and the factor in the brackets, respectively. Consider the case shown in Figure 3-2, where $\underline{q} \in \mathbb{R}^2$ and the norm of one replica vector in $\tilde{Q}$ is significantly less than the norm of the other replica vectors in $\tilde{Q}$. In this case, the adaptive replica vector calculator will tend to select $\underline{q}_{eff}$ to lie in the vicinity of $\underline{q}(\phi_2)$, regardless of the direction of $\underline{q}_{act}$. This will greatly limit the adjustment of the direction of $\underline{q}_{eff}$ reducing the ability of the processor to adapt to different actual replica vectors, and will cause mismatch between $\underline{q}_{eff}$ and $\underline{q}_{act}$.

The array gain (3.4) is a function of the direction of the array weight vector rather than its norm. The direction of the array weight vector (Step 3 of the algorithm in Subsection 3.1.1) is a function of the direction of the $\underline{q}_{eff}$ but not its norm. Therefore, the array gain is a function of the direction of the $\underline{q}_{eff}$ rather than its norm. Therefore, it is desirable for the processor to adjust the direction of $\underline{q}_{eff}$ without regard to the resulting norm of $\underline{q}_{eff}$. The ability to adjust $\underline{q}_{eff}$ is improved significantly by modifying the definition of the replica vector so that the replica vectors have a norm of one. That is, define the replica vector as

$$\underline{q}(f,\underline{z},\underline{\phi}) \triangleq \mathrm{E}[\underline{X}(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}] / \mid \mathrm{E}[\underline{X}(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}] \mid .$$

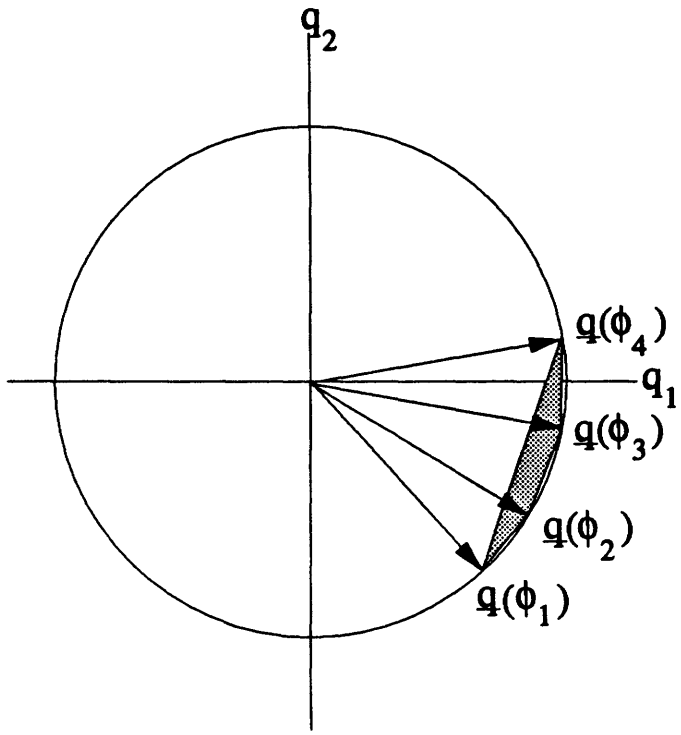


Figure 3-3: The Convex Hull of Replicas with Unit Norms

As a result, $\sigma^2(f, \underline{z})$ no longer has the strict interpretation as the average power in the desired signal as received at the reference sensor. However, in numerical simulations, this modification significantly increased the ability of the adaptive replica vector calculator to adjust $\underline{q}_{eff}$ to track $\underline{q}_{act}$ as the latter changed. There was a smaller mismatch between $\underline{q}_{eff}$ and $\underline{q}_{act}$ and an improved performance of the processor.

A second, related modification, which also improved the performance of the processor is motivated by considering Figure 3-3, which shows the convex hull of four replica vectors, all of which have been normalized to have a norm of one. Clearly, even though the vectors which define the convex hull (i.e., the vectors in $\tilde{Q}(f, \underline{z})$) have a norm of one, the convex hull contains vectors which have norms of less than one. However, from Subsection 3.1.2, the function of $\underline{q}_{eff}$ is to define a rank one signal subspace in which the processor estimates the average power using the MVDR Matched Field Processor. Since the subspace spannedby $\underline{q}_{eff}$ is independent of the norm of $\underline{q}_{eff}$, the resulting estimated average power should be independent of the norm of $\underline{q}_{eff}$. This desired property is accomplished by normalizing $\underline{q}_{eff}$ to have a

norm of one. This normalization is implemented by modifying Steps 2 and 3 of the algorithm. The resulting algorithm is

1. Use the (modified) complementary pivot algorithm to calculate

$$\underline{p}_{1f} = \arg \min_{\substack{\underline{p} \geq \underline{0} \\ \underline{e}^t \underline{p} = 1}} \underline{p}^t \, Q^h(f, \underline{z}) \hat{\mathrm{S}}(f)^{-1} Q(f, \underline{z}) \, \underline{p} \, ,$$

   where all the columns of $Q(f, \underline{z})$ are normalized to have a norm of one.

2. $\hat{\sigma}^2(f, \underline{z}) = \frac{\underline{p}_{1f}^t \, Q^h(f,\underline{z}) \, Q(f,\underline{z}) \, \underline{p}_{1f}}{\underline{p}_{1f}^t \, Q^h(f,\underline{z}) \hat{\mathrm{S}}(f)^{-1} Q(f,\underline{z}) \, \underline{p}_{1f}}$

3. $\underline{w}_{opt}(f, \underline{z}, \hat{\sigma}^2(f, \underline{z})) = \frac{\hat{\sigma}^2(f,\underline{z}) \, \hat{\mathrm{S}}(f)^{-1} \, Q(f,\underline{z}) \, \underline{p}_{1f}}{(\underline{p}_{1f}^t \, Q^h(f,\underline{z}) \, Q(f,\underline{z}) \, \underline{p}_{1f})^{\frac{1}{2}}}$

As with the two modifications detailed previously in this subsection, $Q^h(f, \underline{z}) \hat{\mathrm{S}}(f)^{-1} Q(f, \underline{z})$ can be replaced by $\mathrm{Real}(Q^h(f, \underline{z}) \hat{\mathrm{S}}(f)^{-1} Q(f, \underline{z}))$ in Steps 1 and 2 of the algorithm. For the same reason, $Q^h(f, \underline{z}) \, Q(f, \underline{z})$ can be replaced by $\mathrm{Real}(Q^h(f, \underline{z}) \, Q(f, \underline{z}))$ in Steps 2 and 3.

The preceding two modifications do not entirely eliminate the norm of $\underline{q}_{eff}$ as a factor in selecting the direction of the effective replica vector (i.e. $\frac{\underline{q}_{eff}}{|\underline{q}_{eff}|}$). Referring to Figure 3-3, it is clear that even though the replica vectors in $\tilde{Q}(f, \underline{z})$ have unit norm, vectors with norms less than one will be contained in $\mathcal{H}\left(\tilde{Q}(f, \underline{z})\right)$. Therefore, while the effect of vector norms is greatly reduced by normalizing the replica vectors, the selection of $\underline{q}_{eff}$ according to (3.6) will be biased towards the minimum norm area of the convex hull. This is verified numerically in Subsection 3.2.1.

3.1.4 The Range and Sampling of the Environmental Parameter Set Φ

In Section 2.3, a method of solving the minmax problem in Step 1 of the algorithm was developed which required the range of environmental parameters over which the processor is designed to operate to be sampled to yield a finite set $\Phi = \{\underline{\phi}_1, \ldots, \underline{\phi}_K\}$. The interpretation of Step 1 of the processor as an adaptive signal replica calculator

which selects $\underline{q}_{eff} \in \mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$ where $\tilde{Q}(f,\underline{z}) \triangleq \{\underline{q}(f,\underline{z},\underline{\phi}_i) \mid i = 1,\ldots,K\}$ provides some insight into the desired location of the samples of the operating range which are included in the finite set Φ. Using the term **extreme point** to refer to any point in a set which is not representable as the convex combination of two other points in the set, a convex hull is completely defined by its extreme points. Therefore, nothing is gained by including a point in Φ for which the corresponding replica vector is not an extreme point of $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$. Referring to Figure 3-2, $\underline{q}(\phi_1)$, $\underline{q}(\phi_2)$, and $\underline{q}(\phi_3)$ are the extreme points of the convex hull.

Referring to Figure 3-3, suppose that the operating range of the processor includes the continuous set of points whose replica vectors fill the continuous arc on the circle between $\underline{q}(\phi_1)$ and $\underline{q}(\phi_4)$. Then, the convex hull of these replica vectors includes the region bounded by the arc between $\underline{q}(\phi_1)$ and $\underline{q}(\phi_4)$ and the chord subtended by that arc (i.e., the straight line connecting $\underline{q}(\phi_1)$ and $\underline{q}(\phi_4)$). This convex hull can be approximated closely by including several environmental points in Φ whose corresponding replica vectors are spaced evenly along the arc between $\underline{q}(\phi_1)$ and $\underline{q}(\phi_4)$ (ϕ_2 and ϕ_3 in the example shown). However, with further analysis, it can be seen that this step is unnecessary.

Any point in the convex hull shown in Figure 3-3 which does not fall on the chord connecting $\underline{q}(\phi_1)$ and $\underline{q}(\phi_4)$ can be expressed as a point on that chord multiplied by a real number greater than one. Therefore, referring to (3.6), the term in the brackets will be the same for both the point not on the chord and the corresponding point on the chord. However, the norm-squared term will be smaller for the point on the chord. Therefore, $\underline{q}_{eff}$ will always lie on the chord connecting $\underline{q}(\phi_1)$ and $\underline{q}(\phi_4)$. The solution from within the convex hull of the continuous set of replica vectors will therefore always fall within $\mathcal{H}\left(\{\underline{q}(\phi_1), \underline{q}(\phi_4)\}\right)$. Thus, no performance will be lost if $\Phi = \{\phi_1, \phi_4\}$.

This analysis can be extended to higher dimension replica vectors to determine the number and the location of the samples of the continuous environmental parameter set required to insure that the sampling process does not result in a loss in performance. The requirement which must be met is that the replica vectors at the selected samples

should define a convex hull which is a subset of the convex hull of the continuous set of replica vectors and which contains a set of points which is guaranteed to contain $\underline{q}_{eff}$ (such as the chord in the two-dimensional example above). One problem which may occur in higher dimensions but does not occur in two-dimensions is that the required convex hull may have an infinite number of extreme points. Therefore, it can only be approximated by the convex hull of a finite set of replica vectors. If the actual replica vector falls outside the approximating convex hull, then mismatch between $\underline{q}_{eff}$ and $\underline{q}_{act}$ will occur. The resulting MVDR processor will tend to reject (filter out) the desired signal and will suffer the same performance degradation problems which are characteristic of the MVDR Matched Field Processor in the presence of environmental mismatch [13, 17]. The significance of this degradation will depend on the extent to which $\underline{q}_{act}$ falls outside of $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$. Numerical results illustrating the dependence of processor performance on the sampling density of the set Φ and the performance degradation which results when $\underline{q}_{act}$ falls outside of $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$ are presented in Subsection 3.2.2.

The performance degradaton which will occur if $\underline{q}_{act}$ falls outside of $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$ highlights a conflicting requirement in selecting the range of the environmental parameters over which the processor is designed to operate. As described in Subsection 3.1.2, the processor will treat any portion of the received signal which falls within any rank one subspace spanned by a replica vector in $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$ as desired signal, and will select $\underline{q}_{eff}$ to pass as much of this signal as possible through the resulting spatial filter. If Φ is large, there will tend to be a wide range of rank one subspaces spanned by the replica vectors in $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$. In this case, the processor may select $\underline{q}_{eff}$ to lie in a noisy subspace which contains very little of the power of the desired signal; and the spatial filter will pass a lot of noise and little of the desired signal. If Φ is small, then $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$ will tend to be small and the probability that the actual replica of the desired signal will not fall within $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$ will increase. In this case, mismatch will occur and, as stated in the preceding paragraph, the tendency of the processor will be to reject (filter out) the desired signal resulting in performance degradation. Therefore, it is important to balance the selection of a large operating

range to avoid filtering out the desired signal with the selection of a small operating range to avoid passing a lot of noise through the filter.

The tradeoff in the selection of Φ between filtering out as much noise as possible and avoiding the rejection of the desired signal can be analyzed using the probability of false alarm (P_F) and probability of detection (P_D) performance measures. Assume that a fourth stage is added to the Adaptive Minmax Processor to perform a threshold hypothesis test to estimate whether or not a point source is present at the array focal point. That is, letting H_0 be the hypothesis that there is no point source at the array focal point (i.e., the received signal consists of just noise) and H_1 be the hypothesis that there is a point source at the array focal point (i.e., the received signal consists of the desired signal plus noise), the hypothesis test selects between the two hypotheses using the decision rule

$$H_0: \ \hat{\sigma}^2(f, \underline{z}) < \eta$$
$$H_1: \ \hat{\sigma}^2(f, \underline{z}) > \eta$$

where η is the decision threshold. P_F is defined as the probability that the test will select H_1 when no point source is present at the array focal point and P_D is defined as the probability that the test will select H_1 when there is a point source present at the array focal point.

$\hat{\sigma}^2(f, \underline{z})$ is calculated with the following two steps:

$$\underline{q}_{eff} = \arg \min_{\underline{q} \in \mathcal{H}(Q(f, \underline{z}))} \underline{q}^h \hat{\mathrm{S}}(f)^{-1} \underline{q}$$

and

$$\hat{\sigma}^2(f, \underline{z}) = \frac{\underline{q}^h_{eff} \underline{q}_{eff}}{\underline{q}^h_{eff} \hat{\mathrm{S}}(f)^{-1} \underline{q}_{eff}}.$$

Let S_n be the noise cross-spectral covariance matrix and $\mathrm{S}_{ds}(\underline{\phi})$ be the cross-spectral covariance matrix of the desired signal given the environmental conditions $\underline{\phi}$. Then,

P_F and P_D can be written as

$$P_F = \text{Prob}\left(\frac{\underline{q}_{eff}^h \underline{q}_{eff}}{\underline{q}_{eff}^h \hat{\text{S}}(f)^{-1} \underline{q}_{eff}} > \eta \mid \text{S}(f) = \text{S}_n\right)$$

$$P_D = \text{Prob}\left(\frac{\underline{q}_{eff}^h \underline{q}_{eff}}{\underline{q}_{eff}^h \hat{\text{S}}(f)^{-1} \underline{q}_{eff}} > \eta \mid \text{S}(f) = \text{S}_n + \text{S}_{ds}(\underline{\phi})\right)$$

$\mid \underline{q} \mid = 1 \; \forall \underline{q} \in Q(f, \underline{z})$ implies that $\mid \underline{q} \mid \leq 1 \; \forall \underline{q} \in \mathcal{H}(Q(f, \underline{z}))$. Therefore, defining

$$\mid \underline{q} \mid_{min}^2 \triangleq \min_{\underline{q} \in \mathcal{H}(Q(f,\underline{z}))} \mid \underline{q} \mid^2,$$

the following relationship holds $\forall \underline{q}_{eff}$.

$$\frac{\mid \underline{q} \mid_{min}^2}{\underline{q}_{eff}^h \hat{\text{S}}(f)^{-1} \underline{q}_{eff}} \leq \frac{\underline{q}_{eff}^h \underline{q}_{eff}}{\underline{q}_{eff}^h \hat{\text{S}}(f)^{-1} \underline{q}_{eff}} \leq \frac{1}{\underline{q}_{eff}^h \hat{\text{S}}(f)^{-1} \underline{q}_{eff}}.$$

Therefore, P_F and P_D can be bounded by

$$P_F \leq P_{F_{ub}} \triangleq \text{Prob}\left(\exists \underline{q} \in \mathcal{H}(Q(f, \underline{z})) \text{ s.t. } \underline{q}^h \hat{\text{S}}(f)^{-1} \underline{q} < \eta^{-1} \mid \text{S}(f) = \text{S}_n\right)$$

$$P_F \geq P_{F_{lb}} \triangleq \text{Prob}\left(\exists \underline{q} \in \mathcal{H}(Q(f, \underline{z})) \text{ s.t. } \underline{q}^h \hat{\text{S}}(f)^{-1} \underline{q} < \mid \underline{q} \mid_{min}^2 \eta^{-1} \mid \text{S}(f) = \text{S}_n\right)$$

$$P_D \leq P_{D_{ub}} \triangleq \text{Prob}\left(\exists \underline{q} \in \mathcal{H}(Q(f, \underline{z})) \text{ s.t. } \underline{q}^h \hat{\text{S}}(f)^{-1} \underline{q} < \eta^{-1} \mid \text{S}(f) = \text{S}_n + \text{S}_{ds}(\underline{\phi})\right)$$

$$P_D \geq P_{D_{lb}} \triangleq \text{Prob}\left(\exists \underline{q} \in \mathcal{H}(Q(f, \underline{z})) \text{ s.t. } \underline{q}^h \hat{\text{S}}(f)^{-1} \underline{q} < \mid \underline{q} \mid_{min}^2 \eta^{-1} \mid \text{S}(f) = \text{S}_n + \text{S}_{ds}(\underline{\phi})\right)$$

Let Φ, $Q(f, \underline{z})$, $\mid \underline{q} \mid_{min}^2$, $P_{F_{ub}}$, $P_{F_{lb}}$, $P_{D_{ub}}$, $P_{D_{lb}}$ and Φ', $Q'(f, \underline{z})$, $\mid \underline{q} \mid_{min}^{2'}$, $P'_{F_{ub}}$, $P'_{F_{lb}}$, $P'_{D_{ub}}$, $P'_{D_{lb}}$ be two sets of operating ranges of environmental parameters, replica vectors, and associated bounds. Assume that $\Phi \subset \Phi'$. Then $\mathcal{H}(Q(f, \underline{z})) \subseteq \mathcal{H}(Q'(f, \underline{z}))$ and $\mid \underline{q} \mid_{min}^2 \geq \mid \underline{q} \mid_{min}^{2'}$. Therefore, $P_{F_{ub}} \leq P'_{F_{ub}}$ and $P_{D_{ub}} \leq P'_{D_{ub}}$. However, since $\mid \underline{q} \mid_{min}^2 \geq \mid \underline{q} \mid_{min}^{2'}$, the same conclusion cannot be drawn about the corresponding lower bounds. Therefore, as the operating range of environmental parameters is increased, the upper bounds on the probabilities of false alarm and detection are increased. However, as stated at the end of Subsection 3.1.3, if Φ is increased to the point that $\mid \underline{q} \mid_{min}^2$ becomes small, the mismatch between $\underline{q}_{eff}$ and $\underline{q}_{act}$ will tend to increase when

the SNR is low. This increased mismatch will result in a reduction in ambiguity function of the processor at the source location and therefore reduce the probability of detection. This analysis is supported by the numerical results shown in Figure 3-16 in Subsection 3.2.1. Therefore, at low SNRs, the probability of detection is not likely to increase as much as would be inferred from the increase in $P_{D_{wb}}$ as Φ is increased.

3.2 Numerical Analysis of the Adaptive Minmax Matched Field Processor

The analysis of many characteristics of the Adaptive Minmax Matched Field Processor is improved considerably by the inclusion of numerical results. A principal reason for this is the lack of illustrative analytical solutions to Step 1 of the array processing algorithm. Several numerical examples are included here to assist in the analysis of different characteristics of the algorithm. Subsection 3.2.1 presents results characterizing the performance of the algorithm in a deterministic ideal waveguide. In this case, the uncertain environmental characteristic is the depth of the waveguide. Subsection 3.2.2 presents results which characterize the performance of the algorithm in a deterministic horizontally-stratified ocean with an arctic sound speed profile. Here, the uncertain environmental characteristic is the sound speed profile in the surface layer of the ocean. Finally, Subsection 3.2.3 presents results describing the performance of the algorithm in a randomly time-variant ideal waveguide where, as before, the uncertain environmental characteristic is the depth of the waveguide. In this case, the random perturbations to the ocean are perturbations to the sound speed structure as presented in Chapter 4. These results are presented to allow comparison of the performance of the processor in deterministic and random media.

In all the numerical simulations, an important parameter is the signal to noise ratio. For the purpose of these simulations, the signal to noise ratio is defined as the ratio of the average power per array sensor in the desired signal to the average power per array sensor in the spatially white sensor noise. For the case where there is propagating background noise, the background noise to sensor noise ratio is similarly

defined. Thus, if the SNR is 20 dB and the background noise to sensor noise ratio is 10 dB, then the signal to background noise SNR is 10 dB. In all cases, the average power measurements are taken **after** the temporal processing by the discrete-time Fourier transform.

None of the results presented were generated with Monte Carlo simulations. In all cases, it is assumed that the processor has perfect knowledge of the cross-spectral correlation matrix. That is, $\hat{\mathrm{S}}(f) = \mathrm{S}(f)$. All units of distance measurement used in this section are metric.

3.2.1 The Deterministic Ideal Waveguide

The first set of numerical results were generated using an ideal waveguide model of the ocean and a normal mode representation of the propagating signals. (See Section 4.3 for an introduction to the normal mode representation of signals). In addition, the ocean is assumed to be a deterministic (i.e., time-invariant) medium. The salient parameters of the simulation are that the ocean is assumed to be an isovelocity waveguide with a sound speed of $\overline{C}(z) = 1500\ m/s$ where $\overline{C}(z)$ is the time-invariant depth dependent component of the sound speed defined in Chapter 4, and an unknown but constant depth H between 290 meters and 310 meters. The sea surface is assumed to be a free surface and the ocean bottom is assumed to be infinitely rigid. The source is assumed to be a deterministic complex exponential with a frequency of 20 Hz. In addition, the source is assumed to be sufficiently far from the array so that the horizontal propagation of the modes can be modeled by a complex exponential rather than a Hankel function.

With these assumptions, the waveguide supports eight propagating modes. The mode shapes are a function of depth z and given by

$$\phi_n(z) = \sin(k_{V_n} z) \quad n = 1, \ldots, 8,$$

where $k_{V_n} = \frac{(2n-1)\pi}{2H}$ is the vertical wavenumber of the n^{th} mode. The mode shape $\phi_n(z)$ should not be confused with the environmental parameter $\underline{\phi}_n$. The propagating

signal at the range r, depth z, and time t is given by (4.12)

$$x(r,z,t) = e^{j2\pi ft} \sum_{n=1}^{8} (k_{H_n} r)^{-\frac{1}{2}} e^{-jk_{H_n} r} \sin(k_{V_n} z_{source}) \sin(k_{V_n} z),$$

where $k_{H_n} = \sqrt{(k^2 - k_{V_n}^2)}$ is the horizontal wavenumber of the n^{th} mode and $k = \frac{2\pi f}{C}$ is the wavenumber of the signal.

The receiving array is a nine-element vertical array with the top element at 30 meters depth and the bottom element at 270 meters depth. The inter-element spacing is therefore 30 meters. For all the examples in this subsection, the sources are assumed to be at a range of 50 km from the array and there is assumed to be no propagating background noise (i.e., the received signal consists of spatially white sensor noise and narrowband signals emitted by point sources). In addition, in this section, the localization problem is assumed to be a one-dimensional (depth only) problem. That is, it is assumed that the true range to the source(s) is known and the processor is trying to determine the source depth(s). These assumptions are made to simplify the presentation of the results and the analysis of the salient characteristics of the processor.

For the first results in this subsection, a single source is placed at a depth of 150 meters with an SNR of 10 dB. The purpose of these results is to illustrate the effect which the ocean depth uncertainty has on the MVDR Matched Field Processor, the Bartlett Matched Field Processor, and the Adaptive Minmax Matched Field Processor, and to provide a qualitative comparison of the resolution of each processor. Figures 3-4a through 3-8a show the depth ambiguity functions generated by each processor for the cases where the actual ocean depth is 290 meters; and Figures 3-4b through 3-8b show the depth ambiguity functions generated by each processor for the cases where the actual ocean depth is 310 meters.

The results in Figure 3-4 are generated using a MVDR Processor assuming that the ocean depth is 290 meters. When the ocean depth assumption is accurate (Figure 3-4a), the processor generates a sharp peak in its ambiguity function at the true source depth; when the depth assumption is inaccurate (Figure 3-4b), the peak level of the

ambiguity function drops by approximately 17 dB and there is no single significant peak. The sidelobe suppression of the perfectly matched processor (approximatly 18 dB below the main peak level) is good and the peak in the ambiguity function is sharp. Figure 3-5 shows the complementary results generated using a MVDR Processor assuming that the ocean depth is 310 meters.

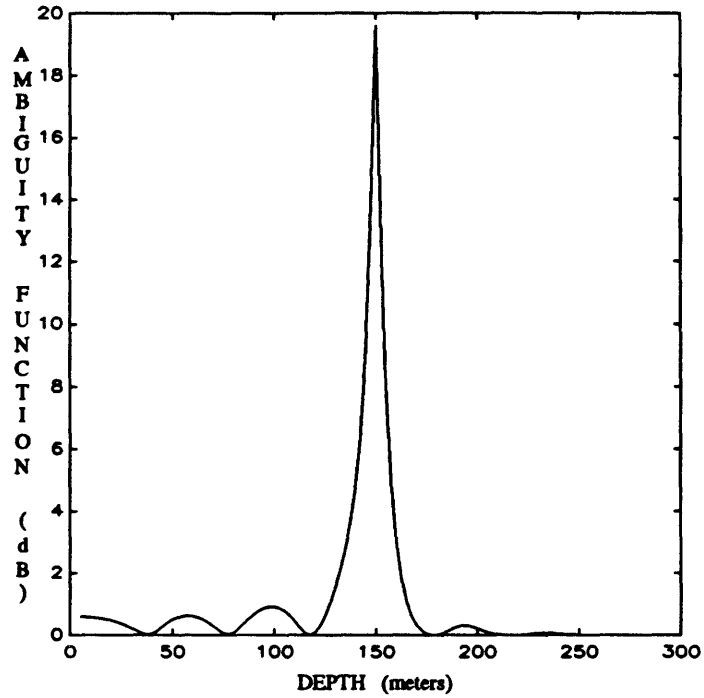
The results in Figure 3-6 are generated using a Bartlett Processor which assume that the ocean depth is 290 meters. As was the case with the MVDR Processor, the perfectly matched Bartlett Processor (Figure 3-6a) generates a single significant peak at the true source depth; the ambiguity function generated by the mismatched Bartlett Processor (Figure 3-6b) has no single significant peak. However, there are three significant differences in the ambiguity functions generated by the processors. First, the sidelobe suppression of the Bartlett Processor is not nearly as good as that of the MVDR Processor with the peak sidelobe level for the perfectly matched Bartlett Processor only 7 dB below the main peak level (compared to 18 dB for the MVDR Processor). Second, the mainlobe width for the perfectly matched Bartlett Processor is much wider than that generated by the MVDR Processor. Finally, the mismatched Bartlett Processor does not experience the significant loss in the peak level of the ambiguity function as was experienced by the mismatched MVDR Processor. These ambiguity function peaks, generated by the mismatched processors, are sidelobes in the source depth/ocean depth plane. Thus, the low peak levels in the ambiguity function of the mismatched MVDR Processor are characteristic of the low sidelobe levels of MVDR Processor and the high peak levels in the ambiguity function of the mismatched Bartlett Processor are characteristic of the high sidelobe levels of the Bartlett Processor. Figure 3-7 shows the complementary results generated by a Bartlett Processor which assumes that the ocean depth is 310 meters.

Finally, the results in Figure 3-8 are generated using the Adaptive Minmax Processor. In this case, the processor assumes that the ocean depth may be between 290 and 310 meters. With only this approximate knowledge of the ocean depth, the ambiguity function generated by the processor shows a single significant peak at the true source depth for the case where the ocean depth is 290 meters and the case where

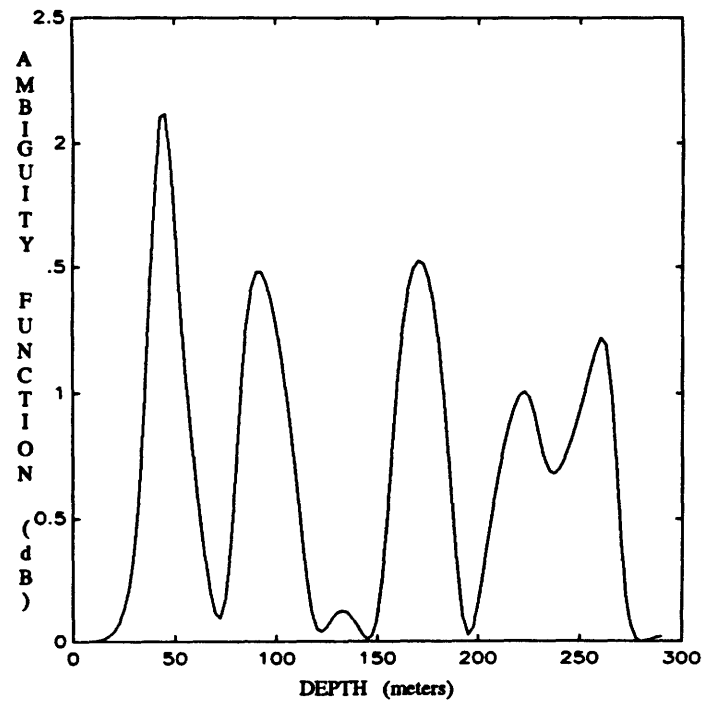
the ocean depth is 310 meters. The sidelobe suppression (approximately 15 dB) is comparable to that of the perfectly matched MVDR Processor and is significantly better than that of the perfectly matched Bartlett Processor. The same observation can be made about the sharpness of the mainlobe. This final observation leads to the qualitative conclusion that the resolution of the Adaptive Minmax Processor is comparable to that of the perfectly matched MVDR Processor and significantly better than than of the perfectly matched Bartlett Processor. This conclusion is supported quantitatively by results contained later in this subsection..

Figure 3-9 shows the ambiguity functions of the perfectly matched MVDR and Bartlett Processors and the Adaptive Minmax Processor for the case where the actual ocean depth is 310 meters, the source is located at 150 meters depth, and the SNR is 0 dB. These results are presented to illustrate that, in general, the ambiguity functions do not change qualitatively when the SNR is lowereed to 0 dB. One significant change is that the peak level of the ambiguity function of the Adaptive Minmax Processor is no longer equal to that of the perfectly matched MVDR and Bartlett Processors as is the case when the SNR is 10 dB. The peak level in this case has dropped to about 2.5 dB below that of the MVDR and Bartlett Processors. This loss in peak level will be analyzed later in this subsection.

As demonstrated with the preceding results, the response of the Adaptive Minmax Processor in the environmental parameter space (in this case ocean depth) is considerably broader than that of the MVDR Processor. That is, the Adaptive Minmax Processor will detect signals from a wider range of environmental parameters than the MVDR Processor. Qualitatively, the signal replica vector is a relatively smooth function of both the source location and the environmental conditions. Therefore, since both processors characterize signals by their replica vectors, it is reasonable to conclude that the response of the Adaptive Minmax Processor in the source location space would be broader than that of the MVDR Processor. The Multiple Constraints Matched Field Processor [17] makes use of this fact and broadens the response of the processor in the environmental parameter space by intentionally broadening the response of the processor in the source location space. The result is a reduction in

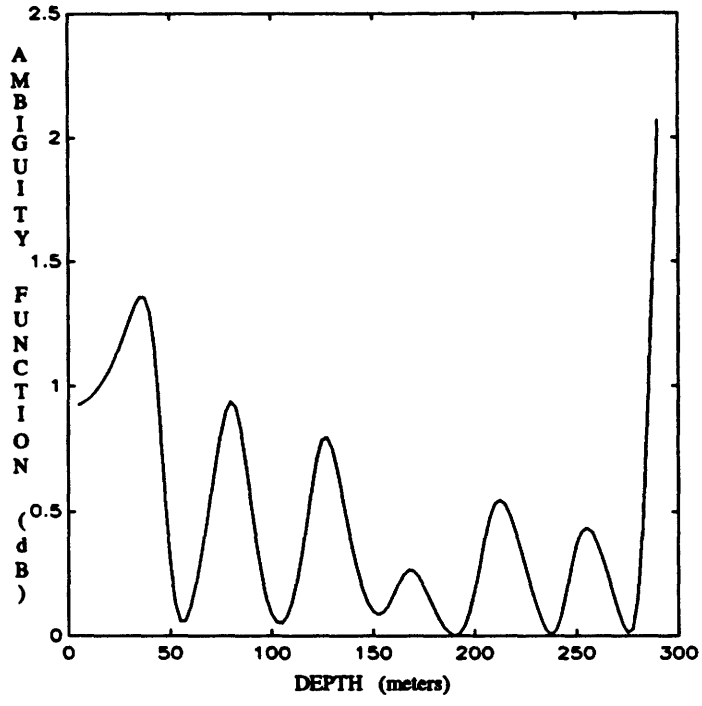


(a) Actual Ocean Depth = 290 meters

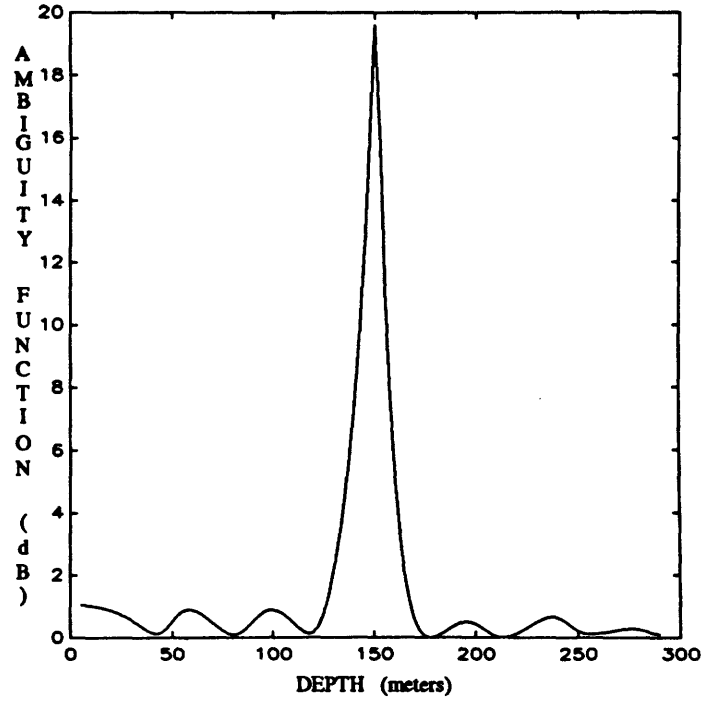


(b) Actual Ocean Depth = 310 meters

Figure 3-4: Ambiguity Function for the MVDR Processor: SNR = 10 dB
Assumed Ocean Depth = 290 meters

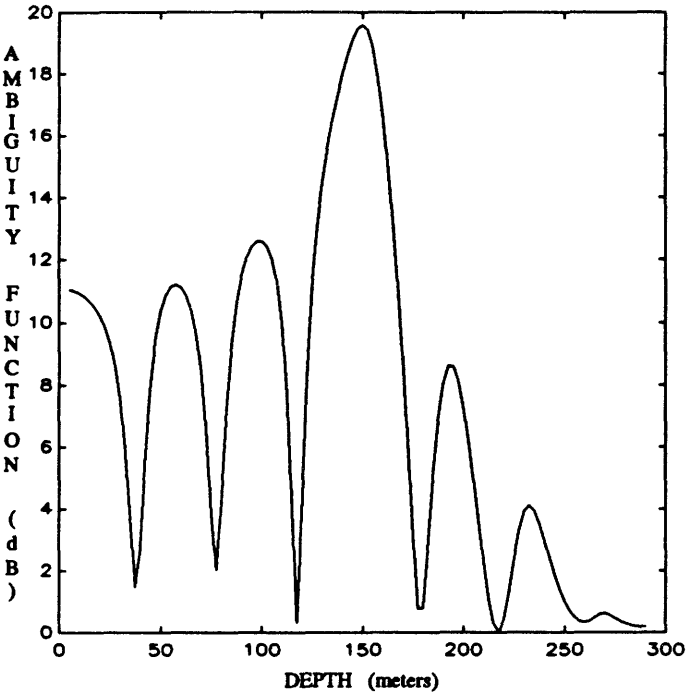


(a) Actual Ocean Depth = 290 meters

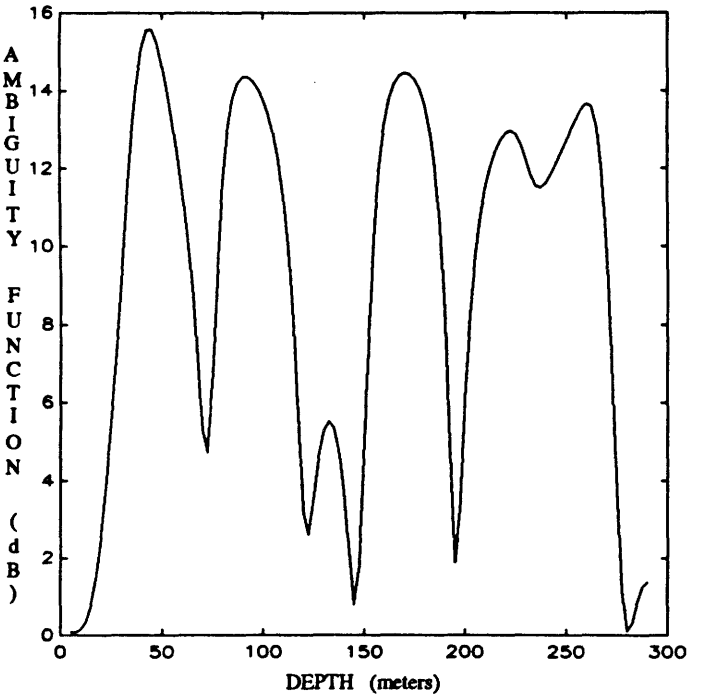


(b) Actual Ocean Depth = 310 meters

Figure 3-5: Ambiguity Function for the MVDR Processor: SNR = 10 dB
Assumed Ocean Depth = 310 meters

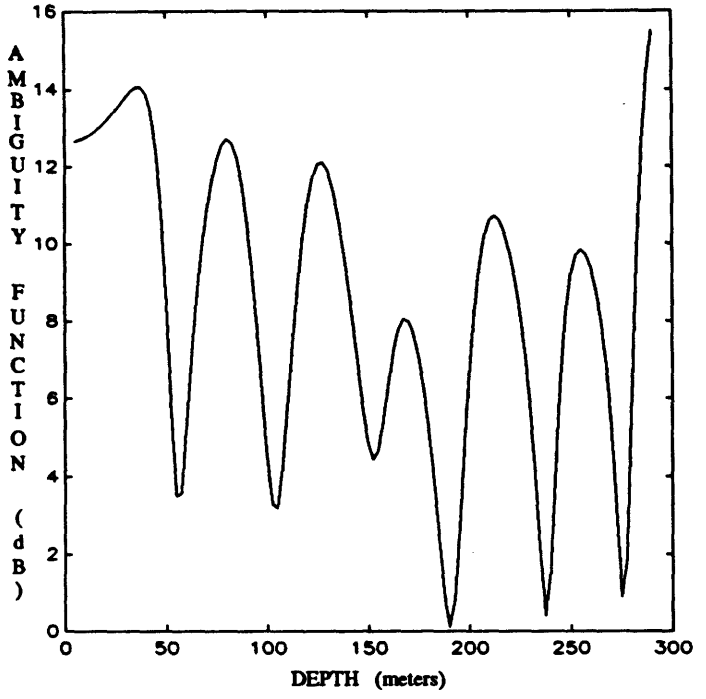


(a) Actual Ocean Depth = 290 meters

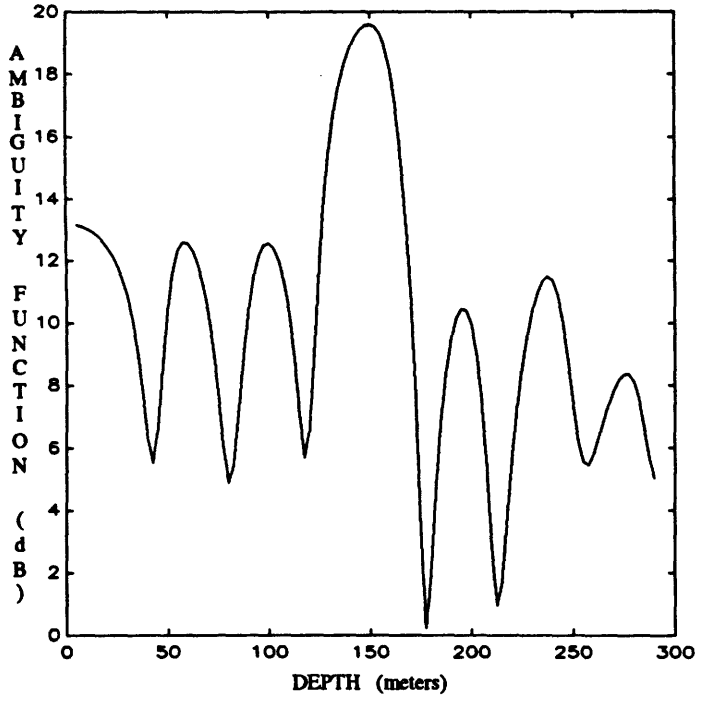


(b) Actual Ocean Depth = 310 meters

Figure 3-6: Ambiguity Function for the Bartlett Processor: SNR = 10 dB
Assumed Ocean Depth = 290 meters

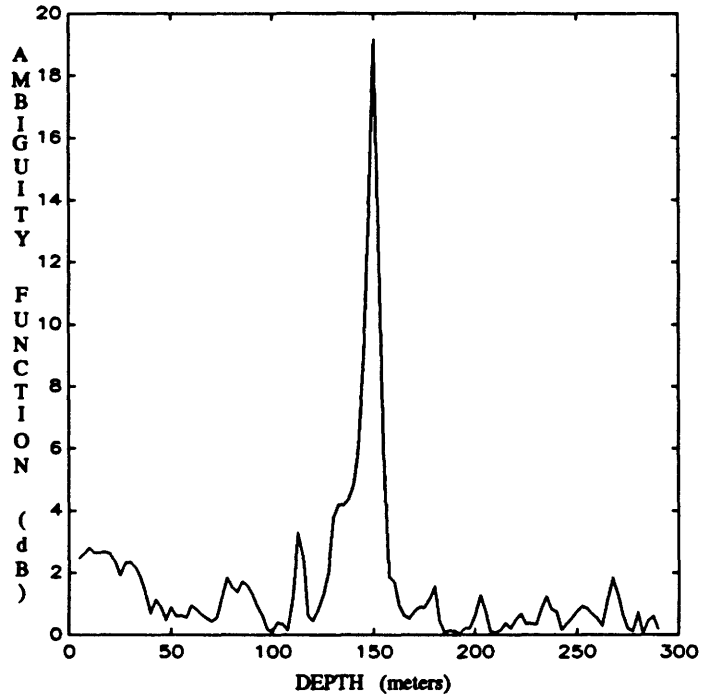


(a) Actual Ocean Depth = 290 meters

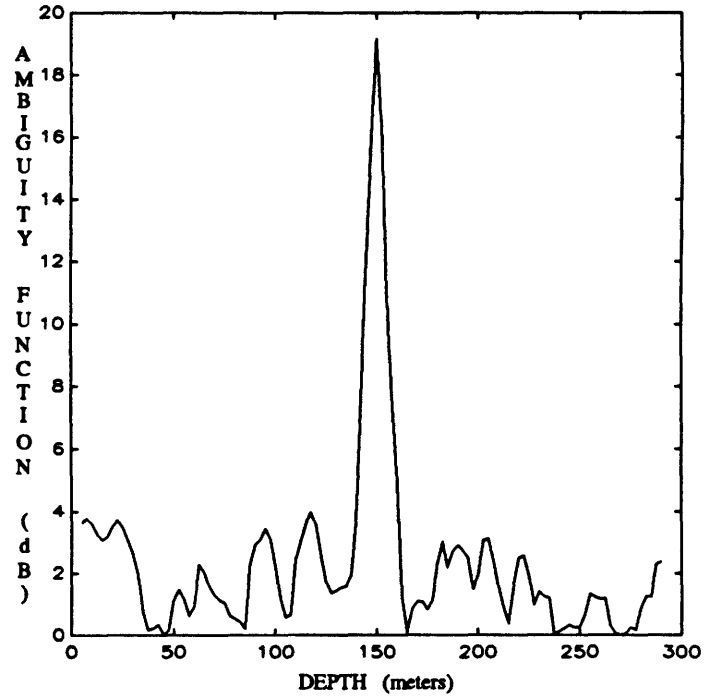


(b) Actual Ocean Depth = 310 meters

Figure 3-7: Ambiguity Function for the Bartlett Processor: SNR = 10 dB
Assumed Ocean Depth = 310 meters

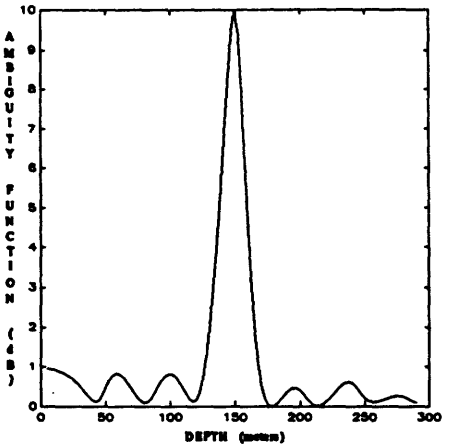


(a) Actual Ocean Depth = 290 meters

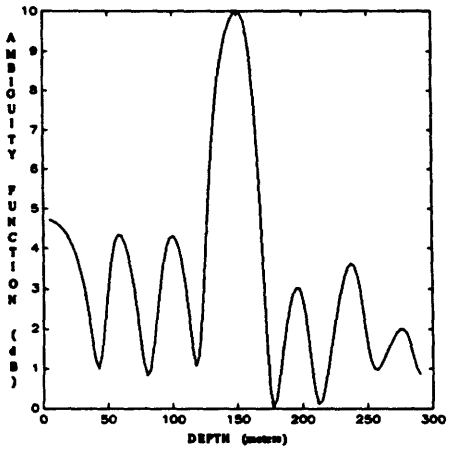


(b) Actual Ocean Depth = 310 meters

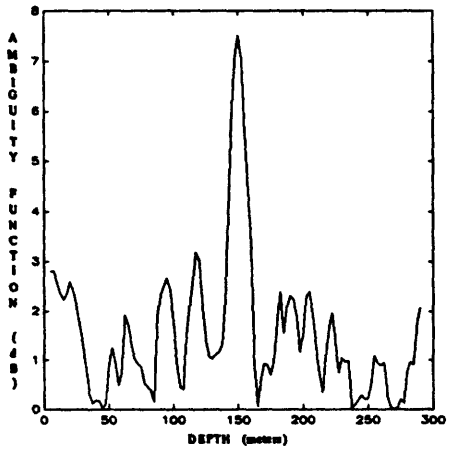
Figure 3-8: Ambiguity Function for the Minmax Processor: SNR = 10 dB
Assumed Range of Ocean Depths = 290 to 310 meters



(a) Perfectly Matched MVDR Processor



(b) Perfectly Matched Bartlett Processor



(c) Adaptive Minmax Processor

Figure 3-9: Ambiguity Functions for SNR = 0 dB and Ocean Depth = 310 meters

the sensitivity of the processor to environmental mismatch and a corresponding reduction in the spatial resolution of the processor. Therefore, an issue of concern is how much spatial resolution must be sacrificed to create the broad response of the Adaptive Minmax Processor in the environmental parameter space.

The next results are presented to allow quantitative comparison of the resolution of the perfectly matched MVDR Processor, the perfectly matched Bartlett Processor, and the Adaptive Minmax Processor. The actual ocean depth for this simulation is 310 meters. Two equal strength sources, each with a 10 dB SNR, are located at a depth of 155 $\pm$ Δ_{depth} meters where $2\Delta_{depth}$ is the depth separation between the sources. Characteristic ambiguity functions generated by the Adaptive Minmax Processor for this two-source case are shown in Figure 3-10. In Figure 3-10a, at a source separation is 8 meters, the processor is unable to resolve the two sources. In Figure 3-10b, at a source separation is 24 meters, the processor is able to resolve the sources.

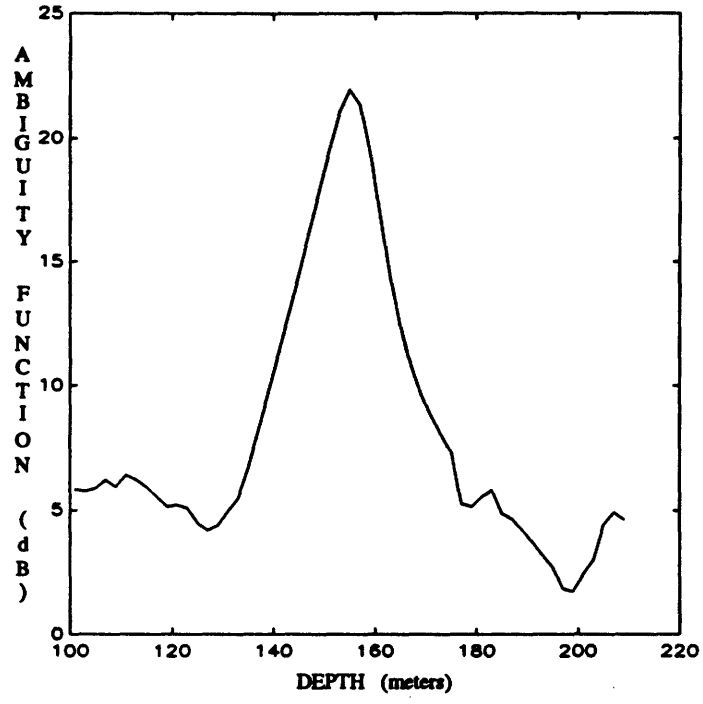
As a measure of a processor's ability to resolve sources, the response ratio, defined as $10\log_{10}\left(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_s^2}\right)$, is used. For the case shown in Figure 3-10a, where there is a single peak in the ambiguity function between the two source locations; $\hat{\sigma}_m^2$ is the peak value of the ambiguity function between the two source locations. For the case shown in Figure 3-10b, where the processor resolves two distinct peaks in the ambiguity function; $\hat{\sigma}_m^2$ is the minimal value of the ambiguity function between the two source locations. For the case shown in Figure 3-10a, where there is a single peak in the ambiguity function between the two source locations; $\hat{\sigma}_s^2$ is the smaller of the values of the ambiguity function evaluated at the two source locations. For the case shown in Figure 3-10b, where the processor resolves two distinct peaks in the ambiguity function; $\hat{\sigma}_s^2$ is the smaller of the two peak values. Therefore, if the response ratio is greater than zero, the processor is unable to resolve separate peaks in the ambiguity function. If the response ratio is less than zero, the processor resolves two distinct peaks in the ambiguity function and the depth of the dip in the ambiguity function between the peaks equals the response ratio in dB. Thus, the lower the response ratio at any source separation, the better the ability of the processor to resolve sources

with that separation.

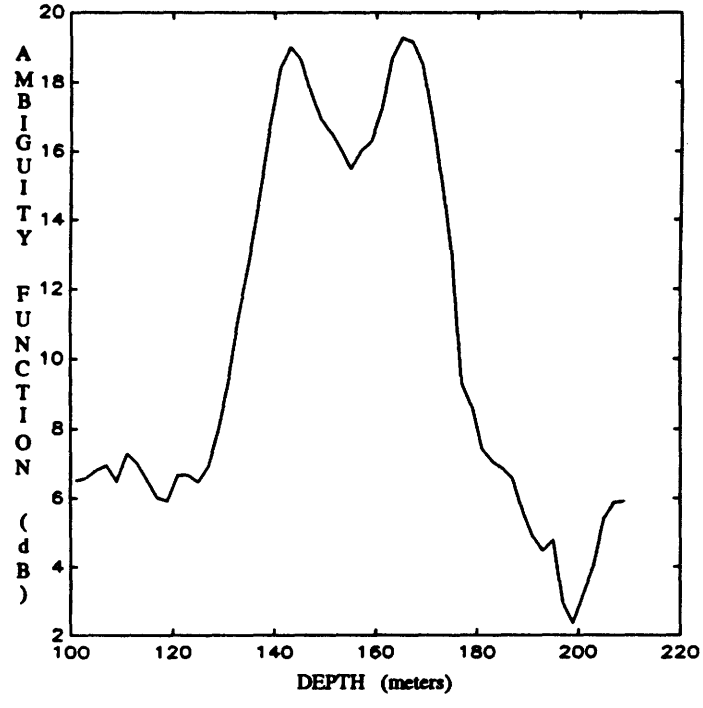
Referring to Figure 3-11, the MVDR and Adaptive Minmax Processors are able to resolve separate peaks for source separations greater than 17 and 18 meters, respectively, and the Bartlett Processor requires approximately 28 meters separation in order to be able to resolve separate peaks. The depth of the dip between source peaks is between .1 and 2 dB less for the Adaptive Minmax Processor than it is for the MVDR Processor. The depth of the dip for the Bartlett Processor is approximately 7 dB less than it is for the Adaptive Minmax and MVDR Processors. Thus, the resolution of the Adaptive Minmax Processor is slightly less than than that of the perfectly matched MVDR Processor but it is significantly greater than that of the perfectly matched Bartlett Processor.

The preceding results provide a means of assessing the capabilities of the Adaptive Minmax Processor. The following two sets of results illustrate characteristics of the processor in order to better understand how the processor functions. The second stage of the processor (Figure 2-4) is a linear spatial filter, the coefficients of which are calculated in Step 3 of the array processing algorithm. For each array focal point, a different set of weights are calculated. Just as a frequecy response is defined for any set of weights of a linear temporal filter, a source location/environmental condition response can be defined for each set of weights of the linear spatial filter. In the ideal waveguide example, where source location is parameterized by the source depth and the environmental conditions are parameterized by the ocean depth, the desired response is the source depth/ocean depth response. Letting $\underline{X}(z,H)$ be the snapshot of a signal emitted by a source at the depth z given the ocean depth H, the source depth/ocean depth response for the weights $\underline{w}$ is given by $W(z,H) \triangleq \frac{|\underline{w}^h \underline{X}(z,H)|}{|\underline{X}(z,H)|}$. This magnitude response is the gain which the spatial filter applies to a signal emitted by a source at depth z and which propagates through an ocean with depth H.

For the results shown in Figures 3-12 through 3-15, a single source is located at 250 meters depth with an SNR of 20 dB. Figure 3-12 shows the source depth/ocean depth response of the array weights generated by the Adaptive Minmax Processor when the array focal point is 175 meters depth and the actual ocean depth is 290 meters. The



(a) Source Depths = 151 and 159 meters



(b) Source Depths = 143 and 167 meters

Figure 3-10: Ambiguity Functions for the Minmax Processor: Two Sources

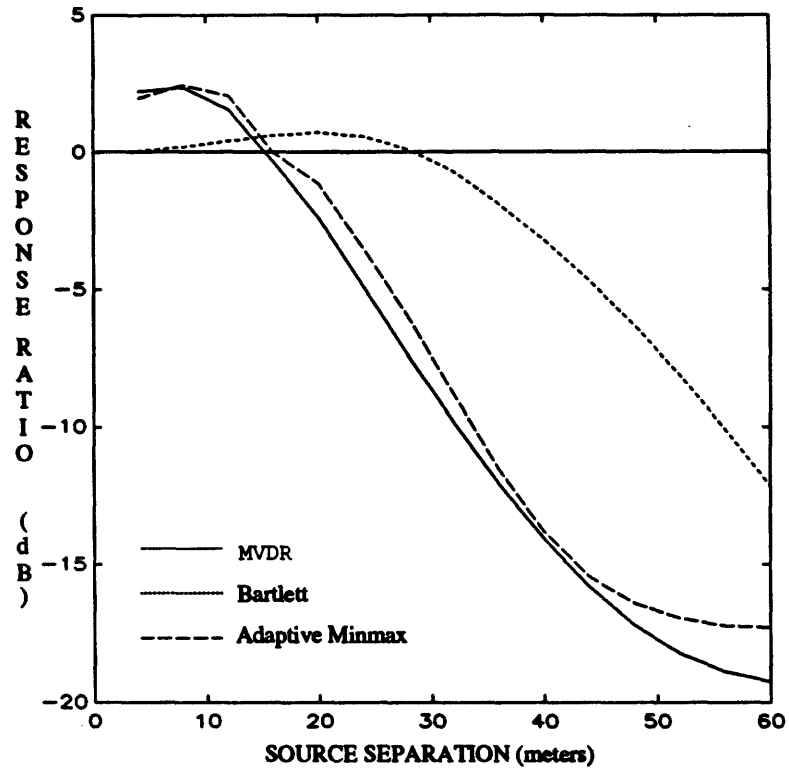


Figure 3-11: Resolution Comparison for Different Processors

magnitude response evaluated at an ocean depth of 290 meters (Figure 3-12a) has a sharp null at the source depth of 250 meters. This is consistent with the processor attempting to null the signal emitted by a source at the depth of 250 meters which propagates through an ocean of depth 290 meters. The magnitude response evaluated at an ocean depth of 310 meters (Figure 3-12b) has no sharp null at the source depth of 250 meters. This is consistent with the fact that the received signal has no component which propagated through an ocean for which depth was 310 meters. Hence, there is no signal with a replica vector characteristic of this point (250 meters source depth, 310 meters ocean depth) to cancel. Figure 3-13 shows a complementary set of results for the case where the actual ocean depth is 310 meters. Here, the sharp null appears in the response evaluated at an ocean depth of 310 meters but does not appear in the response evaluated at an ocean depth of 290 meters.

Figure 3-14 shows the source depth/ocean depth response of the array weights generated by the Adaptive Minmax Processor when the array focal point is 250 meters and the actual ocean depth is 290 meters. The response evaluated at an ocean depth

of 290 meters (Figure 3-14a) is close to one at the source depth of 250 meters. This is consistent with the interpretation of the spatial filter as a Two-Stage MVDR Processor where the filter response at the point corresponding to the effective replica vector equals one, and is also constitent with the conclusion in Subsection 3.1.2 that the effective replica vector will be very close to the actual replica vector at high SNRs.

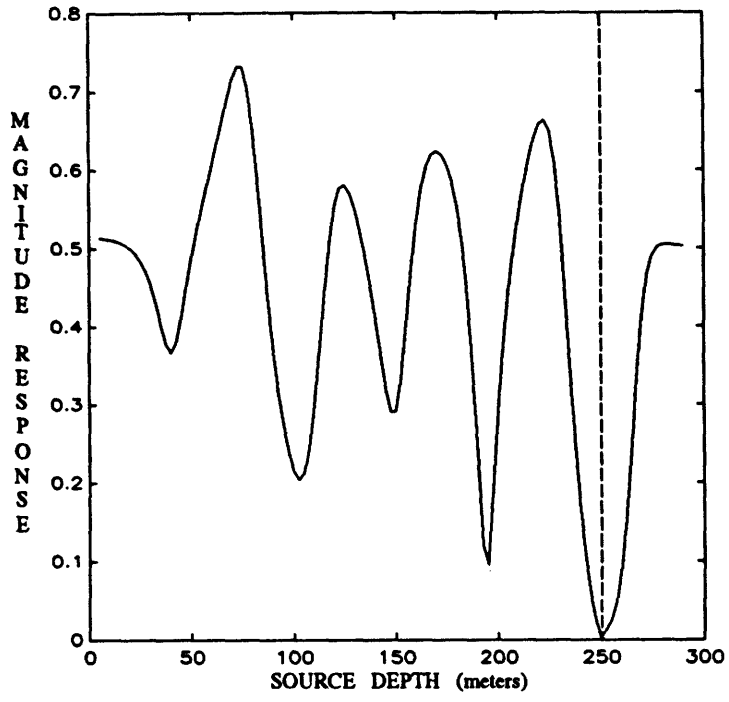
The response evaluated at an ocean depth of 310 meters (Figure 3-14b) is not close to one which is also consistent with the above interpretation. Since the received signal contains no component which propagated through an ocean of depth 310 meters, the signal power which the processor detects at this point is small. Thus $\underline{q}_{eff}$ is not forced to be near the replica vector for this point, and the response at this point is not constrained to be close to one. Figure 3-15 shows a complementary set of results for the case where the actual ocean depth is 310 meters.

The final results in this subsection allow an assessment of how the difference in the norms of the replicas contained in $\mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)$ can adversely affect the performance of the Adaptive Minmax Processor at low SNRs. These results were generated with a source at 250 meters depth and an actual ocean depth of 310 meters. The focal point of the Adaptive Minmax Processor is 250 meters depth. For this particular case, where the noise consists of only spatially white sensor noise, (3.6) can be rewritten as

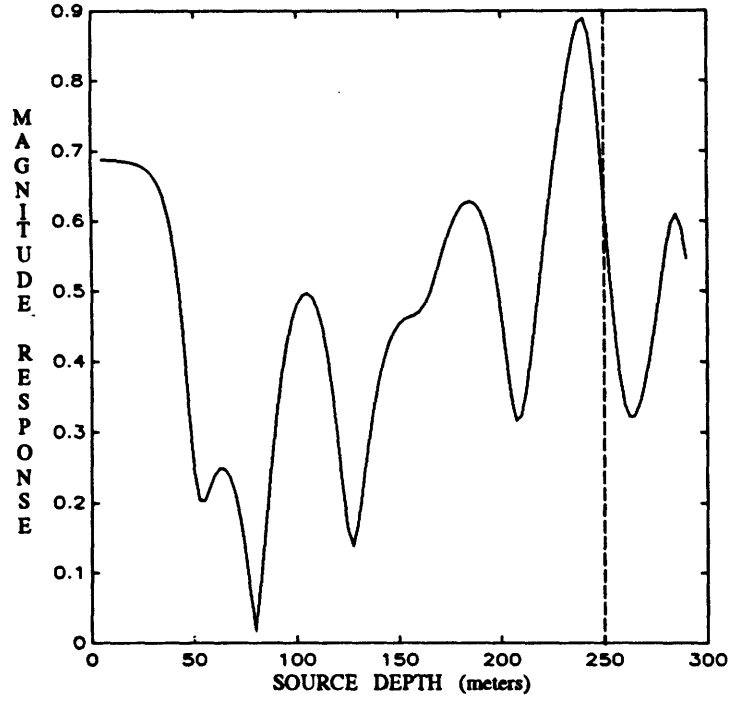
$$\underline{q}_{eff} = \arg \min_{\underline{q} \in \mathcal{H}\left(\tilde{Q}(f,\underline{z})\right)} |\underline{q}|^2 \left(\frac{N\frac{\sigma_s^2}{\sigma_n^2} + 1}{N\frac{\sigma_s^2}{\sigma_n^2}} - \cos^2(\underline{q}_{act}, \underline{q}; \mathrm{I}) \right).$$

As the SNR decreases, the first term in the parenthesis increases, the reduction of the norm of $\underline{q}_{eff}$ is emphasized, and the reduction of the mismatch between $\underline{q}_{eff}$ and $\underline{q}_{act}$ (i.e. increasing $\cos^2(\underline{q}_{eff}, \underline{q}_{act}; \mathrm{I})$) is deemphasized. Therefore, $\cos^2(\underline{q}_{eff}, \underline{q}_{act}; \mathrm{I})$ should decrease as the SNR decreases. Figure 3-16a supports this conclusion.

Figures 3-14a and 3-15b show that, at a high SNR, the source depth/ocean depth response of the weights generated by the Adaptive Minmax Processor equals one when the array focal point is the source location and the response is evaluated at the actual source depth and the actual ocean depth. However, when the SNR is decreased, Figure 3-16b shows that the response evaluated at the actual source depth

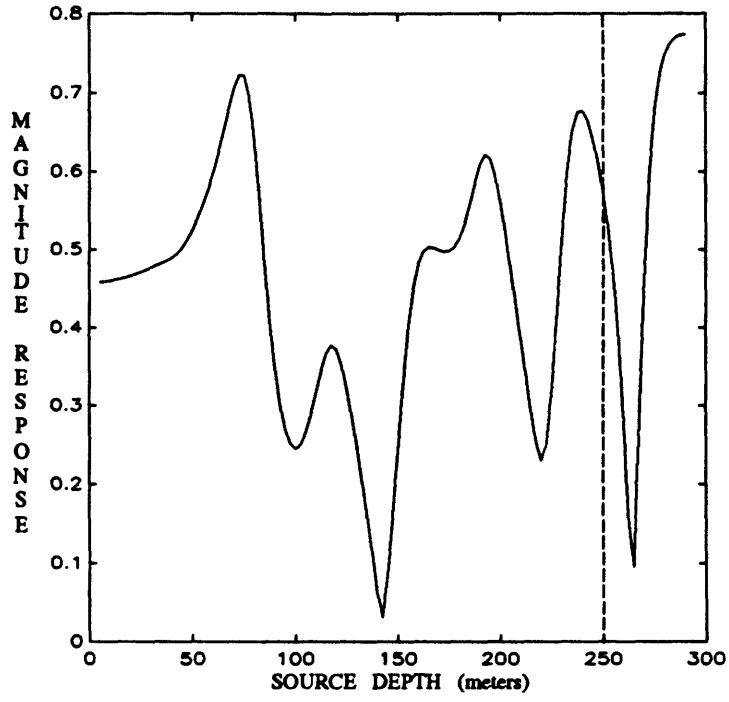


(a) Response at Ocean Depth = 290 meters

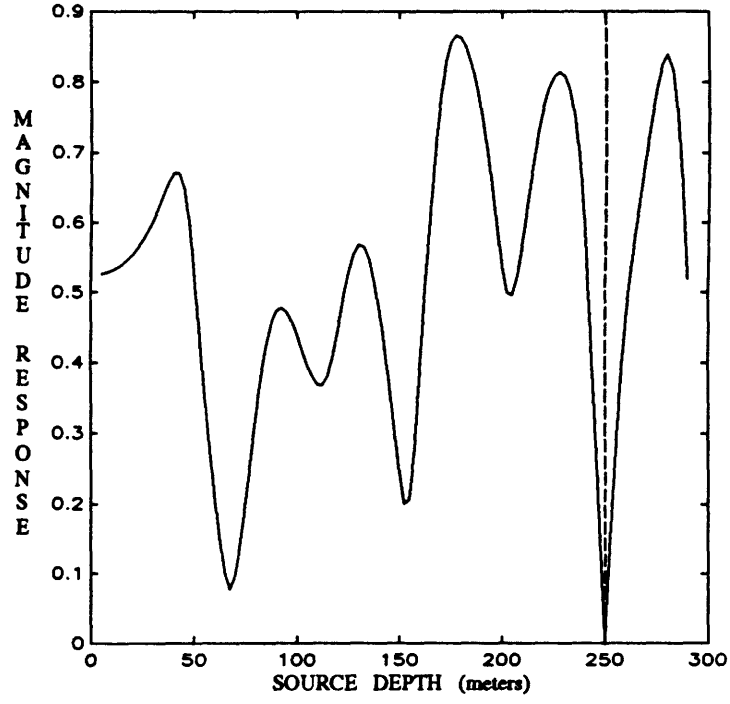


(b) Response at Ocean Depth = 310 meters

Figure 3-12: Source Depth/Ocean Depth Response for Array Focal Point = 175 meters and Actual Ocean Depth = 290 meters

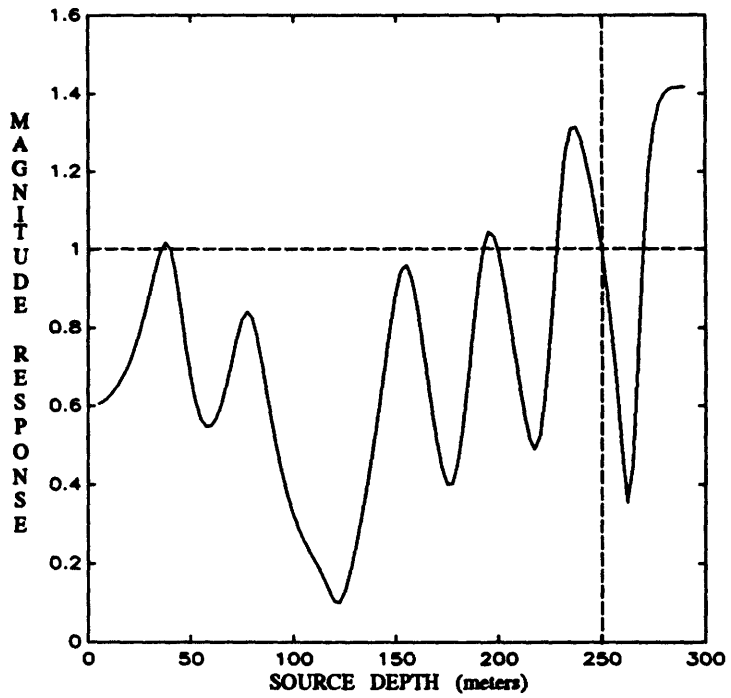


(a) Response at Ocean Depth = 290 meters

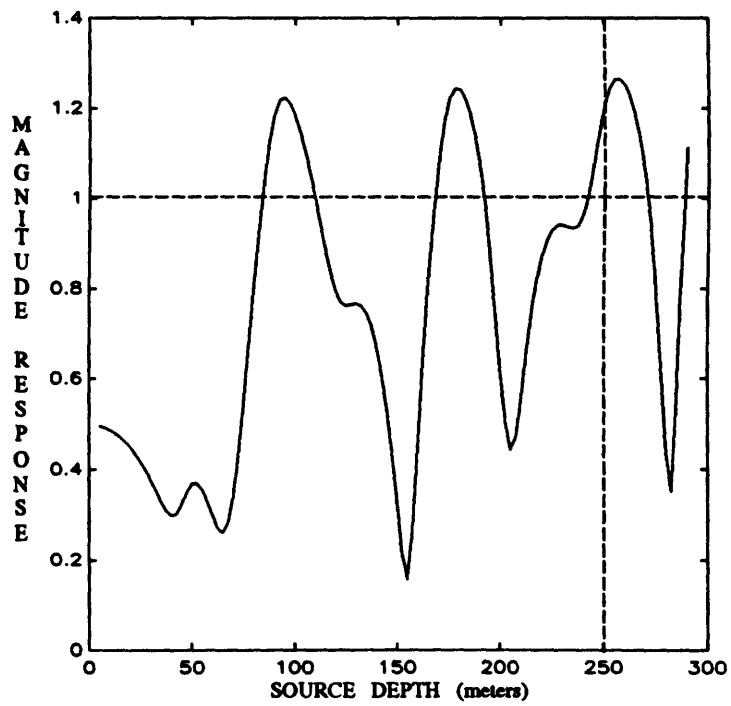


(b) Response at Ocean Depth = 310 meters

Figure 3-13: Source Depth/Ocean Depth Response for Array Focal Point = 175 meters and Actual Ocean Depth = 310 meters

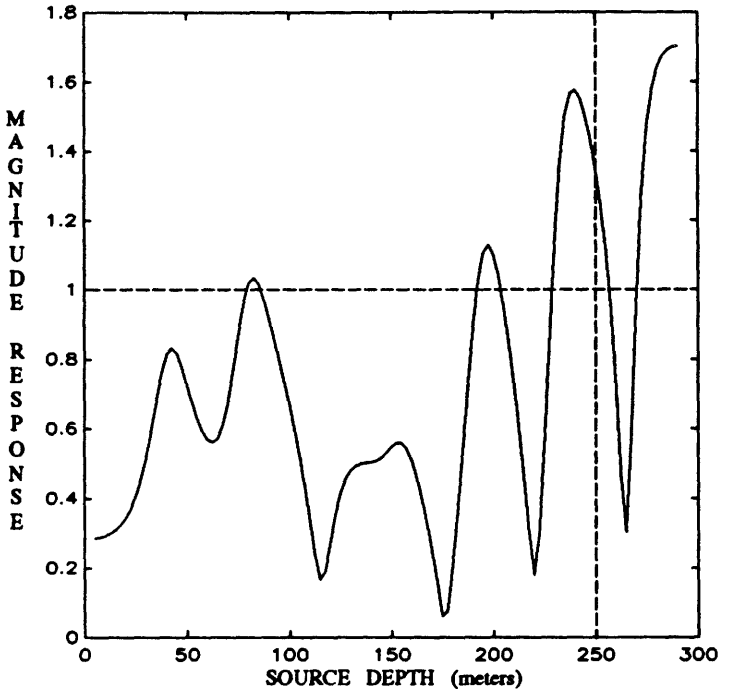


(a) Response at Ocean Depth = 290 meters

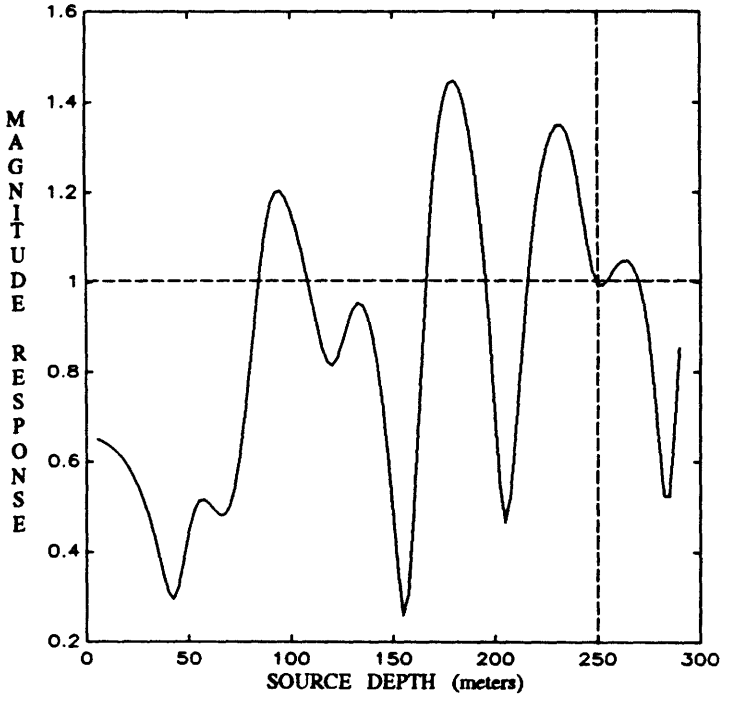


(b) Response at Ocean Depth = 310 meters

Figure 3-14: Source Depth/Ocean Depth Response for Array Focal Point = 250 meters and Actual Ocean Depth = 290 meters



(a) Response at Ocean Depth = 290 meters



(b) Response at Ocean Depth = 310 meters

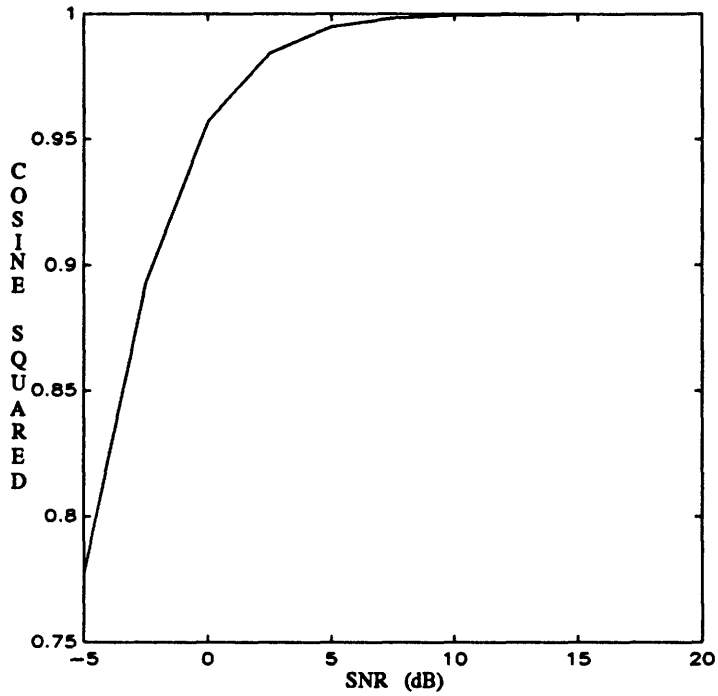
Figure 3-15: Source Depth/Ocean Depth Response for Array Focal Point = 250 meters and Actual Ocean Depth = 310 meters

and actual ocean depth drops below one. This drop can be partially attributed to the increase in processor mismatch as measured by the difference between q_{eff} and q_{act}, which results from a decrease in SNR; this is illustrated in Figure 3-16a. The increased mismatch and resulting drop in the magnitude response partially accounts for the drop in the peak level of the ambiguity function of the Adaptive Minmax Processor which was discussed earlier and illustrated in Figure 3-9.

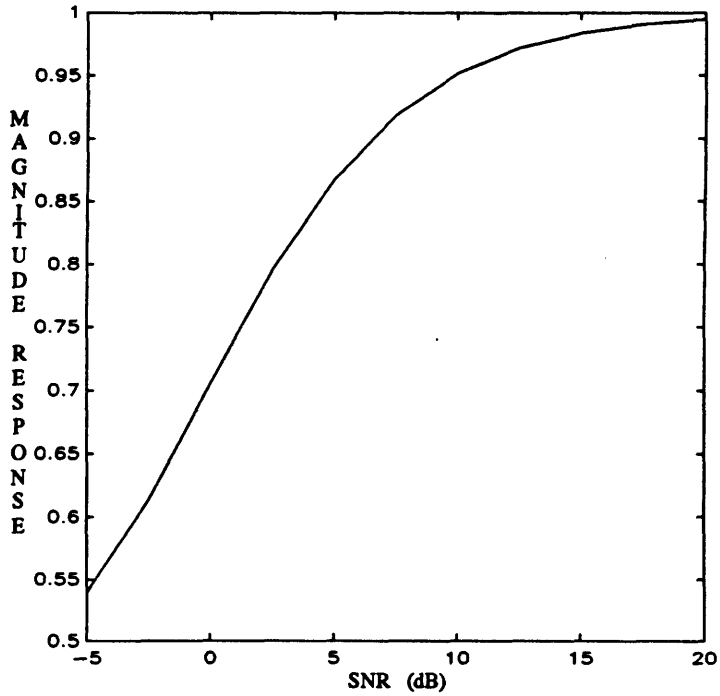
3.2.2 The Arctic Ocean

The second set of numerical results were generated modeling the ocean as a deterministic horizontally-stratified medium with the arctic sound speed profile shown in Figure 3-17. The dominant characteristic of the sound speed profile is the strong surface duct created as a result of the location of the sound speed minimum at the surface. Propagating sound will be refracted to the surface and tend to stay in the duct close to the surface as it propagates. The sound speed profile at depths greater than or equal to 85 meters is assumed to be known to the processor. The sound speed in the top 85 meters of the water column is assumed to vary in a pseudo-linear fashion (i.e., $\frac{1}{\overline{C}(z)^2}$ varies linearly with depth) between an unknown value at the surface and a known value at the depth of 85 meters. Therefore, the sound speed profile is completely parameterized by the surface sound speed $\overline{C}(0)$ and Φ is the set of possible values for $\overline{C}(0)$. The salient parameters of the simulation are that the ocean depth is 3800 meters, the sea surface is assumed to be a free surface, and the ocean bottom is assumed to be a soft bottom so that no bottom reflected waves are able to reach the array sensors. The array is an 18 element vertical array with the top element at a depth of 60 meters and the bottom element at a depth of 570 meters (inter-element spacing is 30 meters). All of the results shown here were generated with a single omnidirectional 20 Hz source located at a depth of 190 meters and a range of 250 km from the array. In all cases, the range of the surface sound speed given to the Adaptive Minmax Processor (Φ) is from $\overline{C}(0) = 1430.75$ meters/second to $\overline{C}(0) = 1432.25$ meters/second.

The noise field consists of both sensor noise and surface-generated background



(a) $\cos^2(\underline{q}_{eff}, \underline{q}_{act}; \mathrm{I})$ vs SNR



(b) Source Depth/Ocean Depth Response vs SNR

Figure 3-16: Adaptive Minmax Processor Characteristics vs SNR

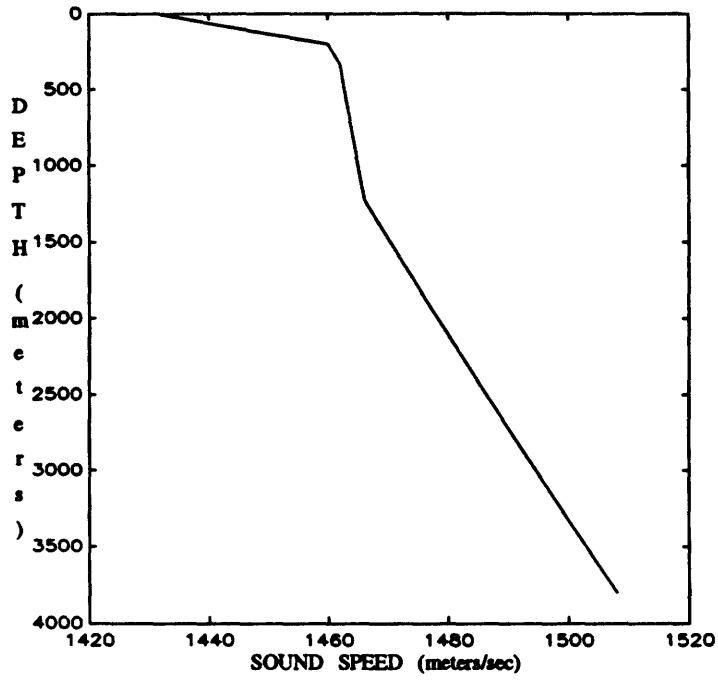


Figure 3-17: Arctic Sound Speed Profile

noise. Consistent with the surface noise model in [35], the surface noise is modeled as generated by a horizontal sheet of stochastic monopole sources which are spatially uncorrelated and located 0.5 meters below the surface.

Figures 3-18 through 3-21 show the ambiguity functions for the Matched and the Mismatched MVDR Processors, the Matched Bartlett Processor, and the Adaptive Minmax Processor. In all cases, the actual surface sound speed is 1431 meters/second, the SNR is 20 dB, and the background noise to sensor noise ratio is 10 dB. Therefore, the signal to background noise SNR is 10 dB. The Mismatched MVDR Processor operates with an assumed surface sound speed of 1432 meters/second resulting in a 1 meter/second mismatch in surface sound speed. The first figure for each processor shows the ambiguity function evaluated on a course grid (45 meters vertical spacing, 2 km horizontal spacing) over the range of 10 to 1000 meters depth and 150 to 300 km range. The second figure for each processor shows the ambiguity function evaluated on a much finer grid (5 meters vertical spacing, 250 meters horizontal spacing) over the range of 140 to 240 meters depth and 241 to 259 km range.

A common feature of all of the ambiguity functions is the range-extended band

of elevated response in the upper 300 meters of the ocean. In other numerical experiments not detailed here, it is indicated that when the source is removed, the ambiguity functions in this upper layer of the ocean for the Matched MVDR and Adaptive Minmax Processors drop approximately 3 to 4 dB in the regions away from the source location and drop approximately 17 dB at the source location. The ambiguity functions for the no-source case shows levels in the upper 250 meters of the ocean to be approximately 4 dB above those in the ocean below 350 meters depth for both the MVDR and the Adaptive Minmax Processors. Therefore, the elevated response in this region is partially due to the inability of the processors to resolve the source location in range and partially due to the presence of surface-generated background noise. The Matched Bartlett Processor shows a comparable ambiguity function in the no-source case in the upper layer of the ocean. In the lower layer of the ocean for the no-source case, the level of the ambiguity function for the Bartlett Processor is only 1 to 2 dB below that in the upper layer, which is consistent with the inferior depth resolution of the Bartlett Processor when compared to the MVDR and the Adaptive Minmax Processors. When the source is added, the ambiguity function in the entire upper layer increases by approximately 14 db; the ambiguity function in the lower layer increases by approximately 4 dB. Therefore, the high level of the ambiguity function of the Bartlett Processor in the upper layer of the ocean is due primarily to the poor range resolution of the Bartlett Processor (i.e., high sidelobes).

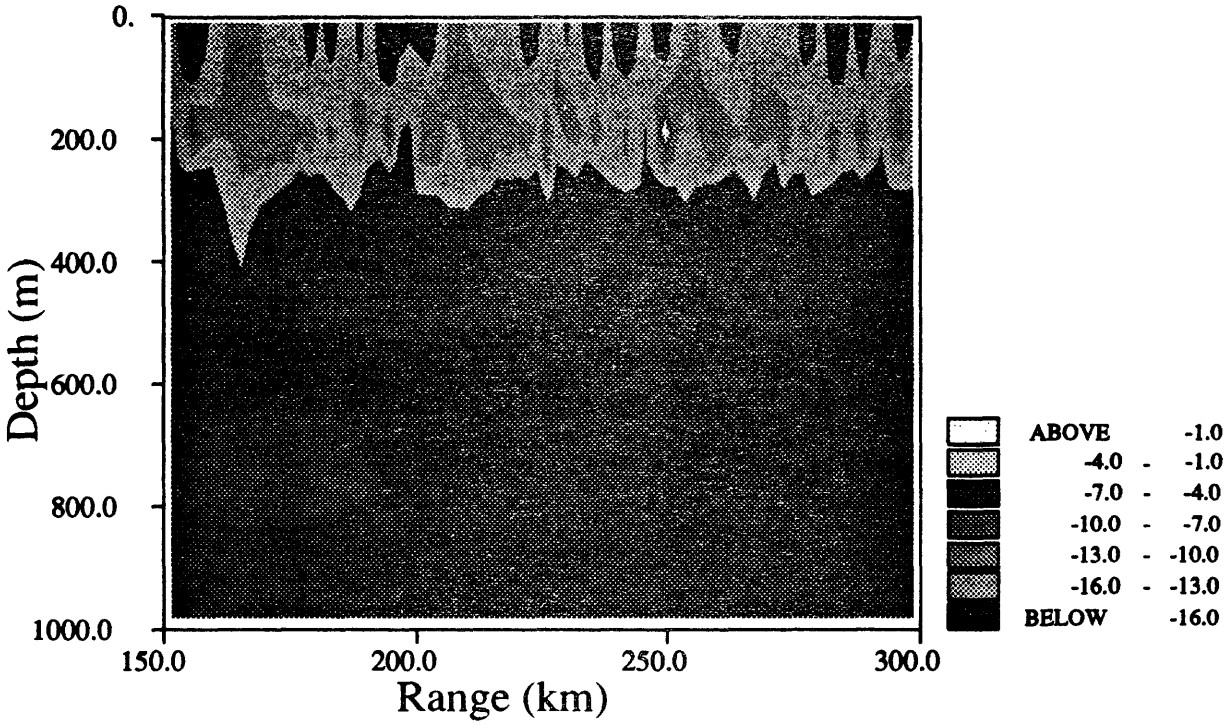
The ambiguity functions for the Matched MVDR Processor (Figure 3-18), the Matched Bartlett Processor (Figure 3-20), and the Adaptive Minmax Processor (Figure 3-21) have global maxima at the true source location at a normalized level of approximately 0 dB. The Mismatched MVDR Processor (Figure 3-19) has a global maximum at a depth of 185 meters and range 250.25 km at a normalized level of -1.15 dB. The slight mismatch of 1 meter/second in surface sound speed results in a slight offset of the peak location and a loss of 1.15 dB in the peak response of the MVDR Processor. Consistent with the spatial resolution results in Subsection 3.2.1, the mainlobe of the Adaptive Minmax Processor is slightly broader than the mainlobe the MVDR Processor and considerably narrower than the mainlobe of the Bartlett

Processor.

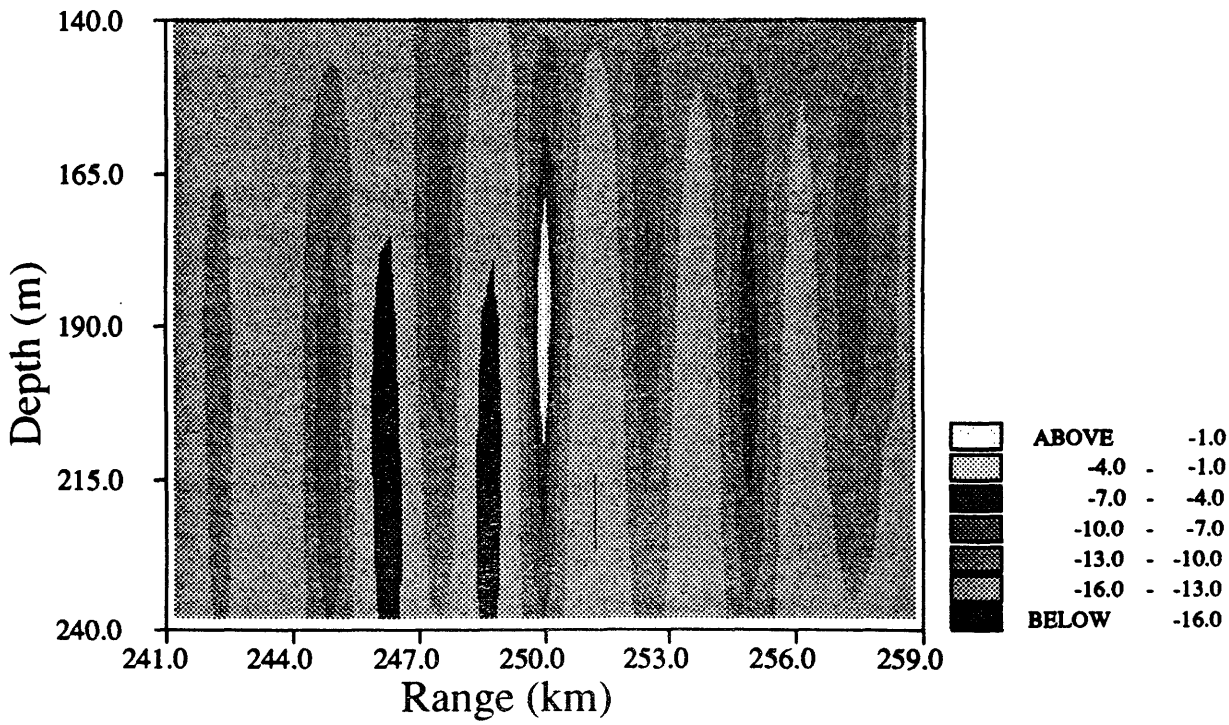
The array gain defined in Subsection 3.1.2 is a good measure of a processor's ability to separate a desired signal from the noise in which it is embedded. Figure 3-22 shows the array gain of the Matched and the Mismatched MVDR Processors, the Matched and the Mismatched Bartlett Processors, and the Adaptive Minmax Processor as a function of SNR for several surface sound speeds when the array focal point is the source location. In all cases, the mismatched processors operate with an assumed surface sound speed of 1432 meters/second and the background noise to sensor noise ratio is 10 dB. In all the figures, the independent variable is source to background noise SNR rather than the source to sensor noise SNR. The array gain shown has been normalized so that the gain of the Matched Bartlett Processor equals one.

For all SNRs and surface sound speeds shown, the array gain of the Adaptive Minmax Processor is less than the array gain the Matched MVDR Processor and greater than the array gain of the Matched MVDR Processor. As the amount of the mismatch between the surface sound speed assumed by the mismatched processors and the actual surface sound speed increases, the gain of the Mismatched MVDR Processor falls when compared to the Matched MVDR Processor and the Adaptive Minmax Processor. As the SNR increases, a loss of array gain in the Adaptive Minmax Processor results from the small mismatch between $\underline{q}_{eff}$ and $\underline{q}_{act}$. An interesting note is that, for small amounts of mismatch between the assumed and the actual surface sound speeds, the gain of the Mismatched Bartlett Processor is greater than that of the Matched Bartlett Processor. This result would be impossible in a spatially white noise field because the Matched Bartlett and the Matched MVDR Processors maximize the array gain when the noise field is spatially white. The equivalence of the Matched Bartlett and the Matched MVDR Processors in this case is based upon the assumption that $\mathrm{S}(f) = \hat{\mathrm{S}}(f)$.

In Subsection 3.1.3, it is concluded that, for the case where the array focal point is the source location, the mismatch between $\underline{q}_{eff}$ and $\underline{q}_{act}$ will decrease as the SNR increases and, as a result, $\cos^2(\underline{q}_{eff}, \underline{q}_{act}; \mathrm{S}_n(f)^{-1})$ will increase as the SNR increases. Figure 3-23, which shows the generalized cosine-squared between $\underline{q}_{eff}$ and $\underline{q}_{act}$ as a

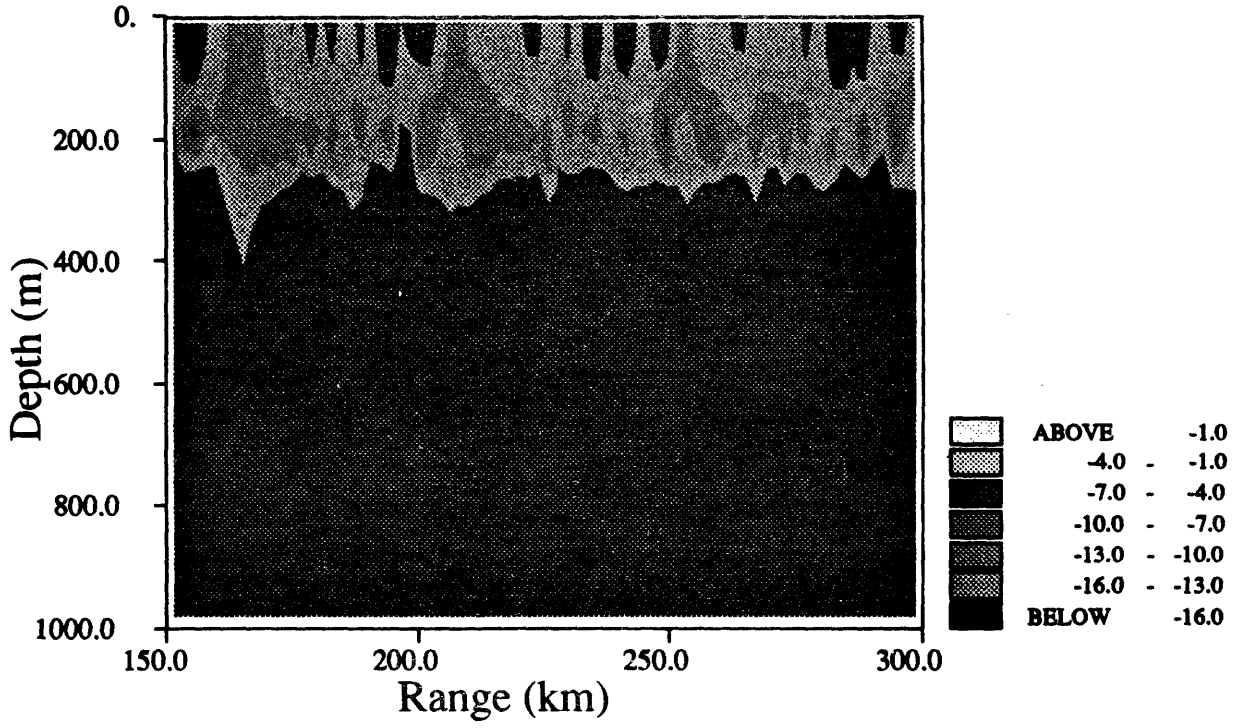


(a) Course Scale

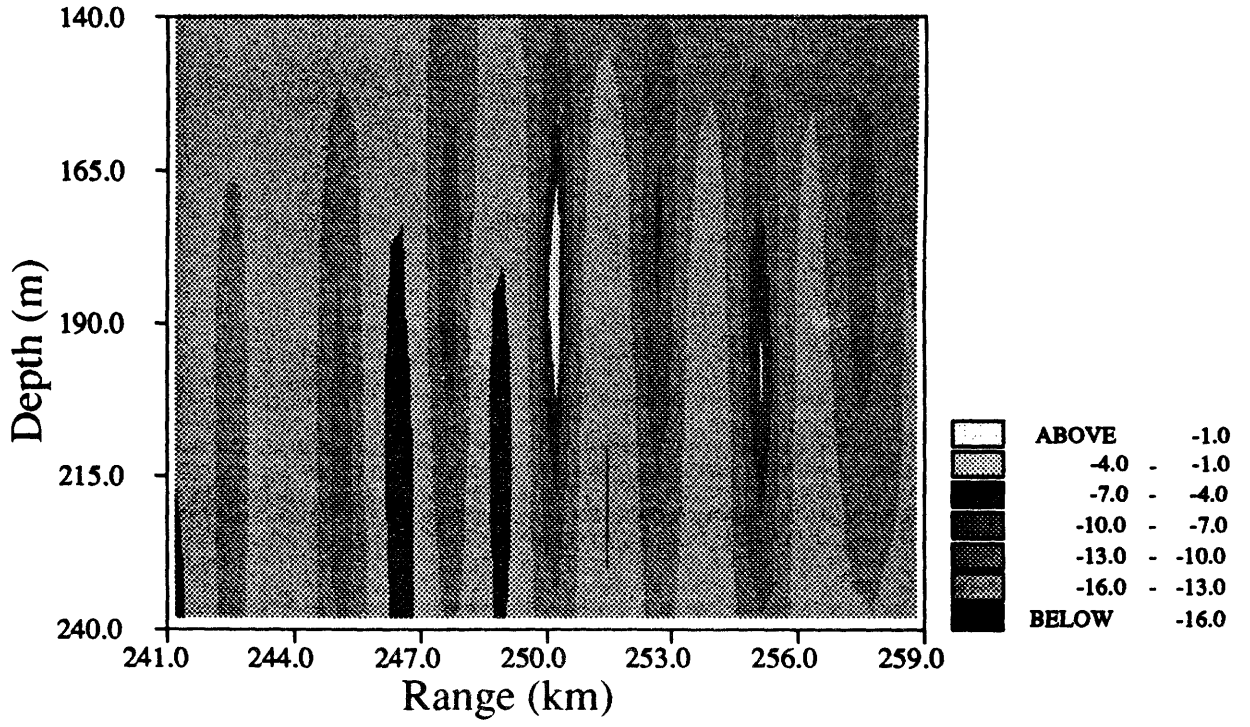


(b) Fine Scale

Figure 3-18: Ambiguity Function for the Matched MVDR Processor

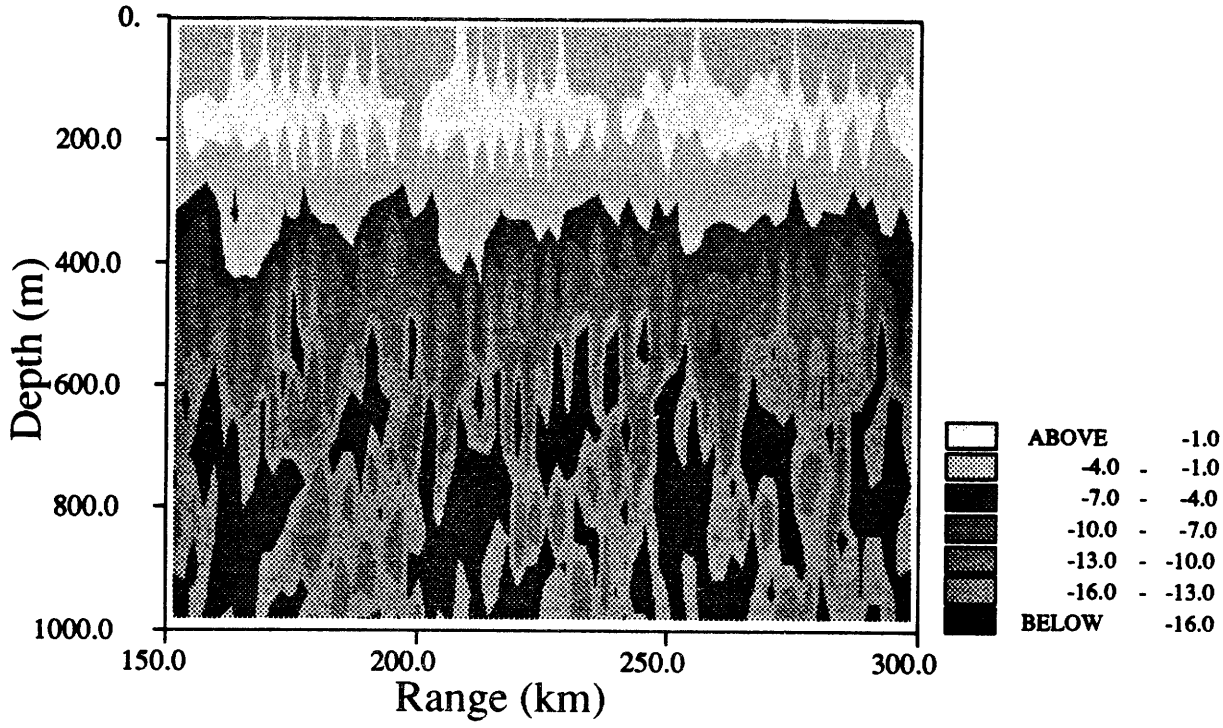


(a) Course Scale

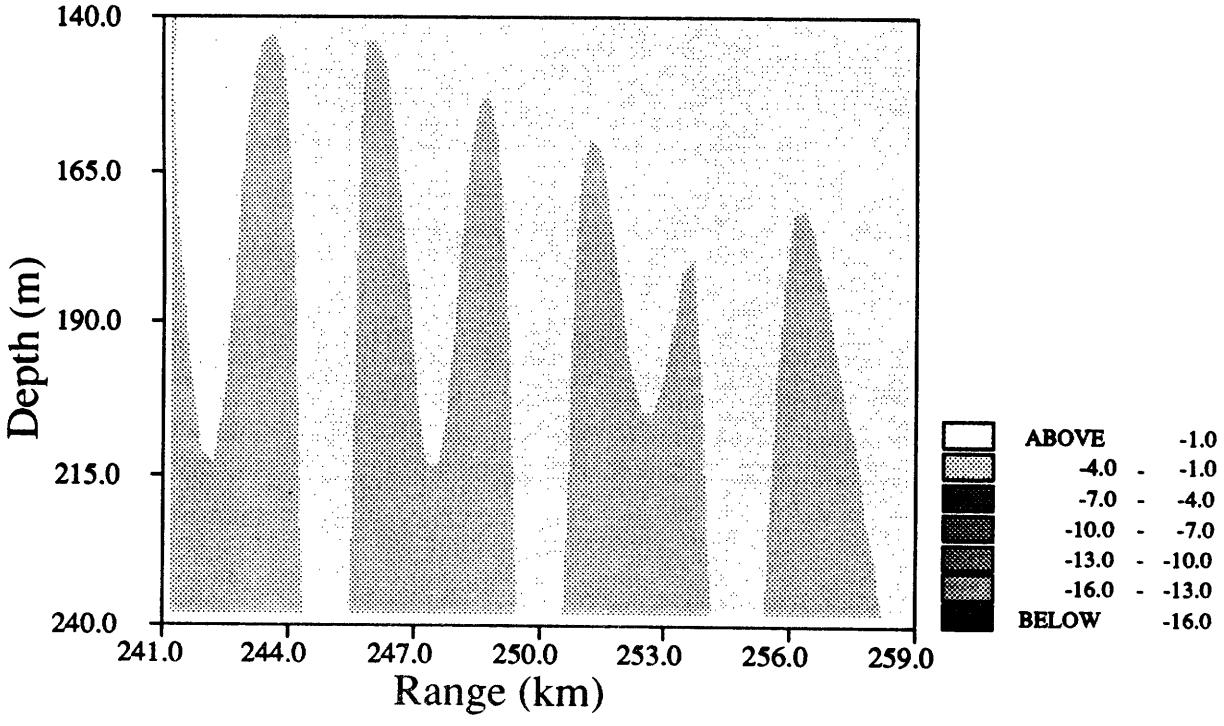


(b) Fine Scale

Figure 3-19: Ambiguity Function for the Mismatched MVDR Processor

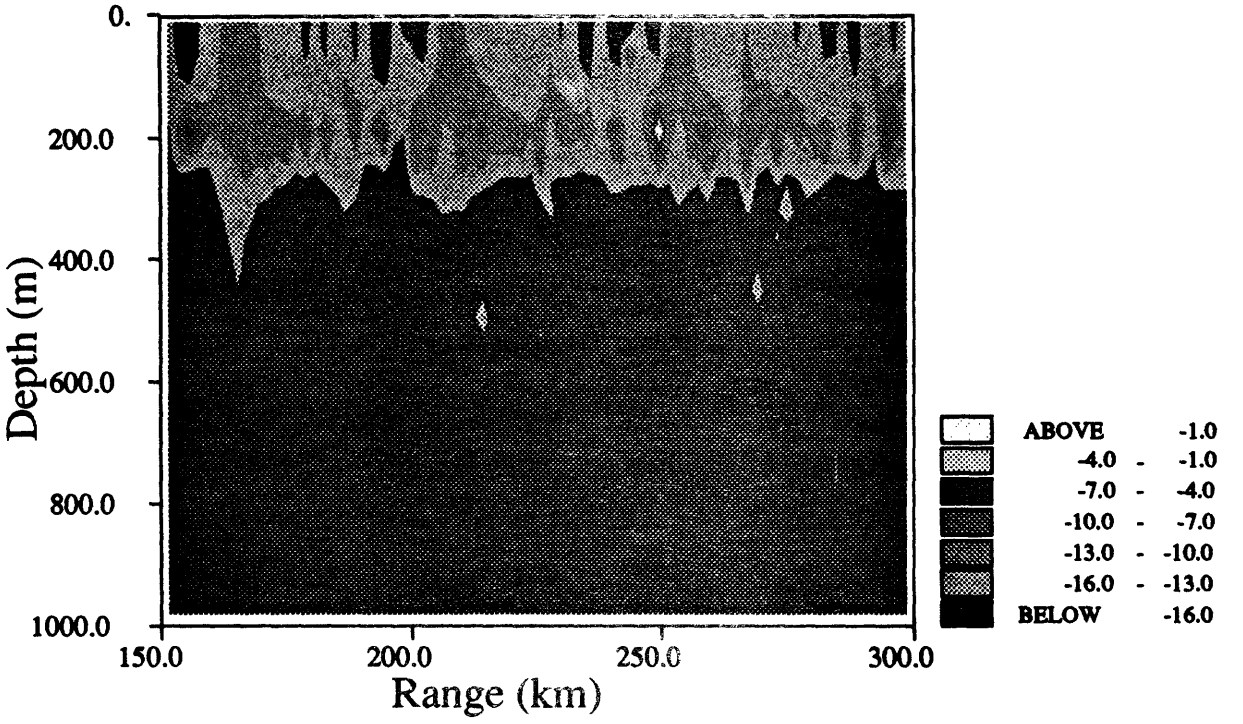


(a) Course Scale

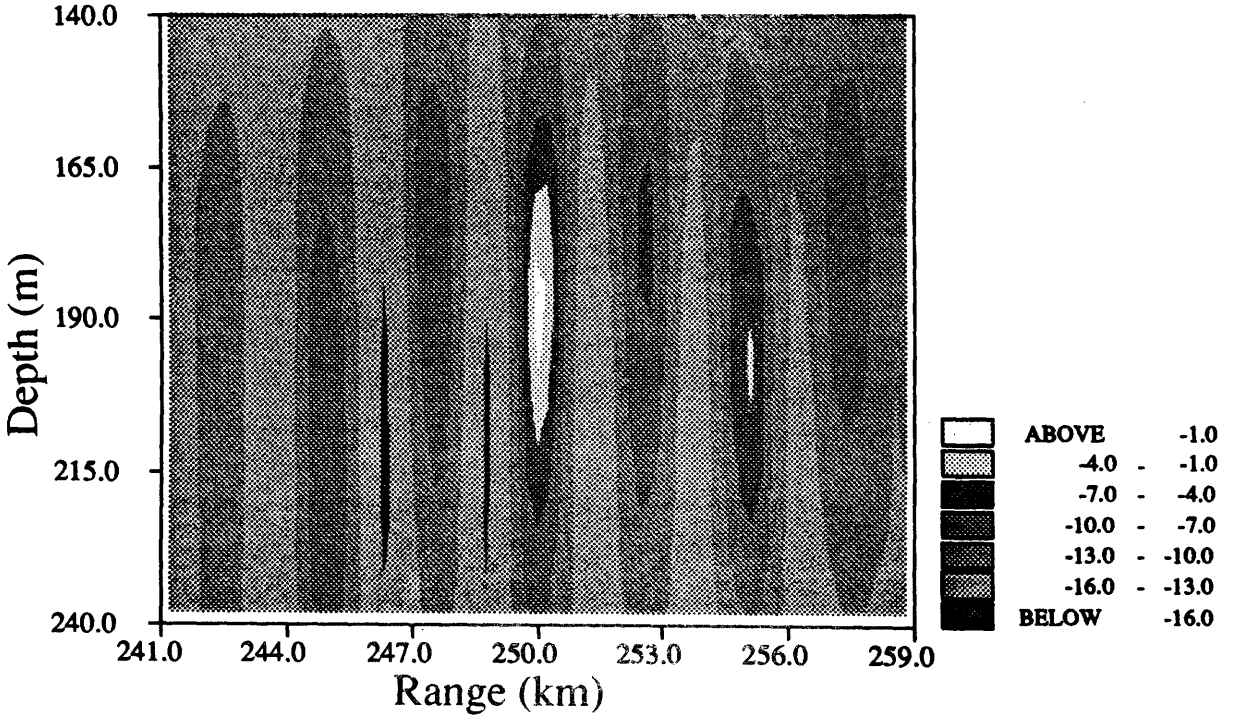


(b) Fine Scale

Figure 3-20: Ambiguity Function for the Matched Bartlett Processor

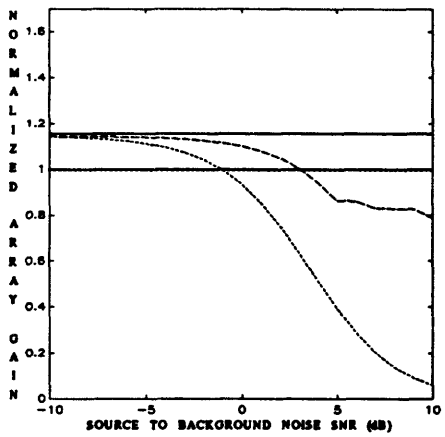


(a) Course Scale

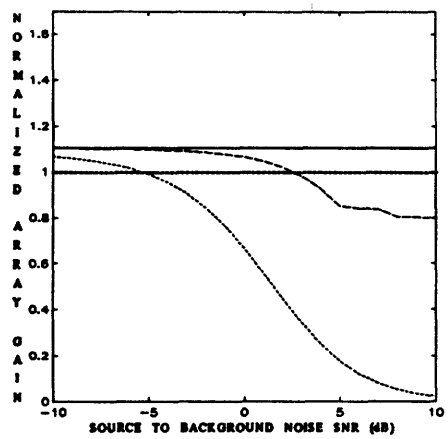


(b) Fine Scale

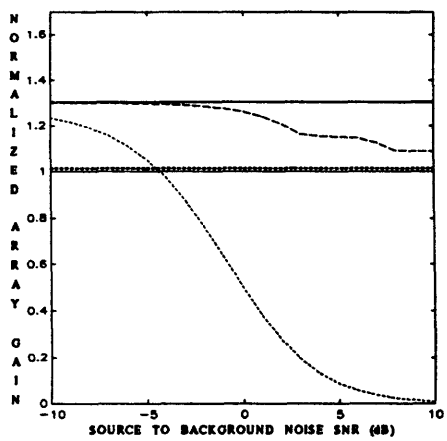
Figure 3-21: Ambiguity Function for the Adaptive Minmax Processor



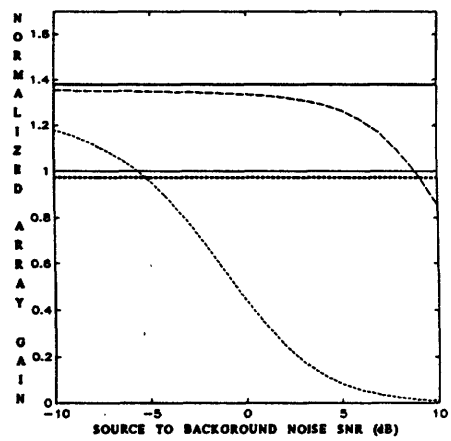
(a) Surface Sound Speed = 1431.75 m/s



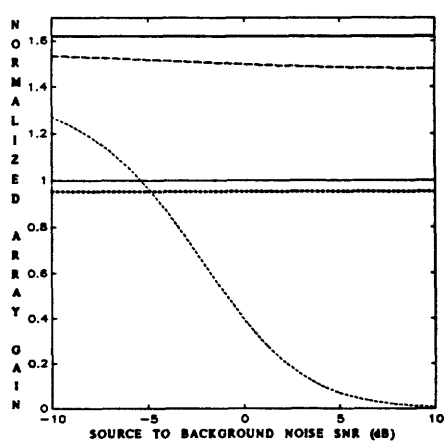
(b) Surface Sound Speed = 1431.50 m/s



(c) Surface Sound Speed = 1431.25 m/s



(d) Surface Sound Speed = 1431.00 m/s



(e) Surface Sound Speed = 1430.75 m/s

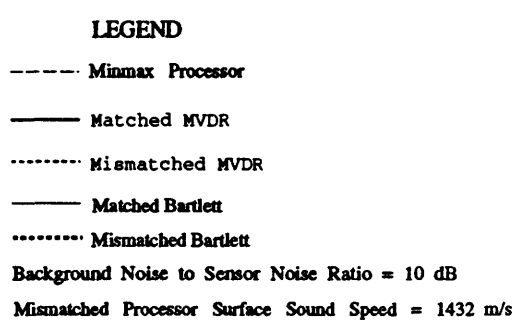
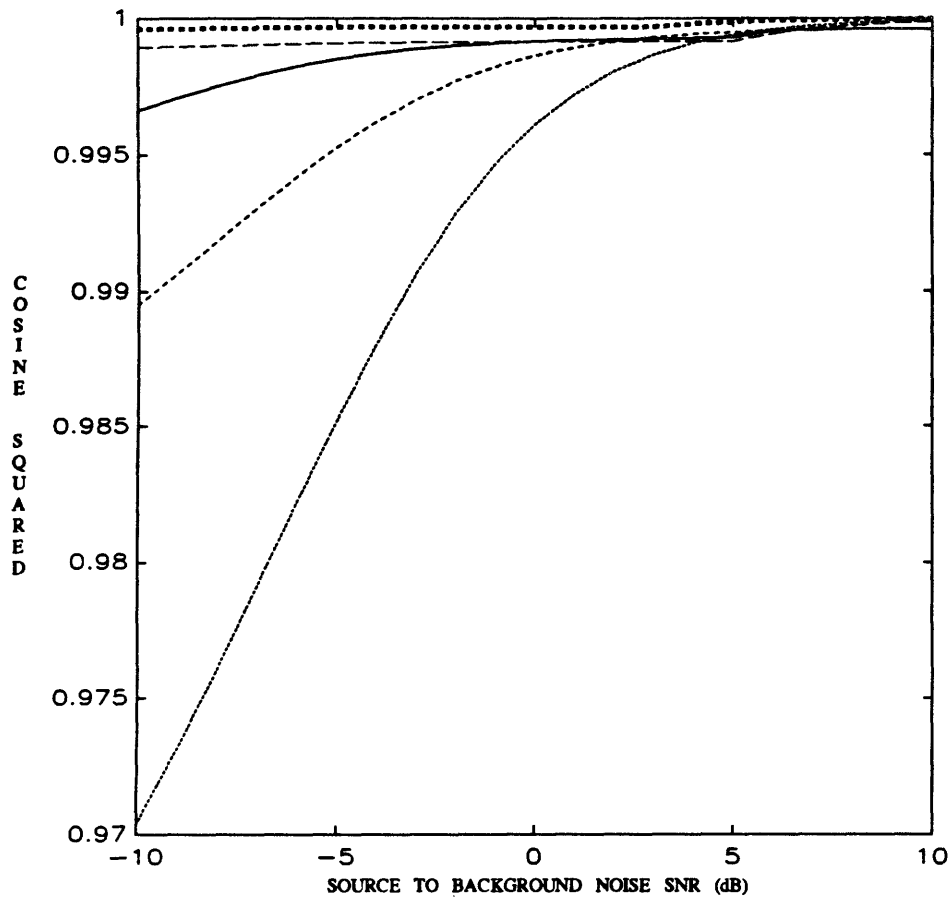


Figure 3-22: Array Gain vs SNR for Various Surface Sound Speeds

function of the source to background noise SNR for various surface sound speeds when the array focal point is the source location, supports this conclusion. However, the results in Subsection 3.2.1 demonstrated the same dependence of $\cos^2(\underline{q}_{eff}, \underline{q}_{act}; S_n(f)^{-1})$ on the SNR for the case where $S_n(f) = I$. Therefore, it is unclear as to whether the cause of this dependence is the processor attempting to reduce $| \underline{q}_{eff} |^2$ or it is the processor attempting to reduce $\frac{\underline{q}_{eff}^h S_n(f)^{-1} \underline{q}_{eff}}{\underline{q}_{eff}^h \underline{q}_{eff}}$. The role of these quantities can be seen in (3.6).

The Adaptive Minmax Processor adjusts $\underline{q}_{eff}$ by adjusting the least favorable pmf ($\underline{p}_{lf}$). As the actual surface sound speed changes, it is reasonable to expect that when the array is focused at the source location, $\underline{p}_{lf}$ will change to exhibit a peak in the neighborhood of the actual surface sound speed. Numerical simulations were run to test whether or not the processor exhibits this behavior. In these tests, there is no propagating background noise, the signal to sensor noise SNR is 10 dB, and the source location and array focal point are both at a range of 250 km and a depth of 190 meters. The environmental parameter set Φ is sampled at seven points (1430.75 meters/second through 1432.25 meters/second in 0.25 meters/second increments) and the actual surface sound speed is varied between 1430.75 meters/second and 1432.25 meters/second in 0.125 meters/second increments. As shown in Figures 3-24 and 3-25, the Adaptive Minmax Processor exhibits the expected behavior while the actual surface sound speed is in the lower part of this range. However, when the actual surface sound speed is in the range between 1431.75 meters/second and 1432.0 meters/second, $\underline{p}_{lf}$ does not peak in the neighborhood of the actual surface sound speed. However, the results presented in Figure 3-26 and discussed in the following paragraph indicate that this failure of $\underline{p}_{lf}$ to peak as expected for some values of the actual surface sound speed does not adversely affect the performance of the Adaptive Minmax Processor.

In Subsection 3.1.4 the effect on the performance of the Adaptive Minmax Processor of the range and sampling density of the environmental parameter set Φ is qualitatively analyzed. A set of simulations were run with different ranges and sampling densities to quantitatively evaluate the effect. The test conditions are a signal to



LEGEND

Surface Sound Speed = 1431.75 m/s

Surface Sound Speed = 1431.50 m/s

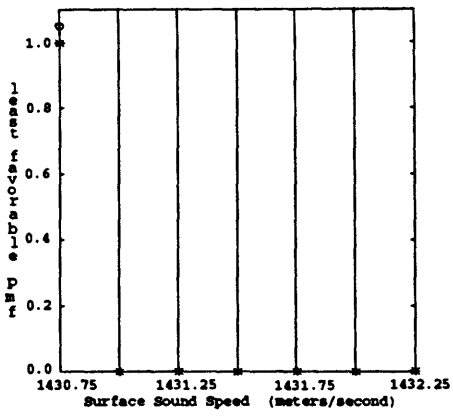
Surface Sound Speed = 1431.25 m/s

Surface Sound Speed = 1431.00 m/s

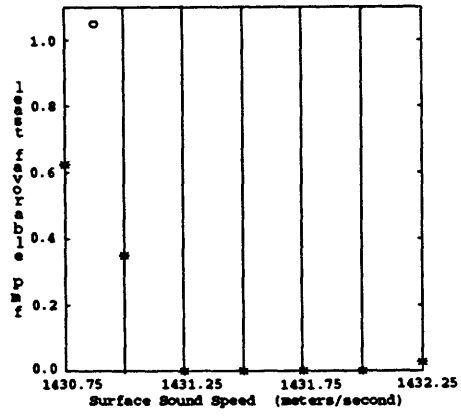
Surface Sound Speed = 1430.75 m/s

Background Noise to Sensor Noise Ratio = 10 dB

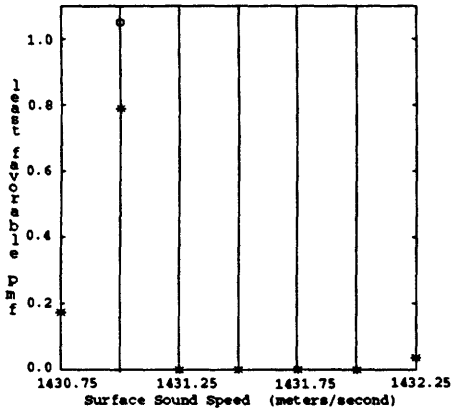
Figure 3-23: $\cos^2(\underline{q}_{eff}, \underline{q}_{act}; S_n(f)^{-1})$ vs SNR for Various Surface Sound Speeds



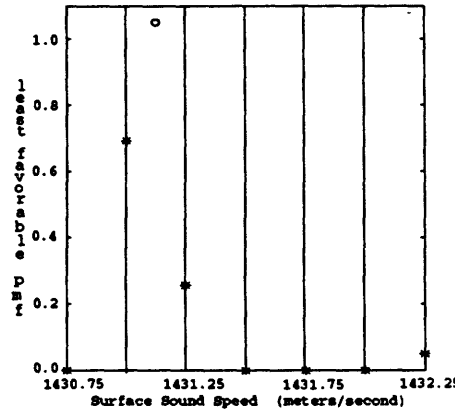
(a) c(0) = 1430.75 m/s



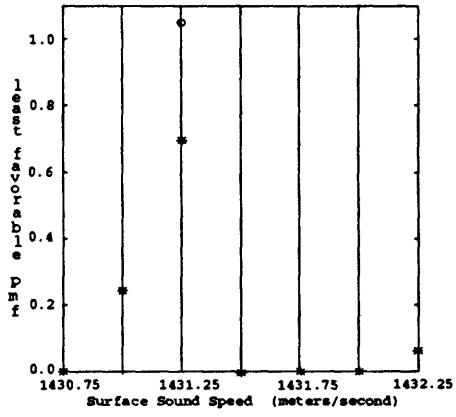
(b) c(0) = 1430.875 m/s



(c) c(0) = 1431.00 m/s



(d) c(0) = 1431.125 m/s

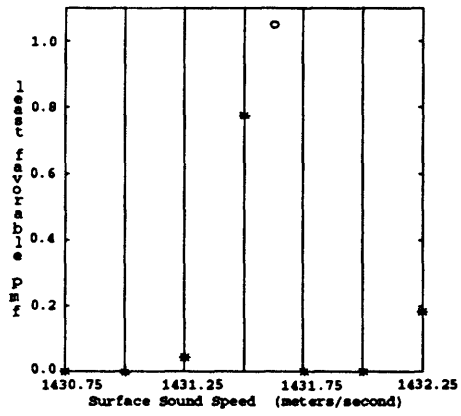


(e) c(0) = 1431.25 m/s

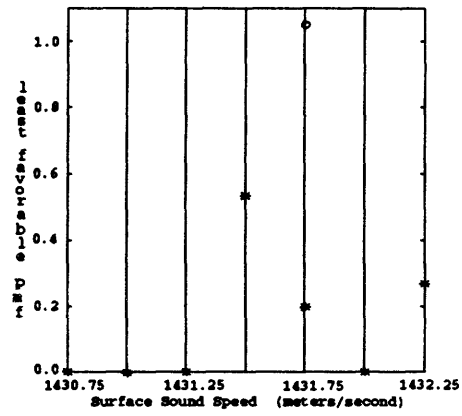
* least favorable pmf

o actual surface sound speed

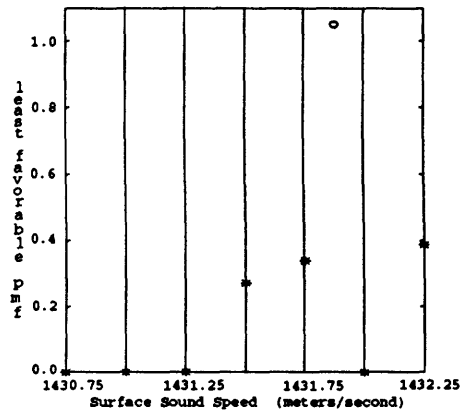
Figure 3-24: Least Favorable PMF for Various Actual Surface Sound Speeds



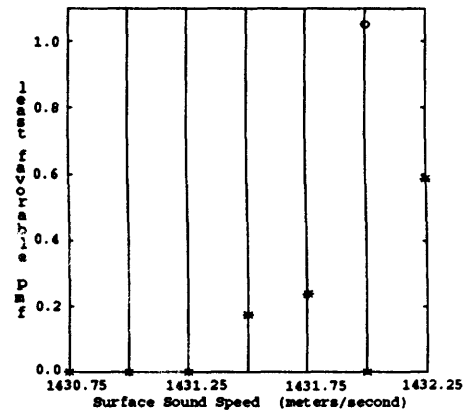
(a) c(0) = 1431.625 m/s



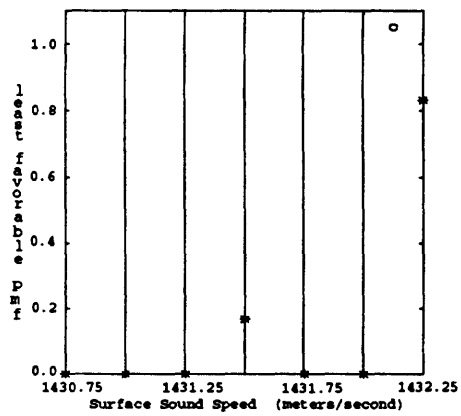
(b) c(0) = 1431.75 m/s



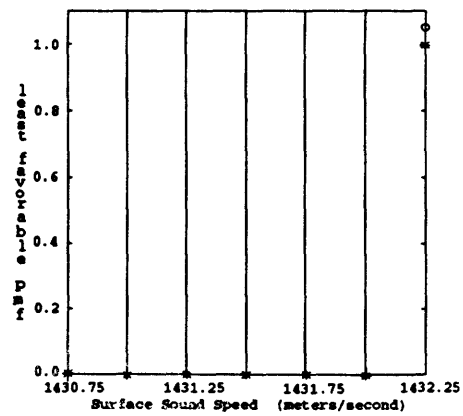
(c) c(0) = 1431.875 m/s



(d) c(0) = 1432.00 m/s



(e) c(0) = 1432.125 m/s



(f) c(0) = 1432.25 m/s

Figure 3-25: Least Favorable PMFs (continued)

surface noise SNR of 10 dB and a signal to sensor noise SNR of 20 dB. A single source is located at a range of 250 km and a depth of 190 meters. Tests were run for actual surface sound speeds of 1431.5 meters/second, 1431.625 meters/second, 1431.75 meters/second, 1431.875 meters/second, 1431.9375 meters/second, 1432.0 meters/second, 1432.125 meters/second, and 1432.25 meters/second and for an array focal point of a range of 250 km and a depth of 190 meters (i.e., the source location). The peak response loss, which is defined as the estimated average signal power generated by the Adaptive Minmax Processor minus that generated by the Matched MVDR Processor, is plotted versus actual surface sound speed in Figure 3-26. The three cases shown are for the sampling of Φ at two points (1431.0 meters/second and 1432.0 meters/second), three points (1431.0 meters/second, 1431.5 meters/second, and 1432.0 meters/second), and seven points (1430.75 meters/second through 1432.25 meters/second in 0.25 meters/second increments). The Adaptive Minmax Processor shows almost no loss in response relative to the MVDR Processor for all cases when the actual surface sound speed falls within the interval spanned by the samples of the surface sound speed parameter set Φ. This indicates that for this class of environmental uncertainties, the performance of the Adaptive Minmax Processor is fairly insensitive to changes in the sampling density of Φ. However, when the actual surface sound speed falls outside the interval spanned by the the samples of the surface sound speed parameter set Φ (i.e., $\overline{C}(0) > 1432.0$ meters/second for the two and three sample cases), the Adaptive Minmax Processor shows a dramatic loss in performance as measured by the peak response loss. For this same range, the Adaptive Minmax Processor utilizing seven sample of Φ shows no loss in performance. Therefore, as predicted in Subsection 3.1.4, the Adaptive Minmax Processof suffers a significant degradation in performance when the actual environmental conditions which are encountered fall outside the range of the sampled environmental parameter set Φ.

The norm-squared of the array weight vector of a linear weight-and-sum beamformer is often used as a measure of the sensitivity of the processor to environmental mismatch [8]. However, the use of this measure depends upon the processor using a replica vector which is calculated a priori rather than adaptively as is done in the

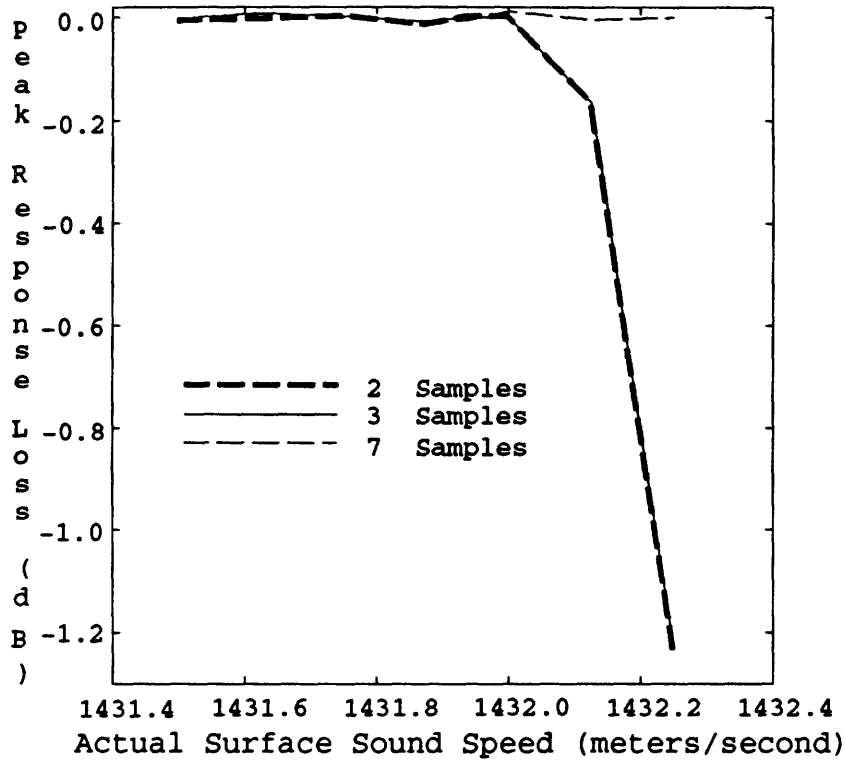


Figure 3-26: Peak Response Loss for Various Numbers of Environmental Samples

Two-Stage MVDR Processor interpretation of the Adaptive Minmax Processor.

A set of tests were run to measure the norm-squared of the weight vectors of the Matched MVDR and Adaptive Minmax Processors under varous conditions. In all cases, the signal to surface noise SNR is 10 db, the signal to sensor noise SNR is 20 db, the source is at a range of 250 km and a depth of 190 meters, and the set Φ is sampled at seven points as described in the preceeding paragraph. Tables 3.1 and 3.2 list the norm-squared of the array weight vectors for the MVDR and Adaptive Minmax Processors when the range of the array focal point is 250 km and the focal point depth and actual surface sound speed take on several values. Note that the norm-squared of the weight vectors for the two processors are comparable under equivalent conditions. The large norm-squared when the depth of the array focal point is 165 meters can be attributed to the processors attempting to null out the source at 190 meters depth while maintaining a reasonable response (equal to one in the MVDR case) to a signal emitted by a source at 165 meters depth. Under the standard criterion, these weight vector norms would indicate that the MVDR

focal depth	140 meters	165 meters	190 meters
c(0) = 1430.75 m/s	11.41	24.67	6.11
c(0) = 1431.375 m/s	13.19	36.61	4.49
c(0) = 1431.9375 m/s	11.37	30.39	1.18
c(0) = 1432.00 m/s	18.93	39.15	3.13

Table 3.1: Norm-Squared of Matched MVDR Processor Weight Vectors

focal depth	140 meters	165 meters	190 meters
c(0) = 1430.75 m/s	12.19	24.66	5.70
c(0) = 1431.375 m/s	14.13	35.64	1.22
c(0) = 1431.9375 m/s	11.03	30.55	1.24
c(0) = 1432.00 m/s	20.29	39.13	1.12

Table 3.2: Norm-Squared of Adaptive Minmax Processor Weight Vectors

and Adaptive Minmax Processors are equally sensitive to environmental mismatch. This is consistent with the loss in performance by the Adaptive Minmax Processor when the actual environmental condition encountered falls outside the range of Φ. However, the results presented in this chapter indicate that when the environmental conditions fall within the range of Φ, the Adaptive Minmax Processor does not suffer a degradation in performance as the environmental conditions change. Therefore, the norm-squared of the array weight vector is an inappropriate measure of the capability of the Adaptive Minmax Processor to adjust to changing environmental conditions within the range of Φ; but is a good measure of the processor's loss of performance when the environmental conditions fall outside the range of Φ.

3.2.3 The Random Ideal Waveguide

This final set of numerical results were generated using an ideal waveguide model of the ocean and a normal mode representation of the propagating signals. The ocean is assumed to be a randomly time-variant medium with the salient parameters of the simulation being the same as in Subsection 3.2.1. That is, the deterministic component of the ocean is assumed to be an isovelocity waveguide with a sound speed

of $\overline{C} = 1500\ m/s$ and an unknown but constant depth H between 290 meters and 310 meters. The sea surface is assumed to be a free surface and the ocean bottom is assumed to be infinitely rigid. The source is assumed to be a deterministic complex exponential with a frequency of 20 Hz. In addition, the source is assumed to be sufficiently far from the array so that the horizontal propagation of the modes can be modeled by a complex exponential rather than a Hankel function.

The receiving array is also the same as in Subsection 3.2.1. It is a nine-element vertical array with the top element at 30 meters depth, the bottom element at 270 meters depth, and an inter-element spacing of 30 meters. In all cases here, there is a single source at a depth of 150 meters and at a range of 50 km from the array. There is assumed to be no propagating background noise and the localization problem is assumed to be a one-dimensional (depth only) problem.

As before, the deterministic component of the waveguide supports eight propagating modes. The mode shapes are a function of depth z and given by

$$\phi_n(z) = \sin(k_{V_n} z) \quad n = 1, \ldots, 8.$$

From (4.10) the deterministic component of the n^{th} mode of the desired signal received at the i^{th} array sensor is, to within a scale factor, given by

$$\tilde{x}_{i_n}(t) = (k_{H_n} R)^{-\frac{1}{2}} e^{j(2\pi f t - k_{H_n} R)} \sin(k_{V_n} z_i),$$

where R is the horizontal range from the source to the array sensors and z_i is the depth of the i^{th} array sensor.

The second order characterization of the reduced wavefunctions ψ_n used here differs slightly from that given in Section 4.3. The key difference is that, in Section 4.3, the modal phases are assumed to be incoherent between modes. Here, it is assumed that the phase of the reduced wavefunction of the n^{th} mode at a range R is given by $\angle\psi_n(R,0) = \theta_n(R)$ where $\theta_n(R)$ is a zero-mean Gaussian random variable with $\mathrm{E}[\theta_n(R)\,\theta_m(R)] = \frac{R}{\beta}\,\delta_{nm}$. β is the modal phase decorrelation range. Then, using the identity $\mathrm{E}[e^{j\theta}] = e^{-\frac{\sigma^2}{2}}$ for θ a zero-mean Gaussian random variable with vari-

ance σ^2, the cross-correlation function of the reduced wavefunction required for the computation of the replica vector is assumed to be given by

$$\mathrm{E}[\psi_n(R,0)\psi_m^*(R,0)] = \left(\frac{\sin(k_{V_n} z_{source})}{|\sin(k_{V_n} z_{source})|}\right)\left(\frac{\sin(k_{V_m} z_{source})}{|\sin(k_{V_m} z_{source})|}\right)(W_n(R)\,W_m(R))^{\frac{1}{2}}\, e^{-\frac{R}{\beta}(1-\delta_{nm})}.$$

$W_n(R)$ is the power in the n^{th} mode at range R and evolves according to the differential equation in (4.20).

With these assumptions and letting

$$v_i(R,n) = \left(\frac{W_n(R)}{R\,k_{H_n}}\right)^{\frac{1}{2}}\left(\frac{\sin(k_{V_n} z_{source})}{|\sin(k_{V_n} z_{source})|}\right) e^{-jk_{H_n}R}\sin(k_{V_n} z_i),$$

the ensemble cross-spectral correlation matrix of the desired signal is given by

$$\mathrm{S}_{x_{il}}(f) = \mathrm{E}[X_i(f)X_l^*(f)] = \sum_{n=1}^{8}\sum_{m=1}^{8} v_i(R,n)\, v_l^*(R,m)\, e^{-\frac{R}{\beta}(1-\delta_{nm})} \qquad (3.7)$$

The i^{th} element of the signal replica vector is, prior to normalization, equal to $\mathrm{S}_{x_{ik}}(f)$ evaluated at the appropriate set of environmental conditions. For the case where $R \gg \beta$, the modal phases are incoherent, $e^{-\frac{R}{\beta}} \approx 0$, and (3.7) reduces to

$$\mathrm{S}_{x_{il}}(f) = \sum_{n=1}^{8} v_i(R,n)\, v_l^*(R,n) = \sum_{n=1}^{8} (k_{H_n}R)^{-1}\sin(k_{V_n} z_i)\sin(k_{V_n} z_l)W_n(R),$$

which, when evaluated at $l = k$ yields (4.22). For the case where $R \ll \beta$, the modal phases are perfectly coherent, $e^{-\frac{R}{\beta}} \approx 1$, and (3.7) reduces to

$$\mathrm{S}_{x_{il}}(f) = \left(\sum_{n=1}^{8} v_i(R,n)\right)\left(\sum_{n=1}^{8} v_l(R,n)\right)^*.$$

The performance of the MVDR, the Bartlett, and the Adaptive Minmax Processors will be shown to suffer when modal phase coherence is reduced. There are several reasons for the performance loss. The first is that, as modal phase coherence is reduced, the correlation among the desired signal as received at each of the array sensors is reduced. As a result, the cross-spectral correlation matrix of the desired

signal will begin to resemble the cross-spectral correlation matrix of the propagating background noise. In the limiting case of a total loss of correlation among the desired signal as received at each of the array sensors, the cross-spectral correlation matrix of the desired signal will be the same as the cross-spectral correlation matrix of sensor noise. In this limiting case (assuming that the noise field consists of only sensor noise), the array gain

$$G(\underline{w}) = \frac{\underline{w}^h S_x(f)\, \underline{w}}{\underline{w}^h S_n(f)\, \underline{w}} \tag{3.8}$$

will be one for all array weight vectors (i.e., the array processing yields no improvement in SNR). In (3.8), the cross-spectral correlation matricies are normalized to have traces equal to one. In the more general case, as signal coherence is lost and $S_x(f)$ becomes more like $S_n(f)$, the maximum array gain achievable by any linear array processor will be reduced. This is a fundamental limitation on the performance of any array processor which utilizes a linear filter to achieve spatial discrimination.

The second cause of performance loss is particular to linear processors using the same signal model as that used by the MVDR, the Bartlett, and the Adaptive Minmax Processors. For the case where $R \ll \beta$, the cross-spectral correlation matrix of the desired signal and the signal subspace will be rank one. In addition, the normalized signal replica vector will be the eigenvector which corresponds to the non-zero eigenvalue of $S_x(f)$. However, for the case where $R \gg \beta$, the cross-spectral correlation matrix and the signal subspace can have a rank of up to M. The MVDR, the Bartlett, and the Adaptive Minmax Processors all model the signal subspace as having rank one and attempt to pass only those signals falling within that subspace. Thus, at ranges where modal coherence is reduced, and more generally for situations where signal coherence across the array is reduced, these processors can miss a significant portion of the power in the desired signal. In this case, the MVDR and the Adaptive Minmax Processors will attempt to adapt and filter out a significant portion of the desired signal. This problem is a model mismatch problem since it results in a mismatch between the assumed and the actual second-order statistics of the desired signal.

To illustrate the performance degradation of the processors as modal coherence is lost, several tests were run with different modal phase decorrelation ranges and without modal coupling. That is,

$$\left(\frac{\sin(k_{V_n} z_{source})}{|\sin(k_{V_n} z_{source})|}\right) W_n(R)^{\frac{1}{2}} = \sin(k_{V_n} z_{source}).$$

While this assumption of no modal coupling in a random waveguide is unrealistic, it is made in order to isolate the cause of the performance degradation as the loss of modal phase coherence. The introduction of modal coupling can be expected to further degrade the depth resolution of the processors because the source depth information is carried in the relative amplitudes of the modal excitations [36]; and when modal coupling is present, the modal powers will tend to an equilibrium state which may not depend on the initial levels of modal excitation. For example, in the case where all of the modal coupling parameters a_{nm} defined in (4.21) equal one, the equilibium state of the modal powers is $W_n(R) = W_m(R)$ for all n and m. In this case, the processor will not have depth resolution once the equilibrium state is reached.

As a measure of the extent to which the modal phase decorrelation affects the distribution of signal power among the eigenvectors of the cross-spectral correlation matrix of the desired signal, Figure 3-27 shows the relationship between the maximum eigenvalue of the normalized cross-spectral correlation matrix expressed in dB (i.e., $10 * \log_{10}(\lambda_{max})$) and the extent of the modal phase decorrelation as measured by $e^{-\frac{R}{\beta}}$. $e^{-\frac{R}{\beta}} = 1$ indicates perfectly correlated modal phases and $e^{-\frac{R}{\beta}} = 0$ indicates uncorrelated modal phases. The cross-spectral correlation matrix is normalized so that the sum of the eigenvalues equals one. This maximum eigenvalue is the maximum signal power which exists in any rank one subspace. As the maximum eigenvalue decreases, the signal power lost by any linear processor which assumes a rank one signal subspace will increase.

Figures 3-28 through 3-31 show the ambiguity functions for the MVDR, the Bartlett, and the Adaptive Minmax Processors for the cases where $\beta = \infty$, $2R$,

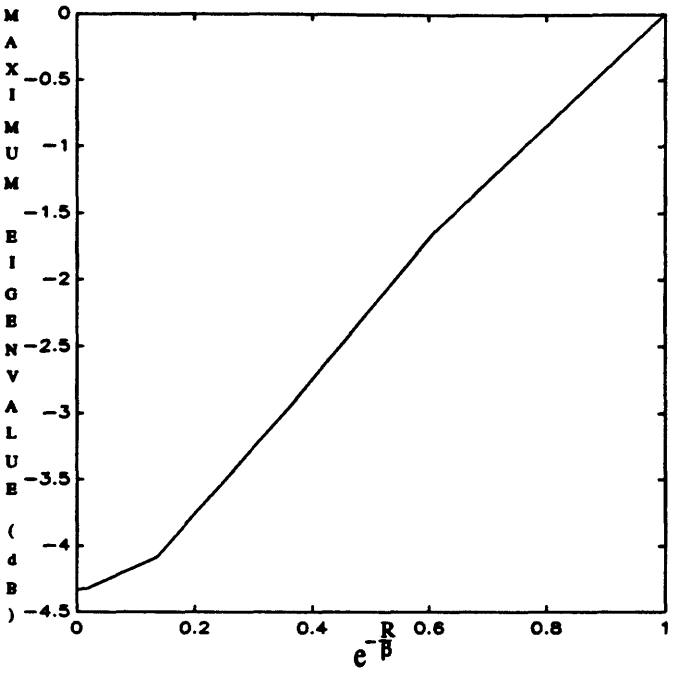


Figure 3-27: Maximum Cross-Spectral Correlation Matrix Eigenvalues vs Modal Phase Decorrelation

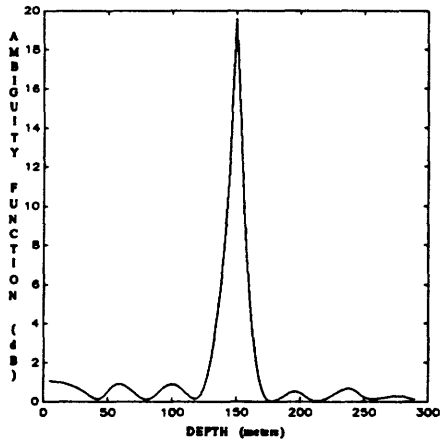
R, and 0 ($e^{-\frac{R}{\beta}} = 1$, .607, .368, and 0, respectively). $\beta = \infty$ corresponds to the deterministic ideal waveguide, the analysis of which is presented in Subsection 3.2.1. $\beta = 0$ corresponds to the modal phase decorrelation model assumed in [32]. All of the processors use the correct values for β. The ocean depth in all cases is 310 meters. In the figures, the terms "Matched" and "Mismatched" refer to the assumed ocean depth used by the processors in calculating the signal replica vectors. The "Matched" processors use the correct value of the ocean depth. The "Mismatched" processors use an assumed ocean depth of 290 meters. The localization performance of the Matched MVDR Processor, the Matched Bartlett Processor, and the Adaptive Minmax Processor shows a significant deterioration as the modal phase correlation decreases.

The evolution of the ambiguity functions of the Matched MVDR, the Matched Bartlett, and the Adaptive Minmax Processors as $\beta \rightarrow 0$ is characteristic of what could be expected with increasing model mismatch. The model mismatch is not caused by an incorrect assumed value of the ocean depth or of β, but by an incorrect

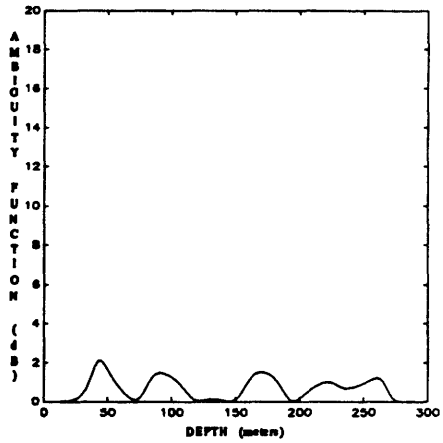
implicit model of the signal subspace as having rank one. The ambiguity functions of these processors in Figures 3-29 through 3-31 show reduced peak levels when compared to those shown in Figure 3-28 and show peak locations which are slightly offset from the true source depth of 150 meters.

The model mismatch problems which the MVDR, the Bartlett, and the Adaptive Minmax Processors have in the random waveguide motivates a modification to the definition of the replica vector which slightly improves the performance of the processors in random media. The modification is motivated by the observation that the rank one signal subspace containing the greatest amount of signal power is that spanned by the eigenvector of the cross-spectral correlation matrix of the desired signal which corresponds to the largest eigenvalue. Thus, the replica vector is redefined to be the eigenvector which corresponds to the largest eigenvalue. The magnitude normalization convention applied to the replica vector is that it has a norm of one, and the phase normalization convention is that the element corresponding to the reference sensor is real and non-negative. In the case where the signal subspace has rank one, this definition of the replica vector is equivalentto the definition of the replica vector contained in Subsection 3.1.3.

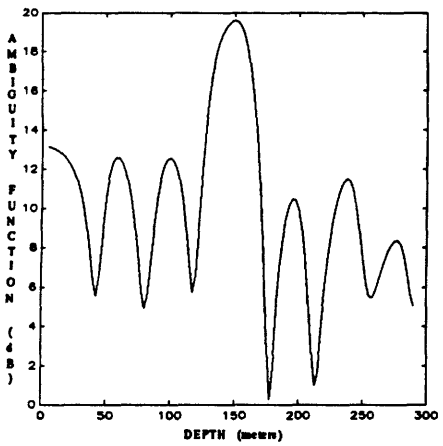
Figures 3-32 through 3-35 show the ambiguity functions generated by processors using this new definition of the replica vector for test conditions which are identical to those used to generate Figures 3-32 through 3-35. As was the case with the previous set of results generated using the definition of the replica vector from Subsection 3.1.3, the ambiguity function of the Adaptive Minmax Processor shows characteristics which are superior to either of the mismatched processors. For the Matched MVDR Processor, the Matched Bartlett Processor, and the Adaptive Minmax Processor, the peak levels of the ambiguity functions are greater than those in the previous set of results. With the exception of the case of the Adaptive Minmax Processor for $\beta = 0$, the peaks of the ambiguity functions occur at the true source depth. The peak in the ambiguity function of the Adaptive Minmax Processor for the case where $\beta = 0$ occurs at a depth of 147.5 meters compared with the true source depth of 150 meters. These results indicate a reduction in the "mismatch" caused by the implicit rank one



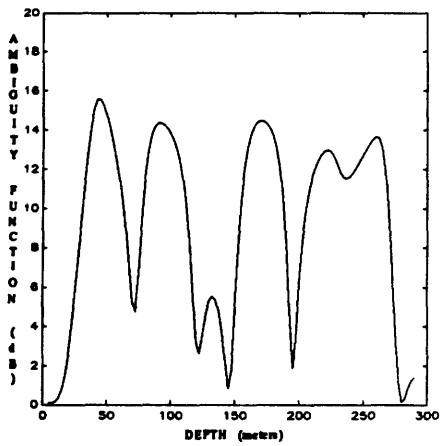
(a) Matched MVDR Processor



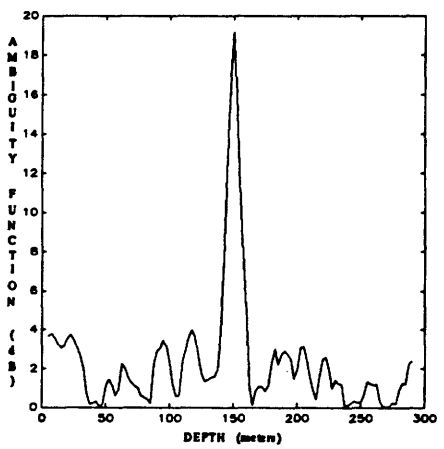
(b) Mismatched MVDR Processor



(c) Matched Bartlett Processor

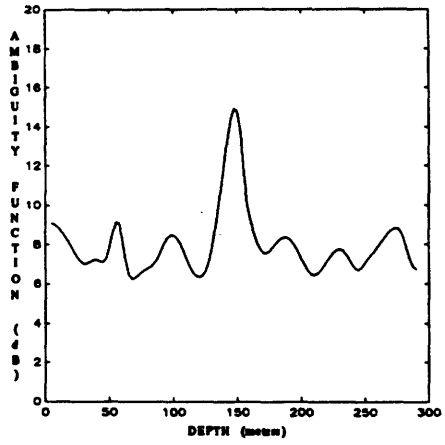


(d) Mismatched Bartlett Processor

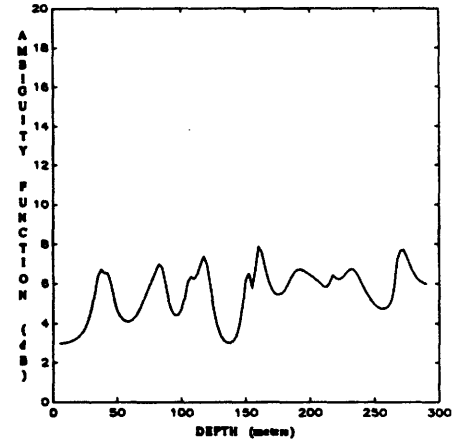


(e) Adaptive Minmax Processor

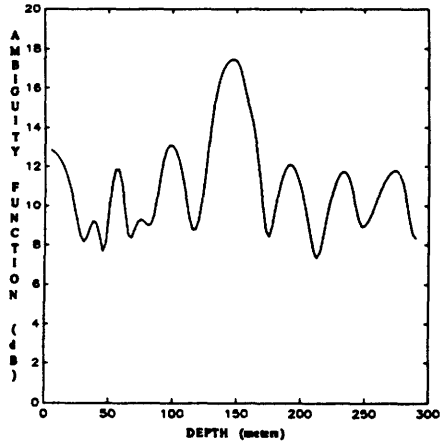
Figure 3-28: Ambiguity Functions for $\beta = \infty$



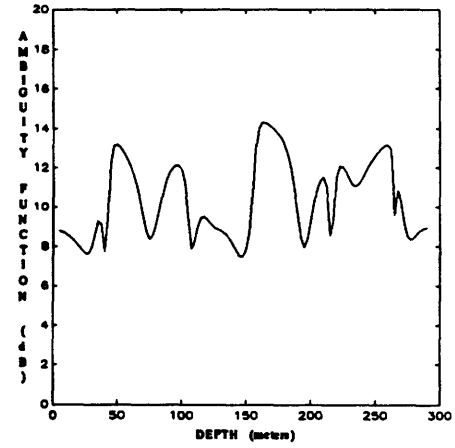
(a) Matched MVDR Processor



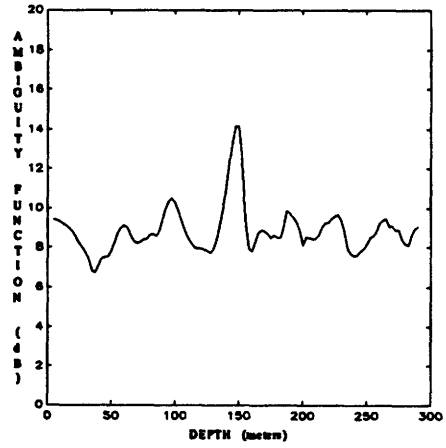
(b) Mismatched MVDR Processor



(c) Matched Bartlett Processor

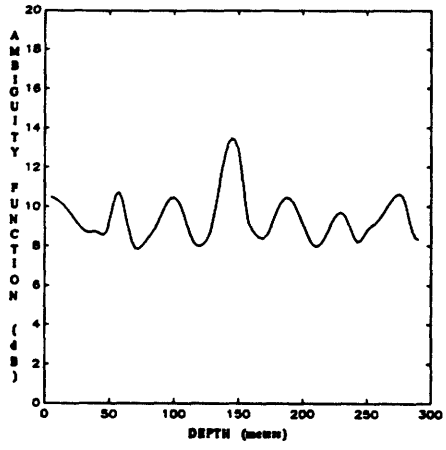


(d) Mismatched Bartlett Processor

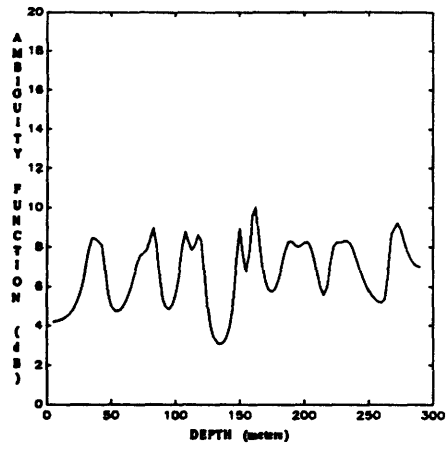


(e) Adaptive Minmax Processor

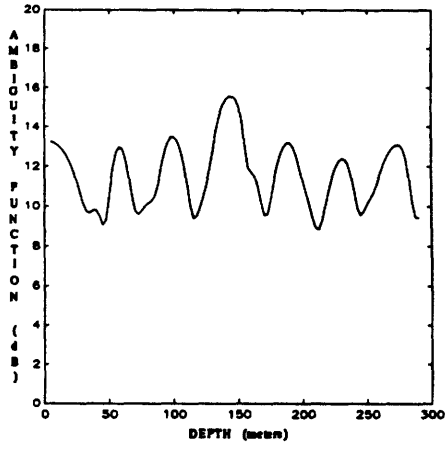
Figure 3-29: Ambiguity Functions for $\beta = 2R$



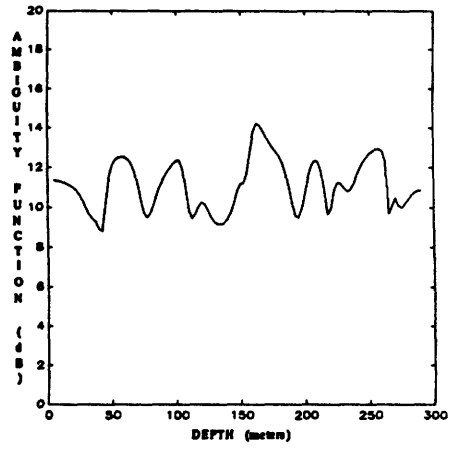
(a) Matched MVDR Processor



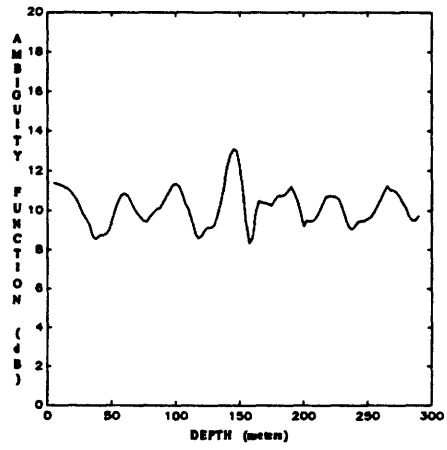
(b) Mismatched MVDR Processor



(c) Matched Bartlett Processor

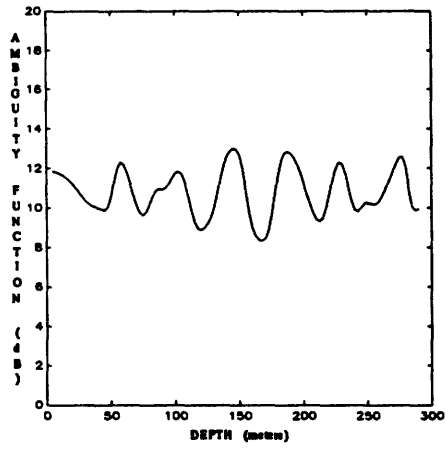


(d) Mismatched Bartlett Processor

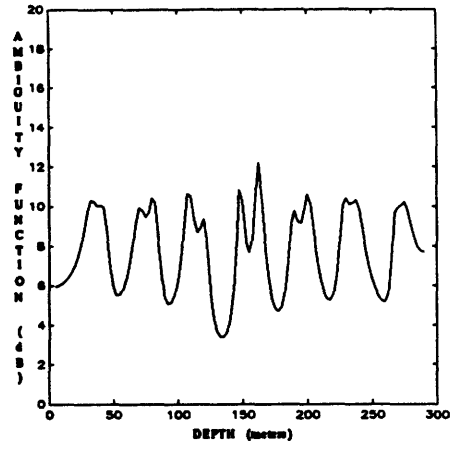


(e) Adaptive Minmax Processor

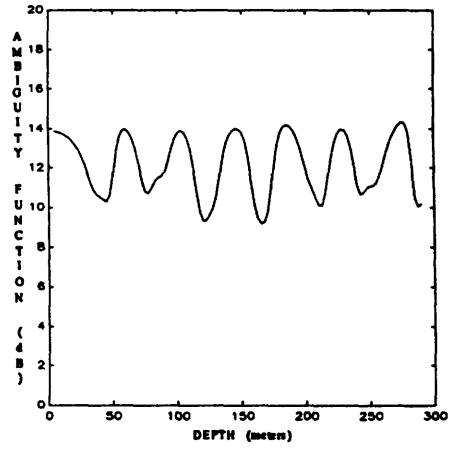
Figure 3-30: Ambiguity Functions for $\beta = R$



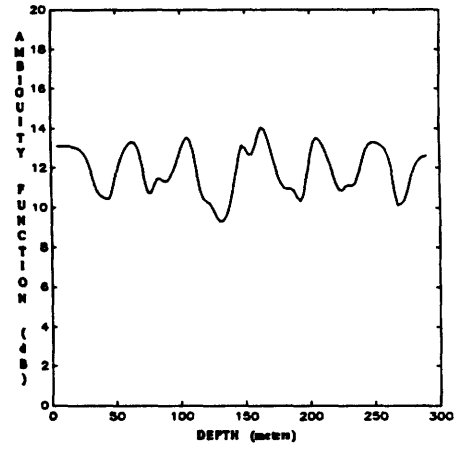
(a) Matched MVDR Processor



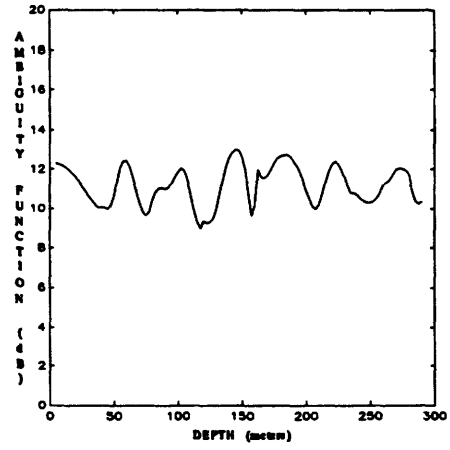
(b) Mismatched MVDR Processor



(c) Matched Bartlett Processor



(d) Mismatched Bartlett Processor



(e) Adaptive Minmax Processor

Figure 3-31: Ambiguity Functions for $\beta \approx 0$

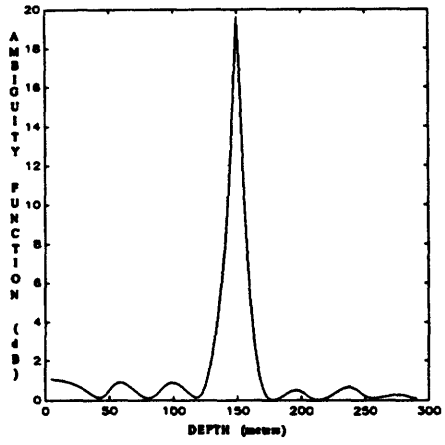
signal subspace assumption.

The peaks of the ambiguity functions of the Matched MVDR and Bartlett Processors for the cases where $\beta < \infty$, when compared to the peak value for the case where $\beta = \infty$, show a loss which is, to within 0.08 dB, equal to the reduction in the magnitude of the maximum eigenvalue of the normalized cross-spectral correlation matrix for the appropriate values of β (Figure 3-27). This indicates that these processors are correctly estimating the power contained in the rank one signal subspace spanned by the replica vector. However, the peak value losses for the Adaptive Minmax Processor are greater than those which would be predicted by the reduction in the maximum eigenvalue of the cross-spectral correlation matrix. This increased loss can be explained using the results from Subsection 3.2.1 for the deterministic ideal waveguide. Figure 3-16 shows the source depth/ocean depth response evaluated at the true source depth and ocean depth of the weights of the Adaptive Minmax Processor when the focal point is the source depth. This response drops in magnitude as the SNR decreases. In this case, the decrease in the power in the rank one signal subspace spanned by the eigenvector corresponding to the maximum eigenvalue and the corresponding increase in the power in the other portions of the signal subspace is interpreted by the processor as a decrease in SNR. Therefore, the magnitude response of the processor at the source location when the processor is focused on the source location drops. The peak in the ambiguity function will exhibit a corresponding drop.

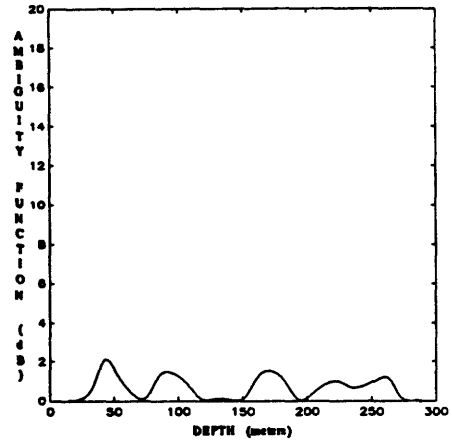
Sidelobe suppression remains a significant problem in all of the processors. This indicates that the eigenvector corresponding to the maximum eigenvalue of the cross-spectral correlation matrix is an ambiguous characterization of the spatial structure of the desired signal as a function of source depth for cases where the signal subspace has a rank greater than one. Future work on this problem is discussed in Chapter 5.

3.3 Algorithm Complexity

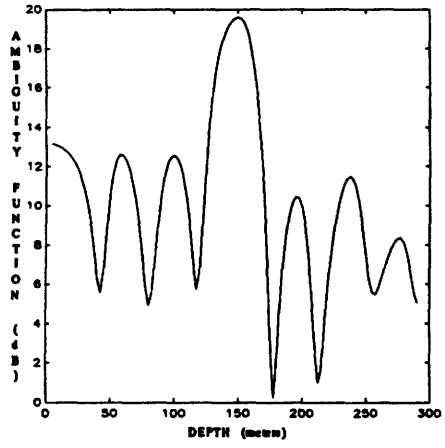
The solution of the minmax problem in Step 1 of the Adaptive Minmax Matched Field Processor is by far the most computationally complex of the tasks which must



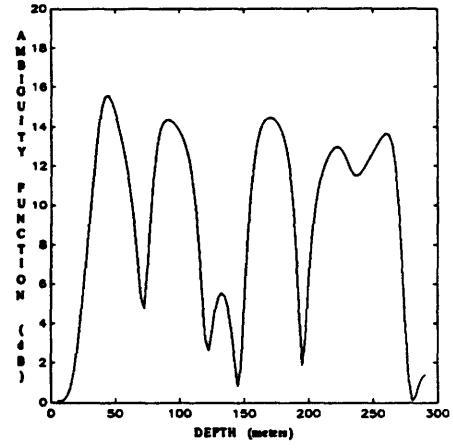
(a) Matched MVDR Processor



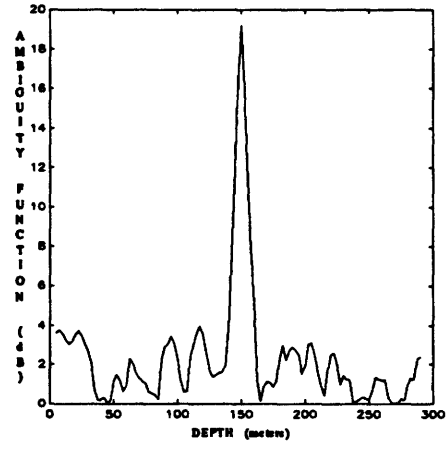
(b) Mismatched MVDR Processor



(c) Matched Bartlett Processor

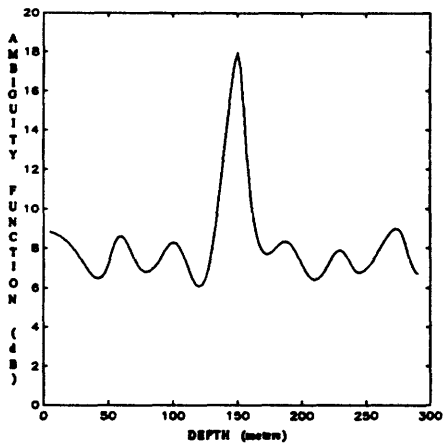


(d) Mismatched Bartlett Processor

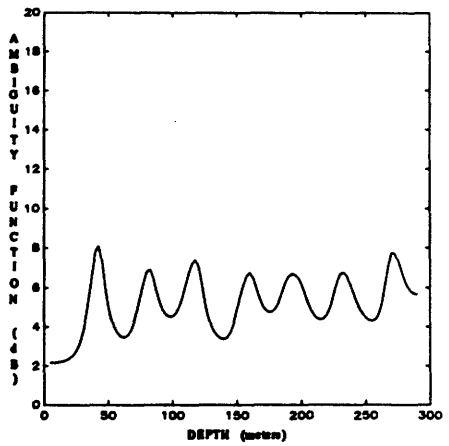


(e) Adaptive Minmax Processor

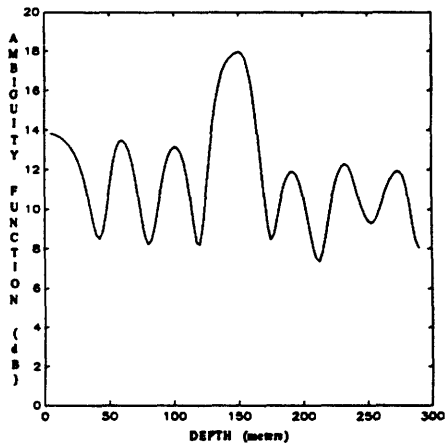
Figure 3-32: Ambiguity Functions for $\beta = \infty$ with redefined Replica Vector



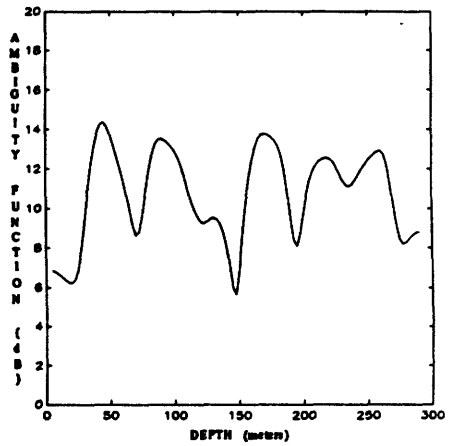
(a) Matched MVDR Processor



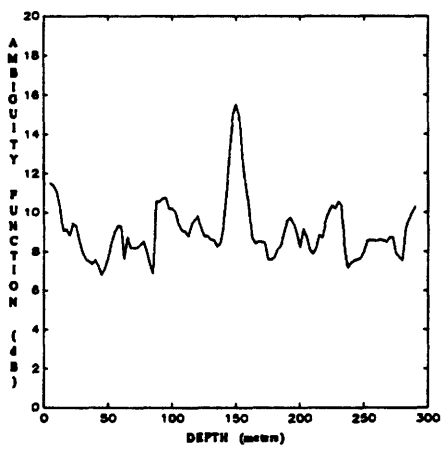
(b) Mismatched MVDR Processor



(c) Matched Bartlett Processor

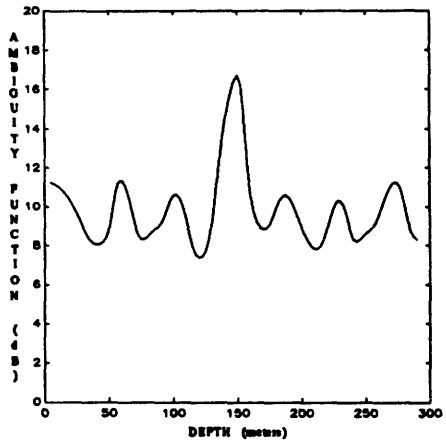


(d) Mismatched Bartlett Processor

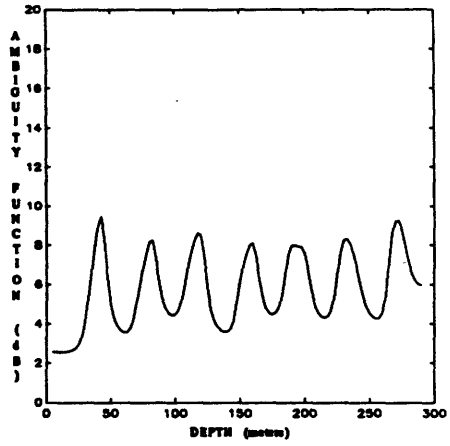


(e) Adaptive Minmax Processor

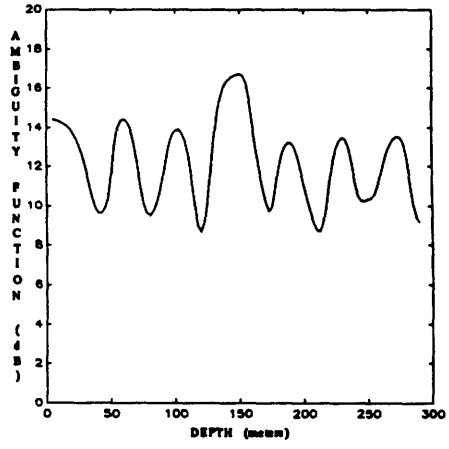
Figure 3-33: Ambiguity Functions for $\beta = 2R$ with redefined Replica Vector



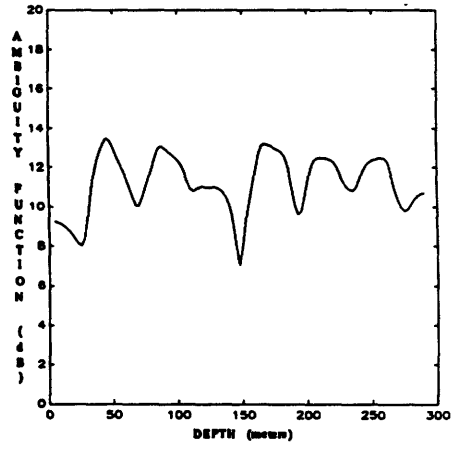
(a) Matched MVDR Processor



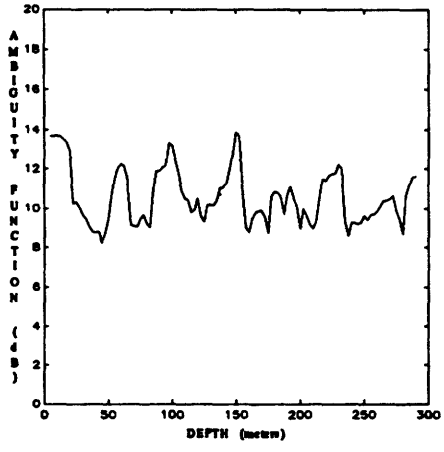
(b) Mismatched MVDR Processor



(c) Matched Bartlett Processor

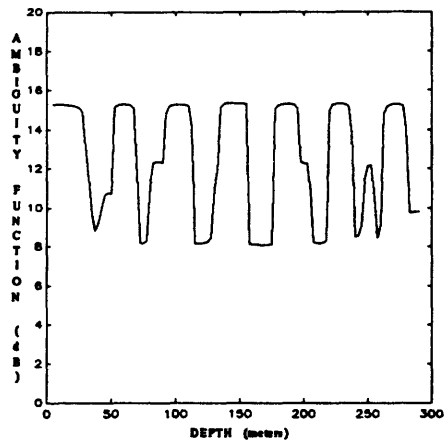


(d) Mismatched Bartlett Processor

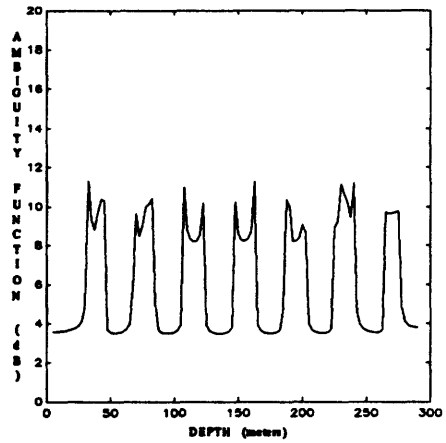


(e) Adaptive Minmax Processor

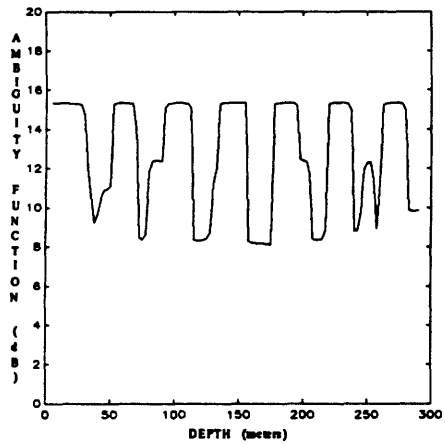
Figure 3-34: Ambiguity Functions for $\beta = R$ with redefined Replica Vector



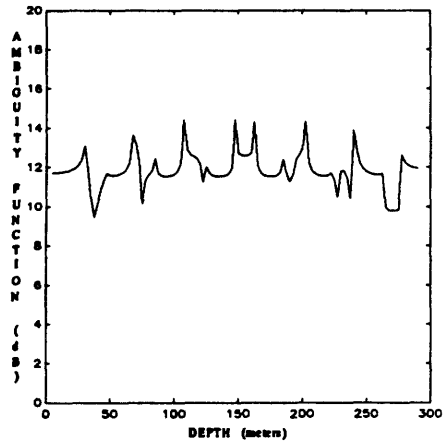
(a) Matched MVDR Processor



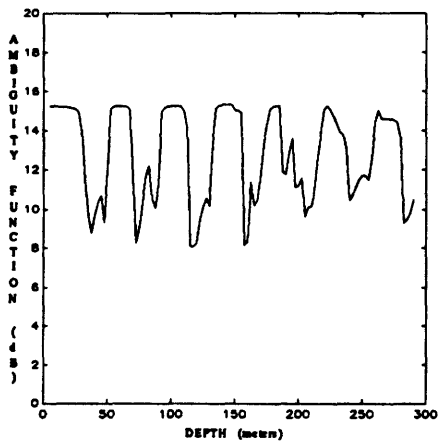
(b) Mismatched MVDR Processor



(c) Matched Bartlett Processor



(d) Mismatched Bartlett Processor



(e) Adaptive Minmax Processor

Figure 3-35: Ambiguity Functions for $\beta \approx 0$ with redefined Replica Vector

K	11	22	33	44
min	2	3	4	5
max	18	32	38	46
ave	8.12	13.62	17.56	20.49

Table 3.3: Iterations of Complementary Pivot Algorithm vs K

be performed. From Section 2.3, this problem is solved by finding the nonnegative vectors $\tilde{p}$ and $\tilde{v}$ contained in $\mathbb{R}^K$ (where K is the number of points in the sampled environmental parameter set Φ) and the real nonnegative real scalar z which satisfy

$$\begin{aligned} &\min z_o \text{ s.t.} \\ z_0,\ \tilde{v},\ \tilde{p} &\geq 0\ , \\ \tilde{v}^t \tilde{p} &= 0\ , \text{ and} \\ \tilde{v} - M(f,\underline{z})\,\tilde{p} - \underline{d}\,z_o &= \underline{r}(f,\underline{z}) + I\,\underline{\delta}. \end{aligned}$$

During numerical tests of the Adaptive Minmax Matched Field Processor using the deterministic and random ideal waveguide acoustic propagation models, the number of iterations required to solve the minmax problem were tabulated. These tests were not Monte Carlo simulations. The tests were conducted for $K = 11,\ 22,\ 33,$ and 44. The results in Table 3.3 show the minimum, the maximum, and the average number of iterations required to find a solution to the minmax problem. These results are based on 4830 trials.

The upper bound on the number of required iterations can be explicitely calculated. By construction of the complementary pivot algorithm, only one variable from each complementary pair can be in the basis at any given time. With the exception of the termination of the algorithm when the variable z leaves the basis, there will be one and only one complementary pair which has neither variable in the basis. At termination, each complementary pair will have one variable in the basis and one variable outside the basis. Therefore, since there are K complementary pairs of variables, there are $K2^{(K-1)}$ possible non-terminating bases.

From Section 2.3, two of the essential points of the convergence proof for the complementary pivot algorithm [26] are that, at each iteration the basis changes and no basis can be visited by the algorithm more than once. Therefore, since there are only $K2^{(K-1)}$ possible non-terminating bases, this number is also an upper bound on the number of iterations required to find a solution.

This number shows an exponential growth in complexity with the number of environmental states. This is an upper bound on the number of iterations, and does not accurately reflect the actual complexity of the algorithm. Treating $M(f, \underline{z})$ and $\underline{r}(f, \underline{z})$ as random variables with some restrictions on their distributions, the results presented in Section 2.5 of [27] place an upper bound on the mean number of iterations required to find a solution. This upper bound is $\frac{K(K+1)}{4}$. This indicates $\mathcal{O}(K^2)$ growth in the expected number of iterations required to find a solution. When compared to results in Table 3.3, the upper bounds on both the maximum number of iterations and the mean number of iterations appear to be very conservative.

Finally, the results shown above use the required number of iterations to find a solution as a measure of the complexity of the complementary pivot algorithm. Of equal importance is the amount of computation required for each iteration. The bulk of the computation in each iteration of the algorithm involves computing K ratios to determine which variable will leave the basis at that iteration, and implementing the Gaussian elimination which adds one variable to the basis and removes another variable from the basis which requires K multiplications and $2K^2$ additions. Therefore, each iteration requires approximately $2K^2$ additions and $2K$ multiplications.

Chapter 4

Matched Field Calculation of the Signal Replica Vector

Matched field calculation of the signal replica vector requires that the cross-correlation functions of the signals received at each of the array sensors be calculated taking into account the propagation characteristics of the ocean environment. These characteristics are parameterized by the statistics of the temporally and spatially varying sound speed structure of the ocean $C(\underline{z},t)$. This chapter does not present original work on the development of methods to calculate the required cross-correlation functions given the statistics of $C(\underline{z},t)$. Rather, it explains how the theory and the methods developed by others can be applied to the calculation of the required cross-correlation functions.

In general, for each region in the oceans, season, and time of day, it is fairly accurate to model $C(\underline{z},t)$ as the sum of a time-invariant equilibrium component and a time-varying stochastic component. The equilibrium component is a function of only the depth. The stochastic component is a function of the three-dimensional location and the time [30, 32]. That is,

$$C(\underline{z},t) = C_o(1 + U_o(z) + \mu(\underline{z},t)),$$

where z denotes depth, $\underline{z}$ denotes three-dimensional location, and C_o is a reference

sound speed. $\overline{C}(z) \triangleq C_o(1 + U_o(z))$ is the equilibrium (or deterministic) component of the sound speed structure. This component is dominated by the seasonally and geographically varying temperature and salinity profiles of the ocean and the constant pressure gradient of the oceans. $C_o\mu(\underline{z}, t)$ is a zero-mean stochastic process capturing the random component of the sound speed structure. In the open oceans, $C_o\mu(\underline{z}, t)$ is dominated by the internal waves propagating in the oceans [30] and the tides [31]. It is assumed that $\mu(\underline{z}, t)$ is a temporally wide-sense stationary process.

The relationship between the signal replica vector and the spatial/temporal cross-correlation function of the desired signal is derived in Section 4.1. There are a number of methods available for calculating the spatial/temporal correlation function of the desired signal given $\overline{C}(z)$ and the second-order statistics of $C_o\mu(\underline{z}, t)$. One of these methods utilizes the ray approximation to the solution of the wave equation and is described in Section 4.2. Another method uses the normal mode approximation to the solution of the wave equation and is described in Section 4.3. A third method, which is to numerically solve the wave equation without approximations, is briefly described in Section 4.4.

4.1 The Spatial/Temporal Cross-Correlation Function

For the three definitions of the signal replica vector contained herein, the signal replica vector depends on only the conditional a priori cross-correlation of the snapshots of the desired signal as received at different array sensors. That is,

$$\mathrm{E}[X_i(f, \underline{z}) X_k^*(f, \underline{z}) \mid \underline{\phi}],$$

where the superscript l denoting the DTFT segment number has been dropped. From Section 2.2

$$X_i(f) = \sum_{m=0}^{M-1} x_i[m, \underline{z}] e^{-j2\pi f m \Delta_t},$$

and $\underline{x}^l[m, \underline{z}]$ is a multi-dimensional temporally wide-sense stationary random process. Ignoring the distortion to the desired signal introduced by the continuous-time anti-

aliasing filter and the temporal sampling process, the cross-correlation function can be expressed as

$$\mathrm{E}[X_i(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}] = \sum_{s=-(M-1)}^{(M-1)} (M- \mid s \mid)e^{-j2\pi f s\Delta_t}\mathrm{E}[x_i(s\Delta_t,\underline{z})x_k^*(0,\underline{z}) \mid \underline{\phi}], \quad (4.1)$$

where $x_i(t,\underline{z})$ is the continuous-time desired signal received at the i^{th} array sensor. Therefore, the calculation of the signal replica vector requires calculating the spatial/temporal correlation function of the desired signal at the array sensors.

From Subsection 2.2.1, the point source at the array focal point is assumed to be a wide-sense stationary random process. Let $\mathrm{E}[x_i(s\Delta_t,\underline{z})x_k^*(0,\underline{z}) \mid \underline{\phi}, f_o]$ be the cross-correlation function on the righthand side of (4.1) for the case where the point source emits a deterministic complex exponential $e^{j2\pi f_o t}$. Let $P_x(f)$ be the power spectral density of the wide-sense stationary point source. Then, assuming that the ocean is a linear acoustic medium, the spatial/temporal cross-correlation function in (4.1) for the wide-sense stationary point source is

$$\mathrm{E}[x_i(s\Delta_t,\underline{z})x_k^*(0,\underline{z}) \mid \underline{\phi}] = \int_{-\infty}^{\infty} P_x(f_o)\,\mathrm{E}[x_i(s\Delta_t,\underline{z})x_k^*(0,\underline{z}) \mid \underline{\phi}, f_o]\,df_o.$$

Therefore, (4.1) can be rewritten as

$$\mathrm{E}[X_i(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}] = \int_{-\infty}^{\infty} P_x(f_o) \sum_{s=-(M-1)}^{(M-1)} (M- \mid s \mid)e^{-j2\pi f s\Delta_t}\mathrm{E}[x_i(s\Delta_t,\underline{z})x_k^*(0,\underline{z}) \mid \underline{\phi}, f_o]df_o. \quad (4.2)$$

That is, the cross-correlation function of the snapshots of the desired signal can be calculated by integrating the individual contributions made by sources emitting deterministic complex exponentials over an appropriate range of frequencies. Therefore, for the remainder of this chapter, it will be assumed that the signal emitted by the point source at the array focal point is the deterministic complex exponential $e^{j2\pi f_o t}$

and the development will focus on calculating

$$\mathrm{E}[X_i(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}, f_o] = \sum_{s=-(M-1)}^{(M-1)} (M- \mid s \mid) e^{-j2\pi f s \Delta_t} \mathrm{E}[x_i(s\Delta_t,\underline{z})x_k^*(0,\underline{z}) \mid \underline{\phi}, f_o]. \tag{4.3}$$

As a practical matter, the frequency selectivity of the DTFT makes it possible to closely approximate (4.2) by considering the frequencies f_o in a small neighborhood of f, the extent of which is roughly inversely proportional to the segment length of the DTFTs and Δ_t. The dependence of $\mathrm{E}[X_i(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}, f_o]$ on the quantity $(f - f_o)$ is explicitly shown in (4.7), (4.8), (4.22), and (4.24). These equations provide some insight into the extent of the neighborhood of f over which it is necessary to evaluate $\mathrm{E}[X_i(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}, f_o]$.

Given that the deterministic ocean component is modeled as being horizontally-stratified, the sound field received at any sensor which would have propagated through the deterministic ocean can be assumed to have propagated completely within the vertical plane containing both the source and the sensor. Any location within this plane will be denoted by its horizontal range from the source r and its depth z. The horizontal distance from the source to the i^{th} array sensor will be denoted by R_i. Throughout this section, the dependence of the desired signal on the source location $\underline{z}$ will be assumed to be understood and the source location argument will be dropped from the desired signal $\underline{x}(t,\underline{z})$. In addition, for the remainder of this section, the dependence of the calculations required to compute the expectations in (4.3) on the assumed value of the environmental parameter $\underline{\phi}$ will be assumed to be understood and the environmental parameter will be dropped from the conditional expectation notation.

4.2 Ray Approximation

The ray treatment of the effects of the deterministic and stochastic components of the sound speed structure on the propagation of sound is developed in [30]. The ray approximation models the behavior of the propagating sound field by assuming that

the direction of propagation at any point is in a direction normal to the wavefront at that point. This approximation is only valid at sufficiently high frequencies, the definition of which depends on the scale size of the sound speed fluctuations in the ocean. In general, the smaller the scale size, the higher the frequency necessary in order for this approximation to be valid.

At frequencies for which the ray approximation is valid, the sound field is modeled as propagating along paths which obey Snell's law. The sound field received at any point can then be expressed as the sum of the fields which propagated along the possibly multiple paths which obey Snell's Law and connect the source to the receiver. Here, the term ray will denote a path from the source to any location of interest, which obeys Snell's Law and is calculated using the deterministic sound speed structure $\overline{C}(z)$. The term path will be used to refer to an actual path followed by the sound as it propagates from the source to any location of interest calculated using the true sound speed structure $C(\underline{z},t)$. A path results from the perturbations to a ray caused by the fluctuating ocean processes modeled by $\mu(\underline{z},t)$.

The paths which result from the perturbations to a particular ray lie within a ray tube surrounding the ray. The size of this ray tube depends upon the strength and scale length of the random process $\mu(\underline{z},t)$. Following the convention, but not the notation of [30], the received signal which propagated through the n^{th} ray tube from the source to the i^{th} array sensor can be expressed as the product of the signal which would have propagated along the n^{th} ray in an unperturbed deterministic ocean and a **reduced wavefunction** which is a stochastic process dependent on $\mu(\underline{z},t)$) along the propagation paths within the n^{th} ray tube. That is,

$$x_i(t) \;=\; \sum_{n=1}^{N_i} x_{i_n}(t),$$

and

$$x_{i_n}(t) \;=\; \tilde{x}_{i_n}(t)\,\psi_{i_n}(t).$$

$\tilde{x}_{i_n}(t)$ is the signal emitted by the source which would have been received at the i^{th} array sensor via the n^{th} ray in a deterministic ocean. N_i is the number of rays

connecting the source with the i^{th} array sensor. $\psi_{i_n}(t)$ is the reduced wavefunction which accounts for random perturbations in $x_{i_n}(t)$ due to the random perturbations of the sound speed structure along the possibly multiple time-varying paths within the n^{th} ray tube.

The ray approximation developed in [30] treats the deterministic ocean as a linear time-invariant system. Therefore, $\tilde{x}_{i_n}(t)$ is the product of $e^{j2\pi f_o t}$ and a complex scaling constant which accounts for the losses and the time delay along the n^{th} ray between the source and the i^{th} sensor. $\psi_{i_n}(t)$ can be considered a modulating signal which accounts for the Doppler spread introduced by the random ocean component.

The spatial/temporal correlation functions required in (4.3) can be expressed as

$$\mathrm{E}[x_i(s\triangle_t)x_k^*(0) \mid f_o] = \sum_{n=1}^{N_i}\sum_{m=1}^{N_k} \tilde{x}_{i_n}(s\triangle_t)\,\tilde{x}_{k_m}^*(0)\,\mathrm{E}[\psi_{i_n}(s\triangle_t)\psi_{k_m}^*(0) \mid f_o]. \qquad (4.4)$$

The expectation operator has been removed from the first two factors of each term in (4.4) because they are not stochastic processes. These factors can be evaluated by calculating the delay and attenuation introduced along each ray by the horizontally-stratified deterministic ocean. The second factor in each term can be evaluated by calculating the statistics of the fluctuations in $x_{i_n}(t)$ caused by the random ocean component $\mu(\underline{z},t)$ along the n^{th} ray.

The calculation of $\tilde{x}_{i_n}(s\triangle_t)$ required in (4.4) is conceptually straightforward. The only rays which are considered are those which are fully refracted within the ocean (i.e.,those which do not have surface or bottom interaction). Therefore, the rays considered are those whose angle with respect to the horizontal is always fairly small (the cutoff of 15^o is used in [30]).

For long-range propagation, the exclusion of bottom-reflected rays is reasonable, because the high losses which usually occur during bottom interaction prevent the rays having such interaction from contributing much energy to the received signal. A slight modification to the techniques developed in [30], which makes possible the inclusion of surface-reflected rays, is presented at the end of this section.

Given the small-angle assumption, Snell's Law relating the depth of the n^{th} ray

connecting the source to the i^{th} sensor at a range r from the source is

$$\frac{\partial^2 z_{i_n}(r)}{\partial r^2} + \frac{\frac{\partial U_o(z_{i_n}(r))}{\partial z}}{1 + U_o(z_{i_n}(r))} = 0,$$
$$z_{i_n}(0) = z_{source}, \text{ and}$$
$$\frac{\partial z_{i_n}(r)}{\partial r}\Big|_{r=0} = \theta_{i_n}(0).$$

$\theta_{i_n}(0)$ is the initial angle of the ray with respect to the horizontal, z_{source} is the source depth, and $z_{i_n}(r)$ is the depth of the n^{th} ray at the range r. If the vertical separation of the source and the i^{th} array sensor is small compared to their horizontal separation R_i, the acoustic path length along the n^{th} ray can be reasonably approximated by $R_i + S_{i_n}$ where

$$S_{i_n} = \int_0^{R_i} \frac{1}{2}\left(\frac{\partial z_{i_n}(r)}{\partial r}\right)^2 - U_o(z_{i_n}(r))\, dr,$$

and S_{i_n} is called the eikonal. Then, $\tilde{x}_{i_n}(s\triangle_t)$ can be expressed as

$$\tilde{x}_{i_n}(s\triangle_t) = R_i^{-\frac{1}{2}} K_{i_n} e^{j(2\pi f_o s\triangle_t - k_o(R_i + S_{i_n}))}, \tag{4.5}$$

where $k_o = 2\pi f_o / C_o$ is the nominal wavenumber of the signal and K_{i_n} is the normalization factor mentioned on page 83 and discussed in Part IV of [30]. This calculation can be carried out for all rays connecting the source to each of the array sensors. Here, the signal attenuation due to absorption by the medium has been ignored. If the difference in the absorption along different rays may be significant, an absorption factor can be included in (4.5).

The calculation of the second-order statistics of $\psi_{i_n}(s\triangle_t)$, which are required to evaluate (4.4), is much more involved. To begin, several parameters used to characterize the behavior of the medium within each ray tube are introduced. The first is the Fresnel-zone radius, $R_F(r)$, at the horizontal distance r from the source. Here, the indices i and n, denoting the endpoint of the ray and the ray number, are dropped. Intuitively, $R_F(r)$ is a measure of the radius of a ray tube through which a path could pass without having its acoustic path length differ from the acoustics path length of the ray by more than half a wavelength. More precisely, $R_F(r)$ is defined as the

maximum distance such that for any point $p_o = (z_o, r)$ for which $|z_o - z_{i_n}(r)| \leq R_F(r)$, the following is condition holds. As the point p is moved from $(z_{i_n}(r), r)$ to p_o, let the n^{th} ray from the source to the i^{th} array sensor be continuously deformed into two rays, one from the source to p and the other from p to the i^{th} array sensor. Then, the sum of the acoustic path lengths of the two rays must always differ from the acoustic path length of the n^{th} ray by less than half a wavelength.

The next parameter is a measure of the strength of the sound speed fluctuations modeled by $\mu(r, z, t)$ along any ray where, as mentioned before, the location parameters r and z denote the horizontal range from the source and the depth in a vertical plane containing the ray of interest. Let $T_{i_n}(r) = 1 + \frac{1}{2}\left(\frac{dz_{i_n}(r)}{dr}\right)^2$. Then, using the small-angle approximation made possible by the exclusion of the rays which are not fully refracted, this strength parameter Φ_{i_n} (not to be confused with the set of environmental conditions Φ defined earlier) is defined as

$$\Phi_{i_n}^2 \triangleq \mathrm{E}\left[\left(\frac{2\pi f_o}{C_o}\int_0^{R_i} T_{i_n}(r)\mu(r, z_{i_n}(r), t)\, dr\right)^2\right],$$

where the dependence of Φ_{i_n} on f_o is not shown in the notation. For the cases where the signal propagates through the ray tube via only a single path, this strength parameter is the mean squared phase fluctuation in $x_{i_n}(t)$.

The final parameter is a measure of the diffraction caused by the spatial extent of sound speed fluctuations along any ray. This diffraction parameter Λ_{i_n} is a weighted average of $(\frac{R_F}{L})^2$ along the ray and can be calculated as described in Chapter 7 of [30]. Here, R_F is the Fresnel-zone radius defined earlier and L is the scale length of the stochastic process μ. For large Λ_{i_n}, the sound field propagating within a ray tube of radius R_F may see very different sound speed structures within any cross-section of the tube. The sound field will be diffracted, resulting in many paths within the ray tube. For small Λ_{i_n} and small Φ_{i_n}, the sound field propagating within a ray tube of radius R_F will pass through roughly the same sound speed structure within any cross-section of the tube and will propagate along only a single path within the tube.

With these dofomed parameters, the $\Lambda_{i_n}\Phi_{i_n}$ space is divided into three regions.

The first, referred to as the saturated region, is defined as the region where $\Lambda_{i_n}\Phi_{i_n} > 1$ and $\Phi_{i_n} > 1$. In this region, the sound speed perturbations are of sufficient strength and with a sufficiently small scale length so that the acoustic field propagating in the ray tube follows multiple paths, called micromultipaths. In addition, the micromultipaths are spread over a sufficiently large area so that the sound speed perturbations along each of the different micromultipaths are uncorrelated. The second region, referred to as the unsaturated region, is defind as the region where $\Lambda_{i_n}\Phi_{i_n}^2 < 1$ or $\Phi_{i_n} < 1$. In this region, the scale length of the sound speed perturbations is large enough or the strength of these perturbations is small enough so that the acoustic field continues to propagate along a single path within the ray tube. The final region, referred to as the partially-saturated region, is defined as the region where $\Lambda_{i_n}\Phi_{i_n}^2 > 1$ and $\Lambda_{i_n}\Phi_{i_n} < 1$. In this region, the sound field follows several different micromultipaths along which the sound speed perturbations are correlated.

The reduced wavefunction for any particular ray can be expressed as

$$\psi_{i_n}(t) = \int d(\text{paths})[e^{-j\int_0^{R_i} \frac{2\pi f_o}{C(r,z_{\text{path}}(r),t_{\text{path}}(r))} T_{\text{path}}(r)\mu(r,z_{\text{path}}(r),t_{\text{path}}(r))\,dr}],$$

where the first integral is over all micromultipaths in the ray tube, $z_{\text{path}}(r)$ is the depth of the particular micromultipath at the horizontal range r from the source, and $t_{\text{path}}(r)$ is the time at which the sound field received at time t at the i^{th} sensor via the particular multipath would have been at a horizontal range r from the source. For conceptual and notational simplicity, it has been assumed here that the paths resulting from perturbation to a particular ray remain in the vertical plane containing that ray. The second integral has the interpretation of the phase difference between $\tilde{x}_{i_n}(t)$ and the signal received at the i^{th} array sensor via the particular path over which the integration is carried out.

In the unsaturated region, the reduced wavefunction accounts for the random phase perturbation to the received signal resulting from the perturbation to the ray which yields a single path within the ray tube. In the partially-saturated and saturated regions, the reduced wavefunction has the interpretation of representing the

random phase and amplitude perturbations to the received signal resulting from the addition of the signals which were subject to possibly uncorrelated random phase perturbations as they propagated along the multiple paths within the ray tube. Under the assumption that the random fluctuations in the sound speed structure do not increase or decrease the energy in the received sound field, $\mathrm{E}[\psi_{i_n}(t)\psi^*_{i_n}(t) \mid f_o] = 1$.

Assuming that the phase perturbations introduced along the individual paths are jointly Gaussian random variables, the quantity $\mathrm{E}[\psi_{i_n}(s\Delta_t)\psi^*_{k_m}(0) \mid f_o]$ can be expressed as

$$\mathrm{E}[\psi_{i_n}(s\Delta_t)\psi^*_{k_m}(0) \mid f_o] \;=\; e^{-\frac{1}{2}D(i_n,k_m,s\Delta_t,f_o)}.$$

$D(i_n, k_m, s\Delta_t, f_o)$ is the phase-structure function for the spatial separation between the i^{th} and k^{th} array sensors and the temporal separation of $s\Delta_t$. $D(i_n, k_m, s\Delta_t, f_o)$ can be interpreted as the variance of the phase difference between signals received at different points in space and time caused by the random fluctuations in the sound speed structure of the ocean. $D(i_n, k_m, s\Delta_t, f_o)$ is defined as

$$D(i_n, k_m, s\Delta_t, f_o) \triangleq (\frac{2\pi f_o}{C_o})^2 \tag{4.6}$$
$$\mathrm{E}\left[\left(\int_0^{R_i} T_{i_n}(r)\mu(r, z_{i_n}(r), t_{i_n}(r))\, dr \;-\; \int_0^{R_k} T_{k_m}(r)\mu(r, z_{k_m}(r), t_{k_m}(r))\, dr\right)^2\right].$$

Here $t_{i_n}(r)$ is the time at which the signal received at the i^{th} array sensor via the n^{th} ray at the time $t = s\Delta_t$ would have been at the range r from the source. Likewise, $t_{k_m}(r)$ is the time at which the signal received at the k^{th} array sensor via the m^{th} ray at the time $t = 0$ would have been at the range r from the source. As defined earlier, $z_{i_n}(r)$ is the depth of the n^{th} ray from the source to the i^{th} array sensor at the range r from the source. The assumption has been made that the radius of any ray tube is small compared to the scale over which the statistics of $\mu(\underline{z}, t)$ undergo significant change. This allows the statistics of $\psi(t)$ for a ray tube to be calculated using the statistics of $\mu(\underline{z}, t)$ along the ray within that tube.

As mentioned earlier, $D(i_n, k_m, s\Delta_t, f_o)$ can be interpreted as the variance of the phase difference between signals received at different points in space and time caused by the random fluctuations in the sound speed structure. As either the temporal

separation, $s\Delta_t$, or the average spatial separation between the n^{th} ray to the i^{th} array sensor and the m^{th} ray to the k^{th} array sensor increase, the correlation between the random fluctuations in the phases of $x_{i_n}(s\Delta_t)$ and $x_{k_m}(0)$ is reduced. Therefore, as either the temporal or the average spatial separation increases, $D(i_n, k_m, s\Delta_t, f_o)$ approaches a limit of the sum of the variance of the random phase fluctuations in each of the two signals. That is, $D(i_n, k_m, s\Delta_t, f_o) \rightarrow \Phi_{i_n}^2 + \Phi_{k_m}^2$. In addition, $D(i_n, k_m, s\Delta_t, f_o)$ is roughly proportional to $\Phi_{i_n}^2$ and $\Phi_{k_m}^2$. Therefore, as either of these parameters increases, the cross-correlation between $\psi_{i_n}(s\Delta_t)$ and $\psi_{k_m}(0)$ goes to zero.

The evaluation of (4.6) can be handled as two separate cases. In the first case, the n^{th} ray from the source to the i^{th} array sensor can be considered a simple shifting of the m^{th} ray from the source to the k^{th} array sensor. The rays in Figure 4-1 illustrate such a case. In this case, the perturbations introduced along each of the two rays may or may not be correlated. In order to allow for the possiblity of this correlation, the phase-structure function must be evaluated as described in Section 7.3 of [30]. Using the method described therein, the integration of the statistics of $\mu(r, z(r), t(r))$ along the n^{th} and m^{th} rays is approximated by an integration along a ray midway between the two rays. In the second case, the n^{th} ray from the source to the i^{th} array sensor has an entirely different form than the m^{th} ray from the source to the k^{th} array sensor. The rays in Figure 4-2 illustrate such a case. In this case, the perturbations introduced along each of the two rays can be assumed to be uncorrelated. As mentioned in the preceding paragraph, the phase-structure function for this case reduces to

$$D(i_n, k_m, s\Delta_t, f_o) \;=\; \Phi_{i_n}^2 + \Phi_{k_m}^2.$$

With appropriate substitutions made, (4.3) can be written as

$$\mathrm{E}[X_i(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}, f_o] \;=\; (R_k R_i)^{-\frac{1}{2}} e^{jk_o(R_k - R_i)} \sum_{n=1}^{N_i} \sum_{m=1}^{N_k} K_{i_n} K_{k_m} e^{jk_o(S_{k_m} - S_{i_n})}$$
$$\left[\sum_{s=-(M-1)}^{(M-1)} e^{-j2\pi(f-f_o)s\Delta_t}(M - \mid s \mid) e^{-\frac{1}{2}D(i_n, k_m, s\Delta_t)}\right]. \qquad (4.7)$$

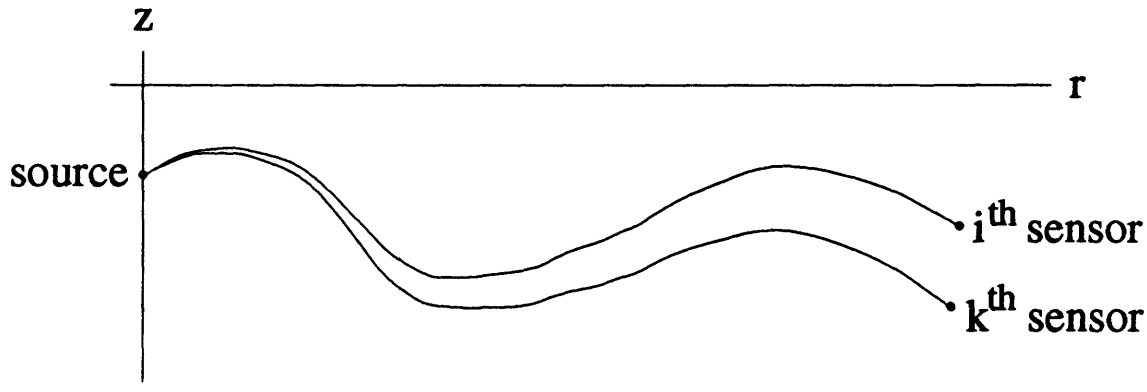


Figure 4-1: Similar Rays

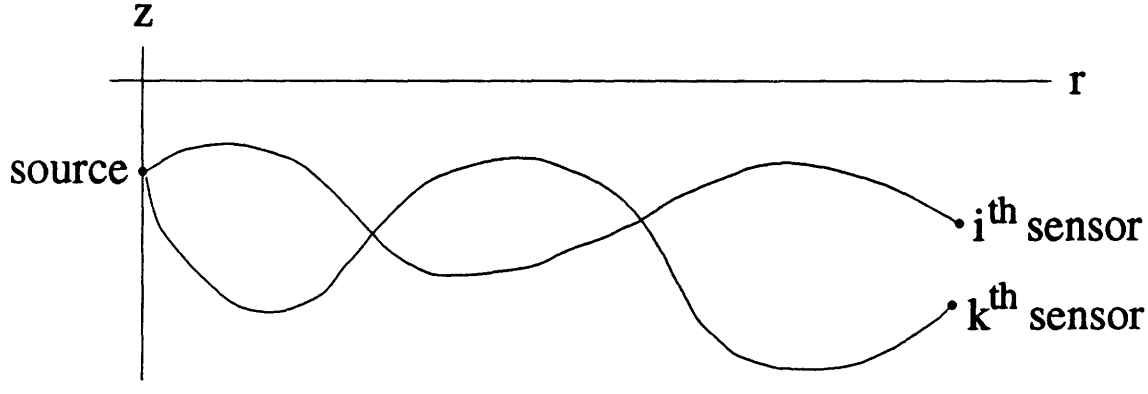


Figure 4-2: Distinctly Different Rays

The term in the square brackets is the windowed DTFT of $e^{-\frac{1}{2}D(i_n,k_m,s\Delta_t)}$ evaluated at the frequency $(f - f_o)$.

If the correlation time (or time-constant) of the random sound speed fluctuations is greater than the segment length used by the DTFT ($(M-1)\Delta_t$), then the temporal incoherence between the signals received at different times separated by less than $(M-1)\Delta_t$ can be ignored. That is, defining $B \triangleq \frac{1}{(M-1)\Delta_t}$ as the bandwidth of the first stage DTFT of the array processor and T_μ as the correlation time (or time-constant) of μ, the temporal incoherence can be ignored when $BT_\mu > 1$. In this case, $D(k_n, k_m, s\Delta_t, f_o) \approx D(k_n, k_m, 0, f_o)$ and (4.7) can be simplified to

$$\mathrm{E}[X_i(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}, f_o] = (R_k R_i)^{-\frac{1}{2}} e^{jk_o(R_k - R_i)} \left(\frac{\sin^2(\pi(f-f_o)\Delta_t M)}{\sin^2(\pi(f-f_o)\Delta_t)}\right) \sum_{n=1}^{N_i}\sum_{m=1}^{N_k} K_{i_n}K_{k_m} e^{jk_o(S_{k_m} - S_{i_n})} e^{-\frac{1}{2}D(i_n,k_m,0,f_o)}. \quad (4.8)$$

The calculation of the strength parameter, the diffraction parameter, and the phase-structure function can be done using the spectral functions of the internal-wave field and the tidal variations. If other sources of perturbation to the sound speed structure are to be considered, the spectral function must include the effects of these processes. Measurements of the amplitude and phase fluctuations of a signal propagating within a single ray tube have indicated that the phase fluctuations are less than, and the amplitude fluctuations are greater than, those predicted by ray-based random propagation theory [31].

It is important to note that only the components of the temporal spectrum corresponding to fluctuations with time constants of on the order of or smaller than the observation interval ($(M-1)\Delta_t$) should be included in these calculations. The processes with time constants much greater than the observation interval will not show strong fluctuations over the interval and are more appropriately treated as time-invariant but unknown characteristics of the medium.

As presented thus far, the ray treatment of the effects of the deterministic and stochastic components of the sound speed structure of the ocean on the propagation of sound considers only fully refracted rays. As mentioned earlier, bottom-reflected

rays can be ignored because of the high losses which usually occur during bottom interaction. However, the exclusion of surface-reflected rays which have no bottom interaction cannot be justified on this basis. Such rays will be those emitted by sources which are much closer to the surface than to the bottom and for which the initial angle of propagation with respect to the horizontal is fairly small. The following minor modification, exploiting the Lloyd mirror effect, extends the treatment to handle these rays [40].

For the purpose of calculating the **rays** from a source at depth z_{source} to each array sensor, an **image** ocean is placed on top of the actual ocean and the ocean surface is ignored. Therefore, letting z_d be the ocean depth, the propagation medium extends from z_d to $-z_d$ rather than from 0 to z_d where $z = 0$ is the actual ocean surface. The sound speed profile of the image ocean is the mirror image of the sound speed profile of the real ocean (i.e., $\overline{C}(z) = \overline{C}(-z)$). In addition, an image source is placed at the depth $-z_{source}$ and an image sensor is placed at the corresponding image position of the real sensor. A ray is then traced from the actual source to the actual sensor with an initial propagation angle of θ and possibly passing through both the actual and the image ocean. A corresponding ray is traced from the image source to the image sensor. The combination of the portions of these two rays which lie in the actual ocean (shown by the dotted line in Figure 4-3) is the actual ray from the actual source to the actual sensor with an initial propagation angle of θ. This modification allows the surface-reflected rays, the corresponding eikonals (S_{i_n}), and the corresponding deterministically-propagated signals ($\tilde{x}_{i_n}(s\Delta_t)$) to be calculated by selecting $z_{i_n}(r)$ to follow the ray contained in the actual ocean.

An additional effect which must be considered is the increased perturbation to the reduced wavefunction caused by the scattering of the reflected signal from the turbulent sea surface. This is accomplished by modifying the definition of the phase-structure function (4.6) to include a term D_{sr} which is proportional to the number of surface reflections and the strength of the surface turbulence and which is inversly-proportional to the spatial correlation distance and the temporal correlation time of the surface turbulence. Experimental results [40] indicate that, for moderate and

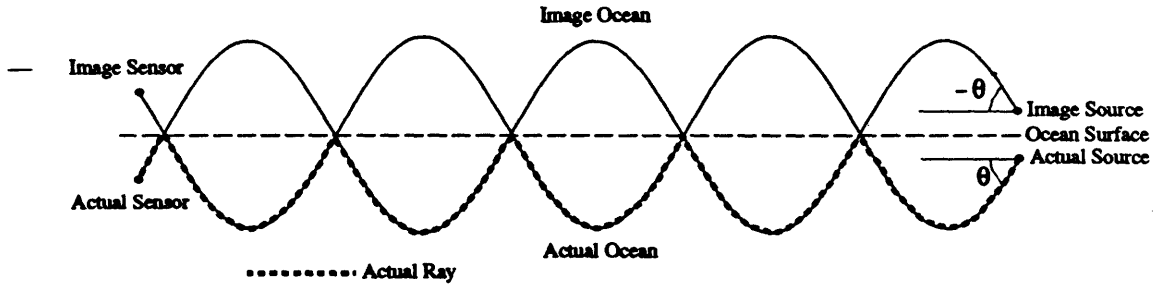


Figure 4-3: The Lloyd Mirror Effect Modification

long-range propagation problems, this term is much smaller than the term accounting for the signal fluctuations induced by the internal wave field and can be safely ignored.

4.3 Normal Mode Approximation

The normal mode treatment of the effects of the deterministic and stochastic components of the sound speed structure on the propagation of sound is developed in [32]. The normal mode approximation models the behavior of the propagating sound field by asuming that the field is the sum of horizontally-propagating vertical standing waves called normal modes. The standing waves are the result of the constructive and destructive interference between successive upward or downward-going wavefronts which have been reflected from the sea surface and bottom. This approximation is valid at sufficiently low frequencies, the definition of which depends on the depth of the ocean. In an ocean with a depth of 4000 meters, the frequency at which the normal mode approximation breaks down is somewhere between 200 Hz and 500 Hz [33]. As the ocean depth decreases, the frequency above which the normal mode approximaton is no longer valid will increase.

Each normal mode is characterized by its modal shape $\phi_n(z)$ (not to be confused with the environmental parameter $\underline{\phi}$ defined earlier) and its horizontal wavenumber k_{H_n}. $\phi_n(z)$ is the vertical standing wave set up by the reflections of the propagating sound field from the sea surface and bottom. Assuming the ideal case where the sea surface is modeled as a free surface and the sea bottom is modeled as a perfectly

rigid bottom, the set of horizontal wavenumbers for any temporal frequency of at which the modes exist will be a discrete set. Therefore, there will exist only a finite number of modes with a vertical wavenumber small enough so that the horizontal wavenumber is a real number ($k_H = \sqrt{(k^2 - k_V^2)}$ where k is the wavenumber and k_V is the vertical wavenumber). Therefore, there are only a finite number of modes for which the vertical standing wave propagates in the horizontal direction. The evanescent (non-propagating) modes decay exponentially in range. Due to this rapid attenuation, they are ignored here.

A more realistic assumption about the sea bottom is that it is not perfectly rigid and is characterized by a finite sound speed C_b and a finite density ρ_b. Then, for wavefronts with propagation angles greater than the critical angle of the water/bottom interface (here the propagation angle is measured with respect the horizontal), the sound field will not be entirely reflected back into the water column and some energy will propagate into the bottom. For these wavefronts, perfect cancellation of successive bottom reflections by destructive interference is not possible. Thus, a continuum of horizontal wavenumbers in the region $k_H < kcos\theta_c$ will correspond to propagating modes. These modes are referred to as the continuous spectrum. [34]

For wavefronts with horizontal wavenumbers in the region $k_H > kcos\theta_c$, the magnitude of the bottom reflection coefficient will be one, and perfect cancellation will occur. Therefore, in this region the set of horizontal wavenumbers corresponding to propagating modes will be a finite discrete set. These modes are referred to as the discrete spectrum.

Due to the leakage of the modes in the continuous spectrum into the bottom, these modes are attenuated rapidly. A good rule of thumb is that at ranges greater than 3 or 4 times the ocean depth, the contribution of the continuous spectrum to the propagating sound field can be ignored [39]. Therefore, it will be assumed that the desired signal received by the array sensors consists of only propagating modes in the discrete spectrum.

Following the convention in [32], it will be assumed that all array sensors lie in a vertical plane which also contains the source, and the signal reaching any of these

sensors will be assumed to have propagated within this plane. Therefore, the spatial coordinate system used will be the same as that used in Section 4.2 with z representing depth and r representing range from the source. z_{source} will denote the depth of the source and the pair (R_i, z_i) will denote the position of the i^{th} array sensor.

As mentioned earlier, the propagating sound field is considered to be the finite sum of propagating modes. Each propagating mode is expressed as the product of two factors. The first accounts for the effect which the deterministic ocean has on the propagating modes. The second factor, the reduced wavefunction, is a stochastic process which accounts for the effects which the random ocean has on the propagating modes. Therefore, the sound field received at the i^{th} array sensor can be expressed as

$$x_i(t) = \sum_{n=1}^{M} x_{i_n}(t),$$

where

$$x_{i_n}(t) = \tilde{x}_{i_n}(t)\, \psi_{i_n}(t).$$

$\tilde{x}_{i_n}(t)$ is, to within a scale factor, the n^{th} mode of the signal emitted by the source which would have been received at the i^{th} array sensor in a deterministic ocean. M is the number of propagating modes in the discrete spectrum. $\psi_{i_n}(t)$ is the reduced wavefunction which accounts for the amplitude of the excitation of the m^{th} mode by the source, as well as the random perturbations in $x_{i_n}(t)$ due to the perturbations of the sound speed structure $(\mu(r, z, t))$ in the region between the source and the i^{th} array sensor. The development in [32] assumes that the array sensors lie at a common range from the source. This restriction is not a problem when computing the products involving $\tilde{x}(t)$ at two different ranges, but is restrictive when computing the cross-correlation of the reduced wavefunction at two different ranges. Therefore, the normal mode approximation presented here is applicable for computing the signal replica vector for only vertical arrays (i.e., $R \triangleq R_i = R_j \ \forall i, j$).

Using the same notation as in Section 4.2, the spatial/temporal correlation func-

tions required in (4.3) can be expressed as

$$\mathrm{E}[x_i(s\Delta_t)x_k^*(0) \mid f_o] = \sum_{n=1}^{M}\sum_{m=1}^{M} \tilde{x}_{i_n}(s\Delta_t)\,\tilde{x}_{k_m}^*(0)\,\mathrm{E}[\psi_{i_n}(s\Delta_t)\psi_{k_m}^*(0) \mid f_o]. \qquad (4.9)$$

The calculation of $\tilde{x}_{i_n}(s\Delta_t)$ is straightforward. At ranges where R is sufficiently large so that the argument $k_{H_n}R$ justifies the use the large argument approximation of the Hankel function of the second kind [5], $\tilde{x}_{i_n}(s\Delta_t)$ can be expressed as

$$\tilde{x}_{i_n}(s\Delta_t) = (k_{H_n}R)^{-\frac{1}{2}}e^{j(2\pi f_o s\Delta_t - k_{H_n}R)}\phi_n(z_i). \qquad (4.10)$$

Substituting (4.10) into the homogeneous wave equation yields the following differential equation which must be satisfied by ϕ_n and k_{H_n}.

$$\frac{\partial^2\phi_n}{\partial z^2} + \left(\frac{2\pi f_o}{\overline{C}(z)}\right)^2 \phi_n(z) = k_{H_n}^2\phi_n(z) \qquad (4.11)$$

The boundary conditions on the solutions to (4.11) are those corresponding to the ideal case of a free surface and a perfectly rigid bottom. These correspond to zero pressure at the surface ($\phi(0) = 0$) and zero particle displacement at the bottom ($\frac{\partial\phi(z)}{\partial z}\mid_{z=H} = 0$ where H is the ocean depth). Given these boundary conditions, the solutions to (4.11) will be orthogonal functions (i.e., $\int_0^H \phi_n(z)\phi_m(z)dz = \delta_{nm}$ where δ_{nm} is the Dirac delta function) [4]. The M solutions to (4.11) corresponding to the discrete propagating modes are generally arranged in descending order of k_{H_n}. Throughout this section, $\phi_n(z)$ will denote the solutions to (4.11), which are calculated using $\overline{C}(z)$ as the sound speed structure.

As was the case with the ray approach to calculating the signal replica vector, the calculation of the second order statistics of the reduced wavefunction is much more involved than the calculation of $\tilde{x}_{i_n}(t)$. The fundamental phenomenon affecting the reduced wavefunction is as modal coupling. To understand modal coupling, first consider the case of the deterministic horizontally-stratified ocean modeled by $\overline{C}(z)$. In this case, $\psi_{i_n}(t) = \phi_n(z_{source})$. The level of modal excitation depends on only the mode shape and the source depth. At the range r_o the propagating sound field is

given by

$$x(r_o, z, t) = \sum_{n=1}^{M} (k_{H_n} r_o)^{-\frac{1}{2}} e^{j(2\pi f_o t - k_{H_n} r_o)} \phi_n(z_{source}) \phi_n(z). \tag{4.12}$$

Then, by the application of Huygens' principle, the field propagating forward from range r_o can be considered to have been excited by a continuum (in depth) of point sources at range r_o, each with a source level equal to $x(r_o, z, t)$. That is, the excitation level of the m^{th} mode in the field propagating forward from range r_o can be expressed as

$$\int_0^H \phi_m(z') x(r_o, z', t) dz'. \tag{4.13}$$

Substituting (4.12) into (4.13) and using the orthogonal relationship between the modal shapes, the excitation level for the m^{th} mode reduces to

$$(k_{H_m} r_o)^{-\frac{1}{2}} e^{j(2\pi f_o t - k_{H_m} r_o)} \phi_m(z_{source}). \tag{4.14}$$

Therefore, the m^{th} mode is excited at range r_o if and only if the m^{th} mode was originally excited by the source and it is excited at range r_o with an amplitude equal to its original amplitude reduced by an amount to account for the loss due to geometric spreading. Modal propagation which obeys this model is referred to as adiabatic mode propagation, implying the independent propagation of the individual modes.

Now, assume that the deterministic sound speed profile changes at the range r_o. Then the modal shapes for $r > r_o$, denoted by $\phi'_n(z)$, will be different from those for $r < r_o$. Thus, substituting (4.12) into (4.13) to yield an expression for the excitation level of the m^{th} mode propagating forward from the range r_o results in

$$\sum_{n=1}^{M} (k_{H_n} r_o)^{-\frac{1}{2}} e^{j(2\pi f_o t - k_{H_n} r_o)} \phi_n(z_{source}) \int_0^H \phi'_m(z') \phi_n(z') dz'. \tag{4.15}$$

Since the modal shapes for $r > r_o$ ($\phi'_m(z)$) are the solutions to a different differential equation than the modal shapes for $r < r_o$ ($\phi_n(z)$), an orthogonal relationship between these mode shapes does not necessarily exist; therefore, (4.15) cannot be reduced in the fashion which yielded (4.14). Therefore, in the situation where the deterministic sound speed profile changes at $r = r_o$, the excitation level of the m^{th} mode propagating

forward from the range r_o is coupled to the excitation level of all of the modes at the range r_o. Consistent with the development in [32], the waves traveling backward in range resulting from the reflection off of the discontinuity in the sound speed profile at the range r_o will be ignored. The resulting propagation model is referred to as the forward scattering approximation to coupled mode propagation.

While the deterministic sound speed profile is range-invariant, the actual sound speed structure changes with range as a result of the effect of the random process $\mu(r,z,t)$. The modal coupling which is introduced by these random fluctuations in the sound speed structure is represented by the reduced wavefunction. For the case where $\mu(r,z,t) \ll 1$ (which holds for almost all ocean acoustics problems [40]), the coupling coefficient between the m^{th} and n^{th} modes is

$$R_{nm}(r,t) = \frac{k_o^2}{(k_{H_m} k_{H_n})^{\frac{1}{2}}} \int_0^H \mu(r,z,t)\phi_m(z)\phi_n(z)dz.$$

The coupling coefficient is a function only of range and time, which results in the reduced wavefunction being a function only of range and time. Under the reasonable assumption [32] that $\frac{\partial \psi_n(r,t)}{\partial t}$ is small, referred to as the quasistatic approximation, the reduced wavefunction evolves according to the following differential equation,

$$\frac{\partial \psi_n(r,t)}{\partial r} = -j \sum_{m=1}^{M} R_{nm}(r,t) e^{j(k_{Hn}-k_{Hm})r} \psi_m(r,t) \quad n \in \{1,\ldots,M\}, \tag{4.16}$$

with the initial conditions $\psi_n(0,t) = \phi_n(z_{source}) \ \forall n$.

By previous assumption, the ranges $R_i = R_j = R$ and as a result, $\psi_{i_n}(t) = \psi_n(R_i,t) = \psi_n(R,t)$. Therefore, the cross-correlation required in (4.9) can be expressed as

$$\mathrm{E}[\psi_{i_n}(s\Delta_t)\psi^*_{k_m}(0) \mid f_o] = \mathrm{E}[\psi_n(R,s\Delta_t)\psi^*_m(R,0) \mid f_o]. \tag{4.17}$$

The modal approximation theory developed in [32] does not yield an expression which accounts for the temporal decorrelation of the reduced wavefunction at a particular range. However, as described in Section 4.2, when $BT_\mu > 1$, (4.17) can be

closely approximated by

$$\mathrm{E}[\psi_{i_n}(s\triangle_t)\psi^*_{k_m}(0) \mid f_o] \approx \mathrm{E}[\psi_n(R,0)\psi^*_m(R,0) \mid f_o]. \tag{4.18}$$

The cross-correlation function of the solutions to the differential equation (4.16) is given by

$$\mathrm{E}[\psi_n(R,0)\psi^*_m(R,0) \mid f_o] = W_n(R)\,\delta_{nm}. \tag{4.19}$$

The assumption which underlies this expression is that, at the ranges for which the normal mode approximation as given in (4.10) is valid, the modal phases are incoherent from mode to mode. Experimental evidence [41] has indicated that the modal phases may be partially correlated at these ranges. Therefore, the validity of this assumption may be in doubt.

$W_n(r)$ is the power in the n^{th} mode at range r and can be shown to be the solution to the following differential equation.

$$\frac{dW_n(r)}{dr} = \sum_{\substack{m=1 \\ m\neq n}}^{M} a_{nm}(W_m(r) - W_n(r)), \tag{4.20}$$

with the initial conditions $W_n(0) = \phi_n^2(z_{source})\ \forall n$. a_{nm} is the non-negative constant described in the following paragraph. The fact that a_{nm} is non-negative indicates that the effect of the modal coupling will be to transfer power from the modes with greater amounts of power to the modes with lesser amounts power.

a_{nm} is a modal coupling parameter given by

$$a_{nm} = \frac{1}{2\pi}\mathrm{E}[|\,\frac{1}{\epsilon}\int_{-\infty}^{\infty} R_{nm}(r)e^{-j(k_{H_n}-k_{H_m})r}dr\,|^2], \tag{4.21}$$

where ϵ is the small parameter described on page 358 of [32]. ϵ is roughly equal to $L\overline{R}$ where L is the scale length in range for changes in the random process ψ and $\overline{R}$ is the scale length in range for changes in the random process R_{nm}. Under the assumption that μ is a wide-sense stationary random process in range (i.e., spatially

homogeneous in range), (4.21) can be expressed as

$$a_{nm} = \frac{k_o^4 \tilde{K}}{2\pi k_{H_n} k_{H_m}} \int_0^H \int_0^H F((k_{H_m} - k_{H_n}), z, z')\, \phi_n(z)\phi_m(z)\, \phi_n(z')\phi_m(z')\, dz\, dz'.$$

Here,

$$F(k, z, z') = \int_{-\infty}^{\infty} \mathrm{E}[\mu(\Delta_r, z', t)\mu(0, z, t)]\, e^{-jk\Delta_r}\, d\Delta_r$$

is the horizontal wavenumber/vertical position spectrum of the random process μ and can be determined from the spectrum of the internal waves, tidal oscillations, and other sources of significant perturbations in the sound speed structure of the ocean; and $\tilde{K}$ is a constant which is roughly equal to $\frac{2}{L^2 R}$. As was the case in Section 4.2, it is important to note that only the components of the temporal spectrum corresponding to fluctuations with time constants of on the order of or smaller than the observation interval ($(M-1)\Delta_t$) should be included in these calculations.

Combining (4.19), (4.18), (4.10), (4.9), and (4.3) yields the following expression for $\mathrm{S}_{ik}(f)$ given the deterministic source $e^{j2\pi f_o t}$.

$$\mathrm{E}[X_i(f, \underline{z}) X_k^*(f, \underline{z}) \mid \underline{\phi}, f_o] = \left(\frac{\sin^2(\pi(f - f_o)\Delta_t M)}{\sin^2(\pi(f - f_o)\Delta_t)} \right) \sum_{n=1}^{M} (R k_{H_n})^{-1} \phi_n(z_i)\phi_n(z_k) W_n(R) \tag{4.22}$$

4.4 Numerical Solution of the Wave Equation

The numerical solution of the wave equation does not require approximation to the wave equation but can handle the dependence of the signal replica vector on only $\overline{C}(z)$. The solution is the signal as received at the i^{th} array sensor assuming that the source emits the deterministic complex exponential $e^{j2\pi f_o t}$ and that the ocean is a deterministic horizontally-stratified medium. Assuming that the ocean is also a linear acoustic medium, the solution to the wave equation can be expressed as

$$x_i(t) = e^{j2\pi f_o t} \tilde{x}_i(f_o),$$

where $\tilde{x}_i(f_o)$ is the transfer function of the ocean between the source and the i^{th} array sensor. With these assumptions, the expression for the spatial/temporal cross-correlation function of the desired signal reduces to

$$\mathrm{E}[x_i(s\Delta_t)x_k^*(0) \mid f_o] \;=\; e^{j2\pi f_o t}\,\tilde{x}_i(f_o)\,\tilde{x}_k^*(f_o). \tag{4.23}$$

Substituting (4.23) into (4.3) and evaluating the DTFT of the triangle function yields

$$\mathrm{E}[X_i(f,\underline{z})X_k^*(f,\underline{z}) \mid \underline{\phi}, f_o] \;=\; \left(\frac{\sin^2(\pi(f-f_o)\Delta_t M)}{\sin^2(\pi(f-f_o)\Delta_t)}\right)\tilde{x}_i(f_o)\tilde{x}_k^*(f_o). \tag{4.24}$$

Efficient algorithms such as the SAFARI program [29] are available to numerically compute solutions to the wave equation. Since this approach cannot take into account the effect of the random perturbations to the sound speed structure of the ocean, it should only be used in situations where the temporal coherence interval of the ocean is longer than the segment length used by the first stage DTFT of the array processor and where perfect spatial coherence can be assumed across the aperture of the array. The first of these two conditions is roughly equivelent to $BT_\mu > 1$.

Chapter 5

Conclusions and Future Work

The problem of matched field processing without precise knowledge of the characteristics of the propagation environment has been addressed herein. Previous work summarized in Section 1.3 has documented and analyzed the performance degredation which both adaptive and non-adaptive matched field processors suffer when a mismatch exists between the environmental conditions assumed by the processor and those which actually exist. A new algorithm for implementing a matched field processor which is capable of operating with only approximate environmental information yet which offers the high resolution and interference rejection capability characteristic of adaptive processors has been developed using the framework of minmax signal processing.

The matched field processor developed consists of discrete-time Fourier Transform, followed by a linear weight-and-sum beamformer, followed by a module which averages the magnitude squared of the output of the beamformer. This output of this final module is an estimate of the average power in a narrowband signal emitted by a source at the array focal point as received at a particular sensor in the array. This signal is referred to as the desired signal and the particular sensor is referred to as the reference sensor. The weights of the beamformer are choosen to minimize the maximum conditional mean-squared error between the output of the beamformer and the desired signal. The error is conditioned on the characteristics of the propagation environment and the maximum is evaluated over a user specified

range of environmental conditions over which the processor is designed to operate. The processor developed using this minmax framework achieves good performance over a range of environmental conditions and does not require precise knowledge of the characteristics of the propagation environment.

The fundamental theorem upon which the minmax signal processing framework is built is the Characterization Theorem. This theorem, which is presented in Section 2.1, provides a nice physical interpretation of the necessary and sufficient conditions satisfied by the solutions to a large class of minmax approximation problems. The practical importance of the theorm lies in the fact that makes possible the development of an efficient algorithm to solve for the optimal minmax array weights when the average signal power is initially unknown. First, in Section 2.2 it is used to prove that the solution to the minmax array weight problem given any assumed average signal power can be expressed as a real non-negative constant multiplied by the solution to the minmax array weight problem given any other non-zero assumed average signal power. Finally, in Section 2.3 it is used to prove that the optimal minmax array weights are equivalent to the minimum mean-squared error array weights when the "least favorable pmf" is assigned to the environmental conditions. This least favorable pmf interpretation of the optimal minmax array weights motivates the development of an efficient algorithm to solve for this particular pmf and makes possible the efficient solution of the minmax array weight problem which is of interest.

The Characterization Theorem has two other important consequences. First, it motivated the work in in Section 2.4 in which the form of the optimal minmax mean-squared error estimator and an approach to developing bounds on the minmax mean-squared error performance achievable by any estimator are derived. Finally, in Section 3.1 the algorithm derived using the Characterization Theorem leads to the interpretation of the Adptive Minmax Matched Field Processor as a Two-Stage Minimum Variance Distortionless Response (MVDR) Matched Field Processor. This relates the minmax processor to a processor whose performance characteristics are well understood, improves the understanding of what the minmax processor does and how it can be expected to perform, and motivates two small modifications to the

minmax processor.

The numerical analysis of the processor's performance shows that in a deterministic medium (Subsections 3.2.1 and 3.2.2), the processor is able to maintain good performance in the presence of environmental uncertainty. However, the results in Subsection 3.2.3 show that when the medium in randomly time-variant, the performance of both the Adaptive Minmax Matched Field Processor and the MVDR Matched Field Processor suffers. Part of this performance loss can be atributed the loss of signal correlation across the aperture of the array which makes it difficult for any processor which incorporates a linear beamformer to perform effectively. However, some of this performance degradation is attributed to the fact that both processors model the desired signal as having a rank one cross-spectral correlation matrix. Therefore, since the cross-spectral correlation matrix has a rank of greater than one in random propagation medium, the processors are not correctly matched to the signal characteristics in this case.

In applications where the random media fluctuations are dominated by the internal waves propagating in the ocean, the processor's poor performance in random media may not be a significant handicap. This can be observed by noting that the spectrum of the internal waves is dominated by components below around 3 cycles per hour [37, 38]. Therefore, for observation intervals on the order of several seconds the media will appear to be effectively time-invariant but unknown during any observation interval. In these cases, the sample cross-spectral correlation matrix will have a low-rank and the rank one model will be applicable.

Further work needs to be done to develop methods appropriate for applications where significant ocean variation is expected over a single observation inverval. Using the notation of Chapter 4, these methods must be able to deal with uncertainty in the deterministic depth-dependent sound speed structure ($\overline{C}(z)$) of the oceans while accommodating the random ocean fluctuations ($C_o\mu(\underline{z},t)$) which may be significant over a single observation interval.

Appendix A

Proofs for General Minmax Problems

The following two definitions will be used in the proofs in Appendicies A and B.

Definition: Open Neighborhood of a Set Let A be a subset of the metric space S. Then, for any $\delta > 0$, the open neighborhood A of radius δ in S is

$$B(A,\delta) \triangleq \{a \in S \mid \exists a_o \in A \text{ s.t. } d(a_o,a) < \delta\}$$

where $d : S \times S \rightarrow \mathbb{R}$ is the distance function for the metric space S.

Definition: Closed Neighborhood of a Set Let A be a subset of the metric space S. Then, for any $\delta > 0$, the closed neighborhood of A of radius δ in S, denoted by $\overline{B}(A,\delta)$, is the closure of $B(A,\delta)$ in S.

A number of theorems taken directly from [2] will be used in the proofs in Appendicies A and B. They will be referred to by their number in [2] and are stated below without proof.

Theorem 2.19 Every neighborhood is an open set.

Theorem 2.23 A set E is open if and only if its complement is closed.

Theorem 2.24(a) For any collection $\{G_\alpha\}$ of open sets, $\cup_\alpha G_\alpha$ is open.

Theorem 2.35 Closed subsets of compact sets are compact.

Corollary to Theorem 2.35 If F is closed and K is compact, then $F \cap K$ is compact.

Theorem 4.6 Suppose X and Y are metric spaces, $E \subset X$, $p \in E$, p is a limit point of E, and f maps E into Y. Then, f is continuous at p if and only if

$\lim_{x \to p} f(x) = f(p)$.

Corollary to Theorem 4.8 A mapping f of a metric space X into a metric space Y is continuous if and only if $f^{-1}(C)$ is closed in X for every closed set C in Y.

Theorem 4.9 Let f and g be complex continuous functions on a metric space X. Then $f+g$, fg, and f/g are continuous on X.

Theorem 4.14 Suppose f is a continuous mapping of a compact metric space X into a metric space Y. Then $f(X)$ is compact.

Theorem 4.16 Suppose f is a continuous real function on a compact metric space X, and

$$M = \sup_{p \in X} f(p), \quad m = \inf_{p \in X} f(p).$$

Then there exist points $p, q \in X$ such that $f(p) = M$ and $f(q) = m$.

Theorem 4.19 Let f be a continuous mapping of a compact metric space X into a metric space Y. Then f is uniformly continuous on X.

Theorem 5.10 If f is a real continuous function on $[a, b]$ which is differentiable in (a, b), then there is a point $x \in (a, b)$ at which $f(b) - f(a) = (b-a)f'(x)$.

In addition, there are two theorems in [3] which will be used in the proofs in Appendicies A and B. They will be referred to by name and are stated below without proof.

Theorem of Carathéodory Let A be a subset of an n-dimensional linear space. Every point in the convex hull of A ($\mathcal{H}(A)$) is expressible as a convex linear combination of $n+1$ or fewer elements of A.

Theorem on Linear Inequalities Let U be a compact subset of $\mathbb{R}^N$. A necessary and sufficient condition that the system of linear inequalities $< u, z > > 0 \ \forall u \in U$ be inconsistent is that $\underline{0} \in \mathcal{H}(U)$.

The following three lemmas will be used in later proofs of theorems.

Lemma A.1 Let A be a subset of the Euclidian metric space E, B be a compact subset of the metric space Γ, and $f : A \times B \to \mathbb{R}$ be a continuous function on A and B. Then

$$\Delta(a) \triangleq \max_{b \in B} f(a, b)$$

is a continuous function on A.

Proof: $f(a, b)$ is a real function continuous on B. Then, by Theorem 4.16, for each $a \in A$, $\exists b_o \in B$ such that $f(a, b_o) = \sup_{b \in B} f(a, b)$ and therefore $\Delta(a) = f(a, b_o)$.

Assume that $\Delta(a)$ is not continuous on A. Then $\exists a_0 \in A$ and $\epsilon_0 > 0$ such that $\forall \delta_0 > 0$, $\exists a_1 \in A$ such that $d(a_0, a_1) < \delta_0$ and $| \Delta(a_0) - \Delta(a_1) | \geq \epsilon_0$.

Assume that $\Delta(a_0) > \Delta(a_1)$. If this is not the case, the roles of a_0 and a_1 can be switched in the remainder of the proof. As established earlier, $\exists b_0 \in B$ such that

$f(a_0, b_0) = \Delta(a_0)$. By definition, $f(a_1, b_0) \leq \Delta(a_1)$. Therefore, $| f(a_0, b_0) - f(a_1, b_0) | \geq \epsilon_0$.

Therefore $\exists a_0 \in A$ and $\epsilon_0 > 0$ such that $\forall \delta_0 > 0$, $\exists a_1 \in A$ such that $d(a_0, a_1) < \delta_0$ and $| f(a_0, b_0) - f(a_1, b_0) | \geq \epsilon_0$. Therefore, f is not continuous at a_0 which contradicts the continuity of f on A. Therefore, the assumption must be wrong.

Therefore, $\Delta(a)$ is continuous on A. ■

Lemma A.2 Let A be a convex subset of the Euclidian metric space E, B be a compact subset of the metric space Γ, and $f : A \times B \rightarrow \mathbb{R}$ be a convex function on A a continuous function on B. Then

$$\Delta(a) \triangleq \max_{b \in B} f(a, b)$$

is a convex function on A.

Proof: $f(a, b)$ is a real function continuous on B. Then, by Theorem 4.16, for each $a \in A$, $\exists b_o \in B$ such that $f(a, b_o) = \sup_{b \in B} f(a, b)$ and therefore $\Delta(a) = f(a, b_o)$.

Choose any $a_0, a_1 \in A$ such that $a_0 \neq a_1$. Then by the convexity of A and of f on A

$$\forall \lambda \in [0, 1], \ \forall b \in B, \ \ f(\lambda a_0 + (1 - \lambda) a_1, b) \leq \lambda f(a_0, b) + (1 - \lambda) f(a_1, b).$$

By definition, $\forall b \in B$, $f(a_0, b) \leq \Delta(a_0)$ and $f(a_1, b) \leq \Delta(a_1)$. Therefore,

$$\forall \lambda \in [0, 1], \ \forall b \in B, \ \ f(\lambda a_0 + (1 - \lambda) a_1, b) \leq \lambda \Delta(a_0) + (1 - \lambda) \Delta(a_1). \qquad \text{(A.1)}$$

As established earlier

$$\forall \lambda \in [0, 1] \ \ \exists b(\lambda) \in B \text{ s.t.}$$
$$f(\lambda a_0 + (1 - \lambda) a_1, b(\lambda)) = \Delta(\lambda a_0 + (1 - \lambda) a_1). \qquad \text{(A.2)}$$

Choosing b in (A.1) to be $b(\lambda)$ and substituting (A.2) into (A.1) yields

$$\forall \lambda \in [0, 1] \ \ \Delta(\lambda a_0 + (1 - \lambda) a_1) \leq \lambda \Delta(a_0) + (1 - \lambda) \Delta(a_1).$$

Since this is true for any $a_0, a_1 \in A$, $\Delta(a)$ is a convex function on A. ■

Lemma A.3 Let X, Y, and Z be metric spaces with distance functions $d_X(\ ,\ )$, $d_Y(\ ,\ )$, and $d_Z(\ ,\ )$, respectively. Suppose $A \subset X$ and $B \subseteq C \subset Y$, and that $g : A \rightarrow B$ and $f : C \rightarrow Z$ are continuous functions. Then, $f \circ g : A \rightarrow Z$ is a continuous function.

Proof: Select any $p \in A$ and $\epsilon > 0$. Then, by the continuity of f, $\exists \delta_0 > 0$ such that $\forall b \in B$, $d_Y(b, g(p)) < \delta_0 \rightarrow d_Z(f(b), f(g(p))) < \epsilon$. By the continuity of g, $\exists \delta_1 > 0$ such that $\forall a \in A$, $d_X(a, p) < \delta_1 \rightarrow d_Y(g(a), g(p)) < \delta_0$. Therefore, $\exists \delta_1 > 0$ such that $\forall a \in A$, $d_X(a, p) < \delta_1 \rightarrow d_Z(f(g(a)), f(g(p))) < \epsilon$. Therefore, $f \circ g$ is a continuous function on A. ■

The following definition of the gradient operator is required in order to prove the Minmax Characterization Theorem which follows the definition.

Definition: Gradient Operator Let Φ be a subset of a metric space denoted by Γ, W be a subset of a Euclidian metric space denoted by E, $\varepsilon : W \times \Phi \rightarrow \mathbb{R}$ be a continuous function on both W and Φ for which, at each $\underline{w} \in W$, a directional derivative with respect to $\underline{w}$ can be defined on Φ. Then the gradient of ε with respect to $\underline{w}$, denoted by $\nabla_{\underline{w}}\varepsilon\left(\underline{w}, \underline{\phi}\right)$, is defined as any function $\nabla_{\underline{w}}\varepsilon : W \times \Phi \rightarrow E$ which is continuous on W and Φ and for which the following is true.

$$\forall \underline{d} \in E,\ \forall \underline{w} \in W,\ \forall \underline{\phi} \in \Phi \quad \frac{\partial \varepsilon(\underline{w} + \lambda \underline{d}, \underline{\phi})}{\partial \lambda} |_{\lambda=0} = k\ < \nabla_{\underline{w}}\varepsilon\left(\underline{w}, \underline{\phi}\right), \underline{d} >,$$

where k is a real positive constant and $< , >$ denotes the inner product defined on E.

Theorem 1 Let Φ be a compact set contained in a metric space denoted by Γ, W be an open set of a Euclidian metric space denoted by E, $\varepsilon : W \times \Phi \rightarrow \mathbb{R}$ be a continuous function on both W and Φ for which, at each $\underline{w} \in W$, a directional derivative with respect to $\underline{w}$ can be defined on Φ, and $\nabla_{\underline{w}}\varepsilon\left(\underline{w}, \underline{\phi}\right)$ be the gradient of ε with respect to $\underline{w}$. Then a necessary condition for $\underline{w}_o \in W$ to be a solution to the following minmax problem

$$\underline{w}_{opt} = \arg \min_{\underline{w} \in W} \max_{\underline{\phi} \in \Phi} \varepsilon(\underline{w}, \underline{\phi}),$$

is that

$$\underline{0} \in \mathcal{H}\left(\{\nabla_{\underline{w}}\varepsilon\left(\underline{w}_o, \underline{\phi}\right) \mid \underline{\phi} \in M(\underline{w}_o)\}\right).$$

If, in addition, ε is a convex function of $\underline{w}$ and W is a convex set, this condition is a necessary and sufficient condition for $\underline{w}_o \in W$ to be the solution solution to the stated minmax problem.

Proof: $\varepsilon(\underline{w}_o, \underline{\phi})$ is continuous on Φ and Φ is compact. Therefore, by Theorem 4.16, $\exists \underline{\phi}_o \in \Phi$ such that $\varepsilon(\underline{w}_o, \underline{\phi}_o) = \sup_{\underline{\phi} \in \Phi} \varepsilon(\underline{w}_o, \underline{\phi}) = \Delta(\underline{w}_o)$. Therefore, $M(\underline{w}_o)$ is not an empty set.

$\varepsilon(\underline{w}_o, \underline{\phi})$ is continuous on Φ, Φ and $\mathbb{R}$ are metric spaces, and the singleton set $\{\Delta(\underline{w}_o)\}$ is closed in $\mathbb{R}$. Therefore, by Theorem 4.8, $M(\underline{w}_o)$ is closed in Φ.

$M(\underline{w}_o)$ is closed in Φ, and Φ is compact. Therefore, by Theorem 2.35, $M(\underline{w}_o)$ is compact.

$\nabla_{\underline{w}}\varepsilon\left(\underline{w}_o, \underline{\phi}\right)$ is a continuous function on Φ and $M(\underline{w}_o)$ is compact. Therefore, by Theorem 4.14, the set $\{\nabla_{\underline{w}}\varepsilon\left(\underline{w}_o, \underline{\phi}\right) \mid \underline{\phi} \in M(\underline{w}_o)\}$ is compact.

The necessary condition portion of the theorem will be proven first.

Assume that

$$\underline{0} \notin \mathcal{H}\left(\{\nabla_{\underline{w}}\varepsilon\left(\underline{w}_o, \underline{\phi}\right) \mid \underline{\phi} \in M(\underline{w}_o)\}\right).$$

$\{\nabla_{\underline{w}}\varepsilon\left(\underline{w}_o, \underline{\phi}\right) \mid \underline{\phi} \in M(\underline{w}_o)\}$ is a compact subset of the Euclidian metric space E.

Therefore, by the Theorem on Linear Inequalities, $\exists \underline{d} \in E$ such that

$$< \nabla_{\underline{w}} \varepsilon \left(\underline{w}_o, \underline{\phi}\right), \underline{d} > \ < 0 \ \forall \underline{\phi} \in M(\underline{w}_o). \tag{A.3}$$

Let $\epsilon_0 = \sup_{\underline{\phi} \in M(\underline{w}_o)} < \nabla_{\underline{w}} \varepsilon \left(\underline{w}_o, \underline{\phi}\right), \underline{d} >$.

$\nabla_{\underline{w}} \varepsilon \left(\underline{w}, \underline{\phi}\right)$ is a continuous function on W and Φ. Therefore, by Theorem 4.9, $< \nabla_{\underline{w}} \varepsilon \left(\underline{w}, \underline{\phi}\right), \underline{d} >$ is a continuous real function on W and Φ.

$< \nabla_{\underline{w}} \varepsilon \left(\underline{w}_o, \underline{\phi}\right), \underline{d} >$ is a continuous real function on Φ and $M(\underline{w}_o) \subseteq \Phi$ is compact. Therefore, by Theorem 4.16, $\epsilon_0 = \max_{\underline{\phi} \in M(\underline{w}_o)} < \nabla_{\underline{w}} \varepsilon \left(\underline{w}_o, \underline{\phi}\right), \underline{d} >$. Therefore, (A.3) implies that $\epsilon_0 < 0$.

$< \nabla_{\underline{w}} \varepsilon \left(\underline{w}_o, \underline{\phi}\right), \underline{d} >$ is a continuous real function on Φ and Φ is compact. Therefore, by Theorem 4.19, $< \nabla_{\underline{w}} \varepsilon \left(\underline{w}_o, \underline{\phi}\right), \underline{d} >$ is a uniformly continuous function on Φ.

$< \nabla_{\underline{w}} \varepsilon \left(\underline{w}_o, \underline{\phi}\right), \underline{d} >$ is a uniformly continuous function on Φ and $\forall \underline{\phi} \in M(\underline{w}_o) \subseteq \Phi$, $< \nabla_{\underline{w}} \varepsilon \left(\underline{w}_o, \underline{\phi}\right), \underline{d} > \ \leq \epsilon_0 < 0$. Therefore, by the definition of a uniformly continuous function and Theorem 4.6,

$$\exists \delta_0 > 0 \ \text{ s.t. } \ \forall \underline{\phi} \in \overline{B}(M(\underline{w}_o), \delta_0) \cap \Phi, \ < \nabla_{\underline{w}} \varepsilon \left(\underline{w}_o, \underline{\phi}\right), \underline{d} > \ < \epsilon_0/2 < 0.$$

$\overline{B}(M(\underline{w}_o), \delta_0)$ is closed and Φ is compact. Therefore, by the Corollary to Theorem 2.35, $\overline{B}(M(\underline{w}_o), \delta_0) \cap \Phi$ is compact.

$< \nabla_{\underline{w}} \varepsilon \left(\underline{w}, \underline{\phi}\right), \underline{d} >$ is a continuous real function on W and $\overline{B}(M(\underline{w}_o), \delta_0) \cap \Phi$ is compact. Therefore, by Lemma A.1,

$$\max_{\underline{\phi} \in \overline{B}(M(\underline{w}_o), \delta_0) \cap \Phi} < \nabla_{\underline{w}} \varepsilon \left(\underline{w}, \underline{\phi}\right), \underline{d} >$$

is a continuous function on W. Therefore,

$$\exists \delta_1 > 0 \ \text{ s.t. } \ \forall \underline{w} \in B(\{\underline{w}_o\}, \delta_1) \cap W, \ \max_{\underline{\phi} \in \overline{B}(M(\underline{w}_o), \delta_0) \cap \Phi} < \nabla_{\underline{w}} \varepsilon \left(\underline{w}, \underline{\phi}\right), \underline{d} > \ < \epsilon_0/4 < 0.$$

The crucial result of the first portion of this proof can therefore be summarized as,

$$\exists \delta_1 > 0 \ \text{ s.t. } \ \forall \underline{w} \in B(\{\underline{w}_o\}, \delta_1) \cap W \ \text{ and } \ \forall \underline{\phi} \in \overline{B}(M(\underline{w}_o), \delta_0) \cap \Phi, \tag{A.4}$$

$$< \nabla_{\underline{w}} \varepsilon \left(\underline{w}, \underline{\phi}\right), \underline{d} > \ < \epsilon_0/4 \ < 0. \tag{A.5}$$

By Theorems 2.19 and 2.24(a), $B(M(\underline{w}_o), \delta_0)$ is an open set. Therefore, by Theorem 2.23, the set $\{\Phi - B(M(\underline{w}_o), \delta_0)\}$ is closed in Φ. Therefore, since Φ is compact, by Theorem 2.35, $\{\Phi - B(M(\underline{w}_o), \delta_0)\}$ is compact.

Note that $\forall \underline{\phi} \in \{\Phi - B(M(\underline{w}_o), \delta_0)\}$, $\varepsilon(\underline{w}, \underline{\phi}) < \Delta(\underline{w}_o)$, and let

$$\epsilon_1 = \sup_{\underline{\phi} \in \{\Phi - B(M(\underline{w}_o), \delta_0)\}} \varepsilon(\underline{w}_o, \underline{\phi}).$$

$\varepsilon(\underline{w}_o, \underline{\phi})$ is a continuous real function on Φ and $\{\Phi - B(M(\underline{w}_o), \delta_0)\}$ is compact. Therefore, by Theorem 4.16,

$$\epsilon_1 = \max_{\underline{\phi} \in \{\Phi - B(M(\underline{w}_o), \delta_0)\}} \varepsilon(\underline{w}_o, \underline{\phi}) < \Delta(\underline{w}_o).$$

In addition, since $\varepsilon(\underline{w}_o, \underline{\phi})$ is a continuous real function on W, by Lemma A.1,

$$\max_{\underline{\phi} \in \{\Phi - B(M(\underline{w}_o), \delta_0)\}} \varepsilon(\underline{w}, \underline{\phi})$$

is a continuous function on W.

Let $\epsilon_2 = (\Delta(\underline{w}_o) + \epsilon_1)/2$. Then $\epsilon_1 < \epsilon_2 < \Delta(\underline{w}_o)$. Therefore, since

$$\max_{\underline{\phi} \in \{\Phi - B(M(\underline{w}_o), \delta_0)\}} \varepsilon(\underline{w}, \underline{\phi})$$

is a continuous function on W,

$$\exists \delta_2 > 0 \text{ s.t. } \forall \underline{w} \in B(\{\underline{w}_o\}, \delta_2) \cap W, \quad \max_{\underline{\phi} \in \{\Phi - B(M(\underline{w}_o), \delta_0)\}} \varepsilon(\underline{w}, \underline{\phi}) < \epsilon_2.$$

Then, since $\{\Phi - \overline{B}(M(\underline{w}_o), \delta_0)\} \subseteq \{\Phi - B(M(\underline{w}_o), \delta_0)\}$, the crucial result from the second portion of this proof can be expressed as

$$\exists \delta_2 > 0 \text{ s.t. } \forall \underline{w} \in B(\{\underline{w}_o\}, \delta_2) \cap W \text{ and } \forall \underline{\phi} \in \{\Phi - \overline{B}(M(\underline{w}_o), \delta_0)\} \cap \Phi, \quad \text{(A.6)}$$

$$\varepsilon(\underline{w}, \underline{\phi}) < \epsilon_2 < \Delta(\underline{w}_o). \quad \text{(A.7)}$$

W is open. Therefore, $\exists \delta_3 > 0$ such that $B(\underline{w}_o, \delta_3) \subseteq W$. Then let $\lambda_0 = \min(\delta_1, \delta_2, \delta_3)/(2 \mid \underline{d} \mid) > 0$, and define $\underline{w}(\lambda) \triangleq \underline{w}_o + \lambda \underline{d}$. Then, $\forall \lambda \in [0, \lambda_0]$, $\underline{w}(\lambda) \in B(\underline{w}_o, \delta_1) \cap W$ and $\underline{w}(\lambda) \in B(\underline{w}_o, \delta_2) \cap W$. Therefore, by (A.4) and (A.5),

$$\lambda \in [0, \lambda_0] \rightarrow \forall \underline{\phi} \in \overline{B}(M(\underline{w}_o), \delta_0) \cap \Phi, \ < \nabla_{\underline{w}} \varepsilon \left(\underline{w}(\lambda), \underline{\phi}\right), \underline{d} > \ < \epsilon_0/4 < 0.$$

By definition,

$$\frac{\partial \varepsilon(\underline{w} + \lambda \underline{d}, \underline{\phi})}{\partial \lambda} \mid_{\lambda=0} = k < \nabla_{\underline{w}} \varepsilon \left(\underline{w}, \underline{\phi}\right), \underline{d} >,$$

where k is a real positive constant. Therefore,

$$\lambda \in [0, \lambda_0] \rightarrow \forall \underline{\phi} \in \overline{B}(M(\underline{w}_o), \delta_0) \cap \Phi, \ \frac{\partial \varepsilon(\underline{w}(\lambda), \underline{\phi})}{\partial \lambda} < k\epsilon_0/4 < 0. \quad \text{(A.8)}$$

$\underline{w}(\lambda)$ is a continuous function on $\mathbb{R}$ and $\varepsilon(\underline{w}, \underline{\phi})$ is a continuous function on W. Therefore, by Lemma A.3, $\varepsilon(\underline{w}(\lambda), \underline{\phi})$ is a continuous function on $\mathbb{R}$. Therefore, by

(A.8) and Theorem 5.10,

$$\forall \underline{\phi} \in \overline{B}(M(\underline{w}_o), \delta_0) \cap \Phi, \;\; \varepsilon(\underline{w}(\lambda_0), \underline{\phi}) \; < \; \varepsilon(\underline{w}_o, \underline{\phi}) + \lambda_0 k \epsilon_0 / 4.$$

By definition, $\varepsilon(\underline{w}(\lambda_0), \underline{\phi}) \; \leq \Delta(\underline{w}_o)$, and $\lambda_0, k \; > 0$ and $\epsilon_0 \; < 0$. Therefore, letting $\epsilon_3 \; = \; \Delta(\underline{w}_o) + \lambda_0 k \epsilon_0 / 4$,

$$\forall \underline{\phi} \in \overline{B}(M(\underline{w}_o), \delta_0) \cap \Phi, \;\; \varepsilon(\underline{w}(\lambda_0), \underline{\phi}) \; < \; \epsilon_3 \; < \; \Delta(\underline{w}_o). \tag{A.9}$$

Let $\epsilon_4 \; = \; \max(\epsilon_2, \epsilon_3) \; < \Delta(\underline{w}_o)$. Then, since $\forall \lambda \in [0, \lambda_0], \;\; \underline{w}(\lambda) \in B(\underline{w}_o, \delta_2) \cap W$, (A.6), (A.7), and (A.9) can be combined to yield

$$\forall \underline{\phi} \in \Phi, \; \varepsilon(\underline{w}(\lambda_0), \underline{\phi}) \; < \; \epsilon_4 \; < \; \Delta(\underline{w}_o).$$

Therefore, $\Delta(\underline{w}(\lambda_0)) \; < \; \Delta(\underline{w}_o)$.

Therefore, $\underline{w}_o$ is not a minimizing point of $\Delta(\underline{w})$ on W. Therefore, a necessary condition for $\underline{w}_o \in W$ to be a solution to the following minmax problem

$$\underline{w}_{opt} = \arg \min_{\underline{w} \in W} \max_{\underline{\phi} \in \Phi} \varepsilon(\underline{w}, \underline{\phi}),$$

is that

$$\underline{0} \in \mathcal{H}\left(\{ \nabla_{\underline{w}} \varepsilon \left(\underline{w}_o, \underline{\phi} \right) \mid \underline{\phi} \in M(\underline{w}_o) \} \right).$$

The sufficient condition portion of the theorem will now be proven.

Assume that W is a convex set, $\varepsilon(\underline{w}, \underline{\phi})$ is a convex function on W, and that

$$\underline{0} \in \mathcal{H}\left(\{ \nabla_{\underline{w}} \varepsilon \left(\underline{w}_o, \underline{\phi} \right) \mid \underline{\phi} \in M(\underline{w}_o) \} \right).$$

$\{ \nabla_{\underline{w}} \varepsilon \left(\underline{w}_o, \underline{\phi} \right) \mid \underline{\phi} \in M(\underline{w}_o) \}$ is a compact subset of the Euclidian metric space E. Therefore, by the Theorem on Linear Inequalities, $\forall \underline{d} \in E, \; \exists \underline{\phi}_0 \in M(\underline{w}_o)$ such that $< \nabla_{\underline{w}} \varepsilon \left(\underline{w}_o, \underline{\phi}_0 \right), \underline{d} > \; \geq 0$.

Select any $\underline{w}_1 \in W$. Let $\underline{d} = \underline{w}_1 - \underline{w}_o \in E$ and define $\underline{w}(\lambda)$ as before. Then $\underline{w}_o = \underline{w}(\lambda) \mid_{\lambda=0}$ and $\underline{w}_1 = \underline{w}(\lambda) \mid_{\lambda=1}$. Then, by the definition of the gradient operator, $\exists \underline{\phi}_0 \in M(\underline{w}_o)$ such that $\frac{\partial \varepsilon(\underline{w}(\lambda), \underline{\phi}_0)}{\partial \lambda} \mid_{\lambda=0} \geq 0$.

By the definition of convex functions and convex sets,

$$\forall \lambda \in [0,1], \; \varepsilon(\underline{w}(\lambda), \underline{\phi}_0) \; \leq \; (1-\lambda) \varepsilon(\underline{w}_o, \underline{\phi}_0) \; + \; \lambda \varepsilon(\underline{w}_1, \underline{\phi}_0).$$

This can be rewritten as

$$\forall \lambda \in [0,1], \; \varepsilon(\underline{w}(\lambda), \underline{\phi}_0) \; \leq \; \varepsilon(\underline{w}_o, \underline{\phi}_0) \; + \; \lambda (\varepsilon(\underline{w}_1, \underline{\phi}_0) \; - \; \varepsilon(\underline{w}_o, \underline{\phi}_0)). \tag{A.10}$$

By the definition of the directional derivative

$$\frac{\partial \varepsilon(\underline{w}(\lambda), \underline{\phi}_0)}{\partial \lambda} \mid_{\lambda=0} = \lim_{\lambda \to 0^+} \frac{\varepsilon(\underline{w}(\lambda), \underline{\phi}_0) \; - \; \varepsilon(\underline{w}_0, \underline{\phi}_0)}{\lambda} \; \geq \; 0. \tag{A.11}$$

Combining (A.10) and (A.11) and carrying out some algebraic manipulation yields $\varepsilon(\underline{w}_1, \underline{\phi}_0) \geq \varepsilon(\underline{w}_o, \underline{\phi}_0)$. Since $\underline{\phi}_0 \in M(\underline{w}_o)$, $\varepsilon(\underline{w}_o, \underline{\phi}_0) = \Delta(\underline{w}_o)$. Furthermore, $\Delta(\underline{w}_1) \geq \varepsilon(\underline{w}_1, \underline{\phi}_0)$. Therefore, $\Delta(\underline{w}_1) \geq \Delta(\underline{w}_o)$. This is true for any $\underline{w}_1 \in W$. Therefore, $\underline{w}_o$ is the minimizing point of $\Delta(\underline{w})$ on W.

Therefore, if $\varepsilon(\underline{w}, \underline{\phi})$ is convex on W and W is a convex set, then

$$\underline{0} \in \mathcal{H}\left(\{\nabla_{\underline{w}} \varepsilon\left(\underline{w}_o, \underline{\phi}\right) \mid \underline{\phi} \in M(\underline{w}_o)\}\right)$$

is a sufficient condition that $\underline{w}_o$ is a solution to

$$\underline{w}_{opt} = \arg \min_{\underline{w} \in W} \max_{\underline{\phi} \in \Phi} \varepsilon(\underline{w}, \underline{\phi}).$$

■

Theorem 5 Let $\varepsilon(g, \phi)$ be any performance measure for the estimator g given the environmental condition ϕ. Let $\Phi = \{\phi_1, \ldots, \phi_K\}$ and let $\underline{p} \in P = \{\underline{p} \in \mathbb{R}^K \mid \underline{p} \geq 0 \text{ and } \underline{e}^t \underline{p} = 1\}$ be any pmf assigned to ϕ on Φ. Let $\varepsilon(\underline{p}, g) = \sum_{i=1}^K p_i \varepsilon(g, \phi_i)$ and let $\beta(\underline{p})$ be any global lower bound on $\varepsilon(\underline{p}, g)$. That is,

$$\beta(\underline{p}) \leq \min_{g: X \to \mathbb{R}} \varepsilon(\underline{p}, g).$$

Then, $\beta(\underline{p})$ is also a lower bound on $\Delta(g)$. That is,

$$\Delta_o \triangleq \min_{g: X \to \mathbb{R}} \Delta(g) \geq \beta(\underline{p}).$$

Proof:

$$\varepsilon(\underline{p}, g) = \sum_{i=1}^K p_i \varepsilon(g, \phi_i) \leq \sum_{i=1}^K p_i (\max_{\phi \in \Phi} \varepsilon(g, \phi)) = \max_{\phi \in \Phi} \varepsilon(g, \phi) = \Delta(g)$$

This hold true $\forall g : X \to \mathbb{R}$ and $\forall \underline{p} \in P$. Therefore,

$$\beta(\underline{p}) \leq \min_{g: X \to \mathbb{R}} \varepsilon(\underline{p}, g) \leq \min_{g: X \to \mathbb{R}} \Delta(g) = \Delta_o.$$

This holds true for any $\underline{p} \in P$. Therefore,

$$\forall \underline{p} \in P, \ \beta(\underline{p}) \leq \Delta_o.$$

■

Theorem 6 Let $\Phi = \{\phi_1, \ldots, \phi_K\}$, $P = \{\underline{p} \in \mathbb{R}^K \mid \underline{p} \geq 0 \text{ and } \underline{e}^t \underline{p} = 1\}$, and $g_o : X \times P \to \mathbb{R}$ be given by

$$g_o(x, \underline{p}) = \frac{\sum_{i=1}^K p_i \, p_{x|\phi}(x \mid \phi_i) \mathrm{E}[\theta \mid x, \phi_i]}{\sum_{i=1}^K p_i \, p_{x|\phi}(x \mid \phi_i)}, \tag{A.12}$$

where $p_{x|\phi}(x \mid \phi_i)$ is the conditional pdf or pmf of the observation x given that the environmental parameter $\phi = \phi_i$. Let $\varepsilon(g, \phi) = \mathrm{E}[(\theta - g(x))^2 \mid \phi]$.

Let the least favorable pmf $\underline{p}_{lf} \in P$ be defined as

$$\underline{p}_{lf} \triangleq \arg\max_{\underline{p} \in P} \sum_{i=1}^{K} p_i \, \varepsilon(g_o(x, \underline{p}), \phi_i).$$

Then

$$g_o(x, \underline{p}_{lf}) = \arg \min_{g: X \to \mathbb{R}} \max_{\phi \in \Phi} \varepsilon(g, \phi),$$

and the quantity

$$\sum_{i=1}^{K} p_{lf_i} \, \varepsilon(g_o(x, \underline{p}_{lf}), \phi_i)$$

is an achievable lower bound on

$$\Delta(g) = \max_{\phi \in \Phi} \varepsilon(g, \phi).$$

Proof: This proof closely parallels that of Theorem 4.

Consider ϕ to be a random parameter and let $\underline{p}$ be any assigned pmf. Then

$$\varepsilon(\underline{p}, g) = \sum_{i=1}^{K} p_i \, \varepsilon(g, \phi_i)$$

is the mean squared estimation error for the estimator g and algabraic manipulation of (A.12) shows that $g_o(x, \underline{p}) = \mathrm{E}[\theta \mid x]$ and is therefore the minimum mean squared error estimator of θ given the observation x. Therefore,

$$\forall g : X \to \mathbb{R} \text{ and } \forall \underline{p} \in P, \ \varepsilon(\underline{p}, g_o(x, \underline{p})) \leq \varepsilon(\underline{p}, g).$$

Therefore, $\varepsilon(\underline{p}, g_o(x, \underline{p}))$ is a global lower bound on $\varepsilon(\underline{p}, g)$ and, by Theorem 5,

$$\forall \underline{p} \in P, \ \varepsilon(\underline{p}, g_o(x, \underline{p})) \leq \Delta_o \triangleq \min_{g: X \to \mathbb{R}} \Delta(g). \tag{A.13}$$

It will now be shown that $\varepsilon(\underline{p}_{lf}, g_o(x, \underline{p}_{lf})) \geq \Delta_o$. Recall that

$$\varepsilon(\underline{p}, g_o(x, \underline{p})) = \sum_{i=1}^{K} p_i \, \mathrm{E}[(\theta - g_o(x, \underline{p}))^2 \mid \phi_i].$$

Therefore,

$$\frac{\partial \varepsilon(\underline{p}, g_o(x, \underline{p}))}{\partial p_j} = \mathrm{E}[(\theta - g_o(x, \underline{p}))^2 \mid \phi_j] + \sum_{i=1}^{K} p_i \, \frac{\partial \mathrm{E}[(\theta - g_o(x, \underline{p}))^2 \mid \phi_i]}{\partial p_j}, \tag{A.14}$$

and

$$\frac{\partial \mathrm{E}[(\theta - g_o(x,\underline{p}))^2 \mid \phi_i]}{\partial p_j} = -2\mathrm{E}[(\theta - g_o(x,\underline{p}))\frac{\partial g_o(x,\underline{p})}{\partial p_j} \mid \phi_i]. \quad \text{(A.15)}$$

Substituting (A.15) into (A.14) and carrying out the summation yields

$$\frac{\partial \varepsilon(\underline{p}, g_o(x,\underline{p}))}{\partial p_j} = \mathrm{E}[(\theta - g_o(x,\underline{p}))^2 \mid \phi_j] - 2\mathrm{E}[(\theta - g_o(x,\underline{p}))\frac{\partial g_o(x,\underline{p})}{\partial p_j}]. \quad \text{(A.16)}$$

Recall that $g_o(x,\underline{p}) = \mathrm{E}[\theta \mid x]$ is the minimum mean squared error estimator of θ given x. Therefore,

$$\mathrm{E}[(\theta - g_o(x,\underline{p}))\frac{\partial g_o(x,\underline{p})}{\partial p_j}] = \mathrm{E}[(\theta - \mathrm{E}[\theta \mid x])\frac{\partial g_o(x,\underline{p})}{\partial p_j}]. \quad \text{(A.17)}$$

Since $\frac{\partial g_o(x,\underline{p})}{\partial p_j}$ is a function of x and is not a function of θ, the application of the orthogonality principle yields

$$\mathrm{E}[(\theta - \mathrm{E}[\theta \mid x])\frac{\partial g_o(x,\underline{p})}{\partial p_j}] = 0. \quad \text{(A.18)}$$

Substituting (A.17) and (A.18) into (A.16) yields

$$\frac{\partial \varepsilon(\underline{p}, g_o(x,\underline{p}))}{\partial p_j} = \mathrm{E}[(\theta - g_o(x,\underline{p}))^2 \mid \phi_j] = \varepsilon(g_o(x,\underline{p}), \phi_j). \quad \text{(A.19)}$$

Let $m \in \{1,\ldots,K\}$ be any index such that $p_{lf_m} > 0$ and $n \in \{\{1,\ldots,K\} - \{m\}\}$ be any other index. For $\lambda \in [0, p_{lf_m}]$ define $\underline{p}(\lambda)$ as follows: $p_i(\lambda) = p_{lf_i}$ for $i \in \{\{1,\ldots,K\} - \{m,n\}\}$, $p_m(\lambda) = p_{lf_m} - \lambda$, and $p_n(\lambda) = p_{lf_n} + \lambda$.

Using the chain rule,

$$\frac{\partial \varepsilon(\underline{p}(\lambda), g_o(x,\underline{p}(\lambda)))}{\partial \lambda} = \sum_{i=1}^{K} \frac{\partial \varepsilon(\underline{p}(\lambda), g_o(x,\underline{p}(\lambda)))}{\partial p_i(\lambda)} \frac{\partial p_i(\lambda)}{\partial \lambda}. \quad \text{(A.20)}$$

Substituting (A.19) into (A.20) and noting that

$$\frac{\partial p_i(\lambda)}{\partial \lambda} = 0 \; \forall i \in \{\{1,\ldots,K\} - \{m,n\}\}, \quad \frac{\partial p_m(\lambda)}{\partial \lambda} = -1, \text{ and } \frac{\partial p_n(\lambda)}{\partial \lambda} = 1$$

yields

$$\frac{\partial \varepsilon(\underline{p}(\lambda), g_o(x,\underline{p}(\lambda)))}{\partial \lambda} = \varepsilon(g_o(x,\underline{p}(\lambda)), \phi_n) - \varepsilon(g_o(x,\underline{p}(\lambda)), \phi_m). \quad \text{(A.21)}$$

Note that by the definition of $\underline{p}_{lf}$,

$$\forall \lambda \in [0, p_{lf_m}], \quad \varepsilon(g_o(x,\underline{p}(\lambda)), \underline{p}(\lambda)) \leq \varepsilon(g_o(x,\underline{p}_{lf}), \underline{p}_{lf});$$

and by the definiton of $\underline{p}(\lambda)$,

$$\varepsilon(g_o(x,\underline{p}(\lambda)),\underline{p}(\lambda))\mid_{\lambda=0} = \varepsilon(g_o(x,\underline{p}_{lf}),\underline{p}_{lf})$$

Therefore,

$$0 = \arg\max_{\lambda\in[0,p_{lf_m}]} \varepsilon(g_o(x,\underline{p}(\lambda)),\underline{p}(\lambda)).$$

Therefore,

$$\frac{\partial\varepsilon(\underline{p}(\lambda),g_o(x,\underline{p}(\lambda)))}{\partial\lambda}\mid_{\lambda=0} \leq 0. \tag{A.22}$$

Substituting (A.21) into (A.22) and noting that $g_o(x,\underline{p}(\lambda))\mid_{\lambda=0}= g_o(x,\underline{p}_{lf})$ yields

$$\varepsilon(g_o(x,\underline{p}_{lf}),\phi_n) \leq \varepsilon(g_o(x,\underline{p}_{lf}),\phi_m). \tag{A.23}$$

Since (A.23) holds for any index m such that $p_{lf_m} > 0$ and for any other index n, it can be concluded that

$$p_{lf_m} > 0 \rightarrow \varepsilon(g_o(x,\underline{p}_{lf}),\phi_m) = \Delta(g_o(x,\underline{p}_{lf})).$$

Using this result and noting that the following sum can be carried out over only those indecies for which $\underline{p}_{lf} > 0$ and that $\underline{e}^t\,\underline{p}_{lf} = 1$ yields

$$\varepsilon(\underline{p}_{lf},g_o(x,\underline{p}_{lf})) = \sum_{i=1}^{K} p_{lf_i}\varepsilon(g_o(x,\underline{p}_{lf}),\phi_i) = \Delta(g_o(x,\underline{p}_{lf})).$$

Therefore,

$$\Delta_o \triangleq \min_{g:X\rightarrow\mathbb{R}} \Delta(g) \leq \Delta(g_o(x,\underline{p}_{lf})) = \varepsilon(\underline{p}_{lf},g_o(x,\underline{p}_{lf})). \tag{A.24}$$

Combining (A.13) and (A.24) and noting that (A.13) holds for any pmf yields

$$\Delta_o \triangleq \min_{g:X\rightarrow\mathbb{R}} \Delta(g) = \Delta(g_o(x,\underline{p}_{lf})) = \varepsilon(\underline{p}_{lf},g_o(x,\underline{p}_{lf})).$$

Therefore,

$$\varepsilon(\underline{p}_{lf},g_o(x,\underline{p}_{lf})) = \sum_{i=1}^{K} p_{lf_i}\,\varepsilon(g_o(x,\underline{p}_{lf}),\phi_i)$$

is an achievable lower bound on $\Delta(g)$ and this bound is achieved by $g_o(x,\underline{p}_{lf})$. ■

Appendix B

Proofs Specific to the Adaptive Minmax Array Processor

In the adaptive minmax array processing problem, the set W and the Euclidian metric space E defined in Appendix A are both the N-dimensional complex plane denoted $\mathcal{C}^N$. This is a $2N$-dimensional Euclidian metric space with the inner product of $\underline{u}, \underline{v} \in \mathcal{C}^N$ defined as

$$< \underline{u}, \underline{v} > \triangleq \text{Real}(\underline{u}^h \underline{v}),$$

the norm of $\underline{u} \in \mathcal{C}^N$ defined as

$$| \underline{u} | \triangleq < \underline{u}, \underline{u} >^{\frac{1}{2}},$$

and the distance between $\underline{u} \in \mathcal{C}^N$ and $\underline{v} \in \mathcal{C}^N$ defined as

$$d_{\mathcal{C}^N}(\underline{u}, \underline{v}) \triangleq | \underline{u} - \underline{v} | .$$

For the error function $\varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2)$ as expressed in (2.11), the gradient operator is given by [28]

$$\nabla_{\underline{w}} \varepsilon \left(\underline{w}, \underline{\phi} \right) = \hat{\mathrm{S}}(f) \underline{w} - \sigma_o^2 \underline{q}(f, \underline{z}, \underline{\phi}). \tag{B.1}$$

Under the assumption that $\underline{q}(f, \underline{z}, \underline{\phi})$ is continuous on Φ and with the constant $k = 2$,

this gradient operator meets all of the requirements set forth in the definition of the gradient operator contained in Appendix A.

The following array weight characterization theorem sets out the necessary and sufficient conditions for a set of array weights to be the solution to the minmax array weight problem.

Theorem 3 Let Φ be a compact set contained in a metric space and $\underline{q}(f, \underline{z}, \underline{\phi})$ be a continuous function on Φ. Then a sufficient condition for $\underline{w}_o$ to be a solution to the following minmax problem

$$\underline{w}_{opt}(f, \underline{z}, \sigma_o^2) = \arg \min_{\underline{w} \in \mathcal{C}^N} \max_{\underline{\phi} \in \Phi} \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2),$$

is that

$$\exists J > 0\,, \tag{B.2}$$

and

$$\exists \tilde{M}(\underline{w}_o) = \{\underline{\phi}_1, \ldots, \underline{\phi}_J\} \subseteq M(\underline{w}_o), \tag{B.3}$$

such that

$$\underline{0} \in \mathcal{H}\left(\{(\hat{\mathrm{S}}(f)\underline{w}_o - \sigma_o^2 \underline{q}(f, \underline{z}, \underline{\phi})) \mid \underline{\phi} \in \tilde{M}(\underline{w}_o)\}\right). \tag{B.4}$$

A necessary condition for $\underline{w}_o$ to be a solution to the following minmax problem

$$\underline{w}_{opt}(f, \underline{z}, \sigma_o^2) = \arg \min_{\underline{w} \in \mathcal{C}^N} \max_{\underline{\phi} \in \Phi} \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2),$$

is that

$$\exists J \in \{1, \ldots, 2N+1\} \tag{B.5}$$

for which (B.3) and (B.4) are satisfied.

Proof: $\varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2)$ is a continuous function of $\underline{w}$ and $\underline{q}(f, \underline{z}, \underline{\phi})$; and $\underline{q}(f, \underline{z}, \underline{\phi})$ is a continuous function on Φ. Therefore, by Lemma A.3, $\varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2)$ is a continuous function on $\mathcal{C}^N$ and Φ. Φ is compact and at each $\underline{w} \in \mathcal{C}^N$, a directional derivative with respect to $\underline{w}$ can be defined on Φ. In addition, $\hat{\mathrm{S}}(f)$ is positive definite and therefore $\varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2)$ is convex on $\mathcal{C}^N$. $\mathcal{C}^N$ is a convex set. Therefore, by Theorem 1 and the definition of the gradient operator at the beginning of this appendix, a necessary and sufficient condition for $\underline{w}_o$ to be a solution to the following minmax problem

$$\underline{w}_{opt}(f, \underline{z}, \sigma_o^2) = \arg \min_{\underline{w} \in \mathcal{C}^N} \max_{\underline{\phi} \in \Phi} \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2),$$

is that

$$\underline{0} \in \mathcal{H}\left(\{(\hat{\mathrm{S}}(f)\underline{w}_o - \sigma_o^2 \underline{q}(f, \underline{z}, \underline{\phi})) \mid \underline{\phi} \in M(\underline{w}_o)\}\right). \tag{B.6}$$

If (B.6) is satisfied on a countable subset of $M(\underline{w}_o)$ then it is also satisfied on $M(\underline{w}_o)$. Therefore, replacing $M(\underline{w}_o)$ in (B.6) by an arbitrary countable subset $\tilde{M}(\underline{w}_o)$ and

noting that (B.6) is a sufficient condition yields the sufficient conditions, (B.2) through (B.4), for $\underline{w}_o$ to be the desired solution.

$\mathcal{C}^N$ is a $2N$-dimensional linear space. Therefore, if (B.6) is satisfied on $M(\underline{w}_o)$, by the Theorem of Carathéodory it must also be satisfied on a subset of $M(\underline{w}_o)$ containing $2N+1$ or fewer elements. Therefore, noting that (B.6) is a necessary condition yields the necessary conditions, (B.5), (B.3) and (B.4), for $\underline{w}_o$ to be the desired solution. ■

Using Theorem (3), the following theorem can be proven.

Theorem 2 Let σ_o^2 be any real positive number, Φ be a compact set contained in a metric space, $\underline{q}(f, \underline{z}, \underline{\phi})$ be a continuous function on Φ, and

$$\underline{w}_{opt}(f, \underline{z}, \sigma_o^2) = \arg \min_{\underline{w} \in \mathcal{C}^N} \max_{\underline{\phi} \in \Phi} \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2).$$

Then for any real non-negative σ^2, the solution to the problem

$$\underline{w}_{opt}(f, \underline{z}, \sigma^2) = \arg \min_{\underline{w} \in \mathcal{C}^N} \max_{\underline{\phi} \in \Phi} \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma^2)$$

is given by

$$\underline{w}_{opt}(f, \underline{z}, \sigma^2) = (\sigma^2/\sigma_o^2)\, \underline{w}_{opt}(f, \underline{z}, \sigma_o^2).$$

Proof: The following notation, which is more explicit than that used in the body of this thesis and in the proofs of the other theorems, will be used in this proof. The extremal value will be denoted by $\Delta(f, \underline{z}, \underline{w}, \sigma^2)$ and the extremal point set will be denoted by $M(f, \underline{z}, \underline{w}, \sigma^2)$ where

$$\Delta(f, \underline{z}, \underline{w}, \sigma^2) \triangleq \max_{\underline{\phi} \in \Phi} \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma^2)$$

and

$$M(f, \underline{z}, \underline{w}, \sigma^2) \triangleq \{\phi \in \Phi \mid \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma^2) = \Delta(f, \underline{z}, \underline{w}, \sigma^2)\}.$$

For any positive real σ_o^2, let

$$\underline{w}_{opt}(f, \underline{z}, \sigma_o^2) = \arg \min_{\underline{w} \in C^N} \max_{\underline{\phi} \in \Phi} \varepsilon(f, \underline{z}, \underline{w}, \underline{\phi}, \sigma_o^2).$$

For any positive real σ^2, define

$$\underline{w}(\sigma^2) \triangleq (\sigma^2/\sigma_o^2)\, \underline{w}_{opt}(f, \underline{z}, \sigma_o^2). \tag{B.7}$$

By Theorem 3,

$$\underline{0} \in \mathcal{H}\left(\{(\hat{\mathrm{S}}(f)\, \underline{w}_{opt}(f, \underline{z}, \sigma_o^2) - \sigma_o^2\, \underline{q}(f, \underline{z}, \underline{\phi})) \mid \underline{\phi} \in M(f, \underline{z}, \underline{w}_{opt}(f, \underline{z}, \sigma_o^2), \sigma_o^2)\}\right). \tag{B.8}$$

Substituting (B.7) into (B.8) yields

$$\underline{0} \in \mathcal{H}\left(\{(\sigma_0^2/\sigma^2)\,(\hat{\mathrm{S}}(f)\,\underline{w}(\sigma^2) - \sigma^2\,\underline{q}(f,\underline{z},\underline{\phi}))\;|\;\underline{\phi} \in M(f,\underline{z},\underline{w}_{opt}(f,\underline{z},\sigma_o^2),\sigma_o^2)\}\right).$$

This implies that

$$\underline{0} \in \mathcal{H}\left(\{(\hat{\mathrm{S}}(f)\,\underline{w}(\sigma^2) - \sigma^2\,\underline{q}(f,\underline{z},\underline{\phi}))\;|\;\underline{\phi} \in M(f,\underline{z},\underline{w}_{opt}(f,\underline{z},\sigma_o^2),\sigma_o^2)\}\right). \qquad \text{(B.9)}$$

Substituting (B.7) into (2.11) and carrying out some algebraic manipulation yields the following expression for $\varepsilon(f,\underline{z},\underline{w}(\sigma^2),\underline{\phi},\sigma^2)$.

$$\varepsilon(f,\underline{z},\underline{w}(\sigma^2),\underline{\phi},\sigma^2) = (\sigma^2/\sigma_o^2)^2\,\varepsilon(f,\underline{z},\underline{w}_{opt}(f,\underline{z},\sigma_o^2),\underline{\phi},\sigma_o^2) + (\sigma^2 - (\sigma^2)^2/\sigma_o^2).$$

Note that $\varepsilon(f,\underline{z},\underline{w}(\sigma^2),\underline{\phi},\sigma^2)$ is a equal to $\varepsilon(f,\underline{z},\underline{w}_{opt}(f,\underline{z},\sigma_o^2),\underline{\phi},\sigma_o^2)$ multiplied by a positive number which is independent of $\underline{\phi}$ and added to a number which is independent of $\underline{\phi}$. This implies that

$$M(f,\underline{z},\underline{w}(\sigma^2),\sigma^2) = M(f,\underline{z},\underline{w}_{opt}(f,\underline{z},\sigma_o^2),\sigma_o^2) \qquad \text{(B.10)}$$

Substituting (B.10) into (B.9) yields

$$\underline{0} \in \mathcal{H}\left(\{(\hat{\mathrm{S}}(f)\underline{w}(\sigma^2) - \sigma^2\,\underline{q}(f,\underline{z},\underline{\phi}))\;|\;\underline{\phi} \in M(f,\underline{z},\underline{w}(\sigma^2),\sigma^2)\}\right).$$

Therefore, $\underline{w}(\sigma^2)$ satisfies the requirements of Theorem 3. This implies that

$$\underline{w}(\sigma^2) = \arg\min_{\underline{w}\in C^N}\max_{\underline{\phi}\in\Phi}\varepsilon(f,\underline{z},\underline{w},\underline{\phi},\sigma^2).$$

Therefore,

$$\underline{w}_{opt}(f,\underline{z},\sigma^2) = (\sigma^2/\sigma_o^2)\underline{w}_{opt}(f,\underline{z},\sigma_o^2).$$

■

Theorem 4 Let P be the set of all possible pmfs which may be assigned to Φ and define the least favorable pmf as

$$\underline{p}_{lf} \triangleq \arg\max_{\underline{p}\in P}\min_{\underline{w}\in C^N}\varepsilon(f,\underline{z},\underline{w},\underline{p},\sigma_o^2) \;=\; \arg\max_{\underline{p}\in P}\varepsilon(f,\underline{z},\underline{w}_{mmse}(f,\underline{z},\sigma_o^2,\underline{p}),\underline{p},\sigma_o^2)\,. \qquad \text{(B.11)}$$

Then

$$\underline{w}_{opt}(f,\underline{z},\sigma_o^2) \triangleq \arg\min_{\underline{w}\in C^N}\max_{\underline{\phi}\in\Phi}\varepsilon(f,\underline{z},\underline{w},\underline{\phi},\sigma_o^2) \;=\; \underline{w}_{mmse}(f,\underline{z},\sigma_o^2,\underline{p}_{lf})\,.$$

Proof: From (2.22), $\varepsilon(f,\underline{z},\underline{w}_{mmse},\underline{p},\sigma_o^2)$ can be expressed as

$$\varepsilon(f,\underline{z},\underline{w}_{mmse},\underline{p},\sigma_o^2) \;=\; \sum_{i=1}^{K} p_i\;\varepsilon(f,\underline{z},\underline{w}_{mmse}(\underline{p}),\underline{\phi}_i,\sigma_o^2).$$

Therefore,

$$\frac{\partial \varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}, \sigma_o^2)}{\partial p_j} = \varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}), \underline{\phi}_j, \sigma_o^2) + \sum_{i=1}^{K} p_i \frac{\partial \varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}), \underline{\phi}_i, \sigma_o^2)}{\partial p_j}. \tag{B.12}$$

Combining (2.5) and (2.24) yields

$$\varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}), \underline{\phi}_i, \sigma_o^2) = \sigma_o^2 - 2(\sigma_o^2)^2 \text{Real}(\underline{q}^h(f, \underline{z}, \underline{\phi}_i)\hat{\text{S}}(f)^{-1} \sum_{l=1}^{K} p_l \underline{q}(f, \underline{z}, \underline{\phi}_l)) + (\sigma_o^2)^2 \sum_{l=1}^{K} \sum_{m=1}^{K} p_l \, \underline{q}^h(f, \underline{z}, \underline{\phi}_l) \, \hat{\text{S}}(f)^{-1} \, \underline{q}(f, \underline{z}, \underline{\phi}_m) \, p_m.$$

Therefore,

$$\frac{\partial \varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}), \underline{\phi}_i, \sigma_o^2)}{\partial p_j} = 2(\sigma_o^2)^2 (\sum_{l=1}^{K} p_l \, \text{Real}(\underline{q}^h(f, \underline{z}, \underline{\phi}_l) \, \hat{\text{S}}(f)^{-1} \, \underline{q}(f, \underline{z}, \underline{\phi}_j)) - \text{Real}(\underline{q}^h(f, \underline{z}, \underline{\phi}_i) \, \hat{\text{S}}(f)^{-1} \, \underline{q}(f, \underline{z}, \underline{\phi}_j))).$$

Noting that the first term is independent of the index i and carrying out the summation over i in (B.12) yields

$$\sum_{i=1}^{K} p_i \frac{\partial \varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}), \underline{\phi}_i, \sigma_o^2)}{\partial p_j} = 0.$$

Therefore,

$$\frac{\partial \varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}, \sigma_o^2)}{\partial p_j} = \varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}), \underline{\phi}_j, \sigma_o^2). \tag{B.13}$$

Let $m \in \{1, \ldots, K\}$ be any index such that $p_{lf_m} > 0$ and $n \in \{\{1, \ldots, K\} - \{m\}\}$ be any other index. For $\lambda \in [0, p_{lf_m}]$ define $\underline{p}(\lambda)$ as follows: $p_i(\lambda) = p_{lf_i}$ for $i \in \{\{1, \ldots, K\} - \{m, n\}\}$, $p_m(\lambda) = p_{lf_m} - \lambda$, and $p_n(\lambda) = p_{lf_n} + \lambda$.

Using the chain rule,

$$\frac{\partial \varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}(\lambda), \sigma_o^2)}{\partial \lambda} = \sum_{i=1}^{K} \frac{\partial \varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}(\lambda), \sigma_o^2)}{\partial p_i(\lambda)} \frac{\partial p_i(\lambda)}{\partial \lambda}. \tag{B.14}$$

Substituting (B.13) into (B.14) and noting that

$$\frac{\partial p_i(\lambda)}{\partial \lambda} = 0 \; \forall i \in \{\{1, \ldots, K\} - \{m, n\}\}, \quad \frac{\partial p_m(\lambda)}{\partial \lambda} = -1, \text{ and } \frac{\partial p_n(\lambda)}{\partial \lambda} = 1$$

yields

$$\frac{\partial \varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}(\lambda), \sigma_o^2)}{\partial \lambda} = \varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}(\lambda)), \underline{\phi}_n, \sigma_o^2) - \varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}(\lambda)), \underline{\phi}_m, \sigma_o^2). \tag{B.15}$$

Note that by the definition of $\underline{p}_{lf}$,

$$\forall \lambda \in [0, p_{lf_m}], \;\; \varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}(\lambda), \sigma_o^2) \;\leq\; \varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}_{lf}, \sigma_o^2);$$

and by the definiton of $\underline{p}(\lambda)$,

$$\varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}(\lambda), \sigma_o^2) \mid_{\lambda=0} = \; \varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}_{lf}, \sigma_o^2).$$

Therefore,

$$0 \;=\; \arg \max_{\lambda \in [0, p_{lf_m}]} \varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}(\lambda), \sigma_o^2).$$

Therefore,

$$\frac{\partial \varepsilon(f, \underline{z}, \underline{w}_{mmse}, \underline{p}(\lambda), \sigma_o^2)}{\partial \lambda} \mid_{\lambda=0} \leq \; 0. \tag{B.16}$$

Substituting (B.15) into (B.16) and noting that $\underline{w}_{mmse}(\underline{p}(\lambda)) \mid_{\lambda=0} = \underline{w}_{mmse}(\underline{p}_{lf})$ yields,

$$\varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}_{lf}), \underline{\phi}_n, \sigma_o^2) \;\leq\; \varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}_{lf}), \underline{\phi}_m, \sigma_o^2). \tag{B.17}$$

Since (B.17) holds for any index m such that $p_{lf_m} > 0$ and for any other index n, it can be concluded that

$$p_{lf_m} \;>\; 0 \;\rightarrow\; \varepsilon(f, \underline{z}, \underline{w}_{mmse}(\underline{p}_{lf}), \underline{\phi}_m, \sigma_o^2) \;=\; \Delta(\underline{w}_{mmse}(\underline{p}_{lf})).$$

Therefore

$$p_{lf_m} \;>\; 0 \;\rightarrow\; \phi_m \in M(\underline{w}_{mmse}(f, \underline{z}, \sigma_o^2, \underline{p}_{lf})).$$

Using the notation of Corollary 1 to Theorem 3, the desired subset of the extremal point set can be selected as

$$\tilde{M}(\underline{w}_{mmse}(f, \underline{z}, \sigma_o^2, \underline{p}_{lf})) \;=\; \{\underline{\phi}_i \in \Phi \mid p_{lf_i} > 0\}.$$

Recall from (2.24) that

$$\underline{w}_{mmse}(f, \underline{z}, \sigma_o^2, \underline{p}_{lf}) \;=\; \sigma_o^2 \, \hat{\mathrm{S}}(f)^{-1} \sum_{i=1}^{K} p_{lf_i} \, \underline{q}(f, \underline{z}, \underline{\phi}_i).$$

Since all environmental conditions for which the pmf is non-zero are contained in the selected subset of the extremal point set, this summation can be carried out over only this subset Therefore, $\underline{w}_{mmse}(f, \underline{z}, \sigma_o^2, \underline{p}_{lf})$ and $\tilde{M}(\underline{w}_{mmse}(f, \underline{z}, \sigma_o^2, \underline{p}_{lf}))$ as selected satisfy the sufficient conditions of Corollary 1 to Theorem 3. Therefore,

$$\underline{w}_{opt}(f, \underline{z}, \sigma_o^2) \;=\; \underline{w}_{mmse}(f, \underline{z}, \sigma_o^2, \underline{p}_{lf}).$$

■

Claim 1 $M(f, \underline{z})$ is a co-positive-plus matrix.

Proof: From [24]

Co-positive matricies are (square) matricies M such that

$$\underline{x} \geq 0 \rightarrow \underline{x}^t M \underline{x} \geq 0.$$

Co-positive-plus matricies are co-positive matricies such that

$$\underline{x} \geq 0 \text{ and } \underline{x}^t M \underline{x} = 0 \rightarrow (M + M^t)\underline{x} = 0.$$

... The class of co-positive-plus matricies also includes positive-semidefinite matricies, ...

Therefore, it is sufficient to show that $M(f, \underline{z})$ is a positive-semidefinite matrix. Recall that

$$T(f, \underline{z}) \triangleq \text{Real}(Q(f, \underline{z})^h \hat{S}(f)^{-1} Q(f, \underline{z})),$$

that $T(f, \underline{z})$ can be broken into the following components

$$T(f, \underline{z}) = \begin{bmatrix} T_{11}(f, \underline{z}) & \underline{\tilde{t}}^t(f, \underline{z}) \\ \underline{\tilde{t}}(f, \underline{z}) & \tilde{T}(f, \underline{z}) \end{bmatrix}, \tag{B.18}$$

and that $\hat{S}(f)$ is a positive-definite matrix. Therefore, $T(f, \underline{z})$ is a positive-semidefinite matrix.

Recall that $M(f, \underline{z})$ is the following matrix

$$M(f, \underline{z}) = \begin{bmatrix} 0 & -\underline{e}^t \\ \underline{e} & \tilde{M}(f, \underline{z}) \end{bmatrix}$$

where the minor

$$\tilde{M}(f, \underline{z}) = \tilde{T}(f, \underline{z}) + T_{11}(f, \underline{z})\, \underline{e}\, \underline{e}^t - \underline{e}\, \underline{\tilde{t}}^t(f, \underline{z}) - \underline{\tilde{t}}(f, \underline{z})\, \underline{e}^t. \tag{B.19}$$

Select any vector $\underline{x} = [x_1, \underline{\tilde{x}}^t]^t \in \mathbb{R}^K$. Then

$$\underline{x}^t\, M(f, \underline{z})\, \underline{x} = \underline{\tilde{x}}^t\, \tilde{M}(f, \underline{z})\, \underline{\tilde{x}}.$$

Substituting (B.19) into this expression yields

$$\underline{x}^t\, M(f, \underline{z})\, \underline{x} = \underline{\tilde{x}}^t\, \tilde{T}(f, \underline{z})\, \underline{\tilde{x}} - 2(\underline{e}^t\, \underline{\tilde{x}})\, (\underline{\tilde{t}}^t(f, \underline{z})\, \underline{\tilde{x}}) + T_{11}(f, \underline{z})\, (\underline{e}^t\, \underline{\tilde{x}})^2. \tag{B.20}$$

Let $\underline{y} = [-\underline{e}^t\, \underline{\tilde{x}}, \underline{\tilde{x}}^t]^t \in \mathbb{R}^K$. Then, since $T(f, \underline{z})$ is positive-semidefinite, the quadratic product $\underline{y}^t T(f, \underline{z}) \underline{y}$ is non-negative. Using (B.18) to expand this quadratic product yields

$$\underline{y}^t T(f, \underline{z}) \underline{y} = \underline{\tilde{x}}^t\, \tilde{T}(f, \underline{z})\, \underline{\tilde{x}} - 2(\underline{e}^t\, \underline{\tilde{x}})\, (\underline{\tilde{t}}^t(f, \underline{z})\, \underline{\tilde{x}}) + T_{11}(f, \underline{z})\, (\underline{e}^t\, \underline{\tilde{x}})^2 \geq 0. \tag{B.21}$$

Note that the righthand side of (B.20) equals the center section of (B.21). Therefore,

$$\underline{x}^t M(f,\underline{z})\,\underline{x} \;=\; \underline{y}^t T(f,\underline{z})\underline{y} \;\geq\; 0.$$

Since this is true for any vector $\underline{x} \in \mathbb{R}^K$, the matrix $M(f,\underline{z})$ is positive-semidefinite. Therefore, $M(f,\underline{z})$ is a co-positive-plus matrix. ∎

Claim 2 The set of equations (B.22) through (B.24) are consistent.

$$\underline{\tilde{p}} \geq 0, \quad \text{and } \exists \underline{\tilde{v}} \geq 0 \text{ s.t.} \tag{B.22}$$

$$\underline{\tilde{v}}^t \underline{\tilde{p}} \;=\; 0 \text{ and} \tag{B.23}$$

$$\underline{\tilde{v}} \;-\; M(f,\underline{z})\,\underline{\tilde{p}} \;=\; \underline{\tilde{q}}(f,\underline{z})\,, \tag{B.24}$$

Proof: Assume that the set of equations (B.22) through (B.24) are inconsistent. Then, since (B.22) through (B.24) are simply the result of the algebraic manipulation of (B.25) through (B.27), (B.25) through (B.27) must also be inconsistent.

$$\underline{p}_o \geq 0, \;\; \underline{e}^t \underline{p}_o = 1, \;\; \exists u \in \mathbb{R}, \;\text{and } \exists \underline{v} \geq 0 \text{ s.t.} \tag{B.25}$$

$$\underline{v}^t \underline{p}_o \;=\; 0 \text{ and} \tag{B.26}$$

$$\underline{v} \;-\; T(f,\underline{z})\,\underline{p}_o \;+ u\,\underline{e} \;=\; 0\,. \tag{B.27}$$

Therefore, since (B.25) through (B.27) are the necessary and sufficient conditions which must be met by any solution to the constrained quadratic minimization problem (B.28), this problem cannot have a solution.

$$\underline{p}_{lf} \;=\; \arg\min_{\underline{p} \in P} \; \underline{p}^t\, T(f,\underline{z})\,\underline{p}. \tag{B.28}$$

However, $\underline{p}^t\, T(f,\underline{z})\,\underline{p}$ is a continuous function of $\underline{p}$ and the set $P = \{\underline{p} \in \mathbb{R}^K \mid \underline{p} \geq 0 \text{ and } \underline{e}^t\,\underline{p} = 1\}$ is compact. Therefore, by Theroem 4.16, $\exists \underline{p}_o \in P$ such that

$$\underline{p}_o^t\, T(f,\underline{z})\,\underline{p}_o \;=\; \inf_{\underline{p} \in P} \; \underline{p}^t\, T(f,\underline{z})\,\underline{p} \;=\; \min_{\underline{p} \in P} \; \underline{p}^t\, T(f,\underline{z})\,\underline{p}.$$

Therefore, $\underline{p}_o$ is a solution to (B.28) which contradicts the statement that the problem cannot have a solution. Therefore, the assumption must be wrong.

Therefore, the set of equations (B.22) through (B.24) are consistent. ∎

Bibliography

[1] V.F. Dem'yanov, V.N. Malozemov, "*Introduction to Minimax.*" John Wiley & Sons, 1974.

[2] W. Rudin, "*Principles of Mathematical Analysis.*", 3rd Ed., McGraw Hill, 1976.

[3] E.W. Cheney, "*Introduction to Approximation Theory.*" McGraw Hill, 1966.

[4] F.B. Hildebrand, "*Methods of Applied Mathematics.*", 2nd Ed., Prentice-Hall, 1965.

[5] M. Abramowitz, I.A. Stegun, "*Handbook of Mathematical Functions.*" National Bureau of Standards Applied Mathematics Series 55, U.S. Department of Commerce, 1972.

[6] W.F. Gabriel, "Using Spectral Estimation Techniques in Adaptive Processing Antenna Systems", *IEEE Trans. on Antennas and Propagation*, vol. AP-34, no. 3, March 1986, pp. 291-300.

[7] S.P. Applebaum, D.J. Chapman, "Adaptive Arrays with Main Beam Constraints", *IEEE Trans. on Antennas and Propagation*, vol. AP-24, no. 5, September 1976, pp. 650-662.

[8] H. Cox, R.M. Zeskind, M.M. Owen, "Robust Adaptive Beamforming", *IEEE Trans. on ASSP*, vol. ASSP-35, no. 10, October 1987, pp. 1365-1376.

[9] J. Capon, "High Resolution Frequency-Wavenumber Spectrum Analysis", *Proc of the IEEE*, vol. 57, no. 8, August 1969, pp. 1408-1418.

[10] S.P. Applebaum, "Adaptive Arrays", *IEEE Trans. on Antennas and Propagation*, vol. AP-24, no. 5, September 1976, pp. 585-598.

[11] A.B. Baggeroer, W.A. Kuperman, H. Schmidt, "Matched Field Processing; Source localization in correlated noise as an optimum parameter estimation problem", *JASA*, vol. 83, no. 2, February 1988, pp. 571-587.

[12] H.P. Bucker, "Use of calculated sound fields and matched-field detection to locate sound sources in shallow water", *JASA*, vol. 59, no. 2, February 1976, pp. 368-373.

[13] H. Cox, "Resolving Power and Sensitivity to Mismatch of Optimum Array Processors", *JASA*, vol. 54, no. 2, March 1973, pp. 771-785.

[14] D.F. Gingras, "Methods for predicting the sensitivity of matched-field processors to mismatch", *JASA*, vol. 86, no. 5, November 1989, pp. 1940-1949.

[15] E.C. Shang, Y.Y. Wang, "Environmental mismatching effects on source localization processing in mode space", *JASA*, vol. 89, no. 5, May 1991, pp. 2285-2290.

[16] P.M. Velardo Jr., "*Robust Matched Field Source Localization*", MS Thesis, Massachusetts Institute of Technology, January 1989.

[17] H. Schmidt, A.B. Baggeroer, W.A. Kuperman, E.K. Scheer, "Environmentally tolerant beamforming for high resolution match field processing; deterministic mismatch", *JASA*, vol. 88, no. 4, October 1990, pp. 1851-1862.

[18] J. L. Krolik, "Matched Field Minimum Variance Beamforming in a Random Ocean Channel", submitted to *JASA*.

[19] K.M. Ahmed, R.J. Evans, "An Adaptive Array Processor with Robustness and Broad-Band Capabilities", *IEEE Trans. on Antennas and Propagation*, vol. AP-32, no. 9, September 1984, pp. 944-950.

[20] A.M. Richardson, L.W. Nolte, "*A posteriori* probability source localization in an uncertain sound speed, deep ocean environment", *JASA*, vol. 89, no. 5, May 1991, pp. 2280-2284.

[21] H.V. Poor, "Uncertainty Tolerance in Underwater Acoustic Signal Processing", *IEEE Journal of Oceanic Eng.*, OE-12, no. 1, January 1987, pp. 48-65.

[22] C.L. Dolph, "A Current Distribution for Broadsize Array Which Optimizes the Relationship Between Beam Width and Side-Lobe Level", *Proc. of the IRE and Waves and Electrons*, June 1946, pp. 335-348.

[23] L.R. Rabiner, J.H. McClellan, T.W. Parks, "FIR Digital Filter Design using Weighted Chebyshev Approximation", *Proc. of the IEEE*, vol. 63, no. 5, April 1975, pp. 595-610.

[24] C.E. Lemke, "On Complementary Pivot Theory", *Mathematics of the Decision Sciences, Part 1*, Dantzig and Veinott, Ed., American Mathematical Society, Providence, RI, 1968, pp. 95-114.

[25] M.S. Bazaraa, C.M. Shetty, "*Nonlinear Programming, Theory and Applications.*" Wiley, 1979.

[26] B.C. Eaves, "The Linear Complementarity Problem", *Management Science*, vol. 11, no. 9, May 1971, pp. 612-634.

[27] K.G. Murty, "*Linear Complementarity, Linear and Nonlinear Programming.*" Heldermann Verlag, Berlin, 1988.

[28] D.H. Brandwood, "A complex gradient operator and its application in adaptive array theory", *IEE Proc.*, vol. 130, parts F and H, no. 1, February 1983, pp. 11-16.

[29] H. Schmidt, "SAFARI. Seismo-Acoustic Fast Field Algorithm for Range Independent Environments. User's Guide," SACLANT ASW Research Centre, La Spezia, Italy, Rep. SR (1987).

[30] S.M. Flatte, "*Sound Transmission Through a Fluctuating Ocean.*" Cambridge University Press, 1979.

[31] T.E. Ewart, "Acoustic fluctuations in the open ocean-A measurement using a fixed refracted path", *JASA*, vol. 60, no. 1, July 1976, pp. 46-59.

[32] L.B. Dozier, F.D. Tappert, "Statistics of normal mode amplitudes in a random ocean. I. Theory", *JASA*, vol. 63, no. 2, February 1978, pp. 353-365.

[33] L.B. Dozier, F.D. Tappert, "Statistics of normal mode amplitudes in a random ocean. II. Computations", *JASA*, vol. 64, no. 2, August 1978, pp. 533-547.

[34] I. Tolstoy, C.S. Clay, "*Ocean Acoustics, Theory and Experiment in Underwater Sound.*" McGraw-Hill, 1966.

[35] W.A. Kuperman, F. Ingenito, "Spatial correlation of surface generated noise in a stratified ocean", *JASA*, vol. 67, no. 6, June 1980, pp. 1988-1996.

[36] T.C. Yang "A method of range and depth estimation by modal decomposition", *JASA*, vol. 82, no. 5, November 1987, pp. 1736-1745.

[37] C. Garrett, W. Munk, "Space-Time Scales of Internal Waves", *Geophysical Fluid Dynamics*, vol. 3, 1972, pp. 225-264.

[38] C. Garrett, W. Munk, "Space-Time Scales of Internal Waves: A Progress Report", *Journal of Geophysical Research*, vol. 80, no. 3, January 1975, pp. 291-297.

[39] Class notes from MIT Course #13.861, Ocean and Seabed Acoustics I, Fall 1989.

[40] Personal communications with Timothy F. Duda.

[41] Personal communications with Arthur B. Baggeroer.