National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Canada

A Multiple Time-Scale Linearly Constrained Adaptive Beamformer

George Henry Niezgoda

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy at

Concordia University

Montréal, Québec, Canada

March 1994

ISBN 0-315-90922-6

ABSTRACT

A Multiple Time-Scale Linearly Constrained Adaptive Beamformer

George Henry Niezgoda, Ph.D.
Concordia University, 1994

In the ocean environment source dynamics and a changing propagation medium determine the stationarity of wavefield measurements collected by a sensor array. As well, under realistic conditions radiating energy from independent acoustic events that overlap in time will generally differ in temporal duration. Large differences in temporal extent between wavefield components present a problem for minimum variance distortionless response (MVDR) beamformer realizations restricted to computing a fixed number of weights simultaneously on a single time-scale. The difficulty arises when adaptation time is greater than the time-scale of an interferer. Once the interference event has disappeared, the mismatch in time-scale between the beamformer and interferer leads to an unnecessary degradation in detection performance. In the hope of improving performance one possible approach involves reducing adaptation time. Unfortunately, the increase in weight noise variance accompanying the change in adaptation time may introduce a greater loss in performance than is gained through a reduction in adaptation time. For this reason a MVDR implementation limited to simultaneously computing all adaptive weights on one time-scale is not an ideal candidate for mixed time-scale interference environments.

This dissertation proposes a multiple time-scale version of the cascaded MVDR

realization. A unique feature of the modular MVDR (M^2VDR) beamformer is the flexibility of distributing adaptive weights among stages arranged in series. This property enables adaptation over different time-scales. To cancel short duration interference a multiple stage beamformer computes a small number of weights assigned to a stage on a time-scale approximating the duration of the interferer. Since the stage has fewer weights to compute than the total number available to the beamformer, a greater reduction in adaptation time is possible without incurring performance degradation from weight noise variance. To combat long duration interference events other beamformer stages apply the remaining adaptive weights.

Closed form expressions relating adaptation time and adaptive degrees of freedom to mean output power, mean-squared error, and output signal-to-noise ratio are derived for the M^2VDR beamformer. Expressions for mean output power show the necessity of constraining the adaptation time of each stage to ensure that the beamformer output is not ill-defined. Analysis indicates that adequate cancellation of short duration interference occurs when a least-squares solution for the stage weight vector depends on current input data samples. In addition, it is shown that the minimum mean-squares error occurs when the adaptation time assigned each stage is greater than or equal to preceding stages. To maximize signal detection when wavefield measurements include short duration interference events, the tradeoff between adaptation time and weight noise variance is examined. Assuming a Gaussian approximation of the output power statistic, a closed form expression determines the optimal adaptation time; otherwise optimization requires numerically evaluating receiver operating curves over different adaptation times. Simulation experiments demonstrate the applicability of a multiple time-scale beamformer when wavefield measurements contain short duration interference components.

I dedicated this thesis to my wife Micheline Girard, and to my parents Henry and Elizabeth Niezgoda.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to Dr. Jeffrey Krolik for both his patience and expert guidance. His understanding and encouragement eased my transition back into academic life.

I am indebted to Dr. David Swingler for our many stimulating discussions. On numerous occasions Dr. Swingler provided insight and critical review that influenced the direction of this thesis.

I wish to acknowledge the support of Defense Research Establishment Atlantic, the National Research Council of Canada, and Concordia University for their generous support. Special thanks to Mr. Steve Davies of Defense Research Establishment Atlantic for his help and understanding.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND SYMBOLS

AG	Array Gain
CSDM	Cross Spectral Density Matrix
DI	Detection Index
DFT	Discrete Fourier Transform
DOF	Degrees Of Freedom
DOA	Direction Of Arrival
erfc	Complementary Error Function
GSC	Generalized Sidelobe Canceller
i.i.d.	Identical and Independently Distributed
INR	Interference-to-Noise Ratio
LMS	Least-Mean-Square
LS	Least-Squares
MSE	Mean-Squared-Error
MMSE	Minimum Mean-Squares Error
MVDR	Minimum Variance Distortionless Response
M^2VDR	Modularized MVDR Beamformer
pdf	Probability Density Function
PFA	Probability of False Alarm
PD	Probability of Detection
ROC	Receiver Operating Curve
RLS	Recursive Least Squares
SNR	Signal-to-Noise Ratio
SVD	Singular Value Decomposition
SNR_{out}	Output Signal to Noise Ratio
I_M	Order M identity matrix
$0_{M,N}$	$M \times N$ matrix of zeros

$\mathbf{1}_{M,N}$	$M \times N$ matrix of ones
$\mathbf{0}_M$	$M \times 1$ vector of zeros
$\mathbf{1}_M$	$M \times 1$ vector of ones
χ^2_n	Complex Chi-squared distribution
$\mathtt{cov}\{\ x\ \}$	Covariance
$\mathtt{cov}\{\ x \mid y\ \}$	Conditional covariance
†	Conjugate transpose
$\mathtt{diag}(\mathbf{x})$	Diagonal matrix constructed from $\mathbf{x}$
$\mathrm{E}\ \{x\}$	Statistical expectation
$\mathrm{E}\ \{x \mid y\}$	Conditional expectation
$\in$	Contained in the set
$\mathtt{I}^+$	All positive non-zero integers
$\mathtt{N}_M(\mu, \mathbf{R})$	Complex Gaussian distribution
$\mathtt{rank}(\mathbf{X})$	Rank of $\mathbf{X}$
$\mathtt{span}(\mathbf{X})$	Linear Span of $\mathbf{X}$
$\subset$	Subset
$^\mathtt{T}$	Matrix transpose
$\mathtt{tr}(\mathbf{X})$	Trace of $\mathbf{X}$
$\mathtt{W}_M(n, \mathbf{R})$	Complex Wishart distribution
$\mathbf{X}$	Boldface upper-case symbols denote matrices
$\mathbf{x}$	Boldface lower-case symbols denote vectors
$(\mathbf{X})_{kj}$	The $(k,j)^{\mathbf{th}}$ element of matrix $\mathbf{X}$
$(\mathbf{x})_k$	The $k^{\mathbf{th}}$ element of vector $\mathbf{x}$

Chapter 1

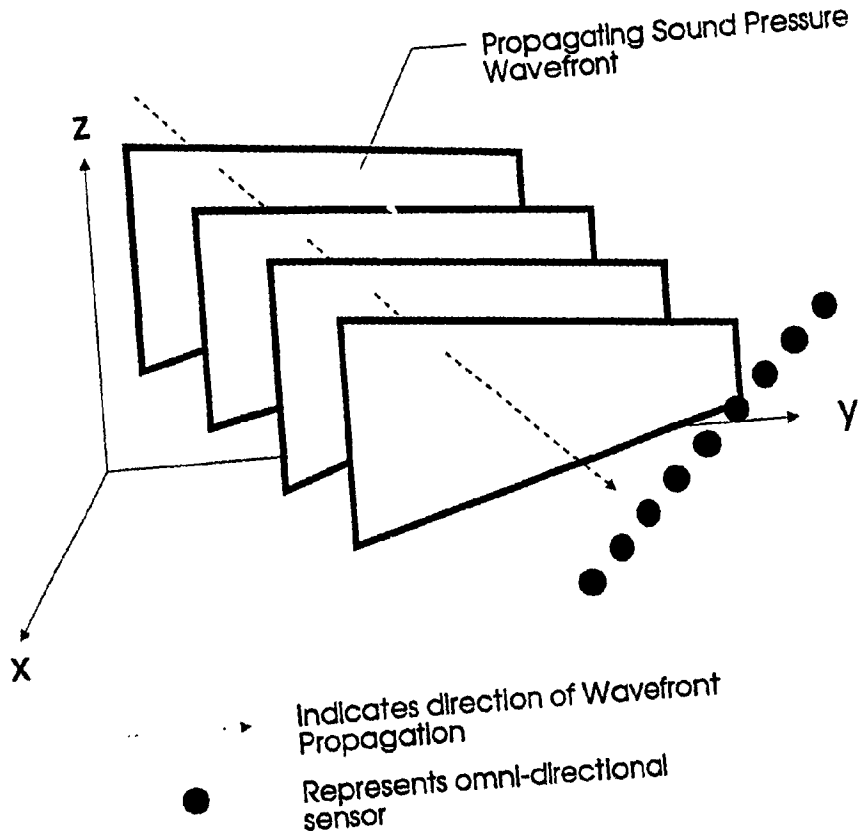
Introduction

For the past several decades the problem of detecting and localizing acoustic energy propagating in the ocean medium from data collected by a passive array of sensors has been extensively investigated. Early examples of research in array processing focus on non-adaptive detection and localization methods [1]-[8]. In [1]-[3] concern centers on the application of Bayesian and Neyman-Pearson decision criteria in detecting a signal in noise. Bryn [4] and Edelblute [5] derive the optimal detector for a signal propagating in a non-dispersive medium. McDonough [6] generalized the optimal detector to an arbitrary medium. MacDonald [8] and Bangs [7] develop the maximum likelihood estimator for signal parameters including source location. The preceding methods assume prior knowledge of the spatial and temporal correlation properties of the signal and noise components of the wavefield to achieve optimal performance. In general though, only estimates of the combined signal plus noise correlation function of the wavefield are available. Further complicating the detection and localization problem is the time-varying nature of the wavefield. This implies that detection and localization operations must be performed by adapting computations to sensor measurements.

In time-varying conditions an important factor affecting the ability to detect

and localize a signal is the appearance of interference or other continuous and intermittently radiating sources of acoustic energy. Such sources are a common feature of the wavefield which appear in sensor measurements as spatially correlated noise. Of particular concern are strong sources of interference that exhibit a high degree of directionality. Such interferers tend to obscure the presence of weaker signals. Because of wavefield dynamics, an extensive number of adaptive array processing techniques have been proposed to suppress directional interference [9]-[19]. One commonly taken approach is adaptive beamforming [9]-[14].

Before discussing adaptive beamforming, we briefly review the non-adaptive or conventional beamformer. For a detailed analysis of conventional beamforming see [20]. The basis for a conventional beamformer is a model of the physical wavefield that only considers the propagation delay between a source and the sensor array. Such a model assumes that the shape of the incident wavefront is a known function of the source location. Assuming an idealized free-space non-dispersive medium, where the location of a sensor array is in the far-field of the sources, the planar wavefront shown in Figure 1.1 provides an appropriate propagation model. To spatially discriminate between different sources of acoustic energy, the beamformer applies weights to sensor outputs that compensate for delays in a wavefront propagating from a hypothesized look direction. Summing the weighted sensor outputs and squaring the output of the beamformer yields a power estimate of the wavefield in the look direction. When this look direction coincides with the direction of a propagation wavefront, spatial samples of the wavefront sum coherently at the output of the beamformer. We refer to energy propagating in the look direction as the signal component. Energy propagating outside the look direction is referred to as interference. Since spatial samples of interference are not summed coherently by the beamforming operation, signal power gain relative to interference and the diffuse noise background results. Repeating the beamforming operation across a spatial region of interest, power estimates made at each look direction form a power



Plane Wave Propagation In a Homogenous Medium

Figure 1.1

directionality map of the wavefield. Estimates of source power and location follow by appropriately interpreting the peaks of the power directionality map.

For the conventional beamformer, accuracy in locating and detecting sources depends on the spatial window formed by the sensor array and weights applied to individual sensor outputs. Such a window imposes limits on the ability of a beamformer to resolve closely spaced sources and reject noise. Suppose 12 equi-spaced sensors are arranged along a straight line. Inter-sensor spacing is set to $\mathtt{d} = \mathtt{c}/(2f)$, where $\mathtt{c}$ equals propagation speed and f frequency. Letting $f = 1\mathtt{Hz}$, the narrow-band directional magnitude response, or beampattern, of this array to a plane wave at a $f = 1\mathtt{Hz}$ is shown in Figures 1.2a and 1.2b when sensor outputs are weighted by uniform shading, ie:

$$\mathtt{w}_m = 1. \qquad \text{for } m = 1, 2, \ldots, M$$

and by the Hamming weights

$$\mathtt{w}_m = \begin{cases} 0.54 - .046\mathtt{cos}\left(\frac{2\pi m}{M}\right) & \text{for } 0 \leq m \leq M-1 \\ 0 & \text{otherwise} \end{cases}$$

where M equals the number of sensors. $\mathtt{w}_m$ corresponds to the weight applied to the $m^{\mathtt{th}}$ sensor output. θ as shown on Figures 1.2a and 1.2b specifies bearing or the angular direction of the plane wave relative to a coordinate reference perpendicular to the array axis. We refer to this reference as the look direction. The magnitude responses shown in Figures 1.2a and 1.2b illustrate the limitations placed on resolution and interference suppression. Two attributes describe the beampattern; the mainlobe width and sidelobe levels. For a conventional beamformer, changing sensor weights to reduce the mainlobe width and improve resolution results in elevated sidelobe levels. By reducing sidelobe attenuation the ability to detect a weak signal degrades because of the increase in output power caused by the leakage of noise from outside the look direction. Conversely, lowering sidelobe levels to improve detection performance causes a reduction in resolution performance.

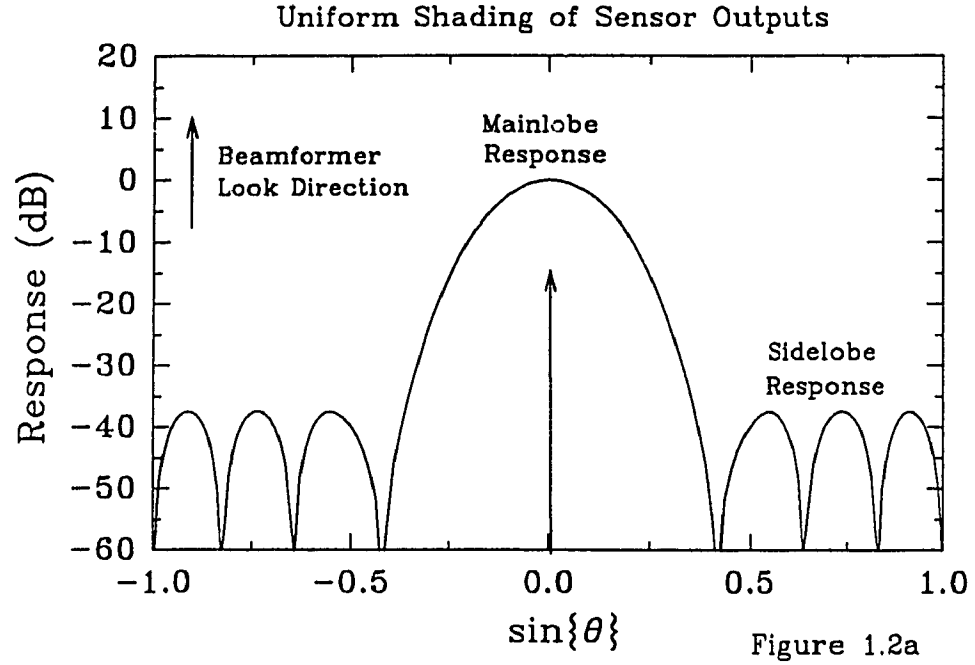


Figure 1.2a

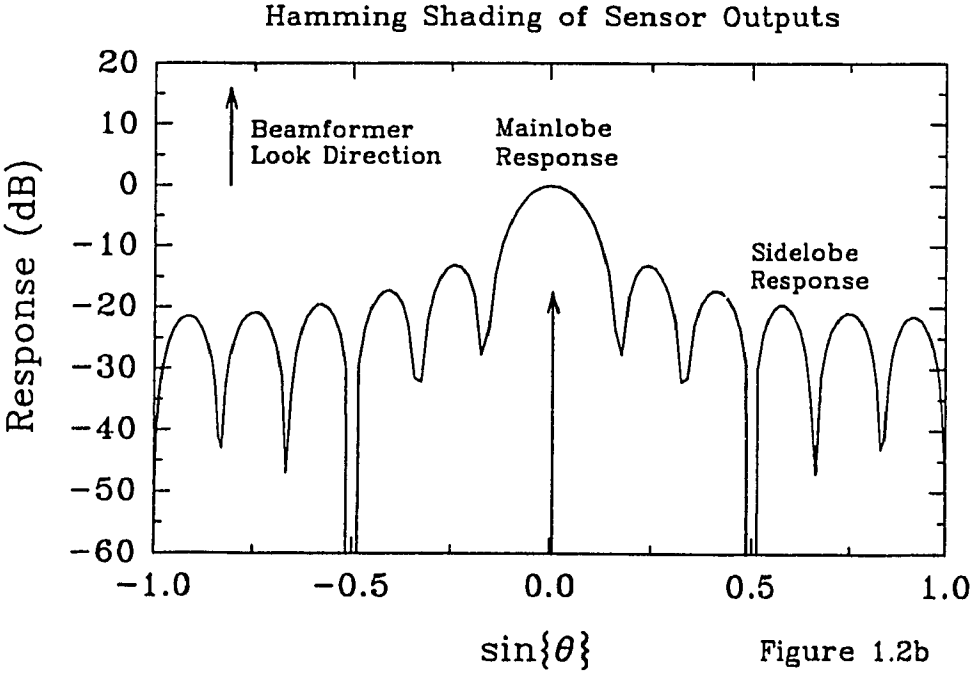


Figure 1.2b

Conventional Beamformer: Magnitude Response

Figure 1.2

In the ocean medium, signal, interference, and noise dynamics restrict the optimality of a conventional beamformer to a fixed set of conditions. For instance, a uniformly weighted linear array of sensors maximizes signal detection only when the measured wavefield consists of a single signal embedded in diffuse noise. In weak signal conditions, data independent beamforming becomes particularly susceptible to performance degradation in the presence of strong time-varying directional interference [15, 21]. This problem motivates the application of adaptive beamforming techniques that provide a way of trading detection performance in diffuse noise for interference suppression.

Historically, the methods proposed in [9]-[12] represent some of the first examples of adaptive beamforming. The adaptive noise cancellation solution developed by Howells and Applebaum [9,10] applies in situations where a reference of the noise process interferering with the desired signal is available. A relevant case for this noise canceller is machinery induced noise [22]. Acquiring a reference involves placing accelerometers in proximity to the machinery. To remove interference, the noise canceller adaptively combines the reference with the output of a conventional beamformer in such a way as to minimize output power. Developed independently of the Howells-Applebaum noise canceller, Widrow [11] proposes an adaptive beamformer based on a reference generated by a wavefront model of a signal emitted from a postulated location. The beamformer computes weights that minimize the mean-squared-error (MSE) between the beamformer output and reference. The principal drawback of Widrow's approach becomes apparent when the reference is an inaccurate model of actual wavefront propagation. Modeling errors arising from this mismatch cause signal distortion and attenuation at the output of the beamformer [23]-[26]. A generalization of the preceding techniques, proposed in [13, 14], assumes separate measurements of the signal and interference plus noise covariance matrices. This leads to a beamformer solution that maximizes output signal-to-noise ratio (SNR) by formulating sensor weights as the solution to a generalized eigenvalue

problem. However, for most practical situations, the signal and interference plus noise covariance matrices cannot be measured separately.

In [9]-[11] beamformer weights depend on a reference of the interferer or the desired signal. Such a reference effectively places soft constraints on the beamformer weights with the intent of preventing signal cancellation. Unfortunately, signal distortion arises under a variety of conditions. For instance, in the case of the noise canceller [9, 10], a signal in the presence of an interferer of comparable strength causes the beamformer to attenuate the signal component of the wavefield [25, 26]. A similar effect occurs with the reference signal beamformer [11]. To overcome this difficulty Levin and Green [27, 28] apply hard constraints to the beamformer response. Frost [12] adopts a hard constraint strategy to form the linearly constrained minimum variance distortionless response (MVDR) beamformer. For the MVDR beamformer, imposing hard constraints entails computing weight that minimize total beamformer output power subject to a unity gain response at selected spatial and spectral points. Assuming plane wave propagation and the same sensor array as in Figure 1.2, Figure 1.3a illustrates the directional magnitude response of a MVDR beamformer to diffuse noise. The beamformer uses a single point constraint. We refer to Figure 1.3a as the quiescent response of the beamformer. In Figure 1.3b, the appearance of an interferer causes adaptive weights to modify the beamformer response by raising sidelobe levels and placing a null in the direction of the interferer. As a result of this action, beamformer output power increases because of diffuse noise leaking through the sidelobes. With the appearance of an interferer, the beamformer trades detection performance in diffuse noise for adaptive interference suppression.

As an alternative to point constraints [12], derivative constraints [15, 29] impose constraints on the beamformer response across a continuous spatial region. Eigenvector constraints introduced by Buckley [30] constrain the beamformer response over spatial and temporal regions.

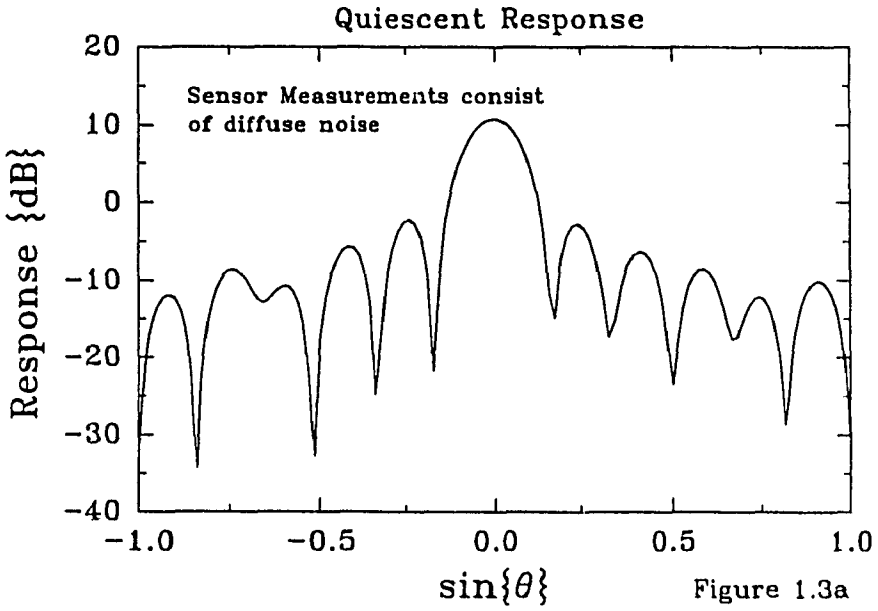


Figure 1.3a

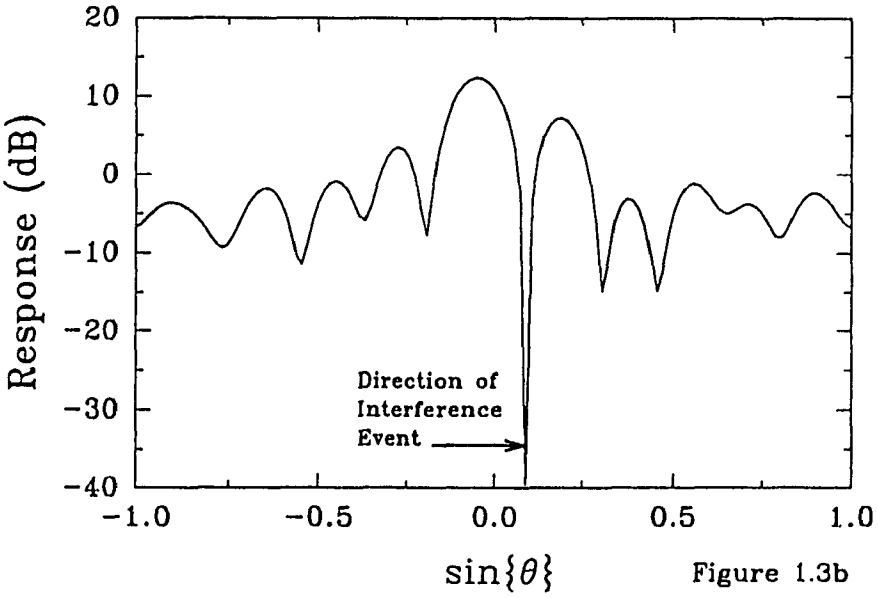


Figure 1.3b

MVDR Beamformer : Magnitude Response

Figure 1.3

The motivation of imposing hard constraints on the response of an adaptive beamformer is to prevent signal distortion by maintaining a response over spatial and temporal regions of interest. These spatial and temporal regions embody the extent of the desired signal as perceived by the beamformer. Unfortunately, hard constraints do not completely preclude distortion of the signal [23]-[26]. For instance, distortion may arise when multi-path propagation conditions cause partial correlation between the signal and interference components of the wavefield. Consequently, cancellation of the interferer results in attenuation of the desired signal [26]. Signal distortion also occurs as a result of a mismatch between the incident signal wavefront and the wavefront model used by the beamformer. Because of this mismatch, components of the measured signal are mistaken for interference [23, 24]. Sensor position uncertainty has a similar effect on the signal [23, 24]. To increase the robustness of an adaptive linearly constrained beamformer against signal distortion, [31] applies an additional set of quadratic inequality constraints of the beamformer weights.

Whereas the resolving ability of a conventional beamformer depends on array geometry and data independent sensor weights, an adaptive beamformer does provide some improvement in resolution performance. The ability of the adaptive beamformer to resolve two closely spaced signals is asymptotically limited by the SNR at the output of the beamformer [22]. To improve upon resolution performance high resolution techniques are applied [16]-[19]. Through an accurate modeling of the wavefield, resolution performance of these techniques depends only on observation time and the accuracy of the wavefield model [22]. Although offering better resolution performance than adaptive beamforming, high resolution methods depend on long observation periods, precluding their application when components of the wavefield are rapidly changing.

1.1 Motivation

In the ocean environment, source dynamics and a changing propagation medium determine the stationarity of wavefield measurements collected by a sensor array. As well, under realistic conditions radiating energy from independent acoustic events that overlap in time will generally differ in temporal duration. Large differences in temporal extent between wavefield components present a problem for minimum variance distortionless response (MVDR) beamformer realizations restricted to computing a fixed number of weights simultaneously on a single time-scale. The difficulty arises when the adaptation time is greater than the time-scale of an interferer.

In responding to an interferer, the MVDR beamformer trades detection performance in diffuse noise for adaptive interference cancellation. When adaptation time exceeds the time-scale of an interferer the tradeoff between diffuse noise and interference cancellation no longer applies once the interference event has vanished from the wavefield. Consequently, the beamformer faces an increase in output power that unnecessarily degrades detection performance in diffuse noise. In the hope of improving performance, one possible approach involves reducing adaptation time. Unfortunately, the increase in weight noise variance accompanying this change in adaptation time may introduce a greater loss in performance than is gained through reducing adaptation time. Reducing the number of adaptive beamformer weights, by either increasing the constraints placed on the beamformer response or applying a partially adaptive design technique, (eg: [32]-[36]), provides one method of reducing adaptation time without incurring performance degradation. However, with fewer adaptive weights the ability of the beamformer to respond to multiple interferers decreases. Because of these limitations, a MVDR beamformer restricted to simultaneously computing a fixed number of adaptive weights on one time-scale is not an ideal candidate for a mixed time-scale environment. This is particularly the case when impulsive noise induced by biological, seismic or explosive phenomena, overlap in time with longer duration wavefield components such as noise-like signals emitted

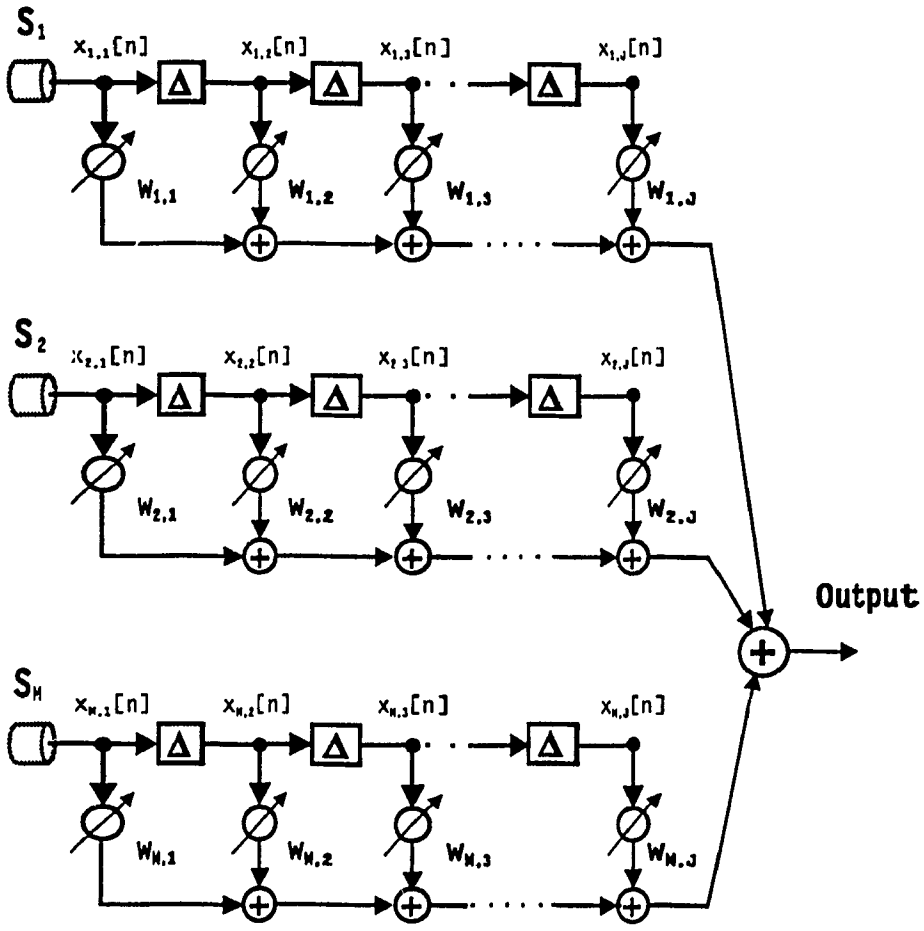
by a distant ship.

Consider now two realizations of the single time-scale MVDR beamformer: the direct implementation [12], Figure 1.4, and the generalized sidelobe canceller (GSC) [37], Figure 1.5. The structure of the GSC follows from a decomposition of the direct form weights $\mathbf{w}$ into data independent and unconstrained adaptive weights, $\mathbf{w_q}$ and $\mathbf{w_a}$. We refer to $\mathbf{w_q}$ as the quiescent weight vector, since the response of the beamformer depends solely on $\mathbf{w_q}$ when sensor measurements consist of diffuse noise. To maintain response constraints, the signal blocking matrix $\overline{\mathbf{C}}$ prevents $\mathbf{w_a}$ from being applied to wavefield components in the spatial and spectral region constrained by the beamformer.

The equivalence between a direct form MVDR realization and GSC depends on the number of adaptive degrees of freedom (DOF) available to each beamformer. We use this measure of beamformer adaptivity, instead of the number of adaptive weights, to eliminate any ambiguity. Unlike adaptive weights, adaptive DOF provide a precise and quantitative definition of beamformer adaptivity that refers directly to the independently adaptable parameters in either $\mathbf{w}$ or $\mathbf{w_a}$.

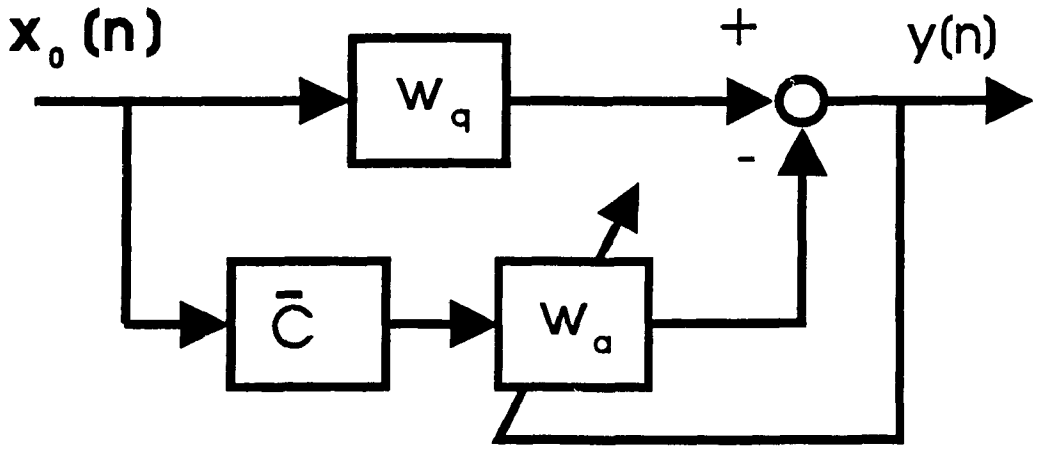
To illustrate the dilemma faced by a GSC in mixed time-scale interference conditions we turn to an example consisting of a horizontal linear array of equi-spaced sensors. Along with diffuse noise, suppose wavefield measurements contain acoustic energy from two independent directional sources located in the far-field of the array. Assume that the temporal duration of energy radiating from one source is very short in comparison with the second event. We apply a GSC to generate a power directionality map of the wavefield at a single frequency. The GSC maintains unity gain response, through a single point constraint, at a look direction that corresponds to a wavefront propagating at an angle perpendicular to the sensor array. At this look direction the two directional sources present in the wavefield appear as interference.

We first examine the case when adaptation time and the time-scale of the



Direction Form Realization of the MVDR Beamformer

Figure 1.4

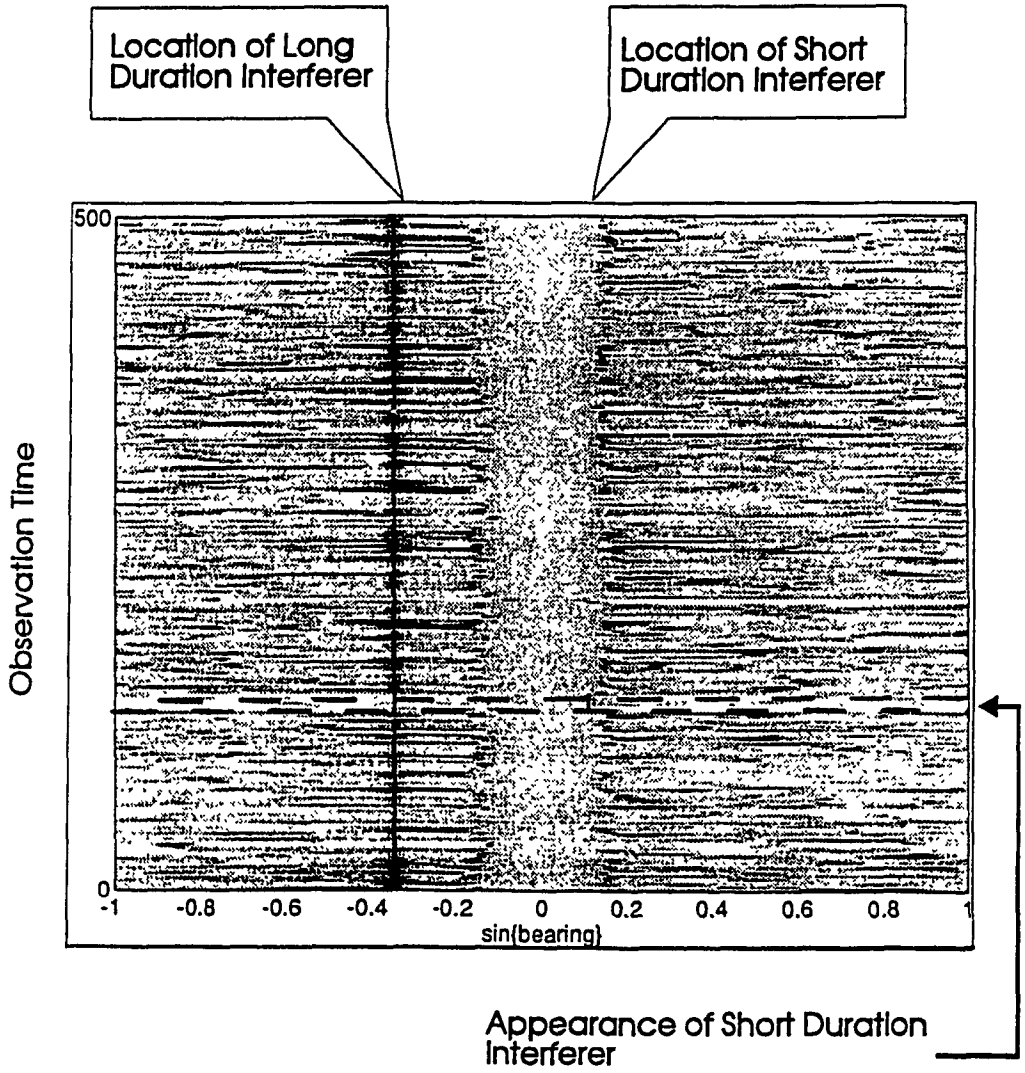


Generalized Sidelobe Canceller

Figure 1.5

shorter duration interference event match. To investigate the effect of adaptation time and adaptive interference suppression on the output power estimate, consider the evolution of the beamformer magnitude response in time. Figure 1.6 shows the evolution of the response. Attenuation in the response corresponds to a grey-level scale, where increased shading represents increasing attenuation. The two dashed horizontal lines on the figure indicate the effective duration of the short duration interferer. The high level of variability suggested by Figure 1.6 indicates that weight noise variance dominates detection performance. Increasing adaptation time, Figure 1.7 shows the beamformer response maintaining a null beyond the temporal extent of the short duration interferer. Although weight noise variance appears to have little impact on the response, sidelobe levels remain elevated well after the short duration interference event has disappear. Over the time interval sidelobe levels remain elevated, increased leakage of diffuse noise into the output power estimate dominates detection performance. As suggested by Figures 1.6 and 1.7, it is difficult for a single time-scale beamformer, adapting a fixed number of DOF, to maximize detection performance in the presence of short duration interference. Since the beamformer cannot reduce adaptation time without incurring an increase in weight noise variance, detection performance suffers because of increased beamformer output power due to diffuse noise. When the occurrence of strong short duration interference is frequent, as is the case in under-ice applications, detection performance of the single time-scale beamformer may be severely affected over long periods of time.

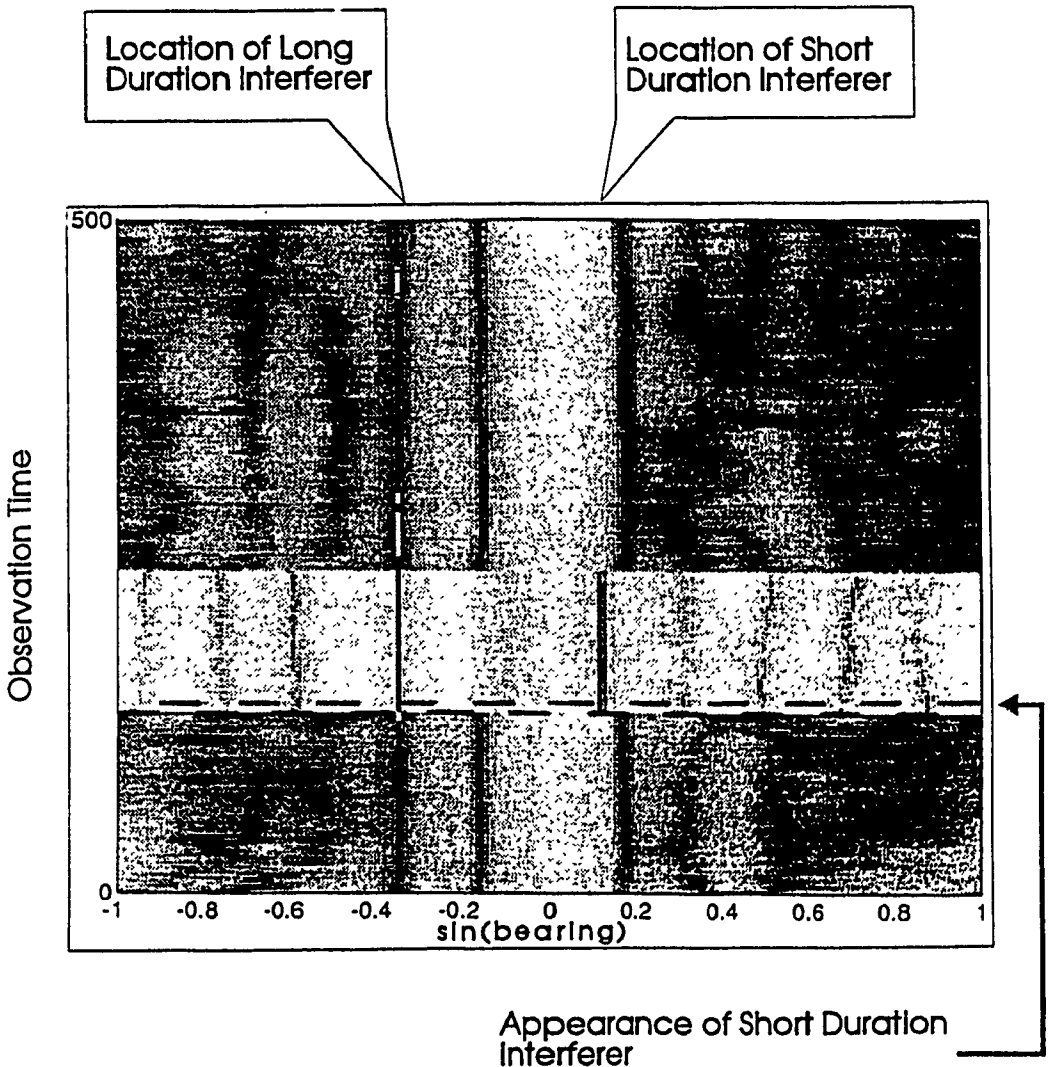
Recently, Liu and Van Veen [38] reported a cascaded version of the linearly constrained MVDR beamformer. A unique feature of this modular MVDR (M^2VDR) beamformer is the flexibility of distributing adaptive DOF among stages arranged in cascade. This property suggests the possibility of adapting beamformer DOF over different time-scales. Motivated by the M^2VDR realization, we propose a multiple time-scale adaptive beamformer that combats interference in mixed time-scale conditions. To cancel short duration interference, the beamformer computes a small



- - - Indicates beginning and end of short duration interferer

MVDR Beamformer Response - Short Adaptation Time

Figure 1.6



- - - Indicates beginning and end of short duration interferer

MVDR Beamformer Response - Long Adaptation Time

Figure 1.7

number of weights assigned to a stage on a time-scale approximating the duration of the interferer. With fewer weights, the beamformer stage introduces the possibility of larger reductions in adaptation time that could be possible for a GSC implementation of the MVDR beamformer. Furthermore, the M^{2}VDR realization does not curtail the ability to respond to multiple interferers since the beamformer applies the remaining DOF, retained in other stages, against long duration interference.

1.2 Contributions and Outline

This thesis addresses the problem of adaptive beamforming in mixed time-scale interference conditions. To maximize detection performance when the temporal duration between wavefield components differs substantially, we propose a multiple time-scale M^{2}VDR beamformer. Previous MVDR realizations, [12] and [37], preclude the simultaneous adaptation of DOF on different time-scales. The M^{2}VDR beamformer is the only MVDR realization that permits multiple time-scale adaptation. This unique feature provides increased flexibility in maximizing detection performance under mixed time-scale interference conditions without sacrificing interference cancellation capability in multiple interferer settings.

The derivation of the M^{2}VDR presented in [38] assumes the exact covariance matrix of the measured wavefield. This leads to a set of stage weights that achieve the minimum MSE at the output of the beamformer. However, under time-varying conditions the ability to compute beamformer stage on different time-scales has not been previously investigated. The principle contribution of this thesis is the development and analysis of a multiple time-scale MVDR beamformer.

The number of adaptive weight coefficients loosely serve as a measure of beamformer adaptivity [39]. By decomposing the GSC into a modular realization, beamformer weights appear in a form that naively suggests a dramatic increase in adaptivity. To correctly analyze this case we introduce a precise interpretation of adaptive

DOF in Section 2.3. We show that a correct assessment of adaptive DOF is the rank of the signal blocking matrix.

Chapter 3 derives closed form expressions relating the adaptation time and number of adaptive DOF assigned each M^{2}VDR beamformer stage, to mean output power, MSE, and output SNR. To simplify this analysis beamformer stage weights are computed by non-overlapping block adaptation [40]. Two types of block adaptation form the basis of our analysis. Type I adaptation uses the same input data block to compute weights and update the output. Type II adaptation computes stage weights based on a past input data block. Van Veen [41] examines in detail the statistical properties of a Type I implementation for the GSC. Krolik and Swingler [42] examine Type I and II GSC implementations. Niezgoda [43] derives the mean output power and MSE expressions for the Type I M^{2}VDR beamformer. Chapter 3 extends the analysis in [41]-[43] to derive the statistical properties of the M^{2}VDR Type I and II implementations. Expressions for mean output power show the necessity of constraining the adaptation time of each stage to ensure that the beamformer output is not ill-defined. For the beamformer output power to be a positive quantity we show that the effective number of statistical DOF in an input data block must exceed the total number of adaptive DOF adapted over the length of the input data block.

The first part of Chapter 4 examines the implementation of Type I adaptation. The important feature of Type I adaptation is the correlation between the beamformer weights and the current input data block. This allows the beamformer to respond to a short duration interferer over the same time interval interference samples appear at the input to the beamformer. In applying this adaptation scheme to a multiple stage M^{2}VDR beamformer the ordering of stages becomes important. Analysis indicates that the minimum mean-squares error (MMSE) at the final beamformer output only occurs when the adaptation time assigned each stage is greater than or equal to preceding stages. The second half of Chapter 4 investigates the effect signal blocking matrix design has on detection performance. Each stage of a

multiple time-scale M^{2}VDR beamformer operates as a partial adaptive beamformer. Partial adaptivity implies that only a subset of available DOF are used by the stage. This has the same effect as increasing the number of constraints imposed on the beamformer response. Spatially, partial adaptivity restricts the adaptive response of the beamformer through the choice of signal blocking matrix. In a short duration interference environment it is important to prevent interference components from leaking into stages adapting on a longer time-scale than the interferer. Otherwise residual components of the interference retained in these stages may cause a null in the magnitude response to be maintained after the short duration interferer has vanished. Any accompanying reduction in sidelobe level attenuation leads to an increase in output power and a loss in detection performance. Chapter 4 derives an approximate expression for array gain degradation due to signal blocking mismatch.

Chapter 5 investigates the maximization of detection performance during short duration interference events for a two-stage beamformer. To maximize signal detection when wavefield measurements include short duration interference events, the tradeoff between adaptation time and weight noise variance is examined. Assuming a Gaussian approximation of the output power statistic, a closed form expression determines the optimal adaptation time; otherwise optimization requires numerically evaluating receiver operating curves over different adaptation times.

Chapter 6 continues the investigation of M^{2}VDR beamformer performance in an impulsive interference environment by simulation experiments. The effects of design mismatch are re-examined. An alternative to Type I block adaptation is a modified version of recursive least squares (RLS). Unlike block adaptation, RLS applies an exponentially weighted window to input data. Although the effective memory of RLS and block adaptation can be equated, the response of RLS to an impulsive interference is prolonged by exponential weighting. On the positive side, RLS results in smaller levels of weight noise variance than block adaptation. As well Chapter 6 provides a qualitative comparison of detection and resolution performance for a GSC

and a two stage M^{2}VDR beamformer in an impulsive interference environment.

Chapter 2

The Multiple Stage Linearly Constrained Adaptive Beamformer

This chapter reviews the relationship between the direct form, GSC, and M^{2}VDR realizations of the linearly constrained MVDR beamformer. Because of the M^{2}VDR decomposition, adaptive beamformer weights appear in a form suggesting an increase in the number of adaptive DOF. To analyze this case correctly we introduce a precise interpretation of adaptive DOF. Notationally, boldfaced lower-case and upper-case symbols denote vectors and matrices, respectively.

2.1 Background

To simplify our discussion, issues arising from propagation effects are de-coupled from the problem of interference suppression by modeling the ocean as a non-dispersive, homogeneous medium. Figure 1.1 provides an example of plane wave propagation in this environment. Information pertaining to the location of an acoustic event is retained in the phase coherence of points along the wavefront radiating

from the source. Suppose that an omni-directional sensor is located at

$$\mathbf{s} = (x, y, z)^{\mathsf{T}}$$

from an arbitrary reference point to observe the wavefield at time t and frequency f, where superscript $^{\mathsf{T}}$ is transpose. At the sensor, sound pressure from a wavefront induces a voltage represented by the complex exponential

$$x(t, f) = \alpha(t, f)\, e^{j(2\pi f t - \mathbf{k}^{\mathsf{T}}(f)\mathbf{s})} \tag{2.1}$$

In this model, the wavenumber vector described by

$$\mathbf{k}(f) = 2\pi \frac{f}{c} \boldsymbol{\xi} \tag{2.2}$$

determines source location, where $\boldsymbol{\xi} = (\xi_x, \xi_y, \xi_z)^{\mathsf{T}}$ is a unit vector aligned in the direction of propagation, and c is propagation velocity. Sensor positions where the product $\mathbf{k}^{\mathsf{T}}(f)\mathbf{s}$ is constant, correspond to points of phase coherence. Completing the description of (2.1), we model the source signal $\alpha(t, f)$ as a zero mean complex random variable on the assumption that the amplitude and phase of $\alpha(t, f)$ are unknown.

We now turn briefly to the relationship between the conventional and MVDR narrow-band beamformers. Suppose that the wavefield consists of two statistically independent directional sources and a noise component. A linear array of M equi-spaced sensors measures the wavefield. The output of the $m^{\mathbf{th}}$ sensor, located at

$$\mathbf{s}_m = (x_m, y_m, z_m)^{\mathsf{T}}$$

is described by

$$x_m(t, f) = \alpha_a(t, f)\, e^{j(2\pi f t + \mathbf{k}_a(f)^{\mathsf{T}}\mathbf{s}_m)} + \alpha_b(t, f)\, e^{j(2\pi f t + \mathbf{k}_b(f)^{\mathsf{T}}\mathbf{s}_m)} + \eta_m(t, f) \tag{2.3}$$

where $\mathcal{E}\{\alpha_a(t, f)\alpha_b(t, f)^*\} = 0$, and superscript * is complex conjugate. Statistical expectation is denoted by $\mathcal{E}\{\cdot\}$. Diffuse noise, represented by $\eta_m(t, f)$, is modeled

as a zero mean spatially and temporally white process, independent of $\alpha_a(t,f)$ and $\alpha_b(t,f)$.

Let $\mathbf{d}(f)$ represent a wavenumber vector of a hypothesized wavefront. For a conventional narrow-band beamformer, that applies uniform weighting to each sensor, output power equals

$$\mathsf{P}_{out}(f) = \frac{1}{M}\mathcal{E}\left\{\left|\sum_{m=1}^{M} x_m(t,f)\, e^{j\mathbf{d}^{\mathsf{T}}(f)\mathbf{s}_m}\right|^2\right\} \tag{2.4}$$

In the process of forming a power directionality map of the wavefield, suppose $\mathbf{d}(f) = \mathbf{k}_a(f)$. Hence, components of sensor measurements associated with $\mathbf{k}_a(f)$ sum coherently in (2.4). The ability of the beamformer to detect a source at the wavenumber location $\mathbf{k}_a(f)$ is affected by the second directional component of the wavefield at $\mathbf{k}_b(f)$ and diffuse noise. Rewriting (2.4) as

$$\begin{aligned}\mathsf{P}_{\mathbf{out}}(f) &= \mathcal{E}\left\{|\alpha_a(t,f)|^2\right\}\frac{1}{M}\left|\sum_{m=1}^{M} e^{j\left(2\pi ft-\mathbf{k}_a^{\mathsf{T}}(f)\mathbf{s}_m+\mathbf{k}_a^{\mathsf{T}}(f)\mathbf{s}_m\right)}\right|^2 \\ &\quad +\mathcal{E}\left\{|\alpha_b(t,f)|^2\right\}\frac{1}{M}\left|\sum_{m=1}^{M} e^{j\left(2\pi ft-\mathbf{k}_b^{\mathsf{T}}(f)\mathbf{s}_m+\mathbf{k}_a^{\mathsf{T}}(f)\mathbf{s}_m\right)}\right|^2 \\ &\quad +\frac{1}{M}\sum_{m=1}^{M}\mathcal{E}\left\{|\eta_m(t,f)|^2\right\} \\ &= M\mathsf{P}_a(f)+M\zeta\mathsf{P}_b(f)+\mathsf{P}_\eta(f)\end{aligned} \tag{2.5}$$

where

$$\begin{aligned}\mathsf{P}_a(f) &= \mathcal{E}\left\{|\alpha_a(t,f)|^2\right\} \\ \mathsf{P}_b(f) &= \mathcal{E}\left\{|\alpha_b(t,f)|^2\right\} \\ \mathsf{P}_\eta(f) &= \mathcal{E}\left\{|\eta_m(t,f)|^2\right\}\end{aligned}$$

The factor $0 \le \zeta \le 1$ accounts for the attenuation in the beamformer response in the direction corresponding to $\mathbf{k}_b(f)$. Note that attenuation of the response decreases as $\mathbf{k}_b(f)$ approaches $\mathbf{k}_a(f)$. SNR at the output of the beamformer is

$$\mathsf{SNR}_{\mathbf{out}}(f) = \frac{M\mathsf{P}_a(f)}{M\zeta\mathsf{P}_b(f)+\mathsf{P}_\eta(f)} \tag{2.6}$$

With SNR at the input to a sensor given by

$$\mathtt{SNR}_{\mathtt{in}}(f) \;=\; \frac{\mathtt{P}_a(f)}{\mathtt{P}_b(f)+\mathtt{P}_\eta(f)}$$

the relative improvement, or array gain (AG) of the conventional beamformer over a single sensor is

$$\mathtt{AG}_{\mathtt{CB}}(f) \;=\; \frac{\mathtt{SNR}_{\mathtt{out}}(f)}{\mathtt{SNR}_{\mathtt{in}}(f)} \;=\; M\left(\frac{1+\mathtt{INR}(f)}{1+M\zeta\mathtt{INR}(f)}\right) \tag{2.7}$$

where

$$\mathtt{INR}(f) \;=\; \frac{\mathtt{P}_b(f)}{\mathtt{P}_\eta(f)}$$

equals interference-to-noise ratio (INR). As indicated by (2.7), $\mathtt{AG}_{\mathtt{CB}}$ decreases as the two directional sources move closer together and INR increases. However, for a narrowband MVDR beamformer, the relative improvement in AG over the conventional beamformer is

$$\frac{\mathtt{AG}_{\mathtt{MVDR}}(f)}{\mathtt{AG}_{\mathtt{CB}}(f)} \;=\; 1+\frac{M^2\mathtt{INR}^2(f)}{1+M\mathtt{INR}(f)}\zeta(1-\zeta) \tag{2.8}$$

[15]. Observe that the stronger the interferer, the larger the relative gain of the MVDR beamformer.

2.2 Review of the M²VDR Beamformer

The first part of this discussion describes the direct form realization of the linearly constrained adaptive wide-band beamformer as shown in Figure 1.4. To provide an adaptive response over a region of frequencies a tapped-delay line of length J follows each sensor output. In contrast, a narrow-band beamformer only requires a single weight per sensor output. For a M sensor wide-band array let

$$\mathbf{x}[n] \;=\; \left(\,\mathbf{x}_1^{\mathsf{T}}[n]\;\mathbf{x}_2^{\mathsf{T}}[n]\;\ldots\;\mathbf{x}_J^{\mathsf{T}}[n]\,\right)^{\mathsf{T}} \tag{2.9}$$

describe the $MJ\times 1$ vector of sensor measurements samples held by the beamformer tapped-delay line structure at sampling instant n, where

$$\mathbf{x}_\prime[n]$$

is the $M \times 1$ vector of measurements contained in the ℓ^{th} tap of each sensor channel. We refer to $\mathbf{x}[n]$ as one snapshot of data. To complete the description, let

$$\mathbf{w} = \left(\mathbf{w}_1^{\mathsf{T}} \; \mathbf{w}_2^{\mathsf{T}} \; \ldots \; \mathbf{w}_J^{\mathsf{T}} \right)^{\mathsf{T}} \tag{2.10}$$

be the $MJ \times 1$ adaptive weight vector of the beamformer, where the $M \times 1$ vector $\mathbf{w}_\ell$ contains weights applied at the ℓ^{th} tap of each sensor channel.

To optimize $\mathbf{w}$, the beamformer minimizes output power

$$\mathsf{P}_{\text{out}} = \mathcal{E}\left\{ \left| \mathbf{w}^\dagger \mathbf{x}[n] \right|^2 \right\} \tag{2.11}$$

subject to a set of linear constraints placed on the spatial and temporal response of the beamformer. Superscript † denotes complex conjugate transpose. As previously mentioned, linear constraints ensure that the desired signal passes through to the beamformer output with unity gain. We express this constrained minimization problem by

$$\min_{\mathbf{w}} \left(\mathbf{w}^\dagger \mathbf{R} \mathbf{w} \right) \quad \text{subject to} \quad \mathbf{C}^\dagger \mathbf{w} = \mathbf{f} \tag{2.12}$$

where

$$\mathbf{R} = \mathcal{E}\left\{ \mathbf{x}[n] \mathbf{x}^\dagger[n] \right\}$$

equals the covariance matrix of $\mathbf{x}[n]$. To maintain constraints on beamformer weights the solution for $\mathbf{w}$ must satisfy

$$\mathbf{C}^\dagger \mathbf{w} = \mathbf{f}$$

where $\mathbf{C}$ is the $MJ \times K$ constraint matrix, and $\mathbf{f}$ the $K \times 1$ response vector $\mathbf{f}$. K equals the number of linear constraints imposed on the beamformer response. The solution to (2.12) equals [12]

$$\mathbf{w} = \mathbf{R}^{-1} \mathbf{C} \left(\mathbf{C}^\dagger \mathbf{R}^{-1} \mathbf{C} \right)^{-1} \mathbf{f} \tag{2.13}$$

The generalized sidelobe canceller, shown in Figure 1.5, provides an alternative realization of the direct form MVDR beamformer. This implementation of the MVDR beamformer decomposes the weight vector $\mathbf{w}$ into two terms: $\mathbf{w}_{\mathsf{q}}$ the $MJ \times 1$

quiescent weight vector, and $\mathbf{w_a}$ the $(MJ-K)\times 1$ adaptive weight vector. To relate the direct form beamformer and the GSC, we express $\mathbf{w}$ in terms of $\mathbf{w_q}$ and $\mathbf{w_a}$ as

$$\mathbf{w} = \mathbf{w_q} - \overline{\mathbf{C}}\mathbf{w_a} \tag{2.14}$$

[37], where

$$\mathbf{w_q} = \mathbf{C}\left(\mathbf{C}^{\dagger}\mathbf{C}\right)^{-1}\mathbf{f}$$

The adaptive weight vector $\mathbf{w_a}$ given by

$$\mathbf{w_a} = \left(\overline{\mathbf{C}}^{\dagger}\mathbf{R}\overline{\mathbf{C}}\right)^{-1}\overline{\mathbf{C}}^{\dagger}\mathbf{R}\mathbf{w_q} \tag{2.15}$$

follows from the solution to the unconstrained minimization, [37], given by

$$\min_{\mathbf{w_a}} \left(\mathbf{w_q} - \overline{\mathbf{C}}\mathbf{w_a}\right)^{\dagger}\mathbf{R}\left(\mathbf{w_q} - \overline{\mathbf{C}}\mathbf{w_a}\right) \tag{2.16}$$

Determined by the the orthogonal complement to $\mathbf{C}$, $\overline{\mathbf{C}}$ defines the $MJ\times Q$, Q $MJ-K$, signal blocking matrix. Since

$$\mathbf{C}^{\dagger}\overline{\mathbf{C}} = \mathbf{0}_{K,Q}$$

the desired intention of the blocking matrix is to prevent adaptive cancellation of the signal. We assume throughout this thesis that the columns of $\overline{\mathbf{C}}$ are independent, ie:

$$\mathtt{rank}\left(\overline{\mathbf{C}}\right) = MJ - K$$

In this instance the beamformer is said to be fully adaptive. Otherwise if

$$\mathtt{rank}\left(\overline{\mathbf{C}}\right) < MJ - K$$

the beamformer is said to be partially adaptive.

To develop the P stage realization of the MVDR beamformer, shown in Figure 2.1, we first express (2.16) in the form

$$\min_{\mathbf{B}} \left(\mathbf{w_q^{\dagger}}\left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}\mathbf{B}\right)^{\dagger}\mathbf{R}\left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}\mathbf{B}\right)\mathbf{w}_q\right) \tag{2.17}$$

where

$$\mathbf{B} = \left(\overline{\mathbf{C}}^{\dagger}\mathbf{R}\overline{\mathbf{C}}\right)^{-1}\overline{\mathbf{C}}^{\dagger}\mathbf{R}$$

equals the solution to (2.17). The relationship between the $(MJ-K)\times MJ$ matrix $\mathbf{B}$ and $\mathbf{w_a}$ is

$$\mathbf{w_a} = \mathbf{B}\mathbf{w_q}$$

As shown in Appendix A.1, central to the development of the multiple stage beamformer is the equivalence between (2.17) and the minimization

$$\min_{\mathbf{B}} \mathtt{tr}\left(\left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}\mathbf{B}\right)^{\dagger}\mathbf{R}\left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}\mathbf{B}\right)\right) \tag{2.18}$$

By equating (2.17) to (2.18) [38] establishes a relationship between the GSC and a cascade of P adaptive beamformer stages through the expression

$$\begin{aligned}\left(\mathbf{w_q} - \overline{\mathbf{C}}\mathbf{w_a}\right) &= \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}\mathbf{B}\right)\mathbf{w}_q \\ &= \prod_{p=1}^{P}\left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_{\ell}\mathbf{B}_{\ell}\right)\mathbf{w}_q\end{aligned} \tag{2.19}$$

For the $\ell^{\mathtt{th}}$ stage of the cascade, described by the expression

$$\mathbf{I}_{MJ} - \overline{\mathbf{C}}_{\ell}\mathbf{B}_{\ell}$$

$\mathbf{B}_{\ell}$ represents the $Q_{\ell}\times MJ$ adaptive weight matrix, and $\overline{\mathbf{C}}_{\ell}$ denotes the $MJ\times Q_{\ell}$ signal blocking matrix formed by column partitioning $\overline{\mathbf{C}}$ as

$$\overline{\mathbf{C}} = \left(\overline{\mathbf{C}}_1\overline{\mathbf{C}}_2\ldots\overline{\mathbf{C}}_{\ell}\ldots\overline{\mathbf{C}}_P\right)$$

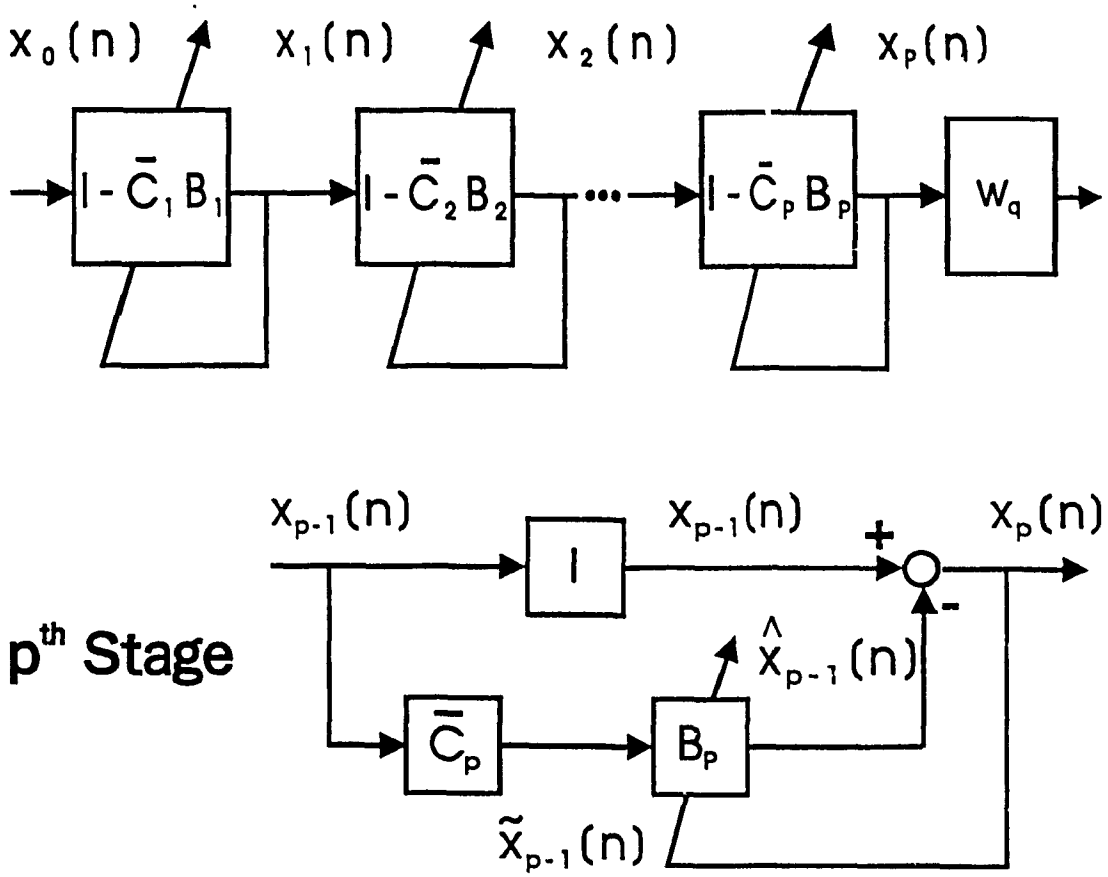
where

$$\sum_{\ell=1}^{P} Q_{\ell} = Q = MJ - K$$

Observe that since

$$\mathtt{rank}\left(\overline{\mathbf{C}}_{\ell}\right) < MJ - K \qquad \text{for } P \geq 2$$

each stage of the M^{2}VDR beamformer is partially adaptive.



M^2VDR Beamformer

Figure 2.1

Letting $\mathbf{x}_p[n]$ denote the $MJ\times 1$ output vector of the p^{th} stage,

$$\mathbf{x}_p[n] = \prod_{\ell=1}^{p} \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_\ell \mathbf{B}_\ell\right)^\dagger \mathbf{x}_0[n] \tag{2.20}$$

where $\mathbf{x}_0[n] = \mathbf{x}[n]$ is the sensor output vector. The adaptive weight matrix $\mathbf{B}_\ell$ follows from the solution to

$$\min_{\mathbf{B}_\ell} \mathtt{tr}\left(\left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_\ell \mathbf{B}_\ell\right)^\dagger \mathbf{R}_{\ell-1} \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_\ell \mathbf{B}_\ell\right)\right) \tag{2.21}$$

where

$$\mathtt{P}_\ell \;=\; \mathbf{w}_{\mathbf{q}}^\dagger \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_\ell \mathbf{B}_\ell\right)^\dagger \mathbf{R}_{\ell-1} \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_\ell \mathbf{B}_\ell\right) \mathbf{w}_{\mathbf{q}}$$

equals the output power of the ℓ^{th} stage and

$$\mathbf{B}_\ell = \left(\overline{\mathbf{C}}_\ell^\dagger \mathbf{R}_{\ell-1} \overline{\mathbf{C}}_\ell\right)^{-1} \overline{\mathbf{C}}_\ell^\dagger \mathbf{R}_{\ell-1}$$

the ℓ^{th} stage weights that minimize $\mathtt{P}_\ell$. Assuming that $\mathcal{E}\left\{\mathbf{x}_{\ell-1}[n]\right\} = \mathbf{0}_{MJ}$, for $\ell = 1, 2, \ldots P+1$, the covariance matrix for $\mathbf{x}_{\ell-1}[n]$ equals

$$\mathbf{R}_{\ell-1} \;=\; \mathcal{E}\left\{\, \mathbf{x}_{\ell-1}[n]\mathbf{x}_{\ell-1}^\dagger[n] \,\right\}$$

2.3 Adaptive Degrees of Freedom

Adaptive degrees of freedom have in the past been typically associated with the number of unconstrained beamformer weights [39]. For the GSC there is an explicit correspondence between $MJ - K$ adaptive DOF and the $(MJ-K)\times 1$ weight vector $\mathbf{w}_{\mathbf{a}}$. However, such a relationship does not exist for the M^{2}VDR beamformer. Notice that the dimensions of $\mathbf{B}_\ell$ exceeds the number of adaptive weight coefficients available to a GSC assigned the blocking matrix $\overline{\mathbf{C}}_\ell$. By relating DOF to the elements defining $\mathbf{w}_{\mathbf{a}}$ or $\mathbf{B}_\ell$, we are lead to the erroneous conclusion that a M^{2}VDR decomposition of the GSC increases the number of adaptive DOF. This section develops a concise interpretation of adaptive DOF by equating the number of DOF with the rank of the signal blocking matrix.

We begin by first considering an example consisting of a 2 sensor, single tap array, where $\mathbf{x}[n]$ is the 2×1 sensor measurement vector. To simplify our discussion we assume that all quantities are real-valued. For the 2 sensor case the GSC controls the response of the array by a single linear constraint and one adaptive DOF. To demonstrate that the number of adaptive DOF equals the rank of the signal blocking matrix $\overline{\mathbf{C}}$, we apply a geometric interpretation to the GSC that describes a random variable as a vector quantity [44]. We begin by examining the scalar weight vector $\mathbf{w_a}$. Suppose

$$y[n] = \mathbf{w_q^T x}[n]$$

and

$$\check{x}[n] = \overline{\mathbf{C}}^{\mathbf{T}}\mathbf{x}[n]$$

are zero mean scalar random variables with standard deviation σ and $\check{\sigma}$, respectively. Then the solution to

$$\min_{w_\mathbf{a}} \mathcal{E}\left\{ |\ y[n] - w_\mathbf{a}\check{x}[n]|^2 \right\} \tag{2.22}$$

in terms of σ and $\check{\sigma}$ is

$$w_\mathbf{a} = \frac{\mathcal{E}\left\{y[n]\check{x}[n]\right\}}{\mathcal{E}\left\{\check{x}[n]^2\right\}} = \rho\frac{\sigma}{\check{\sigma}} \tag{2.23}$$

where the correlation coefficient between $y[n]$ and $\check{x}[n]$ is $0 \leq \rho \leq 1$. From (2.23) the output of the GSC takes the form

$$\begin{aligned} e[n] &= y[n] - \rho\frac{\sigma}{\check{\sigma}}\check{x}[n] \\ &= y[n] - \hat{y}[n] \end{aligned} \tag{2.24}$$

where

$$\hat{y}[n] = \rho\frac{\sigma}{\check{\sigma}}\check{x}[n]$$

$$\mathcal{E}\left\{e[n]\hat{y}[n]\right\} = 0$$

$$\mathcal{E}\left\{e^2[n]\right\} = \sigma^2\left(1-\rho^2\right)$$

$$\mathcal{E}\left\{e[n]y[n]\right\} = \sigma^2\left(1-\rho^2\right)$$

Next, we identify the random variables $y[n]$, $\check{x}[n]$, and $e[n]$ with the 2×1 real vectors $\mathbf{v}_y$, $\mathbf{v}_{\check{x}}$, and $\mathbf{v}_e$, respectively. Assigning to each of these vectors a length equal to the standard deviation of the random variable identified with the vector, set

$$\mathbf{v}_y^{\mathsf{T}}\mathbf{v}_y = \sigma^2$$

$$\mathbf{v}_{\check{x}}^{\mathsf{T}}\mathbf{v}_{\check{x}} = \check{\sigma}^2$$

$$\mathbf{v}_y^{\mathsf{T}}\mathbf{v}_{\check{x}} = \sigma\check{\sigma}\rho$$

As shown in Figure 2.2, a geometric interpretation of the GSC follows from $\mathbf{v}_y$ and $\mathbf{v}_{\check{x}}$, where $\mathbf{span}(\mathbf{v}_{\check{x}})$ denotes the linear span of $\mathbf{v}_{\check{x}}$. To complete the description of the GSC, we project $\mathbf{v}_y$ onto $\mathbf{v}_{\check{x}}$ to form

$$\mathbf{v}_{\hat{y}} = \rho\frac{\sigma}{\check{\sigma}}\mathbf{v}_{\check{x}} \tag{2.25}$$

the optimal linear estimate of $\mathbf{v}_y$, where

$$\mathbf{v}_e = \mathbf{v}_y - \mathbf{v}_{\hat{y}} \tag{2.26}$$

is the error vector between $\mathbf{v}_y$ and $\mathbf{v}_{\hat{y}}$. In this discription of the GSC, the inner-products

$$\mathbf{v}_e^{\mathsf{T}}\mathbf{v}_e = \sigma^2\left(1 - \rho^2\right)$$

$$\mathbf{v}_e^{\mathsf{T}}\mathbf{v}_{\hat{y}} = 0$$

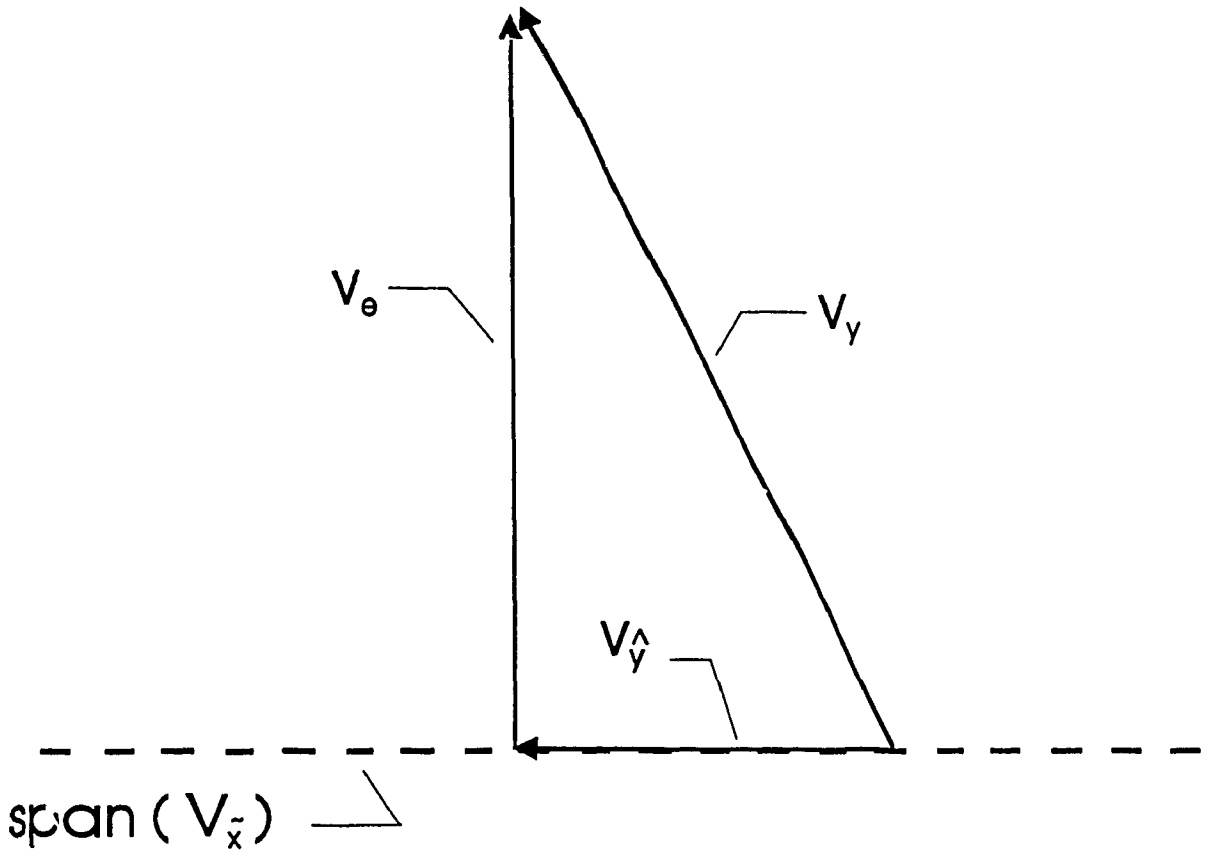
and

$$\mathbf{v}_e^{\mathsf{T}}\mathbf{v}_y = \sigma^2\left(1 - \rho^2\right)$$

correspond to the correlation properties between $e[n]$, $y[n]$, and $\hat{y}[n]$. Returning to (2.25), we see that a single parameter or degree of freedom governs the adaptation of $w_{\mathbf{a}}$ on a one dimensional sub-space determined by the span of $\overline{\mathbf{C}}^{\mathsf{T}}\mathbf{x}[n]$.

Consider now a single stage M²VDR decomposition of the GSC. Expressing $\mathbf{x}[n]$ explicitly as

$$\mathbf{x}[n] = \left(x_1[n],\, x_2[n] \right)^{\mathsf{T}}$$



Geometric Interpretation of the GSC

Figure 2.2

we model the sensor outputs $x_1[n]$ and $x_2[n]$ by two zero mean random variables with standard deviation σ_1 and σ_2, respectively. The correlation coefficient between $x_1[n]$ and $x_2[n]$ is

$$0 \leq \bar{\rho} \leq 1$$

Furthermore, let

$$\check{x}[n] = \overline{\mathbf{C}}^{\mathrm{T}}\mathbf{x}[n]$$

$$\mathbf{R} = \left\{ \mathbf{x}[n]\mathbf{x}^{\mathrm{T}}[n] \right\} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\bar{\rho} \\ \sigma_2\sigma_1\bar{\rho} & \sigma_2^2 \end{pmatrix}$$

Turning to the solution of

$$\min_{\mathbf{B}} \mathcal{E}\left\{ \left| \mathbf{x}[n] - \mathbf{B}^{\mathrm{T}}\check{x}[n] \right|^2 \right\} \tag{2.27}$$

we see that the weight vector

$$\mathbf{B} = (\, b_1,\, b_2\,) = \left(\overline{\mathbf{C}}^{\mathrm{T}}\mathbf{R}\overline{\mathbf{C}}\right)^{-1}\overline{\mathbf{C}}^{\mathrm{T}}\mathbf{R} \tag{2.28}$$

consists of two terms.

Letting $\check{\sigma}^2$ be the variance of $\check{x}[n]$, we express b_1 and b_2 as

$$b_1 = \check{\rho}_1\frac{\sigma_1}{\check{\sigma}} \tag{2.29}$$

and

$$b_2 = \check{\rho}_2\frac{\sigma_2}{\check{\sigma}} \tag{2.30}$$

where $\check{\rho}_1$ and $\check{\rho}_2$ are correlation coefficients determined from

$$\mathcal{E}\left\{x_1[n]\check{x}[n]\right\} = \sigma_1\check{\sigma}\check{\rho}_1$$

$$\mathcal{E}\left\{x_2[n]\check{x}[n]\right\} = \sigma_2\check{\sigma}\check{\rho}_2$$

The geometric description of the beamformer shown by Figure 2.3 follows by assigning a 2×1 real vector to each of the random variables $x_1[n]$, $x_2[n]$, and $\check{x}[n]$, so that

$$\mathbf{v}_{x_1}^{\mathrm{T}}\mathbf{v}_{x_1} = \sigma_1^2$$

$$\mathbf{v}_{x_2}^{\mathsf{T}}\mathbf{v}_{x_2} = \sigma_2^2$$

$$\mathbf{v}_{x_1}^{\mathsf{T}}\mathbf{v}_{x_2} = \sigma_1\sigma_2\bar{\rho}$$

$$\mathbf{v}_{x_1}^{\mathsf{T}}\mathbf{v}_{\hat{x}} = \sigma_1\hat{\sigma}\hat{\rho}_1$$

and

$$\mathbf{v}_{x_2}^{\mathsf{T}}\mathbf{v}_{\hat{x}} = \sigma_2\hat{\sigma}\hat{\rho}_2$$

From Figure 2.3 the optimal linear estimates of $\mathbf{v}_{x_1}$ and $\mathbf{v}_{x_2}$, given $\mathbf{v}_{\hat{x}}$ equal

$$\mathbf{v}_{\hat{x}_\ell} = \hat{\rho}_\ell\frac{\sigma_1}{\hat{\sigma}}\mathbf{v}_{\hat{x}} = b_\ell\mathbf{v}_{\hat{x}} \tag{2.31}$$

for $\ell = 1, 2$, where

$$\mathbf{v}_{e_1} = \mathbf{v}_{x_1} - \mathbf{v}_{\hat{x}_1}$$

$$\mathbf{v}_{e_2} = \mathbf{v}_{x_2} - \mathbf{v}_{\hat{x}_2}$$

represent the output of the beamformer. Since $\mathbf{v}_{x_1}$ and $\mathbf{v}_{x_2}$ both lie entirely on the sub-space spanned by $\mathbf{v}_{\hat{x}}$, the vectors are linearly dependent. Relating $\mathbf{v}_{x_1}$ and $\mathbf{v}_{x_2}$ by

$$\mathbf{v}_{x_2} = \frac{b_2}{b_1}\mathbf{v}_{x_1} \tag{2.32}$$

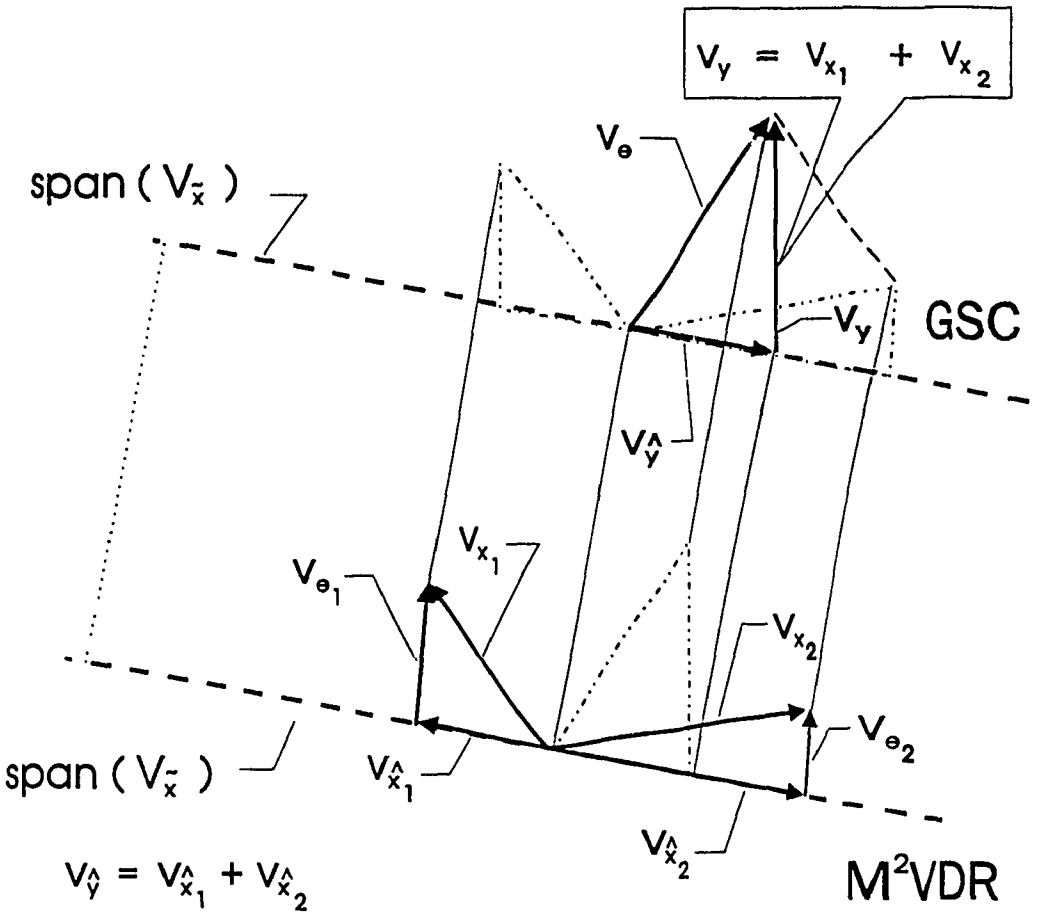
it is apparent that there can only be one independent parameter or degree of freedom governing adaptation. As with the GSC, the single DOF is restricted to a subspace determined by $\mathbf{v}_{\hat{x}}$. Observe further that the GSC first adds $\mathbf{v}_{x_1}$ to $\mathbf{v}_{x_2}$, then projects the sum onto a sub-space of $\mathbf{v}_{\hat{x}}$. The M^{2}VDR beamformer reverses the operation by projecting $\mathbf{v}_{x_1}$ and $\mathbf{v}_{x_2}$ onto a sub-space of $\mathbf{v}_{\hat{x}}$ to form $\mathbf{v}_{\hat{x}_1}$ and $\mathbf{v}_{\hat{x}_2}$, then summing so that

$$\mathbf{v}_{\hat{y}} = \mathbf{v}_{\hat{x}_1} + \mathbf{v}_{\hat{x}_2}$$

Since the summation and projection operations are distributive the GSC and M^{2}VDR beamformer are equivalent.

We now proceed to the case of a M sensor, J tap M^{2}VDR beamformer consisting of P adaptive stages. For the ℓ^{th} stage, the solution to

$$\min_{\mathbf{B}_\ell}\mathcal{E}\left\{\left|\mathbf{x}_{\ell-1}[n] - \mathbf{B}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell[n]\right|^2\right\} \tag{2.33}$$



Geometric Interpretation of the M^2VDR Beamformer

Figure 2.3

takes the form of the $Q_\ell \times MJ$ matrix

$$\mathbf{B}_\ell = \left(\overline{\mathbf{C}}_\ell^{\mathrm{T}} \mathbf{R}_{\ell-1} \overline{\mathbf{C}}_\ell\right)^{-1} \overline{\mathbf{C}}_\ell^{\mathrm{T}} \mathbf{R}_{\ell-1} \tag{2.34}$$

where

$$\tilde{\mathbf{x}}_\ell \;=\; \overline{\mathbf{C}}_\ell^{\mathrm{T}} \mathbf{x}_{\ell-1}$$

and

$$\mathbf{R}_{\ell-1} \;=\; \mathcal{E}\left\{\mathbf{x}_{\ell-1}[n]\mathbf{x}_{\ell-1}^{\mathrm{T}}[n]\right\}$$

By initial inspection the dimension of $\mathbf{B}_\ell$ suggests that the ℓ^{th} stage has $Q_\ell MJ$ adaptive DOF instead of Q_ℓ. To properly analyze $\mathbf{B}_\ell$ we apply an argument similar to that used in the preceding discussion. Let the 2×1 real vector $\mathbf{v}_{x_{\ell-1,j}}$ be associated with the random variable

$$(\mathbf{x}_{\ell-1}[n])_j \;=\; x_{\ell-1,j}[n]$$

such that

$$\mathbf{v}_{x_{\ell-1,j}}^{\mathrm{T}} \mathbf{v}_{x_{\ell-1,k}} \;=\; \begin{cases} \sigma_j \sigma_k \rho_{jk} \;=\; \mathcal{E}\left\{x_{\ell-1,j}[n] x_{\ell-1,k}[n]\right\} & \text{if } j \neq k \\ \sigma_j^2 \;=\; \mathcal{E}\left\{x_{\ell-1,j}^2[n]\right\} & \text{if } j = k \end{cases} \tag{2.35}$$

for $j,k \;=\; 1,2,\ldots,MJ$. Likewise, let $\mathbf{v}_{\tilde{x}_{\ell,m}}$ be identified with

$$(\tilde{\mathbf{x}}_\ell)_m \;=\; \tilde{x}_{\ell,m}$$

where

$$\mathbf{v}_{\tilde{x}_{\ell,m}}^{\mathrm{T}} \mathbf{v}_{\tilde{x}_{\ell,j}} \;=\; \begin{cases} \tilde{\sigma}_m \tilde{\sigma}_j \tilde{\rho}_{mj} \;=\; \mathcal{E}\left\{\tilde{x}_{\ell,m}[n] \tilde{x}_{\ell,j}[n]\right\} & \text{if } m \neq j \\ \tilde{\sigma}_m^2 \;=\; \mathcal{E}\left\{\tilde{x}_{\ell,m}^2[n]\right\} & \text{if } m = j \end{cases} \tag{2.36}$$

and

$$\mathbf{v}_{\tilde{x}_{\ell,m}}^{\mathrm{T}} \mathbf{v}_{x_{\ell-1,j}} \;=\; \tilde{\sigma}_m \sigma_j \rho_{mj} \tag{2.37}$$

for $m,j \;=\; 1,2,\ldots,Q_p$. Hence, the optimal linear estimate of $\mathbf{v}_{x_{\ell-1,j}}$ given $\mathbf{v}_{\tilde{x}_{\ell,m}}$ is

$$\mathbf{v}_{\hat{x}_{\ell,j}} \;=\; \tilde{\rho}_{mj} \frac{\sigma_j}{\tilde{\sigma}_m} \mathbf{v}_{\tilde{x}_{\ell,m}} \;=\; b_{\ell,mj} \mathbf{v}_{\tilde{x}_{\ell,m}} \tag{2.38}$$

where

$$b_{\ell,mj} \;=\; \bar{\rho}_{mj}\frac{\sigma_j}{\check{\sigma}_m} \;=\; (\mathbf{B}_p)_{mj}$$

From (2.38) and

$$\mathbf{v}_{\dot{x}_{\ell,k}} \;=\; \bar{\rho}_{mk}\frac{\sigma_k}{\check{\sigma}_m}\mathbf{v}_{\dot{x}_{\ell,m}} \;=\; b_{\ell,mk}\mathbf{v}_{\dot{x}_{\ell,m}}$$

we see that

$$\mathbf{v}_{\dot{x}_{\ell,j}} \;=\; \frac{b_{\ell,mj}}{b_{\ell,mk}}\mathbf{v}_{\dot{x}_{\ell,k}} \tag{2.39}$$

Because of the dependence between weight coefficients indicated by (2.39), the number of independently adaptable DOF must equal the dimension of $\tilde{\mathbf{x}}_\ell$ or the rank of $\overline{\mathbf{C}}_\ell$.

Chapter 3

Statistical Analysis of the Multiple Time-Scale M^{2}VDR Beamformer

In the preceding chapter the formulation of MVDR beamformer weights assumes prior knowledge of the covariance matrix $\mathbf{R}$. However, because of changes in the propagation medium and source dynamics the covariance matrix of sensor measurements will be time-varying. Hence, weights can only be computed from sensor measurements observed over a finite time observation interval. We assume that the sensor measurement covariance matrix does not change over this time interval. In this discussion we refer to the number of sensor samples used to compute a single beamformer weight estimate as adaptation time.

Suppose $\mathbf{x}[n]$ are independent and identically distributed (i.i.d.), zero mean Gaussian random vectors. In this case the maximum likelihood estimate of $\mathbf{R}$ is Wishart distributed [45]. A Wishart distribution is the multivariate equivalent of the chi-squared distribution. As shown by Goodman [45], the number of statistical degrees of freedom associated with the covariance matrix estimate equals the number of sensor measurement vectors used in forming the estimate. By increasing the

number of statistical DOF, estimator variance decreases.

Statistical properties of the minimum variance beamformer are examined in [14], and [46]-[48]. For a narrow-band MVDR beamformer assuming i.i.d. zero mean Gaussian distributed sensor output vectors, Capon [46] shows that the output power estimate is a Chi-squared distributed random variable. Capon relates the number of statistical degrees of freedom associated with the power estimate to the difference between the adaptation time and number of adaptive DOF. This relationship shows that increasing the number of statistical DOF relative to the number of adaptive DOF reduces the variance of the power estimate. In [47] the distribution of output SNR for the narrow-band beamformer is derived assuming noise only data. Monzingo and Miller [14] analyze output SNR and MSE for several beamformer implementations. Steinhardt [48] derives a closed form expression for the marginal probability density function (pdf) of the narrow-band MVDR beamformer weight vector.

Recently Van Veen [41] examined in detail the statistical properties of the linearly constrained MVDR beamformer for the case when the same input data record is used to compute the weight vector and update the beamformer output. We refer to this adaptation scheme as a Type I implementation. Expressions are derived in [41] that relate the number of adaptive DOF and adaptation time to the mean output power and MSE of the beamformer. Analysis of mean output power reveals that a reduction in signal power occurs in proportion to the ratio of adaptive DOF to adaptation time. Capon and Goodman [46] make a similar observation. As adaptation time decreases in relation to the number of adaptive DOF, total output power of the beamformer approaches zero. One interpretation of this signal cancellation effect views the solution for the beamformer weights in the context of least squares (LS). When the data length equals the number of beamformer DOF, the LS problem forms an exactly determined system of equations. The solution to this problem leads to a zero beamformer output.

An alternative explanation for the reduction in mean output power observed

by [41], is given by Krolik and Swingler [12]. They show that a signal cancellation effect arises because of correlation between $\mathbf{w_a}$ and input data. Correlation between the beamformer weights and input data causes the adaptive path of the GSC to generate a scaled version of the desired signal. Consequently, a partial cancellation of the signal results at the beamformer output. Signal cancellation is avoided by ensuring that beamformer weights are computed from a past input data record, thus de-correlating the weight estimates from the current input data record. We refer to this adaptation scheme as a Type II implementation.

In this chapter we extend the results obtained in [11, 12, 16] to examine the statistical properties of the Type I and II multiple time-scale M^2VDR beamformers. Our discussion focuses on mean output power, MSE and output SNR when beamformer stages are adapted over different time-scales. To simplify analysis we model wavefield measurements as a wide sense stationary process. Weight estimates are computed by block adaptation from non-overlapping, rectangularly windowed, contiguous blocks of data. The length of a data block used in computing a weight estimate defines adaptation time. As an addition measure to ensure a tractable analysis, we assume that the flow of data across the beamfomer takes the following form. Type I adaptation uses the same sequence of data blocks to compute stage weights and update the stage output. This approach requires the succeeding stage to sit idle until all necessary data blocks have been processed by the preceding stage. Type II adaptation uses the past data block to compute weights and update the stage output. Although the weights of a succeeding stage can be computed without delay, our Type II scheme still requires this stage to idle until its current input data block is made available by the preceding stage.

3.1 Mean Output Power (Type I)

Mean output power of the single time-scale MVDR beamformer has been examined in [41, 42, 46]. Each of these investigators showed that a Type I implementation causes output power to be scaled below the level that would be attained had the true data covariance matrix been used in computing beamformer weights. In the following discussion we extend these results to the multiple time-scale beamformer.

Suppose $\mathbf{x}[n]$ are i.i.d., zero mean complex circular Gaussian random vectors. Assuming that beamformer weights are computed over non-overlapping blocks of data, let

$$\mathbf{X}_{p-1}[k] \;=\; (\; \mathbf{x}_{p-1}[1+(k-1)L_p],\; \mathbf{x}_{p-1}[2+(k-1)L_p],\; \ldots\; ,\mathbf{x}_{p-1}[kL_p]\;) \qquad (3.1)$$

for $k \in \mathrm{I}^+$, describe the input data block to the p^{th} stage, where the $MJ \times 1$ vector $\mathbf{x}_{p-1}[n]$ is the output of the $p^{\mathrm{th}} - 1$ stage. I^+ denotes all positive non-zero integers. The input data snapshot to the first stage is $\mathbf{x}_0[n]$. L_p equals the adaptation time assigned to the p^{th} stage. Adaptation time of the p^{th} stage is constrained to be an integer multiple of the adaptation time used in the $(p-1)^{\mathrm{th}}$ stage, so that

$$L_p \;=\; m_p L_{p-1}, \qquad m_p \in \mathrm{I}^+$$

for $p = 2, 3, \ldots, P$. This ensures that the adaptation time of any stage is an integer multiple of L_1, the adaptation time of the first stage. Employing block adaptation we write the output of the p^{th} stage as

$$\mathbf{Z}_p[k] \;=\; (\; \mathbf{x}_p[1+(k-1)L_p],\; \mathbf{x}_p[2+(k-1)L_p],\; \ldots\; ,\mathbf{x}_p[kL_p]\;) \qquad (3.2)$$

Note that $\mathbf{Z}_p[k]$ is constructed from L_p data vectors while the input data block to the $(p+1)^{th}$ stage, $\mathbf{X}_p[k]$, has L_{p+1} column vectors. Letting k correspond to the sampling index of the data block containing the vectors

$$\mathbf{x}_p[1+(k-1)L_1],\; \ldots.\; \mathbf{x}_p[kL_p]$$

we express $\mathbf{X}_p[k]$ as

$$\mathbf{X}_p[k] \;=\; (\; \mathbf{Z}_p[k],\; \mathbf{Z}_p[k+1],\; \ldots,\; \mathbf{Z}_p[k+m_{p+1}-1]\;) \tag{3.3}$$

Given $\mathbf{X}_{p-1}[k]$ an estimate of $\mathbf{B}_p$ follows from the solution to

$$\min_{\hat{\mathbf{B}}_p} \mathtt{tr}\left(\mathbf{X}_{p-1}[k] - \hat{\mathbf{B}}_p^{\dagger}[k]\dot{\mathbf{X}}_p[k]\right)^{\dagger}\left(\mathbf{X}_{p-1}^{\dagger}[k] - \dot{\mathbf{X}}_p^{\dagger}[k]\hat{\mathbf{B}}_p[k]\right) \tag{3.4}$$

where

$$\dot{\mathbf{X}}_p[k] \;=\; \overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1}[k]$$

describes the output of the signal blocking operation. The LS solution to (3.4) equals

$$\hat{\mathbf{B}}_p[k] \;=\; \left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1}[k]\mathbf{X}_{p-1}^{\dagger}[k]\overline{\mathbf{C}}_p\right)^{-1}\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1}[k]\mathbf{X}_{p-1}^{\dagger}[k] \tag{3.5}$$

We then write the output of the $p^{\mathtt{th}}$ stage as

$$\begin{aligned}\mathbf{x}_p[\ell] \;&=\; \left(\mathbf{I}_{MJ} \;-\; \overline{\mathbf{C}}_p\hat{\mathbf{B}}_p[k]\right)^{\dagger}\mathbf{x}_{p-1}[\ell] \\ &=\; \mathbf{x}_{p-1}[\ell] - \hat{\mathbf{B}}_p^{\dagger}[k]\dot{\mathbf{x}}_p[\ell]\end{aligned} \tag{3.6}$$

for

$$\ell \;=\; (1+(k-1)L_p),\; (2+(k-1)L_p),\; \ldots, kL_p$$

where

$$\dot{\mathbf{x}}_p[\ell] \;=\; \overline{\mathbf{C}}_p^{\dagger}\mathbf{x}_{p-1}[\ell]$$

A power estimate based on the output of the first p stages follows by applying the quiescent weight vector $\mathbf{w_q}$ to the sample covariance matrix formed from

$$\mathbf{Z}_p[k] \;=\; (\; \mathbf{x}_p[1+(k-1)L_p]\;\mathbf{x}_p[2+(k-1)L_p]\;\ldots\;\mathbf{x}_p[kL_p]\;) \tag{3.7}$$

From (3.7) the power estimate equals

$$\hat{\mathrm{P}}_p[k] \;=\; \frac{1}{L_p}\mathbf{w}_{\mathbf{q}}^{\dagger}\mathbf{Z}_p[k]\mathbf{Z}_p^{\dagger}[k]\mathbf{w}_{\mathbf{q}} \tag{3.8}$$

The next part of our discussion derives a closed form expression for the expected value of (3.8) in terms of the beamformer parameters

$$Q_\ell, \qquad \ell = 1, 2, \ldots, p$$

and

$$L_\ell, \qquad \ell = 1, 2, \ldots, p$$

The sample index associated with the input and output data blocks is omitted for clarity unless otherwise required by the discussion.

Assuming that input data vectors entering the first beamformer stage are i.i.d. complex Gaussian distributed, the $MJ \times MJ$ sample covariance matrix estimate

$$\hat{\mathbf{R}}_0 \;=\; \frac{1}{L_1}\mathbf{X}_0\mathbf{X}_0^\dagger \tag{3.9}$$

has the complex Wishart distribution

$$L_1\hat{\mathbf{R}}_0 \;\sim\; \mathtt{W}_{MJ}(L_1, \mathbf{R}_0)$$

where L_1 equals the number of statistical degrees of freedom [49]. Substituting (3.9) into (3.8), the power estimate at the input to the first stage, ie:

$$\hat{\mathtt{P}}_0 \;=\; \mathbf{w}_\mathtt{q}^\dagger\hat{\mathbf{R}}_0\mathbf{w}_\mathtt{q} \tag{3.10}$$

has the distribution

$$\mathtt{L}_1\hat{\mathtt{P}}_0 \;\sim\; \mathtt{W}_1(L_1, \mathtt{P}_0)$$

[49, 50], where

$$\mathtt{P}_0 \;=\; \mathbf{w}_\mathtt{q}^\dagger\mathbf{R}_0\mathbf{w}_\mathtt{q}$$

The expected value of (3.10) follows by normalizing $\hat{\mathtt{P}}_0$ by $\mathtt{P}_0$ to form the chi-squared random variable

$$\frac{L_1\hat{\mathtt{P}}_0}{\mathtt{P}_0} \;\sim\; \chi^2(L_1) \tag{3.11}$$

where L_1 equals the statistical degrees of freedom of the chi-squared random variable, and

$$\mathcal{E}\left\{ \chi^2(L_1) \right\} = L_1$$

since

$$\mathcal{E}\left\{\hat{P}_0\right\} = P_0$$

Consider now the output power estimate derived from the first adaptive stage. From (3.4), (3.5) and (3.8)

$$\begin{aligned}
\hat{P}_1 &= \frac{1}{L_1}\mathbf{w}_q^\dagger \mathbf{X}_0\mathbf{X}_0^\dagger \mathbf{w}_q - \frac{1}{L_1}\mathbf{w}_q^\dagger \left(\mathbf{X}_0\mathbf{X}_0^\dagger\overline{\mathbf{C}}_1 \left(\overline{\mathbf{C}}_1^\dagger \mathbf{X}_0\mathbf{X}_0^\dagger\overline{\mathbf{C}}_1\right)^{-1} \overline{\mathbf{C}}_1^\dagger \mathbf{X}_0\mathbf{X}_0^\dagger\right) \mathbf{w}_q \\
&= \frac{1}{L_1}\mathbf{w}_q^\dagger \mathbf{Z}_1\mathbf{Z}_1^\dagger \mathbf{w}_q \\
&= \mathbf{w}_q^\dagger \hat{\mathbf{R}}_1 \mathbf{w}_q \qquad (3.12)
\end{aligned}$$

where the distribution of $\mathbf{X}_0\mathbf{X}_0^\dagger$ is known under our previous assumptions to be a Wishart random matrix. To evaluate the expected value of (3.12) we turn to the following proposition.

Proposition 3.1 *Suppose that* $\mathbf{Y}$ *is a* $N\times L$ *random matrix such that*

$$\mathbf{Y}\mathbf{Y}^\dagger \sim W_N(L, \mathbf{\Sigma})$$

If $\mathbf{E}$ *is a* $N\times Q$ *non-random matrix then the random matrix described by*

$$\mathbf{\Lambda} = \mathbf{Y}\mathbf{Y}^\dagger - \mathbf{Y}\mathbf{Y}^\dagger\mathbf{E}\left(\mathbf{E}^\dagger\mathbf{Y}\mathbf{Y}^\dagger\mathbf{E}\right)^{-1}\mathbf{E}^\dagger\mathbf{Y}\mathbf{Y}^\dagger$$

has the Wishart distribution

$$\mathbf{\Lambda} \sim W_N(\, L-Q, \mathbf{\Sigma}_\Lambda \,)$$

where

$$\mathbf{\Sigma}_\Lambda = \mathbf{\Sigma} - \mathbf{\Sigma}\mathbf{E}\left(\mathbf{E}^\dagger\mathbf{\Sigma}\mathbf{E}\right)^{-1}\mathbf{E}^\dagger\mathbf{\Sigma}$$

for $L > Q$.□

(See Appendix B.2 for a proof).

From Proposition 3.1 we observe that

$$L_1\hat{\mathbf{R}}_1 \;=\; \mathbf{X}_0\mathbf{X}_0^\dagger \;-\; \mathbf{X}_0\mathbf{X}_0^\dagger\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^\dagger\mathbf{X}_0\mathbf{X}_0^\dagger\overline{\mathbf{C}}_1\right)^{-1}\overline{\mathbf{C}}_1^\dagger\mathbf{X}_0\mathbf{X}_0^\dagger$$

where

$$\hat{\mathbf{R}}_1 \;=\; \frac{1}{L_1}\mathbf{Z}_1\mathbf{Z}_1^\dagger$$

has the Wishart distribution

$$L_1\hat{\mathbf{R}}_1 \;\sim\; \mathbf{W}_{MJ}\left(\, L_1 - Q_1, \mathbf{R}_1 \,\right)$$

where

$$\mathbf{R}_1 = \mathbf{R}_0 - \mathbf{R}_0\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^\dagger\mathbf{R}_0\overline{\mathbf{C}}_1\right)^{-1}\overline{\mathbf{C}}_1^\dagger\mathbf{R}_0$$

Hence the first stage power estimate is distributed as

$$L_1\hat{\mathrm{P}}_1 \;\sim\; \mathrm{W}_1\left(\, L_1 - Q_1, \mathrm{P}_1 \,\right)$$

where

$$\mathrm{P}_1 \;=\; \mathbf{w}_\mathbf{q}^\dagger\mathbf{R}_1\mathbf{w}_\mathbf{q}$$

Now evaluating the expected value of $\hat{\mathrm{P}}_1$ we have

$$\mathcal{E}\left\{\hat{\mathrm{P}}_1\right\} \;=\; \frac{1}{L_1}\mathbf{w}_\mathbf{q}^\dagger\mathcal{E}\left\{\mathbf{Z}_1\mathbf{Z}_1^\dagger\right\}\mathbf{w}_\mathbf{q} \;=\; \frac{1}{L_1}\mathbf{w}_\mathbf{q}^\dagger\left(L_1 - Q_1\right)\mathbf{R}_1\mathbf{w}_\mathbf{q}$$

$$= \left(1 - \frac{Q_1}{L_1}\right)\mathrm{P}_1$$

$$= \left(1 - \alpha_1\right)\mathrm{P}_1 \tag{3.13}$$

where

$$\alpha_1 \;=\; \frac{Q_1}{L_1}$$

In order for the power estimate to be a positive definite quantity L_1 must be larger than the number of adaptive DOF assigned to the stage, ie:

$$L_1 > Q_1$$

This condition ensures that the covariance matrix estimate $\hat{\mathbf{R}}_1$ does not have an ill-defined Wishart distribution. Since L_1 is equal to the number of statistical degrees of freedom associated with $\hat{\mathbf{R}}_0$, the inequality $L_1 > Q_1$ suggests that the number of adaptive DOF can not exceed or be equal to the statistical DOF available at the input to the stage. Furthermore, when $L_1 < Q_1$ holds the system is underdetermined. This results in an ambiguous mapping of data by the beamformer.

In the case of the second stage, the output power estimate is formed by

$$\begin{aligned}\hat{\mathsf{P}}_2 &= \frac{1}{L_2}\mathbf{w}_q^\dagger \mathbf{X}_1\mathbf{X}_1^\dagger \mathbf{w}_q - \frac{1}{L_2}\mathbf{w}_q^\dagger \left(\mathbf{X}_1\mathbf{X}_1^\dagger \overline{\mathbf{C}}_2 \left(\overline{\mathbf{C}}_2^\dagger \mathbf{X}_1\mathbf{X}_1^\dagger \overline{\mathbf{C}}_2\right)^{-1} \overline{\mathbf{C}}_2^\dagger \mathbf{X}_1\mathbf{X}_1^\dagger\right)\mathbf{w}_q \\ &= \frac{1}{L_2}\mathbf{w}_q^\dagger \mathbf{Z}_2\mathbf{Z}_2^\dagger \mathbf{w}_q \\ &= \mathbf{w}_q^\dagger \hat{\mathbf{R}}_2 \mathbf{w}_q \end{aligned} \tag{3.14}$$

where

$$\hat{\mathbf{R}}_2 = \frac{1}{L_2}\mathbf{Z}_2\mathbf{Z}_2^\dagger$$

Since $\mathbf{Z}_1[k]\mathbf{Z}_1^\dagger[k]$ is Wishart distributed with $L_1 - Q_1$ DOF, the product $\mathbf{X}_1[k]\mathbf{X}_1^\dagger[k]$, expressed as the sum of i.i.d. matrices

$$\mathbf{X}_1[k]\mathbf{X}_1^\dagger[k] = \sum_{\ell=0}^{m_2-1} \mathbf{Z}_1[k+\ell]\mathbf{Z}_1^\dagger[k+\ell] \tag{3.15}$$

is Wishart distributed with $m_2(L_1 - Q_1)$ DOF [50]. By Proposition 3.1 the sample covariance matrix at the output of the second stage

$$\begin{aligned} L_2\hat{\mathbf{R}}_2 &= \mathbf{Z}_2\mathbf{Z}_2^\dagger \\ &= \mathbf{X}_1\mathbf{X}_1^\dagger - \mathbf{X}_1\mathbf{X}_1^\dagger \overline{\mathbf{C}}_2 \left(\overline{\mathbf{C}}_2^\dagger \mathbf{X}_1\mathbf{X}_1^\dagger \overline{\mathbf{C}}_2\right)^{-1} \overline{\mathbf{C}}_2^\dagger \mathbf{X}_1\mathbf{X}_1^\dagger \end{aligned} \tag{3.16}$$

has the distribution

$$L_2\hat{\mathbf{R}}_2 \sim \mathsf{W}_{MJ}\left(m_2(L_1 - Q_1) - Q_2, \mathbf{R}_2\right) \tag{3.17}$$

where

$$\mathbf{R}_2 \;=\; \mathbf{R}_1 - \mathbf{R}_1\overline{\mathbf{C}}_2\left(\overline{\mathbf{C}}_2^{\dagger}\mathbf{R}_1\overline{\mathbf{C}}_2\right)^{-1}\overline{\mathbf{C}}_2^{\dagger}\mathbf{R}_1$$

The second stage power estimate

$$\hat{\mathtt{P}}_2 \;=\; \mathbf{w}_q^{\dagger}\hat{\mathbf{R}}_2[k]\mathbf{w}_{\mathtt{q}}$$

is then Wishart distributed with $m_2(L_1 - Q_1) - Q_2$ DOF. Taking the expectation $\hat{\mathtt{P}}_2$ we then have

$$\begin{aligned}\mathcal{E}\left\{\hat{\mathtt{P}}_2\right\} \;&=\; \frac{1}{L_2}\mathbf{w}_{\mathtt{q}}^{\dagger}\mathcal{E}\left\{\mathbf{Z}_2\mathbf{Z}_2^{\dagger}\right\}\mathbf{w}_{\mathtt{q}}\\ &=\; \left(1 - \frac{Q_1}{L_1} - \frac{Q_2}{L_2}\right)\mathtt{P}_2 \qquad (3.18)\end{aligned}$$

Since $m_2 L_1 = L_2$ we have

$$\mathcal{E}\left\{\hat{\mathtt{P}}_2\right\} \;=\; (1 - \alpha_1 - \alpha_2)\,\mathtt{P}_2 \qquad (3.19)$$

where

$$\alpha_2 \;=\; \frac{Q_2}{L_2}$$

and

$$\mathtt{P}_2 = \mathbf{w}_{\mathtt{q}}^{\dagger}\mathbf{R}_2\mathbf{w}_{\mathtt{q}}$$

Observe from (3.18) that

$$L_2 > m_2 Q_1 + Q_2$$

must hold in order that output power be a positive definite quantity.

We now use an argument similar to the one used for the first and second stages to show by induction that the output power for the $p^{\mathtt{th}}$ stage will have the distribution

$$L_p\hat{\mathtt{P}}_p \;\sim\; \mathtt{W}_1\left(L_p \;-\; \sum_{k=1}^{p-1}\left(\prod_{\ell=k+1}^{p} m_\ell\right)Q_k \;-\; Q_p,\; \mathtt{P}_p\right) \qquad (3.20)$$

and a mean value equal to

$$\mathcal{E}\left\{\hat{\mathsf{P}}_p\right\} \;=\; \left(1 - \sum_{q=1}^{p} \alpha_q\right) \mathsf{P}_p \tag{3.21}$$

where

$$\mathsf{P}_p \;=\; \mathbf{w}_{\mathbf{q}}^{\dagger} \mathbf{R}_p \mathbf{w}_{\mathbf{q}}$$

$$\mathbf{R}_p \;=\; \mathbf{R}_{p-1} - \mathbf{R}_{p-1}\overline{\mathbf{C}}_p \left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{R}_{p-1}\overline{\mathbf{C}}_p\right)^{-1} \overline{\mathbf{C}}_p^{\dagger}\mathbf{R}_{p-1}$$

and

$$\alpha_q \;=\; \frac{Q_q}{L_q}$$

The form of the $p^{\mathbf{th}}$ stage mean output power given by (3.21) points to a fundamental limitation imposed on the multiple time scale beamformer. As suggested by (3.21), for the power estimate at the output of the $p^{\mathbf{th}}$ stage to be a positive quantity we require

$$\sum_{\ell=1}^{p} \alpha_\ell \;<\; 1 \tag{3.22}$$

to be satisfied, or equivalently

$$\sum_{k=1}^{p-1} \left(\prod_{\ell=k+1}^{p} m_\ell \right) Q_k \;+\; Q_p \;<\; L_p \tag{3.23}$$

This condition implies that the number of `i.i.d.` data blocks used in computing a single power estimate at the $p^{\mathbf{th}}$ stage must be greater than the total number of adaptive `DOF` estimated by the beamformer.

3.2 Mean-Squared-Error (Type I)

In a Type I implementation partial signal cancellation occurs because of correlation between beamformer weights and input data. When signal cancellation is not correctly taken into account in the analysis of `MSE`, as in [41], an overly pessimistic prediction of performance will result [42]. The first part of this discussion focuses on an appropriate measure for `MSE`.

Recall that the signal component of the measured wavefield is orthogonal to the linear span of the beamformer signal blocking matrix. Likewise wavefield components contained in the linear span of $\overline{\mathbf{C}}$ constitute noise. We then rewrite the input data block $\mathbf{X}_0[k]$ as the sum of signal and noise components,

$$\mathbf{X}_0[k] \;=\; \mathbf{S}_0[k] \;+\; \mathbf{X}_{0,\eta}[k] \tag{3.24}$$

where

$$\overline{\mathbf{C}}_\ell^\dagger \mathbf{S}_0[k] \;=\; \mathbf{0}_{Q_\ell, L_1}, \qquad \ell = 1,\ 2,\ \ldots\ ,P$$

$\mathbf{S}_0[k]$ and $\mathbf{X}_{0,\eta}[k]$ represent the signal and noise components of the measured wavefield. We assume that $\mathbf{S}_0[k]$ and $\mathbf{X}_{0,\eta}[k]$ are statistically independent, zero mean complex Gaussian random processes. Extending our notation to subsequent stages, let

$$\mathbf{X}_{p-1}[k] \;=\; \mathbf{S}_{p-1}[k] \;+\; \mathbf{X}_{p-1,\eta}[k] \tag{3.25}$$

where

$$\mathbf{S}_{p-1}[k] \;=\; (\ \mathbf{S}_0[k]\ \mathbf{S}_0[k+1]\ \ldots\ \mathbf{S}_0[k+m_p-1]\ )$$

for $p > 1$. Observe that the signal passes the first beamformer stage undistorted if and only if

$$\mathcal{E}\left\{\mathbf{Z}_1 \mid \mathbf{S}_0\right\} \;=\; \mathbf{S}_0 \tag{3.26}$$

holds. A similar condition applies to subsequent stages. At the p^{th} stage, the signal passes without distortion if and only if

$$\mathcal{E}\left\{\mathbf{Z}_p \mid \mathbf{S}_{p-1}\right\} \;=\; \mathbf{S}_{p-1} \tag{3.27}$$

is satisfied. To quantify the signal cancellation effect $\mathcal{E}\left\{\mathbf{Z}_p \mid \mathbf{S}_{p-1}\right\}$ must be evaluated.

The output of the first beamformer stage is

$$\mathbf{Z}_1 \;=\; \left(\mathbf{I}_M \;-\; \overline{\mathbf{C}}_1 \hat{\mathbf{B}}_1\right)^\dagger \mathbf{X}_0 \tag{3.28}$$

where

$$\hat{\mathbf{B}}_1 = \left(\overline{\mathbf{C}}_1^\dagger \mathbf{X}_0 \mathbf{X}_0^\dagger \overline{\mathbf{C}}_1\right)^{-1} \overline{\mathbf{C}}_1^\dagger \mathbf{X}_0 \mathbf{X}_0^\dagger$$

Substituting (3.24) into the expression for $\hat{\mathbf{B}}_1$, we rewrite $\mathbf{Z}_1$ in the form

$$\mathbf{Z}_1 = \mathbf{S}_0 + \mathbf{X}_{0,\eta} - \mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^\dagger \overline{\mathbf{C}}_1 \left(\overline{\mathbf{C}}_1^\dagger \mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^\dagger \overline{\mathbf{C}}_1\right)^{-1} \overline{\mathbf{C}}_1^\dagger \mathbf{X}_{0,\eta}$$

$$- \mathbf{S}_0 \mathbf{X}_{0,\eta}^\dagger \overline{\mathbf{C}}_1 \left(\overline{\mathbf{C}}_1^\dagger \mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^\dagger \overline{\mathbf{C}}_1\right)^{-1} \overline{\mathbf{C}}_1^\dagger \mathbf{X}_{0,\eta} \tag{3.29}$$

From (3.29) the conditional expectation of $\mathbf{Z}_1$ given $\mathbf{S}_0$ equals

$$\mathcal{E}\{\mathbf{Z}_1 \mid \mathbf{S}_0\} = \mathbf{S}_0 - \mathbf{S}_0 \mathcal{E}\left\{\mathbf{X}_{0,\eta}^\dagger \overline{\mathbf{C}}_1 \left(\overline{\mathbf{C}}_1^\dagger \mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^\dagger \overline{\mathbf{C}}_1\right)^{-1} \overline{\mathbf{C}}_1^\dagger \mathbf{X}_{0,\eta}\right\}$$

$$- \mathcal{E}\left\{\mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^\dagger \overline{\mathbf{C}}_1 \left(\overline{\mathbf{C}}_1^\dagger \mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^\dagger \overline{\mathbf{C}}_1\right)^{-1} \overline{\mathbf{C}}_1^\dagger \mathbf{X}_{0,\eta}\right\} \tag{3.30}$$

where

$$\mathcal{E}\left\{\mathbf{X}_{0,\eta}\right\} = \mathbf{0}_{M,L_1}$$

Given that

$$\overline{\mathbf{C}}_1^\dagger \mathbf{X}_{0,\eta}$$

is Gaussian distributed with zero mean, we may show that

$$\mathcal{E}\left\{\mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^\dagger \overline{\mathbf{C}}_1 \left(\overline{\mathbf{C}}_1^\dagger \mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^\dagger \overline{\mathbf{C}}_1\right)^{-1} \overline{\mathbf{C}}_1^\dagger \mathbf{X}_{0,\eta}\right\}$$

equals zero.

Proposition 3.2 *Let* $\mathbf{V}_1$ *and* $\mathbf{V}_2$ *denote* $Q \times N$ *and* $M \times N$ *random matrices, respectively. Assume that* $\mathbf{V}_1$ *and* $\mathbf{V}_2$ *are correlated. Supposing that* $\mathbf{V}_1$ *and* $\mathbf{V}_2$ *have a jointly Gaussian distribution and*

$$\mathbf{V}_1 \sim \mathrm{N}(\,0, \boldsymbol{\Sigma}_{11}\,)$$

$$\mathbf{V}_2 \sim \mathrm{N}(0, \boldsymbol{\Sigma}_{22})$$

then

$$\mathcal{E}\left\{\mathbf{V}_1^\dagger\left(\mathbf{V}_1\mathbf{V}_1^\dagger\right)^{-1}\mathbf{V}_1\mathbf{V}_2^\dagger\right\} \;=\; \mathbf{0}_{N,M}$$

□

(See Appendix B.2 for a proof).

Letting

$$\mathbf{V}_1 \;=\; \overline{\mathbf{C}}_1^\dagger\mathbf{X}_{0,\eta}$$

and

$$\mathbf{V}_2 \;=\; \mathbf{X}_{0,\eta}$$

we deduce from Proposition 3.2 that

$$\mathcal{E}\left\{\mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^\dagger\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^\dagger\mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^\dagger\overline{\mathbf{C}}_1\right)^{-1}\overline{\mathbf{C}}_1^\dagger\mathbf{X}_{0,\eta}\right\} \;=\; \mathbf{0}_{MJ,L_1} \tag{3.31}$$

To evaluate

$$\mathcal{E}\left\{\mathbf{X}_{0,\eta}^\dagger\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^\dagger\mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^\dagger\overline{\mathbf{C}}_1\right)^{-1}\overline{\mathbf{C}}_1^\dagger\mathbf{X}_{0,\eta}\right\}$$

consider the following.

Proposition 3.3 *Suppose that* $\mathbf{V}$ *describes a* $Q\times L$ *random matrix with* `i.i.d` *columns which are symmetrically distributed about zero, then*

$$\mathcal{E}\left\{\mathbf{V}^\dagger\left(\mathbf{V}\mathbf{V}^\dagger\right)^{-1}\mathbf{V}\right\} \;=\; \frac{Q}{L}\mathbf{I}_L$$

□

(See [42] as well as Appendix B.3 for a proof).

Observing that column vectors of $\overline{\mathbf{C}}_1^\dagger\mathbf{X}_{0,\eta}$ are independent zero mean Gaussian distributed, we apply Proposition 3.3 to the third term in (3.30) and write

$$\mathcal{E}\left\{\mathbf{X}_{0,\eta}^\dagger\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^\dagger\mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^\dagger\overline{\mathbf{C}}_1\right)^{-1}\overline{\mathbf{C}}_1^\dagger\mathbf{X}_{0,\eta}\right\} \;=\; \frac{Q_1}{L_1}\mathbf{I}_{L_1} \tag{3.32}$$

From (3.31) and (3.32) the conditional expectation of $\mathbf{Z}_1$ given $\mathbf{S}_0$ equals a scaled version of the desired signal, ie:

$$\begin{aligned} \mathcal{E}\{\mathbf{Z}_1 \mid \mathbf{S}_0\} &= \left(1 - \frac{Q_1}{L_1}\right) \mathbf{S}_0 \\ &= (1 - \alpha_1)\ \mathbf{S}_0 \end{aligned} \tag{3.33}$$

indicating a partial cancellation of the signal. In the following discussion we examine a similar effect at the output of the second stage.

For the second stage the degree of signal cancellation is determined from

$$\mathcal{E}\{\mathbf{Z}_2 \mid \mathbf{S}_1\}$$

where $\mathbf{S}_1$ is composed of m_2 non-overlapping samples of $\mathbf{S}_0$ and

$$\mathbf{X}_1[k] = (\mathbf{Z}_1[k]\ \mathbf{Z}_1[k+1]\ \ldots\ \mathbf{Z}_1[k+m_2-1])$$

From (3.29) we rewrite $\mathbf{Z}_1$ as

$$\mathbf{Z}_1 = \mathbf{S}_0 + \mathbf{Z}_{1,s} + \mathbf{Z}_{1,\eta} \tag{3.34}$$

so that

$$\mathbf{Z}_{1,s} = -\mathbf{S}_0\mathbf{X}_{0,\eta}^{\dagger}\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^{\dagger}\overline{\mathbf{C}}_1\right)^{-1}\overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,\eta} \tag{3.35}$$

and

$$\mathbf{Z}_{1,\eta} = \mathbf{X}_{0,\eta} - \mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^{\dagger}\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,\eta}\mathbf{X}_{0,\eta}^{\dagger}\overline{\mathbf{C}}_1\right)^{-1}\overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,\eta} \tag{3.36}$$

where $\mathbf{Z}_{1,s}$ is a result of the signal appearing in the estimated beamformer weights, and $\mathbf{X}_{1,\eta}$ corresponds to the noise only component of $\mathbf{Z}_1$. We now express $\mathbf{X}_1$ as the sum

$$\mathbf{X}_1 = \mathbf{S}_1 + \mathbf{X}_{1,s} + \mathbf{X}_{1,\eta} \tag{3.37}$$

where

$$\mathbf{X}_{1,s}[k] = (\mathbf{Z}_{1,s}[k]\ \mathbf{Z}_{1,s}[k+1]\ \ldots\ \mathbf{Z}_{1,s}[k+m_2-1])$$

and

$$\mathbf{X}_{1,\eta}[k] \;=\; \left(\mathbf{Z}_{1,\eta}[k]\;\mathbf{Z}_{1,\eta}[k+1]\;\ldots\;\mathbf{Z}_{1,\eta}[k+m_2-1]\right)$$

Substituting (3.37) into

$$\mathbf{Z}_2 \;=\; \mathbf{X}_1 \;-\; \mathbf{X}_1\mathbf{X}_1^{\dagger}\overline{\mathbf{C}}_2\left(\overline{\mathbf{C}}_2^{\dagger}\mathbf{X}_1\mathbf{X}_1^{\dagger}\overline{\mathbf{C}}_2\right)^{-1}\overline{\mathbf{C}}_2^{\dagger}\mathbf{X}_1 \tag{3.38}$$

the output of the second stage is rewritten as

$$\begin{aligned}\mathbf{Z}_2 \;=\; \mathbf{S}_1 \;+\; \mathbf{X}_{1,s} - \mathbf{S}_1\mathbf{X}_{1,\eta}^{\dagger}\overline{\mathbf{C}}_2\left(\overline{\mathbf{C}}_2^{\dagger}\mathbf{X}_{1,\eta}\mathbf{X}_{1,\eta}^{\dagger}\overline{\mathbf{C}}_2\right)^{-1}\overline{\mathbf{C}}_2^{\dagger}\mathbf{X}_{1,\eta}\\ +\mathbf{X}_{1,\eta} \;-\; \mathbf{X}_{1,\eta}\mathbf{X}_{1,\eta}^{\dagger}\overline{\mathbf{C}}_2\left(\overline{\mathbf{C}}_2^{\dagger}\mathbf{X}_{1,\eta}\mathbf{X}_{1,\eta}^{\dagger}\overline{\mathbf{C}}_2\right)^{-1}\overline{\mathbf{C}}_2^{\dagger}\mathbf{X}_{1,\eta}\end{aligned} \tag{3.39}$$

since

$$\mathbf{X}_{1,s}\mathbf{X}_{1,\eta}^{\dagger} \;=\; \mathbf{0}$$

The conditional expectation of $\mathbf{Z}_2$ equals

$$\begin{aligned}\mathcal{E}\left\{\mathbf{Z}_2 \mid \mathbf{S}_1\right\} \;=\; \mathbf{S}_1 + \mathcal{E}\left\{\mathbf{X}_{1,s} \mid \mathbf{S}_1\right\} - \mathbf{S}_1\mathcal{E}\left\{\mathbf{X}_{1,\eta}^{\dagger}\overline{\mathbf{C}}_2\left(\overline{\mathbf{C}}_2^{\dagger}\mathbf{X}_{1,\eta}\mathbf{X}_{1,\eta}^{\dagger}\overline{\mathbf{C}}_2\right)^{-1}\overline{\mathbf{C}}_2^{\dagger}\mathbf{X}_{1,\eta}\right\}\\ - \mathcal{E}\left\{\mathbf{X}_{1,\eta}\mathbf{X}_{1,\eta}^{\dagger}\overline{\mathbf{C}}_2\left(\overline{\mathbf{C}}_2^{\dagger}\mathbf{X}_{1,\eta}\mathbf{X}_{1,\eta}^{\dagger}\overline{\mathbf{C}}_2\right)^{-1}\overline{\mathbf{C}}_2^{\dagger}\mathbf{X}_{1,\eta}\right\}\end{aligned} \tag{3.40}$$

where

$$\mathcal{E}\left\{\;\mathbf{X}_{1,\eta} \mid \mathbf{S}_1\;\right\} \;=\; \mathbf{0}_{MJ,L_2}$$

To apply Propositions 3.2 and 3.3 to (3.40), as we did in analyzing the first stage, $\mathbf{X}_{1,\eta}$ must be Gaussian distributed. Unfortunately, because of the action of the first stage, output data will not be Gaussian distributed even though

$$\mathbf{X}_{1,\eta}\mathbf{X}_{1,\eta}^{\dagger}$$

is Wishart distributed. As an alternative, observe that $\mathbf{Z}_{1,\eta}$ conditional on $\overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,\eta}$ is Gaussian distributed with zero mean (see Appendix B.4). It then follows that

$\mathbf{X}_{1,\eta}$ conditional on

$$\mathcal{C}_{0,\eta} = \left(\overline{\mathbf{C}}_1^\dagger \mathbf{X}_{0,\eta}[k],\ \overline{\mathbf{C}}_1^\dagger \mathbf{X}_{0,\eta}[k+1],\ \ldots, \overline{\mathbf{C}}_1^\dagger \mathbf{X}_{0,\eta}[k+m_2-1]\right) \tag{3.41}$$

is Gaussian distributed with zero mean. To take advantage of this observation we rewrite $\mathcal{E}\{\mathbf{Z}_2 \mid \mathbf{S}_1\}$ in the equivalent form

$$\begin{aligned}
\mathcal{E}\{\mathbf{Z}_2|\mathbf{S}_1\} &= \mathcal{E}\left\{\mathcal{E}\left\{\mathbf{Z}_2|\mathcal{C}_{0,\eta}, \mathbf{S}_1\right\}|\mathbf{S}_1\right\} \\
&= (1-\alpha_1)\mathbf{S}_1 \\
&\quad -\mathbf{S}_1 \mathcal{E}\left\{\mathcal{E}\left\{\mathbf{X}_{1,\eta}^\dagger \overline{\mathbf{C}}_2 \left(\overline{\mathbf{C}}_2^\dagger \mathbf{X}_{1,\eta}\mathbf{X}_{1,\eta}^\dagger \overline{\mathbf{C}}_2\right)^{-1} \overline{\mathbf{C}}_2^\dagger \mathbf{X}_{1,\eta}|\mathcal{C}_{0,\eta}, \mathbf{S}_1\right\}|\mathbf{S}_1\right\} \\
&\quad - \mathcal{E}\left\{\mathcal{E}\left\{\mathbf{X}_{1,\eta}\mathbf{X}_{1,\eta}^\dagger \overline{\mathbf{C}}_2 \left(\overline{\mathbf{C}}_2^\dagger \mathbf{X}_{1,\eta}\mathbf{X}_{1,\eta}^\dagger \overline{\mathbf{C}}_2\right)^{-1} \overline{\mathbf{C}}_2^\dagger \mathbf{X}_{1,\eta}|\mathcal{C}_{0,\eta}, \mathbf{S}_1\right\}|\mathbf{S}_1\right\}
\end{aligned} \tag{3.42}$$

Evaluating the inner expectation in (3.42), observe that since

$$\mathbf{X}_1|\mathcal{C}_{0,\eta}, \mathbf{S}_1$$

is zero mean Gaussian distributed it follows from Proposition 3.3 that

$$\mathcal{E}\left\{\mathbf{X}_{1,\eta}^\dagger \overline{\mathbf{C}}_2 \left(\overline{\mathbf{C}}_2^\dagger \mathbf{X}_{1,\eta}\mathbf{X}_{1,\eta}^\dagger \overline{\mathbf{C}}_2\right)^{-1} \overline{\mathbf{C}}_2^\dagger \mathbf{X}_{1,\eta}|\mathcal{C}_{0,\eta}, \mathbf{S}_1\right\} = \frac{Q_2}{L_2} = \alpha_2 \tag{3.43}$$

By Proposition 3.2

$$\mathcal{E}\left\{\mathbf{X}_{1,\eta}\mathbf{X}_{1,\eta}^\dagger \overline{\mathbf{C}}_2 \left(\overline{\mathbf{C}}_2^\dagger \mathbf{X}_{1,\eta}\mathbf{X}_{1,\eta}^\dagger \overline{\mathbf{C}}_2\right)^{-1} \overline{\mathbf{C}}_2^\dagger \mathbf{X}_{1,\eta}|\mathcal{C}_{0,\eta}, \mathbf{S}_1\right\} = \mathbf{0}_{M,L_2} \tag{3.44}$$

From (3.42), (3.43), and (3.44), $\mathcal{E}\{\mathbf{Z}_2 \mid \mathbf{S}_1\}$ equals

$$\mathcal{E}\{\mathbf{Z}_2 \mid \mathbf{S}_1\} = (1 - \alpha_1 - \alpha_2)\ \mathbf{S}_1 \tag{3.45}$$

Extending our analysis to succeeding stages of the beamformer we may show by induction that the expectation of $\mathbf{Z}_p$ given $\mathbf{S}_{p-1}$ takes the form

$$\mathcal{E}\{\mathbf{Z}_p \mid \mathbf{S}_{p-1}\} = \left(1 - \sum_{i=1}^{p} \alpha_i\right)\ \mathbf{S}_{p-1} \tag{3.46}$$

where

$$\mathbf{Z}_p = \mathbf{S}_p + \mathbf{X}_{p,s} + \mathbf{X}_{p,\eta}$$

Note further that

$$\mathbf{Z}_{p,s} = -\mathbf{S}_{p-1}\mathbf{X}_{p-1,\eta}^{\dagger}\overline{\mathbf{C}}_p\left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1,\eta}\mathbf{X}_{p-1,\eta}^{\dagger}\overline{\mathbf{C}}_p\right)^{-1}\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1,\eta} \tag{3.47}$$

and

$$\mathbf{Z}_{p,\eta} = \mathbf{X}_{p-1,\eta} - \mathbf{X}_{p-1,\eta}\mathbf{X}_{p-1,\eta}^{\dagger}\overline{\mathbf{C}}_p\left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1,\eta}\mathbf{X}_{p-1,\eta}^{\dagger}\overline{\mathbf{C}}_p\right)^{-1}\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1,\eta} \tag{3.48}$$

for $p \geq 1$. As indicated by (3.46) the output of the p^{th} stage contains a scaled version of the signal. Since this scaling is only affected by known beamformer parameters, (ie: α_ℓ, $\ell = 1, 2, \ldots, p$), it is necessary to account for (3.46) in the definition of MSE to arrive at an accurate measure of performance. an accurate measure of performance.

First, we consider the standard definition of MSE as applied to the p^{th} stage, ie:

$$\begin{aligned}
\varepsilon_p &= \frac{1}{L_p}\mathcal{E}\left\{\left|\, \mathbf{w}_{\mathbf{q}}^{\dagger}\left(\mathbf{Z}_p - \mathbf{S}_{p-1}\right)\right|^2\right\} \\
&= \frac{1}{L_p}\mathcal{E}\left\{\left|\, \mathbf{w}_{\mathbf{q}}^{\dagger}\left(\mathbf{Z}_p - \mathcal{E}\left\{\mathbf{Z}_p \mid \mathbf{S}_{p-1}\right\} + \mathcal{E}\left\{\mathbf{Z}_p \mid \mathbf{S}_{p-1}\right\} - \mathbf{S}_{p-1}\right)\right|^2\right\} \\
&= \frac{1}{L_p}\mathcal{E}\left\{\left|\, \mathbf{w}_{\mathbf{q}}^{\dagger}\left(\mathbf{Z}_p - \mathcal{E}\left\{\mathbf{Z}_p \mid \mathbf{S}_{p-1}\right\}\right)\right|^2\right\} \\
&\quad + \frac{1}{L_p}\mathcal{E}\left\{\left|\, \mathbf{w}_{\mathbf{q}}^{\dagger}\left(\mathcal{E}\left\{\mathbf{Z}_p \mid \mathbf{S}_{p-1}\right\} - \mathbf{S}_{p-1}\right)\right|^2\right\} \\
&= \varepsilon_p^c + \mathsf{b}_p^2
\end{aligned} \tag{3.49}$$

where

$$\begin{aligned}
\mathsf{b}_p^2 &= \tfrac{1}{L_p}\mathcal{E}\left\{\left|\, \mathbf{w}_{\mathbf{q}}^{\dagger}\left(\mathcal{E}\left\{\mathbf{Z}_p \mid \mathbf{S}_{p-1}\right\} - \mathbf{S}_{p-1}\right)\right|^2\right\} \\
&= \tfrac{1}{L_p}\left(\textstyle\sum_{\ell=1}^{p}\alpha_\ell\right)^2 \mathbf{w}_{\mathbf{q}}^{\dagger}\mathcal{E}\left\{\mathbf{S}_{p-1}\mathbf{S}_{p-1}^{\dagger}\right\}\mathbf{w}_{\mathbf{q}}
\end{aligned}$$

equals mean squared conditional bias. Since $\mathbf{b}_p^2$ is dependent only on the number of adaptive weights and adaptation time, and not on noise power, the term

$$\varepsilon_p^c = \frac{1}{L_p}\mathcal{E}\left\{\left|\mathbf{w}_q^\dagger\left(\mathbf{Z}_p - \mathcal{E}\left\{\mathbf{Z}_p \mid \mathbf{S}_{p-1}\right\}\right)\right|^2\right\} \tag{3.50}$$

provides a definition of MSE which is corrected for the signal cancellation effect associated with a Type I implementation. By substituting (3.16) into (3.50) we express the corrected MSE for a p stage network as

$$\varepsilon_p^c = \left(1 - \sum_{\ell=1}^{p}\alpha_\ell\right)\mathsf{P}_{p,\mathrm{n}} + \sum_{q=1}^{p}\alpha_q\left(1 - \sum_{\ell=1}^{p}\alpha_\ell\right)\mathsf{P}_{\mathrm{s}} \tag{3.51}$$

where noise power at the output of the p^{th} stage is

$$\mathsf{P}_{p,\mathrm{n}} = \mathbf{w}_q^\dagger\left(\mathbf{R}_{p-1,\eta} - \mathbf{R}_{p-1,\eta}\overline{\mathbf{C}}_p\left(\overline{\mathbf{C}}_p^\dagger\mathbf{R}_{p-1,\eta}\overline{\mathbf{C}}_p\right)^{-1}\overline{\mathbf{C}}_p^\dagger\mathbf{R}_{p-1,\eta}\right)\mathbf{w}_q$$

and $\mathbf{R}_{p-1,\eta}$ is the covariance matrix of $\mathbf{X}_{p-1,\eta}$, such that

$$\mathbf{R}_{p-1,\eta} = \mathbf{R}_{p-2,\eta} - \mathbf{R}_{p-2,\eta}\overline{\mathbf{C}}_{p-1}\left(\overline{\mathbf{C}}_{p-1}^\dagger\mathbf{R}_{p-2,\eta}\overline{\mathbf{C}}_{p-1}\right)^{-1}\overline{\mathbf{C}}_{p-1}^\dagger\mathbf{R}_{p-2,\eta}$$

for $p > 2$. Signal power equals

$$\mathsf{P}_{\mathrm{s}} = \mathbf{w}_q^\dagger\mathbf{R}_{\mathrm{s}}\mathbf{w}_q$$

where

$$\mathbf{R}_{\mathrm{s}} = \frac{1}{L_p}\mathcal{E}\left\{\mathbf{S}_{p-1}\mathbf{S}_{p-1}^\dagger\right\}$$

As a final point we evaluate (3.49) so that

$$\varepsilon_p = \left(1 - \sum_{\ell=1}^{p}\alpha_\ell\right)\mathsf{P}_{p,\mathrm{n}} + \sum_{q=1}^{p}\alpha_q\mathsf{P}_{\mathrm{s}} \tag{3.52}$$

equals the uncorrected MSE of the beamformer. In comparing this expression with (3.51) we observe that ε_p provides a misleading measure of MSE by the factor

$$\left(\sum_{q=1}^{p}\alpha_q\sum_{\ell=1}^{p}\alpha_\ell\right)\mathsf{P}_{\mathrm{s}}$$

3.3 Signal Cancellation

To illustrate how the choice of

$$Q_\ell \quad \text{and} \quad L_\ell$$

affects the output power of the Type I beamformer consider a simulation experiment consisting of a single source embedded in additive white Gaussian noise. We assume that the vectors $\mathbf{x}[n]$ are `i.i.d.`. Source and noise processes of the wavefield are statistically independent. The spatial field is measured by a 12 sensor single tap equi-spaced array pre-steered to the source location. 11 `DOF` are available to a two stage adaptive beamformer. To verify (3.21) α_1 and α_2 are varied as

$$\alpha_1 = \frac{\ell}{L_1} \quad \text{and} \quad \alpha_2 = \frac{11-\ell}{m_2 L_1} = \frac{11-\ell}{L_2}$$

for $\ell = 1, 2, \ldots, 10$. Based on 150 independent Monte Carlo trials Figure 3.1 shows output power of the two stage beamformer plotted as

$$10\log_{10}\left(\mathcal{E}\left\{\hat{\mathrm{P}}_2\right\}\right) - 10\log_{10}(\mathrm{P}_2)$$

for various values of α_1 and α_2 given ℓ and m_2. Overlaid onto Figure 3.1 are theoretical results obtained from (3.21) A measure of `0dB` on the figure implies zero degradation in output power. Note from Figure 3.1 that as the ratio of `DOF` to adaptation time approaches 1 the signal cancellation effect becomes more pronounced.

To provide a closer examination of signal cancellation recall first that for a finite adaptation time the p^{th} stage weights are computed from

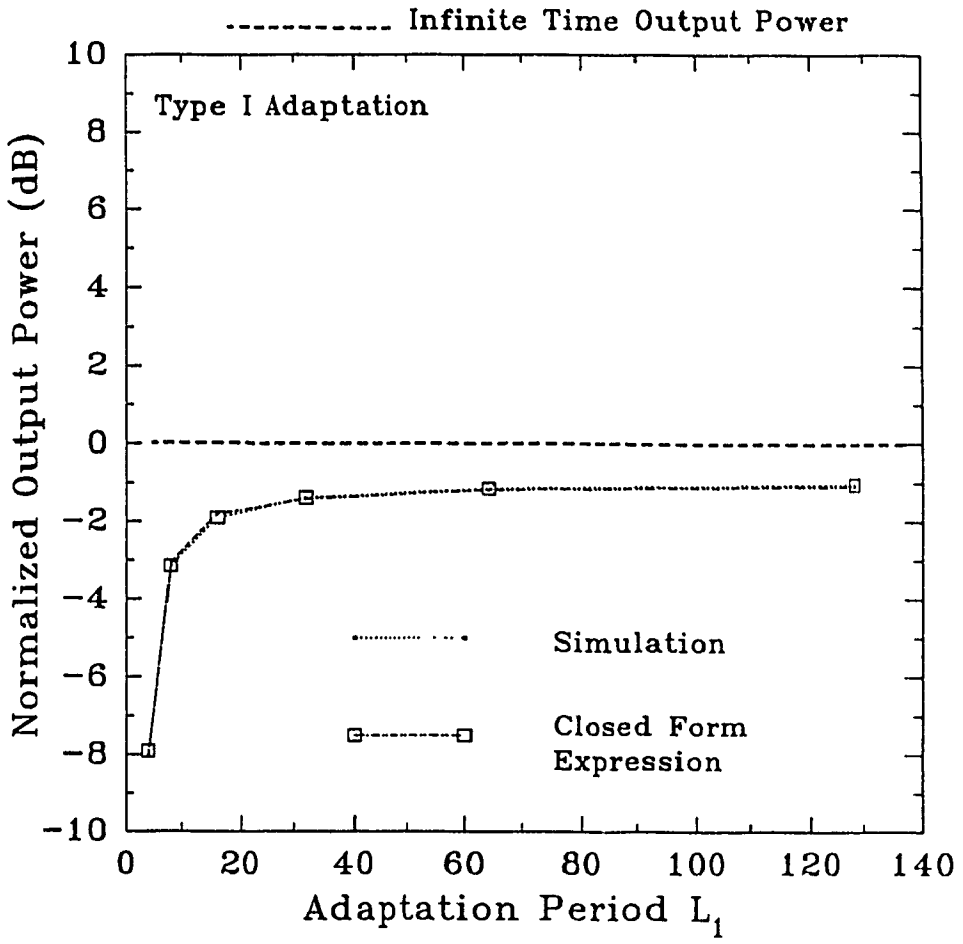
$$\hat{\mathbf{B}}_p[k] = \left(\overline{\mathbf{C}}_p^\dagger \mathbf{X}_{p-1}[k]\mathbf{X}_{p-1}^\dagger[k]\overline{\mathbf{C}}_p\right)^{-1} \overline{\mathbf{C}}_p^\dagger \mathbf{X}_{p-1}[k]\mathbf{X}_{p-1}^\dagger[k] \tag{3.53}$$

Rewriting $\mathbf{X}_{p-1}$ explicitly in terms of a signal and noise component as

$$\mathbf{X}_{p-1} = \mathbf{S}_{p-1} + \mathbf{X}_{p-1,\mathbf{s}} + \mathbf{X}_{p-1,\boldsymbol{\eta}} \tag{3.54}$$

where $\mathbf{S}_{p-1}$ denotes the signal component, such that

$$\overline{\mathbf{C}}_p^\dagger \mathbf{S}_{p-1} = \mathbf{0}_{Q_p, L_p}$$



Mean Output Power – Type I Implementation

Figure 3.1

and $\mathbf{X}_{p-1,\eta}$ is the noise term. Expanding $\hat{\mathbf{B}}_p[k]$ in terms of (3.54) as

$$\hat{\mathbf{B}}_p = \left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1,\eta}\mathbf{X}_{p-1,\eta}^{\dagger}\overline{\mathbf{C}}_p\right)^{-1}\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1,\eta}\mathbf{X}_{p-1,\eta}^{\dagger}$$

$$+\left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1,\eta}\mathbf{X}_{p-1,\eta}^{\dagger}\overline{\mathbf{C}}_p\right)^{-1}\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1,\eta}\mathbf{S}_{p-1}^{\dagger} \tag{3.55}$$

we observe that the signal component appears in the weight estimate. However, if the noise component is removed from the data. (ie: $\mathbf{X}_{p-1,\eta} = 0$), then

$$\hat{\mathbf{B}}_p = \mathbf{0}_{Q_p,MJ} \tag{3.56}$$

suggesting that partial cancellation of output power estimate occurs only in the presence of noise.

3.4 Mean Output Power (Type II)

For comparison purposes the mean output power of the Type II implementation is analyzed in this section. When signal and noise components remain stationary over long periods of time a Type II adaptation scheme is generally preferred over a Type I approach. One reason for this choice is the signal cancellation effect discussed previously. A second reason favoring a Type II implementation is the inherent delay which accompanies the processing of input data by Type I adaptation. Throughout this discussion the notation used in the preceding analysis is retained. Since the Type II implementation does not compute beamformer weights from the current input data block we introduce additional notation to distinguish current input data from data used to compute beamformer weights.

Let $\mathbf{Y}_{p-1}$ describe the $MJ \times L_p$ data block used to compute the p^{th} stage weights. Data vectors contained in $\mathbf{Y}_{p-1}$ are derived from the preceding input data block such that

$$\mathbf{Y}_{p-1}[k] = (\ \mathbf{Z}_{p-1}[k-m_p+1]\ \mathbf{Z}_{p-1}[k-m_p+2]\ \ldots\ \mathbf{Z}_{p-1}[k]\ ) \tag{3.57}$$

where the current input data block is defined as before by

$$\mathbf{X}_{p-1}[k] \;=\; (\; \mathbf{Z}_{p-1}[k]\; \mathbf{Z}_{p-1}[k+1]\; \ldots\; \mathbf{Z}_{p-1}[k+m_p-1])$$

Since there is no data overlap between $\mathbf{Y}_{p-1}[k]$ and $\mathbf{X}_{p-1}[k]$ these matrices are statistically independent. For the Type II beamformer the estimate of $\mathbf{B}_p$ is written as

$$\hat{\mathbf{B}}_p \;=\; \left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{Y}_{p-1}\mathbf{Y}_{p-1}^{\dagger}\overline{\mathbf{C}}_p\right)^{-1}\overline{\mathbf{C}}_p^{\dagger}\mathbf{Y}_{p-1}\mathbf{Y}_{p-1}^{\dagger} \tag{3.58}$$

Consider now the power estimate of the first stage, which we write as

$$\hat{\mathrm{P}}_1 \;=\; \frac{1}{L_1}\mathbf{w}_{\mathrm{q}}^{\dagger}\left(\mathbf{I}_{MJ}-\hat{\mathbf{B}}_1^{\dagger}\overline{\mathbf{C}}_1^{\dagger}\right)\mathbf{X}_0\mathbf{X}_0^{\dagger}\left(\mathbf{I}_{MJ}-\overline{\mathbf{C}}_1\hat{\mathbf{B}}_1\right)\mathbf{w}_{\mathrm{q}} \tag{3.59}$$

Observe that for this adaptation scheme the weight estimate $\hat{\mathbf{B}}_1$ and the input data block $\mathbf{X}_0$ are statistically independent. The expected value of (3.59) can then be written as

$$\mathcal{E}\left\{\hat{\mathrm{P}}_1\right\} \;=\; \mathbf{w}_{\mathrm{q}}^{\dagger}\mathcal{E}\left\{\left(\mathbf{I}_{MJ}-\hat{\mathbf{B}}_1^{\dagger}\overline{\mathbf{C}}_1^{\dagger}\right)\mathbf{R}_0\left(\mathbf{I}_{MJ}-\overline{\mathbf{C}}_1\hat{\mathbf{B}}_1\right)\right\}\mathbf{w}_{\mathrm{q}} \tag{3.60}$$

To proceed further we apply the following proposition.

Proposition 3.4 *Let* $\mathbf{C}$ *denote a non-random* $MJ\times Q$ *matrix and* $\mathbf{w}$ *a* $MJ\times 1$ *vector. Supposing that*

$$\mathbf{Y}\mathbf{Y}^{\dagger} \;\sim\; \mathrm{W}_{MJ}(L,\Sigma)$$

then the conditional distribution of

$$\hat{\mathbf{B}}\mathbf{w} \;=\; \left(\mathbf{C}^{\dagger}\mathbf{Y}\mathbf{Y}^{\dagger}\mathbf{C}\right)^{-1}\mathbf{C}^{\dagger}\mathbf{Y}\mathbf{Y}^{\dagger}\mathbf{w}$$

given

$$\mathbf{C}^{\dagger}\mathbf{Y}\mathbf{Y}^{\dagger}\mathbf{C}$$

is

$$\hat{\mathbf{B}}\mathbf{w} \mid \mathbf{C}^{\dagger}\mathbf{Y}\mathbf{Y}^{\dagger}\mathbf{C} \;\sim\; \mathrm{N}_Q\left(\;\mathbf{B}\mathbf{w},\; \mathrm{P}\left(\mathbf{C}^{\dagger}\mathbf{Y}\mathbf{Y}^{\dagger}\mathbf{C}\right)^{-1}\;\right)$$

where

$$\mathbf{Bw} \;=\; \left(\mathbf{C}^{\dagger}\mathbf{\Sigma C}\right)^{-1}\mathbf{C}^{\dagger}\mathbf{\Sigma w}$$

and

$$\mathtt{P} \;=\; \mathbf{w}^{\dagger}\mathbf{\Sigma w} \;-\; \mathbf{w}^{\dagger}\mathbf{\Sigma C}\left(\mathbf{C}^{\dagger}\mathbf{\Sigma C}\right)^{-1}\mathbf{C}^{\dagger}\mathbf{\Sigma w}$$

□

(See Appendix B.5 for a proof).

Since $\mathbf{Y}_0\mathbf{Y}_0^{\dagger}$ is Wishart distributed, the conditional distribution of $\hat{\mathbf{B}}_1\mathbf{w}_q$ is

$$\hat{\mathbf{B}}_1\mathbf{w}_q \mid \overline{\mathbf{C}}_1^{\dagger}\mathbf{Y}_0\mathbf{Y}_0^{\dagger}\overline{\mathbf{C}}_1 \;\sim\; \mathtt{N}_{Q_1}\left(\mathbf{B}_1\mathbf{w}_q,\; \mathtt{P}_1\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{Y}_0\mathbf{Y}_0^{\dagger}\overline{\mathbf{C}}_1\right)^{-1}\right) \tag{3.61}$$

where

$$\begin{aligned}\mathtt{P}_1 \;&=\; \mathbf{w}_q^{\dagger}\mathbf{R}_0\mathbf{w}_q \;-\; \mathbf{w}_q^{\dagger}\mathbf{R}_0\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{R}_0\overline{\mathbf{C}}_1\right)^{-1}\overline{\mathbf{C}}_1^{\dagger}\mathbf{R}_0\mathbf{w}_q \\ &=\; \mathbf{w}_q^{\dagger}\mathbf{R}_1\mathbf{w}_q\end{aligned}$$

by Proposition 3.4. Based on (3.61), (3.60) is rewritten in the form

$$\begin{aligned}\mathcal{E}\left\{\hat{\mathtt{P}}_1\right\} \;&=\; \mathbf{w}_q^{\dagger}\left(\mathbf{I}_{MJ} \;-\; \mathbf{B}_1^{\dagger}\overline{\mathbf{C}}_1^{\dagger}\right)\mathbf{R}_0\left(\mathbf{I}_{MJ} \;-\; \overline{\mathbf{C}}_1\mathbf{B}_1\right)\mathbf{w}_q \\ &\quad + \mathcal{E}\left\{\mathcal{E}\left\{\mathbf{w}_q^{\dagger}\left(\hat{\mathbf{B}}_1^{\dagger} - \mathbf{B}_1^{\dagger}\right)\overline{\mathbf{C}}_1^{\dagger}\mathbf{R}_0\overline{\mathbf{C}}_1\left(\hat{\mathbf{B}}_1 - \mathbf{B}_1\right)\mathbf{w}_q \mid \overline{\mathbf{C}}_1^{\dagger}\mathbf{Y}_0\mathbf{Y}_0^{\dagger}\overline{\mathbf{C}}_1\right\}\right\} \\ &=\; \mathtt{P}_1 + \mathtt{tr}\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{R}_0\overline{\mathbf{C}}_1\mathcal{E}\left\{\mathtt{cov}\left\{\hat{\mathbf{B}}_1\mathbf{w}_q \mid \overline{\mathbf{C}}_1^{\dagger}\mathbf{Y}_0\mathbf{Y}_0^{\dagger}\overline{\mathbf{C}}_1\right\}\right\}\right)\end{aligned} \tag{3.62}$$

Noting from (3.61) that

$$\mathtt{cov}\left\{\; \hat{\mathbf{B}}_1\mathbf{w}_q \mid \overline{\mathbf{C}}_1^{\dagger}\mathbf{Y}_0\mathbf{Y}_0^{\dagger}\overline{\mathbf{C}}_1 \;\right\} \;=\; \mathtt{P}_1\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{Y}_0\mathbf{Y}_0^{\dagger}\overline{\mathbf{C}}_1\right)^{-1}$$

$\mathcal{E}\left\{\hat{\mathtt{P}}_1\right\}$ simplifies to

$$\mathcal{E}\left\{\hat{\mathtt{P}}_1\right\} = \mathtt{P}_1 + \mathtt{P}_1\mathtt{tr}\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{R}_0\overline{\mathbf{C}}_1\mathcal{E}\left\{\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{Y}_0\mathbf{Y}_0^{\dagger}\overline{\mathbf{C}}_1\right)^{-1}\right\}\right) \tag{3.63}$$

where

$$\mathcal{E}\left\{\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{Y}_0\mathbf{Y}_0^{\dagger}\overline{\mathbf{C}}_1\right)^{-1}\right\} = \frac{1}{L_1 - Q_1}\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{R}_0\overline{\mathbf{C}}_1\right)^{-1}$$

by ([49], pp. 330). The mean output power of stage one can then be expressed in terms of L_1 and Q_1 as

$$\begin{aligned}\mathcal{E}\left\{\hat{\mathsf{P}}_1\right\} &= \mathsf{P}_1 + \frac{1}{L_1 - Q_1}\mathsf{P}_1\mathtt{tr}\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{R}_0\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{R}_0\overline{\mathbf{C}}_1\right)^{-1}\right)\\ &= \mathsf{P}_1 + \frac{Q_1}{L_1 - Q_1}\mathsf{P}_1\\ &= \frac{1}{1-\alpha_1}\mathsf{P}_1 \qquad (3.64)\end{aligned}$$

In deriving (3.64), analysis hinged on

$$\mathbf{Y}_0\mathbf{Y}_0^{\dagger}$$

being Wishart distributed in order for the random vector

$$\hat{\mathbf{B}}_1\mathbf{w}_q \mid \overline{\mathbf{C}}_1^{\dagger}\mathbf{Y}_0\mathbf{Y}_0^{\dagger}\overline{\mathbf{C}}_1$$

to be Gaussian distributed. Turning to the weight estimate of the second stage, ie:

$$\hat{\mathbf{B}}_2 = \left(\overline{\mathbf{C}}_2^{\dagger}\mathbf{Y}_1\mathbf{Y}_1^{\dagger}\overline{\mathbf{C}}_2\right)^{-1}\overline{\mathbf{C}}_2^{\dagger}\mathbf{Y}_1\mathbf{Y}_1^{\dagger} \qquad (3.65)$$

observe that $\mathbf{Y}_1\mathbf{Y}_1^{\dagger}$ is not Wishart distributed. This is a consequence of the input data block being statistically independent of the adaptive stage weights. Hence, the direct application of Proposition 3.1 is precluded. To remedy this situation we consider the approximation

$$\mathbf{Z}_1[k]\mathbf{Z}_1^{\dagger}[k] \mid \overline{\mathbf{C}}_1^{\dagger}\mathbf{Y}_0[k]\mathbf{Y}_0^{\dagger}[k]\overline{\mathbf{C}}_1 \sim \mathtt{W}_{MJ}(L_1,\ \gamma_k\mathbf{R}_1) \qquad (3.66)$$

which we base on an empirical investigation, where

$$\gamma_k = 1 + \mathtt{tr}\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{R}_0\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{Y}_0[k]\mathbf{Y}_0^{\dagger}[k]\overline{\mathbf{C}}_1\right)^{-1}\right)$$

and

$$\mathbf{R}_1 = \frac{1}{L_1}\mathcal{E}\left\{\mathbf{Z}_1[k]\mathbf{Z}_1^\dagger[k]\right\}$$

Note that

$$\mathcal{E}\{\gamma_k\} = 1 + \frac{1}{L_1 - Q_1} = \frac{1}{1-\alpha_1} \tag{3.67}$$

where $\alpha_1 = Q_1/L_1$. Now to extend our approximation into the second stage, the distribution of

$$\hat{\mathbf{R}}_2 = \frac{1}{L_2}\mathbf{Y}_1[k]\mathbf{Y}_1^\dagger[k]$$

$$= \frac{1}{L_2}\sum_{\ell=1}^{m_2}\mathbf{Z}_1[k+\ell-1-m_2]\mathbf{Z}_1^\dagger[k+\ell-1-m_2] \tag{3.68}$$

is approximated by

$$\mathbf{Y}_1\mathbf{Y}_1^\dagger \mid \mathcal{C}_0 \sim \mathrm{W}_{MJ}(L_2,\ \Gamma_0\mathbf{R}_1) \tag{3.69}$$

where

$$\mathcal{C}_0 = \left(\overline{\mathbf{C}}_1^\dagger\mathbf{Y}_0[k-m_2]\mathbf{Y}_0^\dagger[k-m_2]\overline{\mathbf{C}}_1,\ \ldots.\ \overline{\mathbf{C}}_1^\dagger\mathbf{Y}_0[k-1]\mathbf{Y}_0^\dagger[k-1]\overline{\mathbf{C}}_1\right)$$

and

$$\Gamma_0 = \sum_{\ell=1}^{m_2}\gamma_{k-\ell}$$

From

$$\hat{\mathrm{P}}_2 = \frac{1}{L_2}\mathbf{w}_\mathbf{q}^\dagger\left(\mathbf{I}_{MJ} - \hat{\mathbf{B}}_2^\dagger\overline{\mathbf{C}}_2^\dagger\right)\mathbf{X}_1\mathbf{X}_1^\dagger\left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_2\hat{\mathbf{B}}_2\right)\mathbf{w}_\mathbf{q} \tag{3.70}$$

we note that since $\hat{\mathbf{B}}_2$ and $\mathbf{X}_1$ are statistically independent, $\mathcal{E}\{\hat{\mathrm{P}}_2\}$ can be rewritten as

$$\mathcal{E}\{\hat{\mathrm{P}}_2\} = \frac{1}{1-\alpha_1}\mathbf{w}_\mathbf{q}^\dagger\mathcal{E}\left\{\left(\mathbf{I}_{MJ} - \hat{\mathbf{B}}_2^\dagger\overline{\mathbf{C}}_2^\dagger\right)\mathbf{R}_1\left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_2\hat{\mathbf{B}}_2\right)\right\}\mathbf{w}_\mathbf{q} \tag{3.71}$$

where

$$\mathbf{R}_1 = \frac{1}{L_2}\mathcal{E}\left\{\mathbf{X}_1\mathbf{X}_1^\dagger\right\}$$

Under the approximation that

$$\mathbf{Y}_1\mathbf{Y}_1^\dagger \mid \mathcal{C}_0$$

is Wishart distributed, the distribution of

$$\hat{\mathbf{B}}_2\mathbf{w}_\mathbf{q} \mid \overline{\mathbf{C}}_2^\dagger\mathbf{Y}_1\mathbf{Y}_1^\dagger\overline{\mathbf{C}}_2, \mathcal{C}_0$$

will be

$$\hat{\mathbf{B}}_2\mathbf{w}_\mathbf{q}^\dagger \mid \overline{\mathbf{C}}_2^\dagger\mathbf{Y}_1\mathbf{Y}_1^\dagger\overline{\mathbf{C}}_2, \mathcal{C}_0 \sim \mathtt{N}_{Q_2}\left(\mathbf{B}_2\mathbf{w}_\mathbf{q},\ \Gamma_0\mathtt{P}_2\left(\overline{\mathbf{C}}_2^\dagger\mathbf{Y}_1\mathbf{Y}_1^\dagger\overline{\mathbf{C}}_2\right)^{-1}\right) \tag{3.72}$$

where

$$\begin{aligned}\mathtt{P}_2 &= \mathbf{w}_\mathbf{q}^\dagger\mathbf{R}_1\mathbf{w}_\mathbf{q} - \mathbf{w}_\mathbf{q}^\dagger\mathbf{R}_1\overline{\mathbf{C}}_2\left(\overline{\mathbf{C}}_2^\dagger\mathbf{R}_1\overline{\mathbf{C}}_2\right)^{-1}\overline{\mathbf{C}}_2^\dagger\mathbf{R}_1\mathbf{w}_\mathbf{q} \\ &= \mathbf{w}_\mathbf{q}^\dagger\mathbf{R}_2\mathbf{w}_\mathbf{q}\end{aligned}$$

from (3.69) and Proposition 3.4. By applying the same approach taken in the analysis of the first stage, we rewrite (3.71) in the form

$$\begin{aligned}\mathcal{E}\left\{\hat{\mathtt{P}}_2\right\} &= \frac{1}{1-\alpha_1}\left(\mathbf{w}_\mathbf{q}^\dagger\left(\mathbf{I}_{MJ} - \mathbf{B}_2^\dagger\overline{\mathbf{C}}_2^\dagger\right)\mathbf{R}_1\left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_2\mathbf{B}_2\right)\mathbf{w}_\mathbf{q}\right. \\ &\quad \left. + \frac{1}{1-\alpha_1}\left(\mathcal{E}\left\{\mathcal{E}\left\{\mathbf{w}_\mathbf{q}^\dagger\left(\hat{\mathbf{B}}_2^\dagger - \mathbf{B}_2^\dagger\right)\overline{\mathbf{C}}_2^\dagger\mathbf{R}_1\overline{\mathbf{C}}_2\left(\hat{\mathbf{B}}_2 - \mathbf{B}_2\right)\mathbf{w}_\mathbf{q} \mid \overline{\mathbf{C}}_2^\dagger\mathbf{Y}_1\mathbf{Y}_1^\dagger\overline{\mathbf{C}}_2, \mathcal{C}_0\right\}\right\}\right) \\ &= \frac{1}{1-\alpha_1}\left(\mathtt{P}_2 + \mathtt{tr}\left(\overline{\mathbf{C}}_2^\dagger\mathbf{R}_1\overline{\mathbf{C}}_2\mathcal{E}\left\{\mathtt{cov}\left\{\hat{\mathbf{B}}_2\mathbf{w}_\mathbf{q} \mid \overline{\mathbf{C}}_2^\dagger\mathbf{Y}_1\mathbf{Y}_1^\dagger\overline{\mathbf{C}}_2\right\}\right\}\right)\right)\end{aligned} \tag{3.73}$$

From (3.72)

$$\mathtt{cov}\left\{\ \hat{\mathbf{B}}_2\mathbf{w}_\mathbf{q} | \overline{\mathbf{C}}_2^\dagger\mathbf{Y}_1\mathbf{Y}_1^\dagger\overline{\mathbf{C}}_2, \mathcal{C}_0\ \right\} \approx \Gamma_0\mathtt{P}_2\left(\overline{\mathbf{C}}_2^\dagger\mathbf{Y}_1\mathbf{Y}_1^\dagger\overline{\mathbf{C}}_2\right)^{-1}$$

$\mathcal{E}\left\{\hat{\mathtt{P}}_2\right\}$ simplifies further to

$$\mathcal{E}\left\{\hat{\mathtt{P}}_2\right\} \approx \frac{1}{1-\alpha_1}\left(\mathtt{P}_2 + \mathtt{P}_2\mathtt{tr}\left(\overline{\mathbf{C}}_2^\dagger\mathbf{R}_1\overline{\mathbf{C}}_2\mathcal{E}\left\{\Gamma_0\left(\overline{\mathbf{C}}_2^\dagger\mathbf{Y}_1\mathbf{Y}_1^\dagger\overline{\mathbf{C}}_2\right)^{-1}\right\}\right)\right) \tag{3.74}$$

such that

$$\mathcal{E}\left\{\left(\overline{\mathbf{C}}_2^\dagger\mathbf{Y}_1\mathbf{Y}_1^\dagger\overline{\mathbf{C}}_2\right)^{-1} \mid \mathcal{C}_0\right\} \approx \Gamma_0^{-1}\frac{1}{L_2 - Q_2}\left(\overline{\mathbf{C}}_2^\dagger\mathbf{R}_1\overline{\mathbf{C}}_2\right)^{-1} \tag{3.75}$$

[50]. We then write the mean output power of the second beamformer stage as

$$\begin{aligned}\mathcal{E}\{\hat{\mathsf{P}}_2\} &\approx \frac{1}{1-\alpha_1}\left(\mathsf{P}_2 + \frac{1}{L_2-Q_2}\mathsf{P}_2\mathtt{tr}\left(\overline{\mathbf{C}}_2^{\dagger}\mathbf{R}_1\overline{\mathbf{C}}_2\left(\overline{\mathbf{C}}_2^{\dagger}\mathbf{R}_1\overline{\mathbf{C}}_2\right)^{-1}\right)\right) \\ &= \frac{1}{1-\alpha_1}\left(\mathsf{P}_2 + \frac{Q_2}{L_2-Q_2}\mathsf{P}_2\right) \\ &= \frac{1}{(1-\alpha_1)(1-\alpha_2)}\mathsf{P}_2 \qquad (3.76)\end{aligned}$$

As a further approximation consider

$$\mathcal{E}\{\hat{\mathsf{P}}_2\} \approx \frac{1}{(1-\alpha_1-\alpha_2)}\mathsf{P}_2 \qquad (3.77)$$

We extend our results to the $p^{\mathtt{th}}$ stage by applying an argument similar to that used in the preceding analysis of the second stage. We approximate the output power of the p^{th} stage by the expression

$$\mathcal{E}\{\hat{\mathsf{P}}_p\} \approx \frac{1}{(1-\sum_{r=1}^{p}\alpha_r)}\mathsf{P}_p \qquad (3.78)$$

where

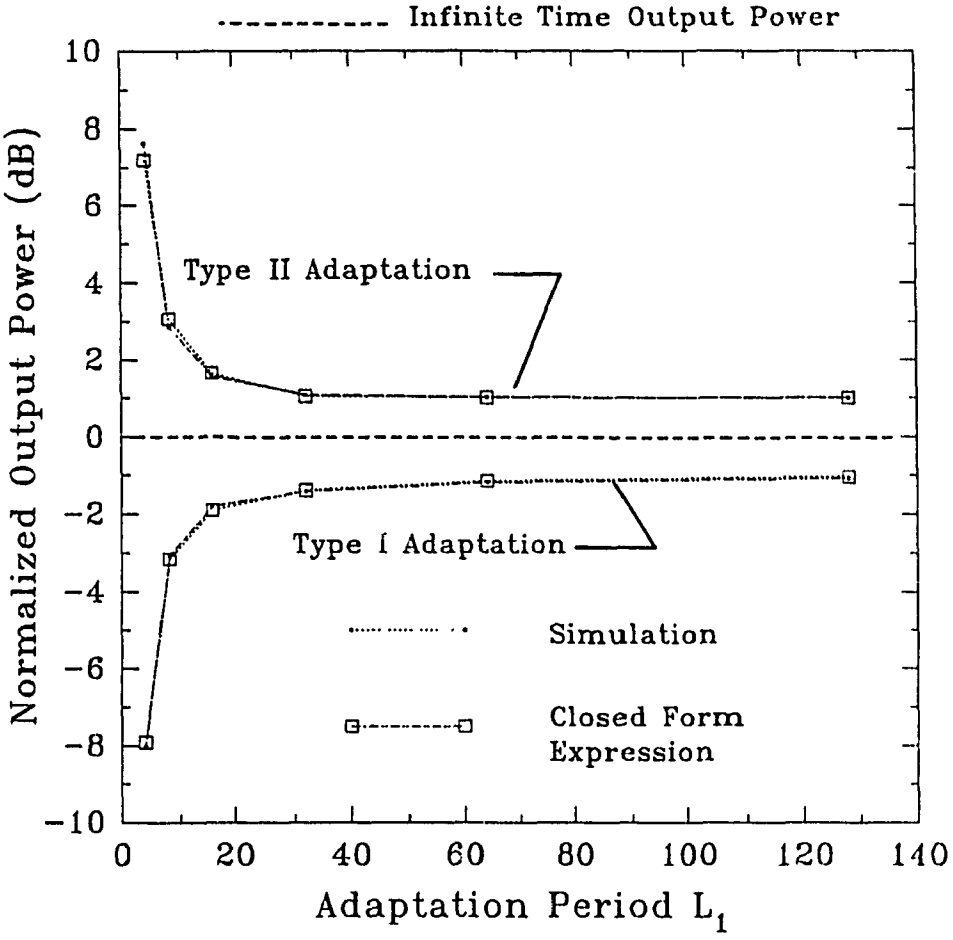
$$\mathsf{P}_p = \mathbf{w}_{\mathsf{q}}^{\dagger}\mathbf{R}_{p-1}\mathbf{w}_{\mathsf{q}} - \mathbf{w}_{\mathsf{q}}^{\dagger}\mathbf{R}_{p-1}\overline{\mathbf{C}}_p\left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{R}_{p-1}\overline{\mathbf{C}}_p\right)^{-1}\overline{\mathbf{C}}_p^{\dagger}\mathbf{R}_{p-1}\mathbf{w}_{\mathsf{q}}$$

and

$$\mathbf{R}_{p-1} = \mathbf{R}_{p-2} - \mathbf{R}_{p-2}\overline{\mathbf{C}}_{p-1}\left(\overline{\mathbf{C}}_{p-1}^{\dagger}\mathbf{R}_{p-2}\overline{\mathbf{C}}_{p-1}\right)^{-1}\overline{\mathbf{C}}_{p-1}^{\dagger}\mathbf{R}_{p-2}$$

for $p > 1$.

To verify the accuracy of (3.78) we numerically compare this expression with the mean output power estimates obtained through Monte Carlo simulation. The simulation experiment is identical to that used previously for the Type I beamformer. Our array consists of 12 sensors followed by a two stage M^2VDR beamformer. Input data vectors to the first stage are taken to be i.i.d.. Measured data consists of a source component scaled to 5dB and spatial white noise uncorrelated with the source. Figure 3.2 shows the mean output power level obtained from (3.78) and by simulation for various values of L_1, L_2, Q_1 and Q_2. As indicated by this figure our approximation of $\mathcal{E}\{\hat{\mathsf{P}}_2\}$ is effectively identical with the experimental results.



Mean Output Power – Type II Implementation

Figure 3.2

3.5 Mean-Squared-Error (Type II)

Unlike a Type I implementation, the Type II M^{2}VDR beamformer is not characterized by a signal cancellation effect. To demonstrate this, recall that the output of the Type II p^{th} stage is equal to

$$\mathbf{Z}_p = \mathbf{X}_{p-1} - \mathbf{Y}_{p-1}\mathbf{Y}_{p-1}^{\dagger}\overline{\mathbf{C}}_p\left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{Y}_{p-1}\mathbf{Y}_{p-1}^{\dagger}\overline{\mathbf{C}}_p\right)^{-1}\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1} \tag{3.79}$$

where $\hat{\mathbf{B}}_p$ and $\mathbf{X}_{p-1}$ are uncorrelated. As a means of verifying that the signal does not experience partial cancellation we evaluate the conditional expectation

$$\mathcal{E}\{\mathbf{Z}_p \mid \mathbf{S}_{p-1}\}$$

where $\mathbf{S}_{p-1}$ is the desired signal component of $\mathbf{X}_{p-1}$. From (3.79) we have

$$\begin{aligned}
\mathcal{E}\{\mathbf{Z}_p \mid \mathbf{S}_{p-1}\} &= \mathcal{E}\{\mathbf{X}_{p-1}|\mathbf{S}_{p-1}\} - \mathcal{E}\left\{\mathbf{Y}_{p-1}\mathbf{Y}_{p-1}^{\dagger}\overline{\mathbf{C}}_p\left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{Y}_{p-1}\mathbf{Y}_{p-1}^{\dagger}\overline{\mathbf{C}}_p\right)^{-1}\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1}|\mathbf{S}_{p-1}\right\} \\
&= \mathbf{S}_{p-1} - \mathcal{E}\left\{\mathbf{Y}_{p-1}\mathbf{Y}_{p-1}^{\dagger}\overline{\mathbf{C}}_p\left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{Y}_{p-1}\mathbf{Y}_{p-1}^{\dagger}\overline{\mathbf{C}}_p\right)^{-1}\overline{\mathbf{C}}_p^{\dagger}\right\}\mathbf{S}_{p-1} \\
&= \mathbf{S}_{p-1} \qquad (3.80)
\end{aligned}$$

since $\overline{\mathbf{C}}_p^{\dagger}\mathbf{S}_{p-1} = \mathbf{0}_{Q_p,L_p}$.

Without the need to correct for signal cancellation, MSE in the p^{th} stage is determined from

$$\begin{aligned}
\varepsilon_p &= \frac{1}{L_p}\mathcal{E}\left\{\left|\,\mathbf{w}_{\mathbf{q}}^{\dagger}(\mathbf{Z}_p - \mathbf{S}_{p-1})\,\right|^2\right\} \\
&= \frac{1}{L_p}\left(\mathcal{E}\left\{\mathbf{w}_{\mathbf{q}}^{\dagger}\mathbf{Z}_p\mathbf{Z}_p^{\dagger}\mathbf{w}_{\mathbf{q}} - \mathbf{w}_{\mathbf{q}}^{\dagger}\mathbf{S}_{p-1}\mathbf{S}_{p-1}^{\dagger}\mathbf{w}_{\mathbf{q}}\right\}\right) \\
&= \mathcal{E}\{\hat{\mathsf{P}}_p\} - \mathsf{P}_s \qquad (3.81)
\end{aligned}$$

Substituting (3.78) into (3.81), an approximate expression for MSE is given by

$$\varepsilon_p \approx \frac{1}{1-\sum_{\ell=1}^{p}\alpha_\ell}\mathsf{P}_{p,\mathsf{n}} + \frac{\sum_{\ell=1}^{p}\alpha_\ell}{1-\sum_{\ell=1}^{p}\alpha_\ell}\mathsf{P}_{\mathsf{s}} \tag{3.82}$$

where $\mathsf{P}_{p,\mathsf{n}}$ and P_{s} are defined as in (3.51).

Notice from (3.82) that as adaptation time scale is reduced, or additional adaptive DOF assigned to a stage, MSE increases. In contrast, a decrease in adaptation time scale or an increase in the number of adaptive DOF leads to a reduction in the desired signal power and a subsequent decrease in MSE for a Type I beamformer. To resolve the apparent difference in performance between the Type I and II beamformer implementations output SNR ($\mathsf{SNR}_{\mathsf{out}}$) is examined in the next section.

3.6 Output SNR

Comparing the MSE expressions derived for the Type I and II beamformers, we are left with the impression that the implementations differ substantially in detection performance for the same assignment of adaptive DOF and adaptation time. To resolve this apparent difference in performance output SNR is examined. For the Type I and II GSC implementations, it is shown in [42] that both adaptation schemes achieve identical $\mathsf{SNR}_{\mathsf{out}}$ performance even though the respective MSE of each beamformer differs. Employing the same definition of $\mathsf{SNR}_{\mathsf{out}}$ as given in [12], we may show as well that the performance of the Type I and II M^{2}VDR implementations are equal with respect to $\mathsf{SNR}_{\mathsf{out}}$.

The $\mathsf{SNR}_{\mathsf{out}}$ of a beamformer is described by the ratio of mean output power of the desired signal to mean output noise power. Following [42], $\mathsf{SNR}_{\mathsf{out}}$ for the p^{th} stage of a Type I beamformer is redefined as

$$\mathsf{SNR}_{\mathsf{out}}(p) = \frac{\mathcal{E}\left\{\left|\mathcal{E}\left\{\mathbf{w}_{\mathsf{q}}^{\dagger}\mathbf{Z}_p \mid \mathbf{S}_{p-1}\right\}\right|^2\right\}}{\mathcal{E}\left\{\left|\mathbf{w}_{\mathsf{q}}^{\dagger}\left(\mathbf{Z}_p - \mathcal{E}\left\{\mathbf{Z}_p \mid \mathbf{S}_{p-1}\right\}\right)\right|^2\right\}} \tag{3.83}$$

to account for the signal cancellation effect. From our analysis of MSE for the Type I structure (3.83) can be expressed in the form

$$\begin{aligned}
\mathtt{SNR}_{\mathtt{out}}(p) &= \frac{\left(1-\sum_{\ell=1}^{p}\alpha_\ell\right)^2 \mathtt{P_s}}{\varepsilon_p} \\
&= \frac{\left(1-\sum_{\ell=1}^{p}\alpha_\ell\right)^2 \mathtt{P_s}}{\left(1-\sum_{\ell=1}^{p}\alpha_\ell\right)\mathtt{P}_{p,\mathtt{n}}+\sum_{\ell}^{p}\alpha_\ell\left(1-\sum_{\ell=1}^{p}\alpha_\ell\right)\mathtt{P_s}} \\
&= \frac{\left(1-\sum_{\ell=1}^{p}\alpha_\ell\right)\overline{\mathtt{SNR}}_{\mathtt{out}}(p)}{1+\sum_{\ell=1}^{p}\alpha_\ell\,\overline{\mathtt{SNR}}_{\mathtt{out}}(p)}
\end{aligned} \tag{3.84}$$

where $\overline{\mathtt{SNR}}_{\mathtt{out}}(p)$ corresponds to the infinite time output SNR as defined by

$$\overline{\mathtt{SNR}}_{\mathtt{out}}(p) = \frac{\mathtt{P_s}}{\mathtt{P}_{p,\mathtt{n}}}$$

Turning to the p^{th} stage of a Type II implementation, from (3.82) output SNR for the stage can be expressed as

$$\begin{aligned}
\mathtt{SNR}_{\mathtt{out}}(p) &= \frac{P_{\mathtt{s}}}{\left(1-\sum_{\ell=1}^{p}\alpha_\ell\right)^{-1}\mathtt{P}_{p,\mathtt{n}}+\sum_{\ell=1}^{p}\alpha_\ell\left(1-\sum_{\ell=1}^{p}\alpha_\ell\right)^{-1}\mathtt{P_s}} \\
&= \frac{\left(1-\sum_{\ell=1}^{p}\alpha_\ell\right)\overline{\mathtt{SNR}}_{\mathtt{out}}(p)}{1+\sum_{\ell=1}^{p}\alpha_\ell\overline{\mathtt{SNR}}_{\mathtt{out}}(p)}
\end{aligned} \tag{3.85}$$

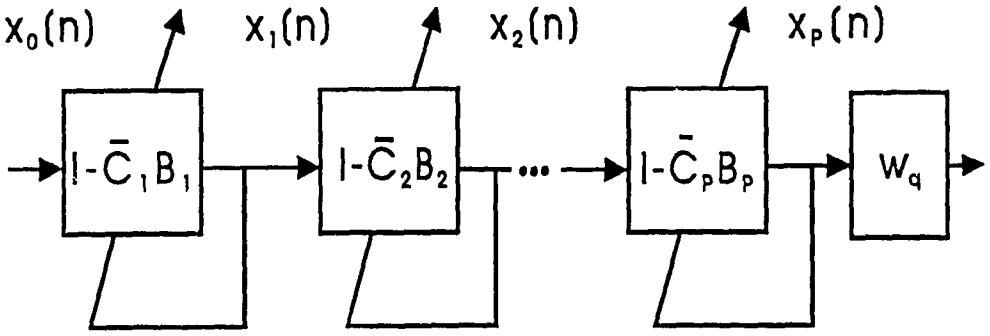
Comparison of (3.84) and (3.85) leads us to conclude, as in the case of the GSC [12], that the Type I and II multiple stage M^{2}VDR beamformer implementations are equivalent with respect to output SNR.

Chapter 4

Beamformer Design and Multiple Time-Scale Adaptation

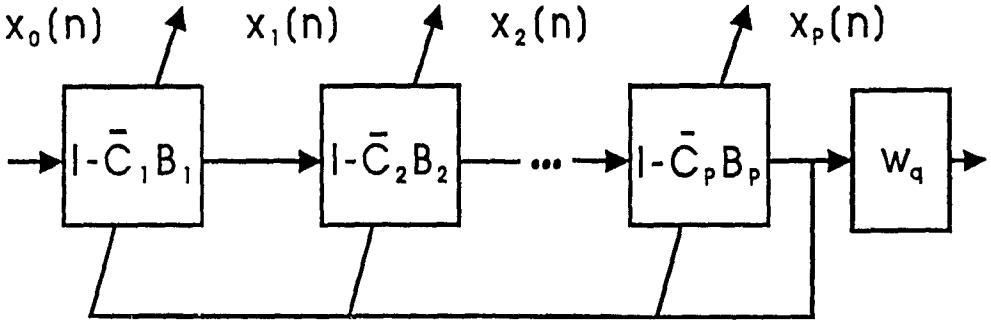
A unique feature of the M^{2}VDR beamformer is the flexibility of distributing adaptive DOF among stages arranged in cascade. This property enables simultaneous adaptation of DOF over different time-scales. The ability to adapt on multiple time scales is a particularily useful feature in combating mixed time-scale interferers. For instance, to cancel short duration interference, a multiple stage beamformer computes a small number of weights assigned to a stage on a time-scale approximating the duration of the interferer. With fewer weights, the beamformer stage introduces the possibility of larger reductions in adaptation time then could be possible for a GSC implementation. Furthermore, the M^{2}VDR realization does not curtail the ability to respond to multiple interferers since the beamformer applies the remaining DOF, retained in other stages, against long duration interference. However, in order for the beamformer to be effective, attention must be paid to the choice of adaptation scheme and the columns of the signal blocking matrices assigned to individual stages.

From the derivation of the M^{2}VDR beamformer, (2.21), we see that adaptive weights are computed from the output of individual stages as shown in Figure 1.1a, and not from the final output of the beamformer, as would be the case in Figure



Stages adapted on individual beamformer stage outputs

FIGURE 4.1a



Beamformer stages adapted on final output

Figure 4.1b

Adaptation of the M²VDR Beamformer
Figure 4.1

4.1b. Assuming an infinite observation period, stage weights derived locally as shown in Figure 4.1a, minimize MSE at the final output of the beamformer [38]. Furthermore, prior knowledge of the input data covariance matrix, $\mathbf{R}_0$ allows stages to be arbitrarily arranged without affecting final MSE performance. This is not the case under time varying conditions. For the adaptive beamformer, the ability to derive weights solely from the output of individual stages is the key factor that allows stages to be adapted on different adaptation periods. Otherwise, if each stage adapts from the final beamformer output, weight estimates become dependent on the longest time scale employed by the beamformer.

When all interference components of the wavefield are slowly time varying, applying either a Type I or II adaptation implementation does not affect final output SNR. A Type II implementation is generally preferred under slowly time varying conditions to prevent a delay in updating the output of the beamformer. The presence of short duration interference causes Type II adaptation to place a null in the beamformer response after the event has passed. This situation motivates us to consider Type I adaptation to synchronize the beamformer response with the current input data sample. In the first part of this chapter we examine the issue of stage ordering under Type I adaptation.

By decomposing a GSC into a M^{2}VDR realization each stage of the network corresponds to a partially adaptive beamformer operating on fewer than $MJ - K$ adaptive DOF. Partially adaptive beamformer realizations have in the past been used to reduce the number of adaptive DOF. This has the effect of offsetting the computational burden imposed by large arrays on computing beamformer weights. (eg: [35, 36, 39]). In reducing the number of adaptive DOF available for interference cancellation care must be exercised to ensure that a significant percentage of interferer power is passed through the signal blocking operation. In other words, the design of the signal blocking matrix should maximize the intersection between the interference subspace and the span of the signal blocking matrix. A similar design problem faces

the multiple time scale beamformer. In the second part of this chapter we consider how performance is affected by the choice of signal blocking matrix.

4.1 Minimum Mean-Squares Error and Stage Adaptation

From the derivation of the M^{2}VDR beamformer, the unconstrained minimization

$$\min_{\mathbf{B}} \left(\mathbf{w}_q^\dagger \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}\mathbf{B} \right)^\dagger \mathbf{R}_0 \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}\mathbf{B} \right) \mathbf{w}_q \right)$$

can be decomposed into P minimizations of the form

$$\min_{\mathbf{B}_\ell} \mathrm{tr} \left(\left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_\ell \mathbf{B}_\ell \right)^\dagger \mathbf{R}_{\ell-1} \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_\ell \mathbf{B}_\ell \right) \right) \tag{4.1}$$

for $\ell = 1, 2, \ldots, P$. As suggested by (4.1), the formulation of $\mathbf{B}_p$ is dependent only on the output of the p^{th} stage and not on the final beamformer output. In this section we re-examine this property of the M^{2}VDR beamformer when stage weights are adaptively estimated.

Consider a deterministic least squares (LS) solution for the adaptive weights of each stage based on Type I adaptation. The global optimality of the p^{th} stage weights with respect to the final output error of the beamformer depends on the orthogonality condition underlying the minimum mean-squares error (MMSE) solution [51, 52]. A definition of the orthogonality condition that is necessary and sufficient for the multiple stage beamformer under Type I adaptation follows.

Let the output data block for the p^{th} stage equal

$$\begin{aligned} \mathbf{Z}_p &= \left(\mathbf{I}_{MJ} - \hat{\mathbf{B}}_p^\dagger \overline{\mathbf{C}}_p^\dagger \right) \mathbf{X}_{p-1} \\ &= \mathbf{X}_{p-1} - \hat{\mathbf{B}}_p^\dagger \tilde{\mathbf{X}}_p \end{aligned} \tag{4.2}$$

where

$$\overline{\mathbf{C}}_p^\dagger \mathbf{X}_{p-1} = \tilde{\mathbf{X}}_p$$

At the p^{th} stage, the local MMSE solution requires the output of the stage to be orthogonal to the noise component $\tilde{\mathbf{X}}_p$, ie:

$$\left(\mathbf{X}_{p-1} - \hat{\mathbf{B}}_p^{\dagger}\tilde{\mathbf{X}}_p \right) \tilde{\mathbf{X}}_p^{\dagger} = \mathbf{Z}_p\tilde{\mathbf{X}}_p^{\dagger} = \mathbf{0} \tag{1.3}$$

To ensure that the computation of $\hat{\mathbf{B}}_p$ is independent of succeeding stages, we further require that

$$\left(\mathbf{X}_{\ell-1} - \hat{\mathbf{B}}_\ell^{\dagger}\tilde{\mathbf{X}}_\ell\right) \tilde{\mathbf{X}}_p^{\dagger} = \mathbf{Z}_\ell\tilde{\mathbf{X}}_p^{\dagger} = \mathbf{0} \tag{1.4}$$

for $\ell > p$, where $\hat{\mathbf{B}}_\ell$ follows from (3.5). When (1.4) is satisfied we say that global MMSE has been achieved. To elaborate, suppose

$$L_{\max} = \max(L_p,\ p = 1,\ 2,\ \ldots\ , P)$$

where

$$\mathbf{X}_{0,\max}[k] = (\mathbf{x}_0[1 + (k-1)L_{\max}],\ \mathbf{x}_0[2 + (k-1)L_{\max}],\ \ldots.\ \mathbf{x}_0[kL_{\max}])$$

represents the total number of sensor output vectors that would be processed by the P stage beamformer over any $L_{\max}$ sample interval. The interference and noise component of $\mathbf{X}_{0,\max}$ follows from

$$\hat{\mathbf{X}}_{0,\max} = \overline{\mathbf{C}}^{\dagger}\mathbf{X}_{\max}$$

where $\overline{\mathbf{C}} = \left(\overline{\mathbf{C}}_1,\ \overline{\mathbf{C}}_2,\ \ldots,\ \overline{\mathbf{C}}_P\right)$. If

$$\mathbf{X}_{P,\max}[k] = (\mathbf{x}_P[1 + (k-1)L_{\max}],\ \mathbf{x}_0[2 + (k-1)L_{\max}],\ \ldots.\ \mathbf{x}_0[kL_{\max}])$$

is the output of the P^{th} stage due to $\mathbf{X}_{P,\max}[k]$ the best linear estimate of the signal component

$$\mathbf{S}_{0,\max} = \mathbf{C}^{\dagger}\mathbf{X}_{\max}$$

is made by the beamformer when

$$\hat{\mathbf{X}}_{0,\max}^{\dagger}[k]\mathbf{X}_{P,\max}[k] = 0$$

Otherwise, if the above condition is not met then the global MMSE solution can only be reached by adapting each stage from the final beamformer output. When this happens all stages are effectively adapting on same time scale, ie: $L_{\max}$.

In this chapter, to simplify notation we remove indices on the zero matrix $\mathbf{0}_{MN}$.

4.1.1 Adaptation Period: $L_\ell = L,\ \ell = 1, \ldots, P$

Correlation between the input data block and the estimated weights characterizes the Type I implementation. It is this feature of Type I adaptation that introduces the possibility of global MMSE. Suppose that the same adaptation period is assigned to all stages of the beamformer,

$$L_\ell = L, \quad \text{for} \quad \ell = 1, 2, \ldots, P$$

We apply (1.3) and (1.4) as requisite conditions for global MMSE.

Observe for the first stage that

$$\begin{aligned} \mathbf{Z}_1 \tilde{\mathbf{X}}_1^\dagger &= \mathbf{X}_0 \mathbf{X}_0^\dagger \overline{\mathbf{C}}_1 - \mathbf{X}_0 \mathbf{X}_0^\dagger \overline{\mathbf{C}}_1 \left(\overline{\mathbf{C}}_1^\dagger \mathbf{X}_0 \mathbf{X}_0^\dagger \overline{\mathbf{C}}_1 \right)^{-1} \overline{\mathbf{C}}_1^\dagger \mathbf{X}_0 \mathbf{X}_0^\dagger \overline{\mathbf{C}}_1 \\ &= \mathbf{0} \end{aligned} \tag{4.5}$$

Hence,

$$\begin{aligned} \left(\mathbf{X}_1 - \hat{\mathbf{B}}_2^\dagger \tilde{\mathbf{X}}_2 \right) \tilde{\mathbf{X}}_1^\dagger &= \left(\mathbf{I}_{MJ} - \hat{\mathbf{B}}_2^\dagger \overline{\mathbf{C}}_2^\dagger \right) \mathbf{X}_1 \tilde{\mathbf{X}}_1^\dagger \\ &= \mathbf{0} \end{aligned} \tag{4.6}$$

since $\mathbf{X}_1 = \mathbf{Z}_1$. Proceeding by induction

$$\begin{aligned} \left(\mathbf{X}_{\ell-1} - \hat{\mathbf{B}}_\ell^\dagger \tilde{\mathbf{X}}_\ell \right) \tilde{\mathbf{X}}_1^\dagger &= \left(\mathbf{I}_{MJ} - \hat{\mathbf{B}}_\ell^\dagger \overline{\mathbf{C}}_\ell^\dagger \right) \mathbf{X}_{\ell-1} \tilde{\mathbf{X}}_1^\dagger \\ &= \mathbf{0} \end{aligned} \tag{4.7}$$

for $\ell = 2, \ldots, P$.

At the second stage we find that

$$\mathbf{Z}_2\tilde{\mathbf{X}}_2^\dagger \;=\; \mathbf{0} \tag{4.8}$$

and

$$\begin{aligned}\left(\mathbf{X}_{\ell-1} - \hat{\mathbf{B}}_\ell^\dagger\tilde{\mathbf{X}}_\ell\right)\tilde{\mathbf{X}}_2^\dagger \;&=\; \mathbf{Z}_\ell\tilde{\mathbf{X}}_2^\dagger \\ &=\; \mathbf{0}\end{aligned} \tag{4.9}$$

for $\ell = 3, \ldots, P$.

Proceeding to the p^{th} stage we have

$$\left(\mathbf{X}_{\ell-1} - \hat{\mathbf{B}}_\ell^\dagger\tilde{\mathbf{X}}_\ell\right)\tilde{\mathbf{X}}_p^\dagger = \mathbf{0} \tag{4.10}$$

for $\ell = p+1, \ldots, P-1$. As a consequence of (4.10), weights computed by Type I adaptation when $L_\ell = L$, $\ell = 1, \ldots, P$ result in global MMSE.

4.1.2 Adaptation Period: $L_\ell \geq L_p$, $\ell \geq p$

This section examines MMSE when stages are adapted over different time-scales. Suppose

$$L_p \leq L_\ell, \qquad \text{for } \ell > p$$

In this case we observe that

$$\mathbf{Z}_1\tilde{\mathbf{X}}_1^\dagger \;=\; \mathbf{0}$$

holds. To proceed further define $\mathbf{X}_{0,\ell}$ as the $MJ \times m_\ell L_1$ block matrix

$$\mathbf{X}_{0,\ell-1} \;=\; (\; \mathbf{X}_0[k]\; \mathbf{X}_0[k+1]\; \ldots\; \mathbf{X}_0[k+m_\ell-1]\;) \tag{4.11}$$

Since

$$\mathbf{X}_1\mathbf{X}_{0,1}^\dagger\overline{\mathbf{C}}_1 \;=\; \sum_{j=1}^{m_\ell}\mathbf{Z}_1[k+j]\mathbf{X}_0^\dagger[k+j]\overline{\mathbf{C}}_1 \;=\; \mathbf{0} \tag{4.12}$$

where

$$\mathbf{X}_1 \;=\; (\; \mathbf{Z}_1[k]\; \mathbf{Z}_1[k+1]\; \ldots\; \mathbf{Z}_1[k+m_p-1]\;)$$

and

$$\mathbf{Z}_1[k+\ell]\mathbf{X}_0^\dagger[k+\ell]\overline{\mathbf{C}}_1 \;=\; \mathbf{0}$$

the relationship between $\tilde{\mathbf{X}}$ and the second stage is

$$\begin{aligned}\left(\mathbf{X}_1 - \hat{\mathbf{B}}_2^\dagger\tilde{\mathbf{X}}_2\right)\mathbf{X}_{0,1}^\dagger\overline{\mathbf{C}}_1 &= \mathbf{X}_1\mathbf{X}_1^\dagger\overline{\mathbf{C}}_2\left(\overline{\mathbf{C}}_2^\dagger\mathbf{X}_1\mathbf{X}_1^\dagger\overline{\mathbf{C}}_2\right)^{-1}\overline{\mathbf{C}}_2^\dagger\mathbf{X}_1\mathbf{X}_{0,2}^\dagger\overline{\mathbf{C}}_1 \\ &= \mathbf{0} \end{aligned} \tag{4.13}$$

A similar argument can be used to show that

$$\left(\mathbf{X}_{\ell-1} - \hat{\mathbf{B}}_\ell^\dagger\tilde{\mathbf{X}}_\ell\right)\mathbf{X}_{0,1}^\dagger\overline{\mathbf{C}}_1 \;=\; \mathbf{0}, \qquad \text{for } \ell > 1 \tag{4.14}$$

Extending our analysis to the p^{th} stage

$$\left(\mathbf{X}_{\ell-1} - \hat{\mathbf{B}}_\ell^\dagger\tilde{\mathbf{X}}_\ell\right)\mathbf{X}_{0,p}^\dagger\overline{\mathbf{C}}_p = \mathbf{0} \tag{4.15}$$

for $\ell > p$, where

$$\mathbf{X}_{p-1,\ell} \;=\; (\;\mathbf{Z}_{p-1}[k]\;\mathbf{Z}_{p-1}[k+1]\;\ldots\;\mathbf{Z}_{p-1}[k+m_\ell - m_p]\;)$$

Thus for $L_\ell \geq L_p$, $\ell \geq p$ stage weights satisfy the global MMSE criteria.

4.1.3 Adaptation Period: $L_\ell < L_p$, $\ell > p$

Assume now that the stages are rearranged so that

$$L_p \;>\; L_\ell, \quad \text{for} \quad \ell > p$$

To adapt our notation to this ordering of stages we partition $\mathbf{X}_{p-1}[k]$ and $\mathbf{Z}_p[k]$ into

$$1 + m_p - m_{p+1}$$

non-overlapping data blocks, so that

$$\mathbf{X}_{p-1}[k] \;=\; \left(\;\mathbf{X}_{(p-1,1)}[k]\;\mathbf{X}_{(p-1,2)}[k]\;\ldots\;\mathbf{X}_{(p-1,1+m_p-m_{p+1})}[k]\;\right) \tag{4.16}$$

and

$$\mathbf{Z}_p[k] = \left(\mathbf{X}_{(p,1)}[k]\ \mathbf{X}_{(p,2)}[k]\ \cdots\ \mathbf{X}_{(p,1+m_p-m_{p+1})}[k] \right) \tag{1.17}$$

where $\mathbf{X}_{(p-1,\ell)}[k]$ and $\mathbf{X}_{(p,\ell)}[k]$ are $MJ \times L_{p+1}$ matrices. Notice that $\mathbf{X}_{(p,1)}[k]$ corresponds to the input data block to the $p^{\text{th}} + 1$ stage. Furthermore, let

$$\begin{aligned} \hat{\mathbf{X}}_p[k] &= \left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{(p-1,1)}[k]\ \overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{(p-1,2)}[k]\ \cdots\ \overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{(p-1,1+m_p-m_{p+1})}[k] \right) \\ &= \left(\hat{\mathbf{X}}_{(p,1)}[k]\ \hat{\mathbf{X}}_{(p,2)}[k]\ \cdots\ \hat{\mathbf{X}}_{(p,1+m_p-m_{p+1})}[k] \right) \end{aligned} \tag{1.18}$$

At the p^{th} stage,

$$\begin{aligned} \mathbf{Z}_p &= \mathbf{X}_p - \hat{\mathbf{B}}_p^{\dagger}\hat{\mathbf{X}}_p \\ &= \mathbf{X}_{p-1} - \mathbf{X}_{p-1}\mathbf{X}_{p-1}^{\dagger}\overline{\mathbf{C}}_p \left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1}\mathbf{X}_{p-1}^{\dagger}\overline{\mathbf{C}}_p \right)^{-1} \overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p-1} \end{aligned}$$

we have

$$\mathbf{Z}_p\hat{\mathbf{X}}_p^{\dagger} = \mathbf{0} \tag{1.19}$$

Consider now the output of the $p^{\text{th}} + 1$ stage, which we write as

$$\mathbf{Z}_{p+1} = \left(\mathbf{Z}_{p+1,1},\ \mathbf{Z}_{p+1,2},\ \ldots,\ \mathbf{Z}_{p+1,1+m_p-m_{p+1}} \right) \tag{1.20}$$

where

$$\mathbf{Z}_{p+1,\ell} = \mathbf{X}_{p,\ell} - \mathbf{X}_{p,\ell}\mathbf{X}_{p,\ell}^{\dagger}\overline{\mathbf{C}}_p \left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p,\ell}\mathbf{X}_{p,\ell}^{\dagger}\overline{\mathbf{C}}_p \right)^{-1} \overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p,\ell}$$

We then express the product $\mathbf{Z}_{p+1}\hat{\mathbf{X}}_p^{\dagger}$ by

$$\mathbf{Z}_{p+1}\hat{\mathbf{X}}_p^{\dagger} = \sum_{\ell=1}^{1+m_p-m_{p+1}} \left(\mathbf{X}_{p,\ell} - \mathbf{X}_{p,\ell}\mathbf{X}_{p,\ell}^{\dagger}\overline{\mathbf{C}}_p \left(\overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p,\ell}\mathbf{X}_{p,\ell}^{\dagger}\overline{\mathbf{C}}_p \right)^{-1} \overline{\mathbf{C}}_p^{\dagger}\mathbf{X}_{p,\ell} \right) \mathbf{X}_{p\ \ 1,\ell}^{\dagger}\overline{C}_p \tag{1.21}$$

Since $\mathbf{X}_{p,\ell}$ is formed from $\mathbf{X}_{p-1}$, and not just $\mathbf{X}_{p-1,\ell}$, we have

$$\mathbf{X}_{p,\ell}\mathbf{X}_{p-1,\ell}^{\dagger}\overline{\mathbf{C}}_p \neq \mathbf{0} \tag{1.22}$$

Hence

$$\mathbf{Z}_{p+1}\hat{\mathbf{X}}_p^{\dagger} \neq \mathbf{0} \tag{1.23}$$

Since $\hat{\mathbf{X}}_p$ is not orthogonal to the output of succeeding stages, as suggested by (4.23), global MMSE cannot be achieved at the final beamformer output. Our analysis of the Type I implementation indicates that the global MMSE solution is conditional on

$$L_\ell \leq L_p, \quad \text{for} \quad \ell > p$$

Intuitively this condition ensures that input data blocks to the p^{th} stage, ie:

$$\mathbf{X}_{p-1}[k],\ \mathbf{X}_{p-1}[k+1],\ \mathbf{X}_{p-1}[k+1],\ \ldots$$

are independent of each other. This allows the stage to compute $\hat{\mathbf{B}}_p[k+\ell]$ using only the L_p column vectors from $\mathbf{X}_{p-1}[k+\ell]$ without detrimentally affecting the final beamformer error.

4.2 Signal Blocking Matrix Design and Mismatch

Distributing adaptive DOF across a cascade of stages enables the M^{2}VDR beamformer to simultaneously adapt DOF on different time-scales. Since individual stages have less DOF than the total available to the beamformer, a stage can apply a larger reduction in adaptation time before weight noise variance causes performance to degrade. Although this feature allows a closer match between adaptation time and the time-scale of an interferer, performance degrades nevertheless with an inappropriate choice of signal blocking matrix. In this section we examine how the choice of signal blocking matrix affects performance.

The signal blocking matrix $\overline{\mathbf{C}}$ acts as a dimension reducing transformation on $\mathbf{x}[n]$, preventing beamformer weight estimates from being applied to the signal component of the measured wavefield. Two factors dictate the choice of signal blocking matrix for a single stage fully adaptive beamformer (ie: $P = 1$ and $Q = MJ - K$). First, to ensure that the desired signal passes undistorted through the beamformer $\overline{\mathbf{C}}$ must satisfy

$$\overline{\mathbf{C}}^{\dagger}\mathbf{C} = \mathbf{0}.$$

Second, the columns of $\overline{\mathbf{C}}$ must be linearly independent so that $\mathtt{rank}\left(\overline{\mathbf{C}}\right)$ corresponds to the number of adaptive DOF. Suppose

$$\mathbf{C} = (1,\ 1,\ \ldots,\ 1)^{\mathsf{T}}$$

and $J, K = 1$. A form typically used for $\overline{\mathbf{C}}$ is

$$\overline{\mathbf{C}} = \begin{pmatrix} 1 & 0 & 0 & 0 & \\ -1 & 1 & 0 & 0 & \\ 0 & -1 & 1 & 0 & \ldots \\ 0 & 0 & -1 & 1 & \\ & & \vdots & & \end{pmatrix} = (\ \bar{\mathbf{c}}_1\, \bar{\mathbf{c}}_2 \ldots \bar{\mathbf{c}}_{M-1}\ ) \qquad (4.21)$$

where each adaptive DOF, represented by a column of $\overline{\mathbf{C}}$, employs information only along two adjacent rows of the data sample covariance matrix, with covariance elements from remaining rows blocked. Because of this choice of blocking matrix each beamformer DOF adapts on data from one sensor pair.

By distributing DOF between P stages, the rank of the signal blocking matrix assigned to a particular stage is less than $MJ - K$. In this sense the p^{th} stage operates as a partially adaptive beamformer over a $Q_p < MJ - K$ dimensional subspace, ie:

$$\mathtt{span}\left(\overline{\mathbf{C}}_p\right) \subset \mathtt{span}\left(\overline{\mathbf{C}}\right)$$

Partial adaptivity implies that components of $\mathbf{X}_{p-1}$ lying only in $\mathtt{span}\left(\overline{\mathbf{C}}_p\right)$ will be regarded as interference by the p^{th} stage. If we were to assume that the same adaptation period is assigned to each stage then the M^{2}VDR beamformer and GSC are identical when

$$MJ - K = \sum_{\ell=1}^{P} Q_\ell$$

holds, and computed weights are associated with the same record of sensor measurements. In this case we can use the columns given by (4.21) in defining the blocking matrix of each stage. When adaptation time-scale differs between stages,

the columns of each blocking matrix operate over a different time-scale. This suggests that the p^{th} stage is tuned to interference components spatially and temporally through $\mathtt{span}\left(\overline{\mathbf{C}}_p\right)$ and the adaptation time assigned to the stage. Since computed stage weights are no longer dependent on the same record of sensor measurements, each stage must be treated as a partially adaptive beamformer.

A key factor affecting the performance of a partially adaptive beamformer is the degree to which an interferer lies in the span of the signal blocking matrix. When the signal blocking matrix is mismatched only partial cancellation of the interferer occurs. In the context of a multiple time-scale M^{2}VDR beamformer, mismatch between the signal blocking matrix of a stage and an interferer results in interference components leaking into the succeeding stage. This situation does not present a problem when the interferer appears over a temporal duration that is greater than or equal to the adaptation time of the succeeding stage. However, performance may degrade if the duration of leaked interference components is shorter than the adaptation of this stage.

Suppose the measured interference component of the wavefield is represented by $\mathbf{v}[n]$. Assume that the adaptation period of the p^{th} stage is matched to the time-scale of the interference. Set

$$\mathbf{v}[n] \in \mathtt{span}(\mathbf{C}) + \mathtt{span}\left(\overline{\mathbf{C}}_p\right) + \mathtt{span}\left(\overline{\mathbf{C}}_\ell\right) \tag{4.25}$$

where $\ell > p$, $L_\ell > L_p$, and

$$\overline{\mathbf{C}}_p^\dagger \overline{\mathbf{C}}_\ell = \mathbf{0}$$

With the appearance of $\mathbf{v}[n]$ the p^{th} stage maintains a null in the beamformer response in the direction of the interferer over the duration of the interference. Since $\mathbf{v}[n]$ is only partially spanned by $\overline{\mathbf{C}}_p$ interference components leak into the ℓ^{th} stage weight estimates. With $L_\ell > L_p$, the ℓ^{th} stage maintains a null in the response longer than the duration of interferer. Maintaining the null past the duration of the interferer may result in an unnecessary elevation of sidelobe levels, leading to an increase in beamformer output power because of diffuse noise.

4.2.1 Array Gain Degradation

As suggested in the preceding discussion, an important factor affecting performance is the match between the measured interference wavefield and the signal blocking matrix assigned to a stage tuned to the temporal duration of the interferer. In this section a lower bound for array gain (AG) degradation is derived when signal blocking matrix mismatch causes an increase in beamformer output power. To derive this bound we assume knowledge of the signal, interference, and noise covariance matrices. To simplify analysis a two stage beamformer assigned a total of Q $MJ - K$ DOF, with $L_1 < L_2$ is considered. We choose $\overline{\mathbf{C}}_2$ to lie in the orthogonal complement of $\overline{\mathbf{C}}_1$,

$$\overline{\mathbf{C}}_1^{\dagger}\overline{\mathbf{C}}_2 = \mathbf{0}$$

where

$$\overline{\mathbf{C}}_1^{\dagger}\overline{\mathbf{C}}_1 = \mathbf{I}_{Q_1}, \quad \text{and,} \quad \overline{\mathbf{C}}_2^{\dagger}\overline{\mathbf{C}}_2 = \mathbf{I}_{Q_2}$$

Suppose wavefield measurements consist of the three components

$$\mathbf{x}_0[n] = \mathbf{s}[n] + \mathbf{v}[n] + \boldsymbol{\eta}[n] \tag{4.26}$$

where $\mathbf{s}[n]$ is the signal, $\mathbf{v}[n]$ interference, and $\boldsymbol{\eta}[n]$ diffuse noise. The temporal duration of $\mathbf{v}[n]$ is represented by t_{d}. Let

$$\mathsf{P}_{\mathrm{s}} = \mathbf{w}_{\mathrm{q}}^{\dagger}\mathbf{R}_{\mathrm{s}}\mathbf{w}_{\mathrm{q}}$$

where $\mathbf{R}_{\mathrm{s}}$ equals the covariance matrix of the signal, and

$$\overline{\mathbf{C}}_1^{\dagger}\mathbf{R}_{\mathrm{s}} = \mathbf{0} \qquad \overline{\mathbf{C}}_2^{\dagger}\mathbf{R}_{\mathrm{s}} = \mathbf{0}$$

The covariance matrix of $\boldsymbol{\eta}[n]$ is

$$\mathbf{R}_{\boldsymbol{\eta}} = \mathsf{P}_{\mathrm{n}}\mathbf{I}_{MJ} \tag{4.27}$$

where P_{n} equals noise power as measured at an individual sensor. Let $\mathbf{R}_{\mathbf{v}}$ represent the covariance matrix associated with the interference component, so that

$$\mathbf{R}_{\mathbf{v}} \in \mathrm{span}\left(\overline{\mathbf{C}}_1\right) + \mathrm{span}\left(\overline{\mathbf{C}}_2\right) + \mathrm{span}(\mathbf{C}) \tag{4.28}$$

To examine the effect a short duration interferer has on beamformer output power when components of $\mathbf{v}[n]$ leak into the second stage we proceed as follows.

With $L_1 < L_2$, the weight matrix of the second stage retains components of $\mathbf{v}[n]$ after the interference event has vanished. Reflecting this condition in the first stage, we set

$$\mathbf{B}_1 = \left(\overline{\mathbf{C}}_1^\dagger \mathbf{R}_{\boldsymbol{\eta}} \overline{\mathbf{C}}_1\right)^{-1} \overline{\mathbf{C}}_1^\dagger \mathbf{R}_{\boldsymbol{\eta}} = \overline{\mathbf{C}}_1^\dagger \tag{4.29}$$

where

$$\mathbf{x}_0[n] = \mathbf{s}[n] + \boldsymbol{\eta}[n]$$

corresponds to sensor outputs after $\mathbf{v}[n]$ has disappeared. From (4.29), the covariance matrix for the output of the first stage equals

$$\left(\mathbf{I}_{MJ} - \mathbf{B}_1^\dagger \overline{\mathbf{C}}_1^\dagger\right)\left(\mathbf{R}_{\mathbf{s}} + \mathbf{R}_{\boldsymbol{\eta}}\right)\left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_1 \mathbf{B}_1\right) = \mathbf{R}_{\mathbf{s}} + \mathsf{P}_{\mathrm{n}}\left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_1 \overline{\mathbf{C}}_1^\dagger\right) \tag{4.30}$$

To account for residual components of $\mathbf{v}[n]$ retained by the second stage weights we set

$$\mathbf{B}_2 = \left(\overline{\mathbf{C}}_2^\dagger \left(\mu \mathbf{R}_{\mathbf{v}} + \mathsf{P}_{\mathrm{n}} \mathbf{I}_{MJ}\right) \overline{\mathbf{C}}_2\right)^{-1} \overline{\mathbf{C}}_2^\dagger \left(\mu \mathbf{R}_{\mathbf{v}} + \mathsf{P}_{\mathrm{n}} \mathbf{I}_{MJ}\right) \tag{4.31}$$

given that

$$\mathsf{P}_{\mathrm{n}} \overline{\mathbf{C}}_2^\dagger \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_1 \overline{\mathbf{C}}_1^\dagger\right) = \mathsf{P}_{\mathrm{n}} \overline{\mathbf{C}}_2^\dagger$$

where

$$\mu = \frac{L_2 - l_{\mathrm{d}}}{L_2}$$

scales interference power retained in $\mathbf{B}_2$ accordingly to the adaptation time of the second stage. Notice that

$$\overline{\mathbf{C}}_2^\dagger \mathbf{R}_{\mathbf{v}} \overline{\mathbf{C}}_2 = \overline{\mathbf{C}}_2^\dagger \mathbf{R}_{\mathbf{v}_2} \overline{\mathbf{C}}_2$$

where $\mathbf{R}_{\mathbf{v}_2}$ embodies the components of $\mathbf{R}_{\mathbf{v}}$ spanned by $\overline{\mathbf{C}}_2$, ie:

$$\mathbf{R}_{\mathbf{v}_2} = \overline{\mathbf{C}}_2 \left(\overline{\mathbf{C}}_2^\dagger \overline{\mathbf{C}}_2\right)^{-1} \overline{\mathbf{C}}_2^\dagger \mathbf{R}_{\mathbf{v}} \qquad \mathbf{R}_{\mathbf{v}_2} \in \mathsf{span}\left(\overline{\mathbf{C}}_2\right)$$

Since

$$\mathsf{P}_{\mathrm{n}} \overline{\mathbf{C}}_2^\dagger \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_1 \overline{\mathbf{C}}_1^\dagger\right) = \mathsf{P}_{\mathrm{n}} \overline{\mathbf{C}}_2^\dagger$$

and

$$P_n \mathbf{w}_q^\dagger \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_1 \overline{\mathbf{C}}_1^\dagger \right) = P_n \mathbf{w}_q^\dagger$$

we express output power of the second stage by

$$\begin{aligned} P_{out} &= \mathbf{w}_q^\dagger \left(\mathbf{I}_{MJ} - \mathbf{B}_2^\dagger \overline{\mathbf{C}}_2^\dagger \right) \left(\mathbf{R}_s + P_n \mathbf{I}_{MJ} \right) \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}_2 \mathbf{B}_2 \right) \mathbf{w}_q \\ &= P_s + P_n \left(1 + \mathbf{w}_q^\dagger \mathbf{B}_2^\dagger \overline{\mathbf{C}}_2^\dagger \overline{\mathbf{C}}_2 \mathbf{B}_2 \mathbf{w}_q \right) \\ &= P_s + P_n \left(1 + \left| \mathbf{w}_q^\dagger \mu \mathbf{R}_v \overline{\mathbf{C}}_2 \left(\overline{\mathbf{C}}_2^\dagger \mu \mathbf{R}_v \overline{\mathbf{C}}_2 + P_n \mathbf{I}_{Q_2} \right)^{-1} \right|^2 \right) \end{aligned} \tag{4.32}$$

Suppose

$$\mathbf{R}_v \in \text{span}\left(\overline{\mathbf{C}}_1 \right) + \text{span}\left(\overline{\mathbf{C}}_2 \right)$$

so that $\mathbf{R}_v$ has no component on the constraint subspace. From (4.32) we then have

$$P_{out} = P_s + P_n$$

since $\mathbf{w}_q^\dagger \mathbf{R}_v = \mathbf{0}$. Consequently when an interferer lies in a null of the beamformer quiescent response, leakage of interference components between stages does not affect output power. In other words, no work is done by the adaptive weights since $\mathbf{w}_q$ provides all interference suppression.

Rewrite (4.32) as

$$P_{out} = P_s + \left(1 + |\gamma|^2 \right) P_n \tag{4.33}$$

where

$$|\gamma|^2 = \left| \mathbf{w}_q^\dagger \mu \mathbf{R}_v \overline{\mathbf{C}}_2 \left(\overline{\mathbf{C}}_2^\dagger \mu \mathbf{R}_v \overline{\mathbf{C}}_2 + P_n \mathbf{I}_{Q_2} \right)^{-1} \right|^2$$

To expand $|\gamma|^2$ further, we rewrite $\mathbf{R}_v$ as

$$\mathbf{R}_v = P_v \, \overline{\mathbf{R}}_v^{1/2} \overline{\mathbf{R}}_v^{1/2} \tag{4.34}$$

where P_v equals interference power measured at a single sensor, and $\overline{\mathbf{R}}_v^{1/2}$ is the matrix square root of

$$\overline{\mathbf{R}}_v = \frac{1}{P_v}\mathbf{R}_v$$

Let

$$\overline{\mathbf{C}}_2^{\dagger}\overline{\mathbf{R}}_v^{1/2} = \mathbf{G}\Lambda^{1/2}\mathbf{H}^{\dagger} \tag{4.35}$$

denote the singular value decomposition of $\overline{\mathbf{C}}_2^{\dagger}\overline{\mathbf{R}}_v^{1/2}$, where

$$\Lambda = \mathtt{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_{n_i} \right)$$

where $\mathtt{rank}(\mathbf{R}_v) = n_v$. The singular values $0 \le \lambda_k \le 1$, $k = 1, 2, \ldots, n_v$, correspond to the fraction of interference power leaked into the second stage adaptive weights. The right and left singular vectors of $\overline{\mathbf{C}}_2^{\dagger}\overline{\mathbf{R}}_v^{1/2}$ are

$$\mathbf{G} = \left(\mathbf{g}_1\ \mathbf{g}_2\ \ldots\ \mathbf{g}_{n_v} \right)$$

and

$$\mathbf{H} = \left(\mathbf{h}_1\ \mathbf{h}_2\ \ldots\ \mathbf{h}_{MJ} \right)$$

From (4.35) we have

$$|\gamma|^2 = \left| \mu\mathtt{INR}\ \mathbf{w}_q^{\dagger}\overline{\mathbf{R}}_v^{1/2}\overline{\mathbf{R}}_v^{1/2}\overline{\mathbf{C}}_2 \left(\mu\mathtt{INR}\ \overline{\mathbf{C}}_2^{\dagger}\overline{\mathbf{R}}_v\overline{\mathbf{C}}_2 + \mathbf{I}_{Q_2}\right)^{-1} \right|^2$$

$$= \left| \mu\mathtt{INR}\ \mathbf{w}_q^{\dagger}\overline{\mathbf{R}}_v^{1/2}\mathbf{H}\Lambda^{1/2}\left(\mu\mathtt{INR}\ \Lambda + \mathbf{I}_{Q_2}\right)^{-1}\mathbf{G}^{\mathrm{T}} \right|^2$$

$$= \sum_{k=1}^{n_v} \frac{\lambda_k \mu^2 \mathtt{INR}^2}{(1 + \lambda_k \mu \mathtt{INR})^2}\ \mathbf{w}_q^{\dagger}\overline{\mathbf{R}}_v^{1/2}\mathbf{h}_k\mathbf{h}_k^{\dagger}\overline{\mathbf{R}}_v^{1/2}\mathbf{w}_q$$

$$= M \sum_{k=1}^{n_v} \frac{\lambda_k \mu^2 \mathtt{INR}^2}{(1 + \lambda_k \mu \mathtt{INR})^2}\ \cos^2(\xi_k) \tag{4.36}$$

where

$$\mathtt{INR} = \frac{P_v}{P_n}$$

and

$$M\ \cos^2(\xi_k)\ =\ \mathbf{w}_q^\dagger \overline{\mathbf{R}}_v^{1/2} \mathbf{h}_k \mathbf{h}_k^\dagger \overline{\mathbf{R}}_v^{1/2} \mathbf{w}_q$$

Observe that ξ_k is the angle between the left singular vector $\mathbf{h}_k$ and the response vector of the quiescent weights to the interferer, ie:

$$\mathbf{w}_q^\dagger \overline{\mathbf{R}}_v^{1/2}$$

Suppose $\overline{\mathbf{C}}_2$ is well matched to $\overline{\mathbf{R}}_v^{1/2}$ and $\mathtt{INR}\ >>\ 1$. Under these assumptions we set

$$\lambda_k\ \approx\ 1. \qquad \forall\ k$$

and

$$\frac{\lambda_k \mathtt{INR}^2}{(1+\lambda_k \mathtt{INR})^2}\ \approx\ 1$$

Hence,

$$\begin{aligned} \mathtt{P_{out}}\ &\approx\ \mathtt{P_s}\ +\ \mathtt{P_n}\left(1+|\ \gamma\ |^2\right) \\ &\leq\ \mathtt{P_s}\ +\ \mathtt{P_n}\left(\ 1+\mathtt{n_v}\ M\ \cos^2(\xi_{\mathtt{max}})\ \right) \end{aligned} \tag{4.37}$$

where

$$\xi_{\mathtt{max}}\ =\ \left(\ \xi_k \,\middle|\, \max\left(\cos^2(\xi_k)\right),\ k=1,2,\ldots,\mathtt{n}_v\ \right)$$

From (4.37), we bound AG by

$$\frac{M}{(\ 1+\mathtt{n_v}\ M\ \cos^2(\xi_{\mathtt{max}})\ )}\ \leq\ \mathtt{AG}\ \leq\ M \tag{4.38}$$

where

$$\cos^2(\xi_{\mathtt{max}})$$

is dependent on the response of the quiescent weights and the signal blocking matrix to the interferer. Notice as well that (4.38) depends on $\mathtt{n_v}$, the dimension of residual interference components retained in the second stage weights.

Chapter 5

The Impulsive Interference Environment

In the ocean medium impulsive sources of energy are induced by biological, seismic, and explosive events. Examples of impulsive noise phenomenon include cracking ice, marine animals, and transient mechanical noise from vessels [53, 51]. We refer to an acoustic event as impulsive when energy radiating from a source is transient and highly concentrated in time. The occurrence of impulsive noise in the presence of slowly time varying interferers poses a difficulty for MVDR realizations restricted to simultaneously adapting a fixed number of DOF on one time-scale. The problem arises when the beamformer computes weights on an adaptation time greater than the duration of the impulsive interferer. As suggested in the preceding chapter, this may lead to an unnecessary degradation in detection performance because beamformer weights retain residual components of the interferer after the event vanishes. The presence of residual interference components in the weight estimate forces the beamformer to maintain a null in the magnitude response against a phantom interferer. This causes an elevation in the sidelobes of the magnitude response that reduces the attenuation of diffuse noise. Consequently, beamformer output power increased because of diffuse noise, degrading the ability to detect weak signals.

To minimize the time interval over which detection performance degrades because of diffuse noise, we are inclined to reduce beamformer adaptation time. Unfortunately, by reducing adaptation time for a fixed number of DOF, weight noise variance increases. From (3.84), output SNR for a single stage beamformer equals

$$\mathsf{SNR}_{\mathsf{out}} = \frac{(1 - Q/L)\ \overline{\mathsf{SNR}}_{\mathsf{out}}}{1 + Q/L\ \overline{\mathsf{SNR}}_{\mathsf{out}}} \tag{5.1}$$

where $\overline{\mathsf{SNR}}_{\mathsf{out}}$ represents the infinite time output SNR of the beamformer. Q and L define the number of adaptive DOF and adaptation time assigned to the beamformer, respectively. For example, reducing L to match the temporal duration of the interference event decreases output SNR as indicated by (5.1). We attributed this effect to an increase in weight noise variance. An exact expression for weight noise variance follows.

Assuming Type I adaptation, we express the GSC weight estimate by

$$\hat{\mathbf{w}}_{\mathsf{a}} = \hat{\mathbf{B}}\mathbf{w}_{\mathsf{q}} = \left(\overline{\mathbf{C}}^{\dagger}\hat{\mathbf{R}}\overline{\mathbf{C}}\right)^{-1}\overline{\mathbf{C}}^{\dagger}\hat{\mathbf{R}}\mathbf{w}_{\mathsf{q}} \tag{5.2}$$

where

$$\hat{\mathbf{R}} = \frac{1}{L}\mathbf{X}\mathbf{X}^{\dagger}$$

and $\mathbf{X}$ is the current $MJ \times L$ input data block. From (3.62), the covariance of $\hat{\mathbf{w}}_{\mathsf{a}}$ equals

$$\begin{aligned}
\mathsf{cov}\left\{\hat{\mathbf{w}}_{\mathsf{a}}\right\} &= \mathcal{E}\left\{\mathcal{E}\left\{\left(\hat{\mathbf{B}} - \mathbf{B}\right)\mathbf{w}_{\mathsf{q}}\mathbf{w}_{\mathsf{q}}^{\dagger}\left(\hat{\mathbf{B}} - \mathbf{B}\right)^{\dagger} \mid \overline{\mathbf{C}}^{\dagger}\hat{\mathbf{R}}\overline{\mathbf{C}}\right\}\right\} \\
&= \mathcal{E}\left\{\mathsf{cov}\left\{\hat{\mathbf{B}}\mathbf{w}_{\mathsf{q}} \mid \overline{\mathbf{C}}^{\dagger}\hat{\mathbf{R}}\overline{\mathbf{C}}\right\}\right\} \\
&= \frac{1}{1 - \frac{Q}{L}}\left(\overline{\mathbf{C}}^{\dagger}\mathbf{R}\overline{\mathbf{C}}\right)^{-1}\mathsf{P}_{\mathsf{out}}
\end{aligned} \tag{5.3}$$

where

$$\mathbf{R} = \frac{1}{L}\mathcal{E}\left\{\mathbf{X}\mathbf{X}^{\dagger}\right\}$$

and

$$\mathsf{P}_{\mathsf{out}} \;=\; \mathbf{w}_{\mathbf{q}}^{\dagger}\left(\mathbf{R}-\mathbf{R}\overline{\mathbf{C}}\left(\overline{\mathbf{C}}^{\dagger}\mathbf{R}\overline{\mathbf{C}}\right)^{-1}\overline{\mathbf{C}}^{\dagger}\mathbf{R}\right)\mathbf{w}_{\mathbf{q}}$$

Observe from (5.3) that when Q is fixed, reducing adaptation time leads to an increase in weight noise variance. However, by decreasing Q it is possible to reduce L to a larger extent without incurring the same increase in variance. For the GSC and direct form MVDR realization, reducing Q degrades the ability of the beamformer in responding to the simultaneous appearance of multiple interferers. This problem does not arise for the multiple stage M^{2}VDR beamformer since the overall number of DOF remains fixed. To reduce adaptation time at an individual stage without incurring an increase in weight noise variance, we decrease the number of DOF assigned to the stage. Excess DOF are redistributed across other parts of the beamformer cascade to preserve the capability of responding to multiple interferers.

The ability of the multiple time-scale beamformer to adjust both the number of adaptive DOF and adaptation time of each stage allows greater flexibility in maximizing detection performance. This enables L_i to be approximately matched to the temporal extent of the interferer, and Q_i adjusted so that

$$L_i - Q_i$$

corresponds to an acceptable level of weight noise variance.

In this chapter a two stage beamformer is analyzed. The discussion focuses on the design of $\overline{\mathbf{C}}_1$ and the optimization of L_1 when wavefield measurements contain impulsive interference components. To deal with short duration impulsive type interference, the first stage is assigned a rapid rate of adaptation and correspondingly a small number of DOF. The succeeding stage functions as a conventional MVDR beamformer in the sense that remaining DOF are adapted over a temporal duration matched to slowly time varying interferers.

5.1 Signal Blocking: The Narrow-band Problem

To allow for the maximum degree of flexibility in selecting L_1 it is imperative that the minimum number of adaptive DOF be assigned to the first stage. This condition minimizes the impact of weight noise variance on performance as L_1 is reduced. In the narrow-band problem we only require a single adaptive DOF for each independent interferer in the wavefield [15]. However, to ensure optimal interference suppression the columns of $\overline{\mathbf{C}}_1$ must be matched to the incident interference wavefront.

Suppose M equi-spaced sensors are aligned in the horizontal plane along the x-axis as shown in Figure 5.1. For the narrow-band problem we set $J = 1$. The planar wavefront depicted in the figure represents interference. Applying (2.1),

$$\mathbf{x}(t,f) \;=\; \alpha(t,f)e^{j2\pi ft}\mathbf{r}_\theta(f) \tag{5.4}$$

where

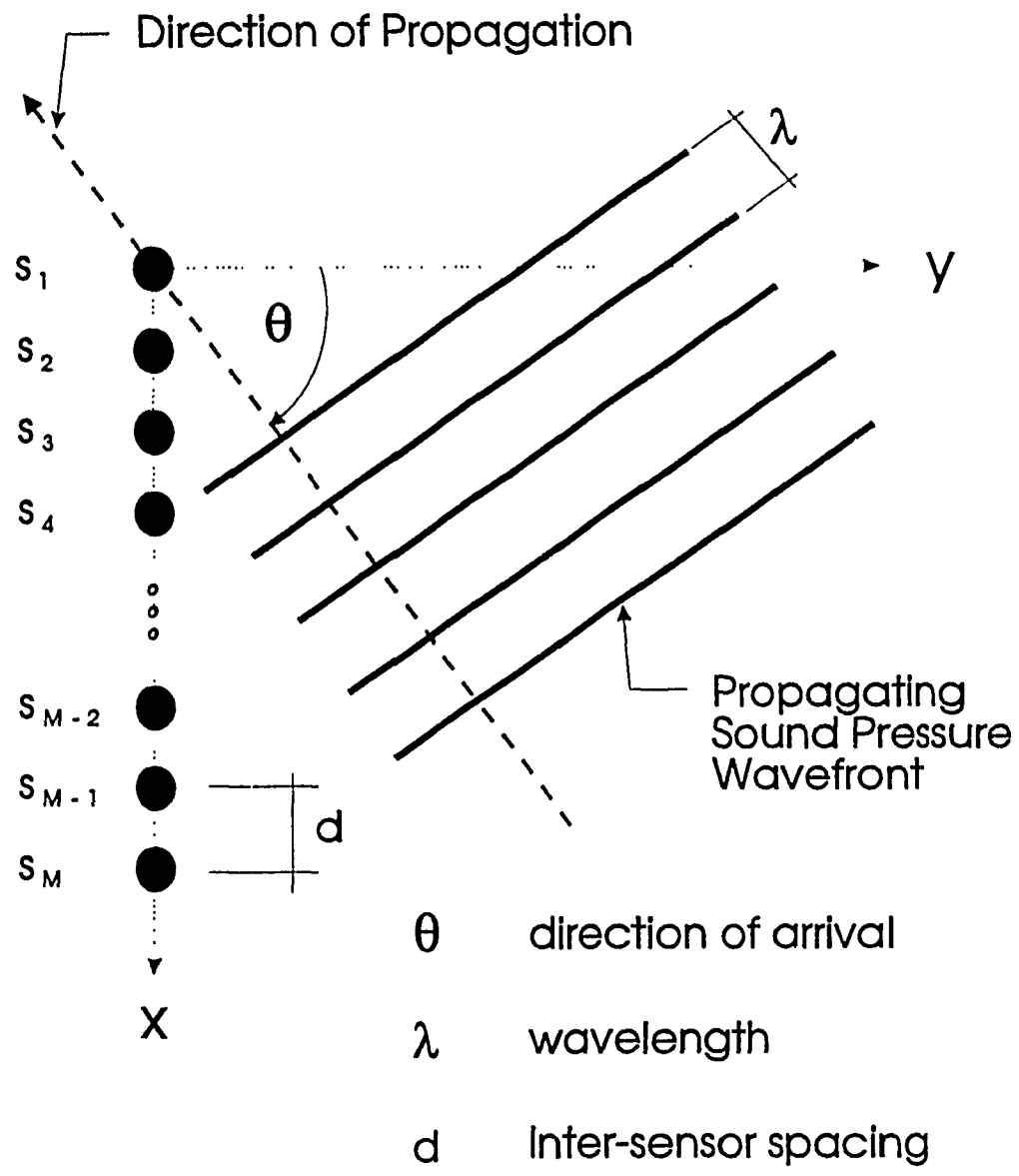
$$\mathbf{r}_\theta(f) \;=\; \left(1,\; e^{j2\pi\frac{f\mathsf{d}}{\mathsf{c}}\sin(\theta)},\; e^{j2\pi\frac{2f\mathsf{d}}{\mathsf{c}}\sin(\theta)},\; \ldots\; e^{j2\pi\frac{(M-1)f\mathsf{d}}{\mathsf{c}}\sin(\theta)}\right)$$

is the response vector of the array to the interferer, d inter-sensor spacing, c wavefront propagation velocity, and $\lambda = \mathsf{c}/f$ wavelength. We equate θ, the angle subtended by the direction of propagation and the y-axis of Figure 5.1, with the wavefront direction of arrival (DOA). If the DOA were known, the optimal choice for the first stage signal blocking matrix corresponds to the projection of $\mathbf{r}_\theta(f)$ onto the orthogonal complement of the signal constraint matrix $\mathbf{C}$, ie:

$$\overline{\mathbf{C}}_1 \;=\; \left(\mathbf{I}_M \;-\; \mathbf{C}\left(\mathbf{C}\mathbf{C}^{\mathsf{T}}\right)^{-1}\mathbf{C}^{\mathsf{T}}\right)\mathbf{r}_\theta(f) \tag{5.5}$$

Since θ is unknown, and time varying, we can only consider maximizing the spatial response of $\overline{\mathbf{C}}_1$ over regions of possible interference activity. Identifying such regions, beam-based designs (eg: [40],[55]-[57]) for partially adaptive beamformers can be applied in forming $\overline{\mathbf{C}}_1$.

Assume $q \leq Q$ possible locations of interference activity, where Q equals the total number of DOF available to the beamformer. We identify each interference



Horizontal Line Array of M equi-spaced Sensors

Figure 5.1

location by a response vector

$$\mathbf{r}_{\theta_\ell}(f) \;=\; \left(1,\; e^{j2\pi\frac{f\mathrm{d}}{\mathrm{c}}\sin(\theta_\ell)},\; e^{j2\pi\frac{2f\mathrm{d}}{\mathrm{c}}\sin(\theta_\ell)},\; \ldots\; e^{j2\pi\frac{(M-1)f\mathrm{d}}{\mathrm{c}}\sin(\theta_\ell)}\right)$$

where θ_ℓ equals the $\ell^{\mathtt{th}}$ hypothesized DOA for an interferer. For q such angles, set

$$\mathbf{A} \;=\; \left(\mathbf{r}_{\theta_1}(f),\; \mathbf{r}_{\theta_2}(f),\; \ldots\; \mathbf{r}_{\theta_q}(f)\right) \tag{5.6}$$

Each column of $\mathbf{A}$, or beam, corresponds to the response of a uniformly weighted conventional beamformer, where the magnitude response of $\mathbf{r}_{\theta_\ell}$ is maximum at θ_ℓ. From (5.5), the projection of $\mathbf{A}$ onto the orthogonal complement of $\mathbf{C}$ equals

$$\mathbf{H}_1 \;=\; \left(\mathbf{I}_M \;-\; \mathbf{C}\left(\mathbf{C}\mathbf{C}^{\mathrm{T}}\right)^{-1}\mathbf{C}^{\mathrm{T}}\right)\mathbf{A} \tag{5.7}$$

$\mathbf{H}_1$ corresponds to the optimal choice for $\overline{\mathbf{C}}_1$ given $\mathbf{A}$. However, to ensure flexibility in selecting L_1 when q is large, we require a low rank ($Q_1 < q$) approximation of $\mathbf{H}_1$. Let

$$\mathbf{H}_1 \;=\; (\;\mathbf{S}_a, \mathbf{S}_b\;) \begin{pmatrix} \mathbf{\Lambda}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_b \end{pmatrix} \begin{pmatrix} \mathbf{U}_a^\dagger \\ \mathbf{U}_b^\dagger \end{pmatrix} \tag{5.8}$$

correspond to the singular value decomposition (SVD) of $\mathbf{H}_1$, where $\mathbf{S}_a$ and $\mathbf{S}_b$ equal $M\times Q_1$ and $M\times(q-Q_1)$ "unitary matrices", respectively. $\mathbf{\Lambda}_a$ and $\mathbf{\Lambda}_b$ are $Q_1\times Q_1$ and $(q-Q_1)\times(q-Q_1)$ diagonal matrices containing the singular values of $\mathbf{H}_1$ arranged in descending order of magnitude. $\mathbf{U}_a$ and $\mathbf{U}_b$ equal $q\times Q_1$ and $q\times(q-Q_1)$ "unitary matrices", respectively. We may then form the rank Q_1 least squares approximation of $\mathbf{H}_1$ by

$$\overline{\mathbf{H}}_1 \;=\; \mathbf{S}_a\mathbf{\Lambda}_a\mathbf{U}_a^\dagger \tag{5.9}$$

such that

$$\left(\mathbf{H}_1 \;-\; \overline{\mathbf{H}}_1\right)\overline{\mathbf{H}}_1^\dagger \;=\; \mathbf{0}$$

$\overline{\mathbf{C}}_1$ follows from

$$\overline{\mathbf{H}}_1 \;=\; \overline{\mathbf{C}}_1\mathbf{T}^\dagger \tag{5.10}$$

where $\mathbf{T}$ is a $q \times Q_1$ transformation. Setting

$$\mathbf{T} = \mathbf{U}_a$$

the first stage signal blocking matrix takes the form

$$\overline{\mathbf{C}}_1 = \mathbf{S}_a \mathbf{\Lambda}_a \tag{5.11}$$

The choice for Q_1 is influenced by two factors, the number of impulsive interference events occurring simultaneously, and the time-scale of interference.

5.2 Signal Blocking: The Wide-band Problem

Unlike the narrow-band beamformer, wide-band processing entails an adaptive response in both the spatial and temporal domains. Hence, the requirement for a tap delay line, $J > 1$. We begin this discussion by describing a model for impulsive interference measurements retained in the beamformer tap delay line.

5.2.1 Background

Suppose the $MJ \times 1$ measurement vector of sensor outputs consists of the three independent components

$$\mathbf{x}[n] = \mathbf{s}[n] + \beta \mathbf{v}[n] + \boldsymbol{\eta}[n] \tag{5.12}$$

where $\mathbf{s}[n]$ and $\boldsymbol{\eta}[n]$ represent the signal and noise components of sensor measurements, respectively. At the output of a sensor

$$\mathsf{P}_\mathbf{s} = \mathbf{w}_\mathbf{q}^\dagger \mathcal{E}\left\{\mathbf{s}[n]\mathbf{s}^\dagger[n]\right\} \mathbf{w}_\mathbf{q}$$

equals average signal power and

$$\mathsf{P}_\mathbf{n} = \mathbf{w}_\mathbf{q}^\dagger \mathcal{E}\left\{\boldsymbol{\eta}[n]\boldsymbol{\eta}^\dagger[n]\right\} \mathbf{w}_\mathbf{q}$$

average noise power. $\beta\mathbf{v}[n]$ represents an impulsive interference component at the output of the sensors. Applying the quiescent weight vector to $\beta\mathbf{v}[n]$ over an observation interval of L samples, interference power equals

$$\mathtt{P}_{\mathtt{v}} \;=\; \frac{|\beta|^2}{L}$$

Suppose that the most significant percentage of interference energy is concentrated within a finite temporal duration of $t_{\mathtt{d}}$ samples. Set

$$(\mathbf{v}[n])_\ell = 0 \quad \text{outside the interval } T_{\mathtt{i}} \leq n \leq T_{\mathtt{i}} + t_{\mathtt{d}} - 1 \tag{5.13}$$

where $T_{\mathtt{i}}$ corresponds to the sample instant when the impulsive interference first appears in the measured data. Characterizing $\mathbf{v}[n]$ in this manner, the entire spatial and temporal extent of the interferer across the array is encompassed within the column matrix

$$\mathbf{V} = (\ \mathbf{v}(T_{\mathtt{i}})\ \ \mathbf{v}(T_{\mathtt{i}}+1)\ldots\mathbf{v}(T_{\mathtt{i}}+t_{\mathtt{d}}-1)\ ) \tag{5.14}$$

We scale $(\ \mathbf{v}[n]\ )_\ell$ so that

$$0 \leq \mathbf{w}_{\mathtt{q}}^{\dagger}\mathbf{V}\mathbf{V}^{\dagger}\mathbf{w}_{\mathtt{q}} \leq M \tag{5.15}$$

To incorporate the spatial/temporal properties of the impulsive interferer in the sample covariance matrix of $\mathbf{x}[n]$ observe that $\beta\mathbf{v}[n]$ impacts

$$\hat{\mathbf{R}} \;=\; \frac{1}{L}\sum_{\ell=0}^{L-1}\mathbf{x}[n+\ell]\mathbf{x}^{\dagger}[n+\ell] \tag{5.16}$$

after the sample instant $n \geq T_{\mathtt{i}}$. Without loss of generality set

$$T_{\mathtt{i}} \;=\; 0$$

We then describe the sample covariance matrix of $\mathbf{x}[n]$ by

$$\hat{\mathbf{R}} \;=\; \frac{1}{L}\sum_{\ell=0}^{L-1}\left(\mathbf{s}[\ell]\mathbf{s}^{\dagger}[\ell] + \boldsymbol{\eta}[\ell]\boldsymbol{\eta}^{\dagger}[\ell]\right) + \frac{|\beta|^2}{L}\mathbf{V}\mathbf{V}^{\dagger} \tag{5.17}$$

when $L \geq t_{\mathtt{d}}$. Over L samples INR equals

$$\mathtt{INR} = \frac{\mathtt{P}_{\mathtt{v}}}{\mathtt{P}_{\mathtt{n}}} \tag{5.18}$$

To complete the description of the impulsive interferer denote

$$\mathtt{n_v} = \mathtt{rank}(\mathbf{V}) \leq t_\mathtt{d}$$

as interference dimensionality. Interference dimensionality increases as the direction of arrival (DOA) approaches endfire, (ie: $|\theta| = 90^\circ$ on Figure 5.1).

5.2.2 Assigning Adaptive Degrees of Freedom

Over the temporal duration of the impulsive interferer the sample covariance matrix of sensor measurements is

$$\hat{\mathbf{R}} = \frac{1}{t_\mathtt{d}} \sum_{\ell=1}^{t_\mathtt{d}} \left(\mathbf{s}[\ell]\mathbf{s}^\dagger[\ell] + \boldsymbol{\eta}[\ell]\boldsymbol{\eta}^\dagger[\ell] \right) + \frac{|\beta|^2}{t_\mathtt{d}} \mathbf{V}\mathbf{V}^\dagger \tag{5.19}$$

Suppose that the average power of the impulsive interferer over $t_\mathtt{d}$ samples is much larger than the power of the signal and noise components of $\mathbf{x}[n]$, ie:

$$\mathtt{P}_\mathtt{I} = \frac{|\beta|^2}{t_\mathtt{d}} >> \mathtt{P}_\mathtt{s} \quad \text{and} \quad \mathtt{P}_\mathtt{i} = \frac{|\beta|^2}{t_\mathtt{d}} >> \mathtt{P}_\mathtt{n}$$

In this instance we approximate $\hat{\mathbf{R}}$ by

$$\hat{\mathbf{R}}_0 \approx \mathtt{P}_\mathtt{v} \mathbf{V}\mathbf{V}^\dagger \tag{5.20}$$

From (5.20), the power estimate of the first stage is then written as

$$\hat{\mathtt{P}}_\mathtt{I} \approx \mathtt{P}_\mathtt{v} \mathbf{w}_\mathtt{q}^\dagger \mathbf{V} \left(\mathbf{I}_{t_\mathtt{d}} - \mathbf{V}^\dagger \overline{\mathbf{C}}_\mathtt{I} \left(\overline{\mathbf{C}}_\mathtt{I}^\dagger \mathbf{V}\mathbf{V}^\dagger \overline{\mathbf{C}}_\mathtt{I} \right)^{-1} \overline{\mathbf{C}}_\mathtt{I}^\dagger \mathbf{V} \right) \mathbf{V}^\dagger \mathbf{w}_\mathtt{q} \tag{5.21}$$

where the level of interference cancellation is shown to depend on the projection of $\mathbf{V}$ onto the subspace described by the linear span $\mathtt{span}\left(\overline{\mathbf{C}}_\mathtt{I}^\dagger \mathbf{V}\right)$. Since

$$\mathtt{rank}\left(\overline{\mathbf{C}}_\mathtt{I}\right) = Q_\mathtt{I} \quad \text{and} \quad \mathtt{rank}(\mathbf{V}) = \mathtt{n_v}$$

the rank of $\overline{\mathbf{C}}_\mathtt{I}^\dagger \mathbf{V}$ is bounded by

$$\mathtt{rank}\left(\overline{\mathbf{C}}_\mathtt{I}^\dagger \mathbf{V}\right) = \mathtt{min}(Q_\mathtt{I}, \mathtt{n_v}) \tag{5.22}$$

Hence,

$$Q_1 \;=\; \mathrm{n_v} \leq t_\mathrm{d}$$

establishes the minimum number of adaptive DOF required by the first stage for maximum interference suppression. Without prior knowledge of $\mathbf{V}$ the possibility of an over or underdetermined estimate of Q_1 is likely. For the overdetermined case $\mathrm{n_v} < Q_1$, weight noise variance may impair the ability to match the adaptation time scale to the temporal duration of the interference. When $n_\mathrm{v} > Q_1$, the leakage of impulsive interference components into the second stage becomes a concern.

Suppose $\mathbf{V}$ is known. Let

$$\mathbf{V} \;=\; \mathbf{F\Omega S}^\mathsf{T} \tag{5.23}$$

describe the SVD of $\mathbf{V}$. Ignore for the moment the constraint

$$\overline{\mathbf{C}}_1^\dagger \mathbf{C} = 0$$

imposed on the first stage signal blocking matrix. Substituting

$$\overline{\mathbf{C}}_1 = \mathbf{F\Omega} \tag{5.24}$$

into (5.21)

$$\begin{aligned}
\hat{\mathrm{P}}_1 \;&\approx\; \mathrm{P_v}\, \mathbf{w}_\mathrm{q}^\dagger\, \mathbf{V} \left(\mathbf{I}_{t_\mathrm{d}} \;-\; \mathbf{V}^\dagger \overline{\mathbf{C}}_1 \left(\overline{\mathbf{C}}_1^\dagger \mathbf{V} \mathbf{V}^\dagger \overline{\mathbf{C}}_1 \right)^{-1} \overline{\mathbf{C}}_1^\dagger \mathbf{V} \right) \mathbf{V}^\dagger\, \mathbf{w_q} \\
&=\; \mathrm{P_v}\, \mathbf{w}_\mathrm{q}^\dagger\, \mathbf{F\Omega S}^\mathsf{T} \left(\mathbf{I}_{t_\mathrm{d}} \;-\; \mathbf{S} \left(\mathbf{S}^\mathsf{T} \mathbf{S} \right)^{-1} \mathbf{S}^\mathsf{T} \right) \mathbf{S\Omega}^\mathsf{T} \mathbf{F}^\mathsf{T}\, \mathbf{w_q} \\
&=\; 0
\end{aligned} \tag{5.25}$$

so that total interference cancellation is achieved. Re-imposing $\overline{\mathbf{C}}_1^\dagger \mathbf{C} = 0$, we project $\mathbf{F\Omega}$ onto the orthogonal complement of $\mathbf{C}$,

$$\overline{\mathbf{C}}_1 = \left(\mathbf{I}_{MJ} - \mathbf{C} \left(\mathbf{C}^\dagger \mathbf{C} \right)^{-1} \mathbf{C}^\dagger \right) \mathbf{F\Omega} \tag{5.26}$$

to find the optimal setting for $\overline{\mathbf{C}}_1$.

Since prior knowledge of $\mathbf{V}$ is precluded, a partially adaptive signal blocking matrix design strategy must be adopted. Van Veen [58] discusses wide-band beam based design. The eigenstructure and power minimization approaches discussed in [33]-[40] provide other examples of wide-band partially adaptive beamformer design. The eigenstructure technique [40] is prone to overdetermined solutions. Since minimization of Q_1 is an important factor in our design criteria this method proves inappropriate. The intractability of the optimization problem associated with the power minimization method [34] results in a suboptimal solution.

5.3 Optimization of Adaptation Time-Scale

Suppose that a wavefield has two candidate models, summarized by the hypotheses

$$\bar{\mathrm{H}}_0 : \quad \mathbf{x}[n] = \boldsymbol{\eta}[n]$$

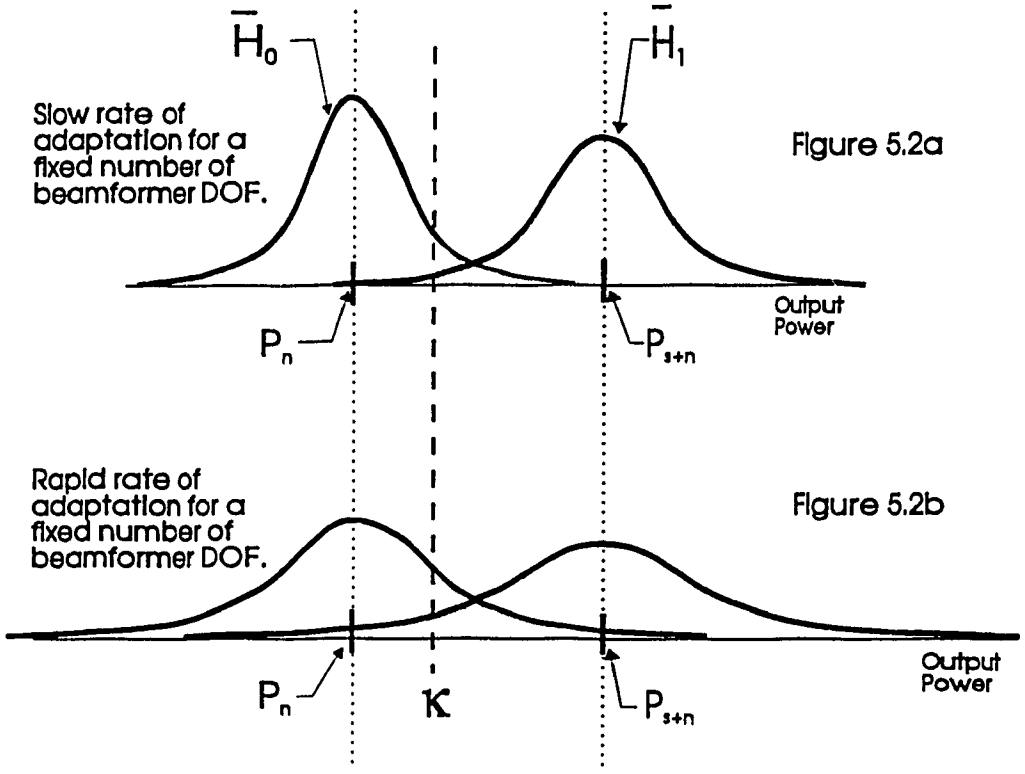
and

$$\bar{\mathrm{H}}_1 : \quad \mathbf{x}[n] = \mathbf{s}[n] + \boldsymbol{\eta}[n].$$

In selecting between $\bar{\mathrm{H}}_0$ and $\bar{\mathrm{H}}_1$, (ie: the presence or absence of a signal), a detector compares the beamformer output power estimate with a threshold. When output power exceeded this threshold we choose $\bar{\mathrm{H}}_1$, otherwise $\bar{\mathrm{H}}_0$ is accepted. The random nature of the beamformer power estimate introduces uncertainty in the decision made by the detector. Assuming a fixed number of DOF, the degree of uncertainty increases as adaptation time decreases, since a reduction in adaptation time translates to an increase in weight noise variance. We summarize this binary detection problem in Figure 5.2. In Figure 5.2a, the probability density functions for $\bar{\mathrm{H}}_0$ and $\bar{\mathrm{H}}_1$ are shown when a slow rate of adaptation is applied in computing beamformer weights, where

$$\mathrm{P_{s+n}} = \mathrm{P_s} + \mathrm{P_n}$$

equals the mean value of output power under $\bar{\mathrm{H}}_1$. The shaded area shown in the figure represents the probability of choosing $\bar{\mathrm{H}}_1$ when $\bar{\mathrm{H}}_0$ is true. We refer to this probability



Signal Detection in Diffuse Noise

Figure 5.2

as the probability of false alarm (PFA). Observe from Figure 5.2b that decreasing adaptation time increases PFA through the increase of weight noise variance. Note that weight variance affects the probability density function (pdf) of beamformer output power under $\bar{\mathrm{H}}_0$ and $\bar{\mathrm{H}}_1$ equally.

Consider now the binary hypothesis problem that arises when an impulsive interferer appears in sensor measurements. Assuming the adaptation time of the first stage is longer than the duration of the interferer, we summarize this detection problem by

$$\mathrm{H}_0: \quad \mathbf{x}[n] = \beta\mathbf{v}[n] + \boldsymbol{\eta}[n]$$

and

$$\mathrm{H}_1: \quad \mathbf{x}[n] = \mathbf{s}[n] + \beta\mathbf{v}[n] + \boldsymbol{\eta}[n]$$

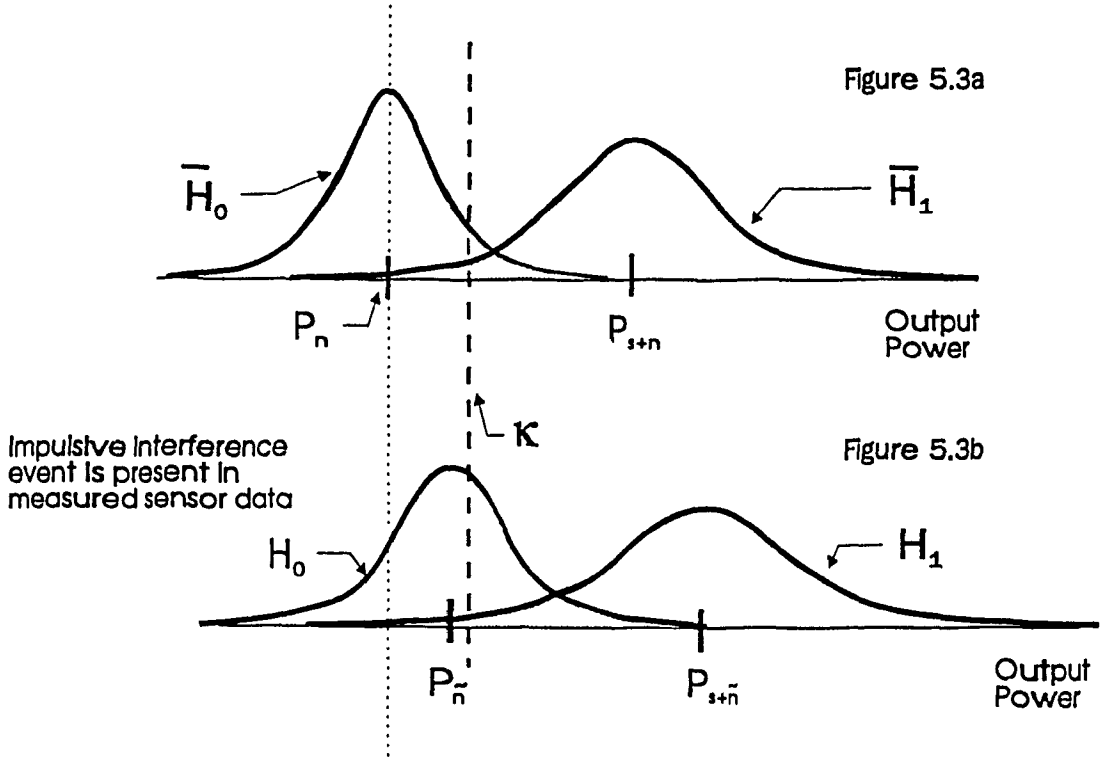
With adaptation time L_1 longer than t_{d}, residual components of the interferer are retained in the first stage beamformer weights after the event has vanished from sensor measurements. As previously mentioned, this gives rise to an increase in output noise power. Figure 5.3b illustrates the detection problem in for this situation, where $\mathrm{P}_{\tilde{\mathbf{n}}}$ now represents output noise power such that

$$\mathrm{P}_{\tilde{\mathbf{n}}} \geq \mathrm{P}_n$$

Note that

$$\mathrm{P}_{\mathbf{s}+\tilde{\mathbf{n}}} = \mathrm{P}_{\mathbf{s}} + \mathrm{P}_{\tilde{\mathbf{n}}}$$

equals mean output power under H_1. Maintaining the threshold decision level, κ, constant, we see from Figure 5.3b that because of a higher output noise power level, PFA increases. To optimally select L_1 we propose decreasing adaptation time to a point where output noise power and weight noise variance results in a minimum PFA. Such a strategy is intended to minimize the affect of impulsive interference on performance by trading weight noise variance for detection performance in diffuse noise.



Signal Detection in Diffuse Noise and Impulsive Interference

Figure 5.3

5.3.1 Closed Form Optimization of L_1

For the purpose of this analysis we assume that

$$Q_1 < Q_2 << L_2 >> 1$$

thus allowing us to approximate $\hat{\mathrm{P}}_2$ by a Gaussian distribution [59].

PFA at the output of the beamformer equals

$$\mathtt{PFA} = \mathtt{erfc}\left(\frac{\kappa - \mathcal{E}\{\hat{\mathrm{P}}_2|\mathrm{H}_0\}}{\sqrt{\mathtt{var}\{\hat{\mathrm{P}}_2|\mathrm{H}_0\}}}\right) \tag{5.27}$$

where κ represents the decision threshold level determined for a fixed PFA, and

$$\mathtt{erfc}(a) = \frac{2}{\sqrt{\pi}}\int_a^\infty \mathrm{e}^{-u^2}\mathrm{d}u$$

is the complementary error function. Setting

$$b = \frac{\kappa - \mathcal{E}\{\hat{\mathrm{P}}_2|\mathrm{H}_0\}}{\sqrt{\mathtt{var}\{\hat{\mathrm{P}}_2|\mathrm{H}_0\}}}$$

probability of detection (PD) equals

$$\mathtt{PD} = \mathtt{erfc}(\ cb - \mathtt{DI}\ ) \tag{5.28}$$

where

$$c = \sqrt{\frac{\mathtt{var}\{\hat{\mathrm{P}}_2|\mathrm{H}_0\}}{\mathtt{var}\{\hat{\mathrm{P}}_2|\mathrm{H}_1\}}}$$

and

$$\mathtt{DI} = \frac{\mathcal{E}\{\hat{\mathrm{P}}_2|\mathrm{H}_1\} - \mathcal{E}\{\hat{\mathrm{P}}_2|\mathrm{H}_0\}}{\sqrt{\mathtt{var}\{\hat{\mathrm{P}}_2|\mathrm{H}_1\}}} \tag{5.29}$$

defines the detection index (DI). Assuming that

$$c \approx 1$$

PD is a monotonic function in DI. Under this assumption maximizing DI is equivalent to maximizing PD. Hence, an optimal choice for L_1 follows from

$$\frac{\partial \mathtt{DI}}{\partial L_1} = 0 \tag{5.30}$$

Suppose $L_1 > t_d$. From (5.26), we set the first stage signal blocking matrix to

$$\overline{\mathbf{C}}_1 = \left(\mathbf{I}_{MJ} - \mathbf{C}\left(\mathbf{C}^\dagger\mathbf{C}\right)^{-1}\mathbf{C}^\dagger\right)\mathbf{F}\boldsymbol{\Omega}.$$

Because of the mismatch between L_1 and t_d output power increases due to diffuse noise. Adapting (4.32) to our analysis, output power of the first stage is approximated by

$$\begin{aligned} \mathrm{H}_0 &: \ \mathrm{P}_1 \approx \left(1 + \overline{\mu}\,|\overline{\gamma}|^2\right)\mathrm{P}_n \\ \mathrm{H}_1 &: \ \mathrm{P}_1 \approx \mathrm{P}_s + \left(1 + \overline{\mu}\,|\overline{\gamma}|^2\right)\mathrm{P}_n \end{aligned} \tag{5.31}$$

where

$$|\overline{\gamma}|^2 = \left(1 + \left|\mathbf{w}_q^\dagger\mathbf{R}_v\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^\dagger\mathbf{R}_v\overline{\mathbf{C}}_1 + \mathrm{P}_n\mathbf{I}_{MJ}\right)^{-1}\right|^2\right)$$

The scaling factor

$$\overline{\mu} = \frac{L_1 - t_d}{L_2}$$

takes into consideration the differences between beamformer adaptation time and the duration of the interference. Taking into account L_1 and Q_1, we approximate the mean output power of the first stage by

$$\begin{aligned} \mathrm{H}_0 &: \ \mathcal{E}\{\hat{\mathrm{P}}_1\} \approx \left(1 - \tfrac{Q_1}{L_1}\right)\left(1 + \overline{\mu}\,|\overline{\gamma}|^2\right)\mathrm{P}_n \\ \mathrm{H}_1 &: \ \mathcal{E}\{\hat{\mathrm{P}}_1\} \approx \left(1 - \tfrac{Q_1}{L_1}\right)\left(\mathrm{P}_s + \left(1 + \overline{\mu}\,|\overline{\gamma}|^2\right)\mathrm{P}_n\right) \end{aligned} \tag{5.32}$$

At the output of the second stage we have

$$\mathcal{E}\{\hat{\mathrm{P}}_2|\mathrm{H}_0\} = \left(1 - \frac{Q_1}{L_1} - \frac{Q_2}{L_2}\right)\left(1 + \overline{\mu}\,|\overline{\gamma}|^2\right)\mathrm{P}_n \tag{5.33}$$

$$\mathcal{E}\{\hat{\mathrm{P}}_2|\mathrm{H}_1\} = \left(1 - \frac{Q_1}{L_1} - \frac{Q_2}{L_2}\right)\left(\left(1 + \overline{\mu}\,|\overline{\gamma}|^2\right)\mathrm{P}_n + \mathrm{P}_s\right) \tag{5.34}$$

and

$$\mathrm{var}\{\hat{\mathrm{P}}_2|\mathrm{H}_0\} = \left(1 - \frac{Q_1}{L_1} - \frac{Q_2}{L_2}\right)\left(\left(1 + \overline{\mu}\,|\overline{\gamma}|^2\right)\mathrm{P}_n + \mathrm{P}_s\right)^2 \tag{5.35}$$

Substituting (5.33), (5.34), and (5.35) into (5.29),

$$\mathtt{DI} = \frac{\overline{\mathtt{SNR}}_{\mathtt{out}}\sqrt{1-\frac{Q_1}{L_1}-\frac{Q_2}{L_2}}}{\left(1+\overline{\mu}\,|\overline{\gamma}|^2\right)+\overline{\mathtt{SNR}}_{\mathtt{out}}} \tag{5.36}$$

where

$$\frac{\partial \mathtt{DI}}{\partial L_1} = \frac{1}{2}\frac{\overline{\mathtt{SNR}}_{\mathtt{out}}Q_2}{\sqrt{1-\frac{Q_1}{L_1}-\frac{Q_2}{L_2}}\left(1+\frac{L_1-t_{\mathtt{d}}}{L_2}|\overline{\gamma}|^2+\overline{\mathtt{SNR}}_{\mathtt{out}}\right)L_1^2} - \frac{\overline{\mathtt{SNR}}_{\mathtt{out}}|\overline{\gamma}|^2\sqrt{1-\frac{Q_1}{L_1}-\frac{Q_2}{L_2}}}{\left(1+\frac{L_1-t_{\mathtt{d}}}{L_2}|\overline{\gamma}|^2+\overline{\mathtt{SNR}}\right)^2} \tag{5.37}$$

Setting (5.37) to zero, the optimal solution for L_1 follows from the positive root of the quadratic expression

$$2|\overline{\gamma}|^2(L_2-Q_2)L_1^2 - 3|\overline{\gamma}|^2Q_1L_2L_1 + L_2Q_1\left(|\overline{\gamma}|^2t_{\mathtt{d}} - L_2(1+\overline{\mathtt{SNR}}_{\mathtt{out}})\right) = 0 \tag{5.38}$$

such that

$$L_1 = \frac{3}{4(1-Q_2/L_2)}Q_1\left(1+\sqrt{1+\frac{8}{9Q_1}\left(\frac{L_2(1+\overline{\mathtt{SNR}}_{\mathtt{out}})}{|\overline{\gamma}|^2}-t_{\mathtt{d}}\right)\left(1-\frac{Q_2}{L_2}\right)}\right) \tag{5.39}$$

As suggested by (5.39), the magnitude of $|\overline{\gamma}|^2$ in relation to $\overline{\mathtt{SNR}}_{\mathtt{out}}$ is an important factor in optimizing L_1. To demonstrate this point L_1, Q_2, and L_2 are fixed, and the temporal duration of the impulsive interferer is set to $t_{\mathtt{d}} = 2$ samples. Figure 5.5 shows a series of curves determined from (5.39), with each curve corresponding to a fixed value of $\overline{\mathtt{SNR}}_{\mathtt{out}}$. Note that as $\overline{\mathtt{SNR}}_{\mathtt{out}}$ increases in relation to $|\overline{\gamma}|^2$ less emphasis is placed on matching L_1 with the temporal duration of the interferer. Included on Figure 5.4 is the point where L_1 equals Q_1. Observe from the figure that (5.39)

yields values of L_1 that are greater than Q_1. This ensures that beamformer power estimates are positive quantities.

Notice that we have assume prior knowledge of the duration of the impulsive interferer. Such information would be available in such applications as active sonar. However, for many phenomena, including seismic events or cracking ice, $\mathtt{t_d}$ will be unknown or behave as a random variable. Under these conditions the usefulness of (5.39) is limited because of the sensitivity of this expression to $t_{\mathtt{d}}$.

5.3.2 Numerical Optimization of L_1

In formulating a closed form expression for L_1 a Gaussian approximation of the beamformer output power statistic is required. As an alternative strategy, we optimize L_1 directly through a comparison of receiver operating curves generated from the true PDF of $\hat{\mathtt{P}}_2$, where each receiver operating curve (ROC) is associated with a set value for L_1. Recall that the PDF of the beamformer output power estimate is defined by the Wishart density function

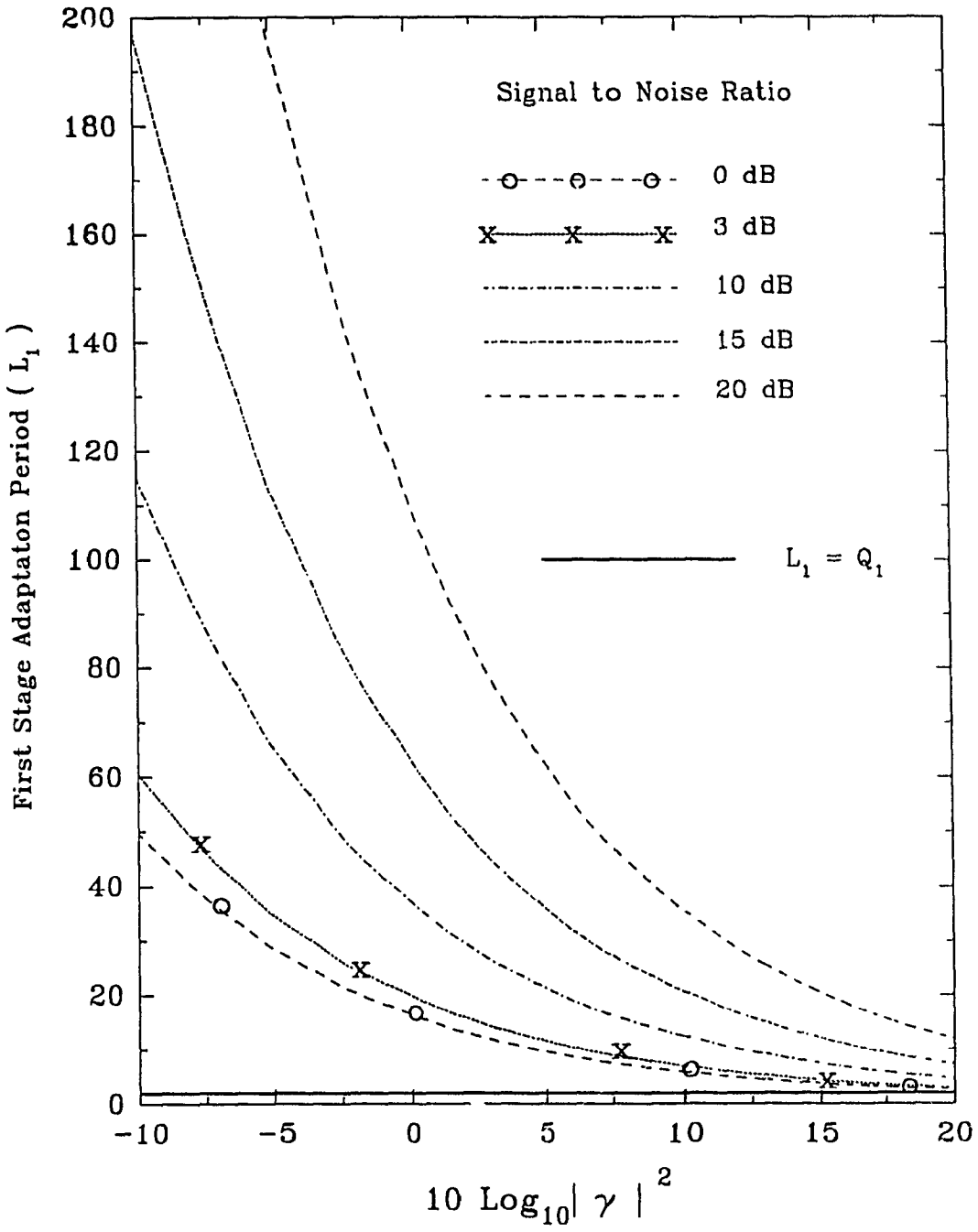
$$\mathtt{f}_{\hat{\mathtt{P}}_2|\mathtt{H}_\ell}(\nu) = \frac{\left(\frac{\nu}{\mathtt{P}_{2,\ell}}\right)^{N/2-1} e^{-\nu/(2\mathtt{P}_{2,\ell})}}{\mathtt{P}_{2,\ell} 2^{N/2} \Gamma(N/2)} \tag{5.40}$$

where $\Gamma(N/2)$ is the gamma function, and

$$\begin{aligned}
\mathtt{P}_{2|\mathtt{H}_0} &= \left(1 - \frac{Q_1}{L_1} - \frac{Q_2}{L_2}\right)\left(1 + \overline{\mu}\,|\,\overline{\gamma}\,|^2\right)\mathtt{P}_{\mathtt{n}} \\
\mathtt{P}_{2|\mathtt{H}_1} &= \left(1 - \frac{Q_1}{L_1} - \frac{Q_2}{L_2}\right)\left(\left(1 + \overline{\mu}\,|\,\overline{\gamma}\,|^2\right)\mathtt{P}_{\mathtt{n}} + \mathtt{P}_{\mathtt{s}}\right)
\end{aligned}$$

Thus for a fixed value of L_1, PFA is given by

$$\mathtt{PFA} = \int_{\lambda}^{\infty} \mathtt{f}_{\hat{\mathtt{P}}_2|\mathtt{H}_0}(\nu)\,\mathrm{d}\nu \tag{5.41}$$



Optimization of First Stage Adaptation Time

Figure 5.4

and

$$\mathtt{PD} = \int_{\kappa}^{\infty} f_{\hat{p}_2|H_1}(\nu)\, d\nu \tag{5.42}$$

where each ROC is generated for a fixed value of L_1 by varying the detection threshold κ. An optimal choice for L_1 which optimizes detection performance follows from the ROC corresponding to the highest PD for the lowest PFA.

To demonstrate the ROC approach, set

$$P_s = .5$$

$$P_n = 1.0$$

$$|\overline{\gamma}|^2 = 2.0$$

and

$$t_d = 2;$$

where Q_1, Q_2, and L_2 are set to

$$Q_1 = 2$$

$$Q_2 = 9$$

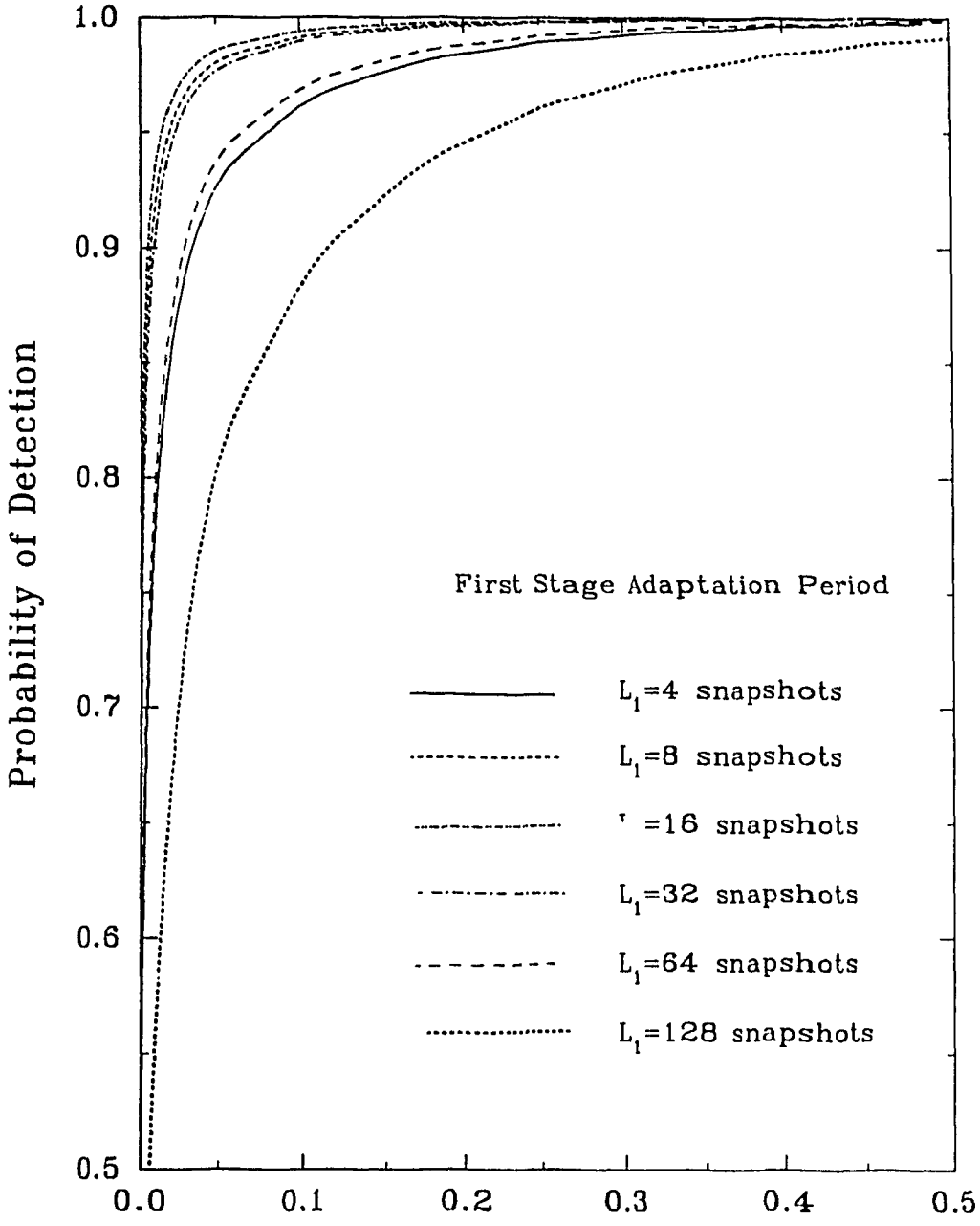
and

$$L_2 = 256$$

The adaptation time of the first stage is varied between 4 and 128. From (5.41) and (5.42) the corresponding set of ROCs are shown in Figure 5.5. Observe from the figure that $L = 16$ optimizes detection performance. Substituting the same scenario into the closed form expression given in (5.39), we find that $L_1 = 15.8$.

5.3.3 Convergence Rate and Dynamic Adaptation

Through our discussion in this and preceding chapters we have assumed block adaptation in the computation of stage weights. Such an approach to adaptation entails two difficulties. First, the beamformer output is delayed according to the number



Numerical Optimization of Adaptation Time

Figure 5.5

of input samples used to form the covariance matrix estimate of input data. Second, block adaptation requires a matrix inverse operation to compute beamformer weights. As an alternative, a large number of dynamic adaptation strategies, that recursively update beamformer weights from input data, have been proposed (eg: [12, 37] and [60]-[62]).

As suggested by Figure 5.4, the range of adaptation times for the first stage makes necessary an adaptation algorithm that is capable of a rapid rate of convergence. Moreover, the convergence rate of an adaptation scheme must be insensitive to the condition number, or ratio of maximum to minimum eigenvalues of the input covariance matrix. This property is critical, since covariance matrices with large condition numbers may result when high energy impulsive interference is present. If the convergence rate of an adaptation algorithm is larger than the scale-time required to suppress an interferer, then detection performance degrades. Consequently least-mean-square (LMS) algorithms, such as those suggested in [12, 37], are not ideal candidates because of condition number sensitivity. Self-orthogonalizing LMS routines provide greater robustness against large condition number situations [61, 62]. Alternatively, recursive least squares (RLS) adaptation is immune to ill-conditioned covariance matrices [52]. Unfortunately, the drawback of RLS is computational intensity. This however does not pose a great problem for the M^2VDR beamformer since RLS need only be applied to stages assigned a small number of adaptive DOF. For a stage assigned one DOF, RLS adaptation translates to scalar operations. In the next chapter we briefly examine the application of RLS adaptation to the M^2VDR beamformer.

Chapter 6

Simulation Results and Discussion

This chapter continues the investigation of impulsive interference suppression and the two stage M^{2}VDR beamformer by simulation experiment. We have in the preceding chapters identified two key aspects of M^{2}VDR beamformer design; adaptation time, and the signal blocking matrix. The first part of this chapter examines, through simulation the impact beamformer design, particularly mismatch in L_1 and $\bar{\mathbf{C}}_1$, has on performance. The second half of the chapter focuses on a comparative simulation study of detection and localization performance in an impulsive interference environment between the M^{2}VDR beamformer and the single time-scale GSC.

6.1 Background

Simulation experiments are based on a two stage M^{2}VDR narrow-band beamformer implemented in the frequency domain. Figure 5.1 depicts the geometry for our array, where $M = 12$ sensors and $J = 1$ taps. Inter-sensor spacing equals

$$\mathsf{d} = \frac{\mathsf{c}}{2f}$$

for $f = 1.0$ Hz. c equals propagation speed. To implement the beamformer we first partition sampled sensor outputs into non-overlapping segments of N samples each. Let

$$\mathbf{X}_0[k] \;=\; (\; \mathbf{x}_0[1+(k-1)N],\; \mathbf{x}_0[2+(k-1)N],\; \ldots\; .\mathbf{x}_0[kN]\;)$$

describe the k^{th} data segment or snapshot, where

$$\mathbf{x}_0[\ell+(k-1)N]$$

represents the $M\times 1$ sensor measurement vector at sampling instant $\ell+(k-1)N$. N is chosen to be much larger than the maximum propagation delay of a signal across the array. This ensures that the frequency coefficients, found by taking the discrete Fourier transform (DFT) of $\mathbf{X}_0[k]$, are effectively uncorrelated. Non-overlapping snapshots are used in this investigation to eliminate correlation between successive frequency estimates. In practice though, snapshots are overlapped to reduce variance in spectral power estimates. Taking the DFT along the rows of $\mathbf{X}_0[k]$, let

$$\boldsymbol{x}_0[k,f]$$

represent the $M\times 1$ vector of frequency coefficients of sensor outputs at frequency f. Over an observation period of L data segments, the covariance matrix or cross spectral density matrix (CSDM) sample estimate of $\boldsymbol{x}_0[k,f]$ takes the form

$$\hat{\mathbf{R}}_0(f) \;=\; \frac{1}{L}\sum_{k=1}^{L}\boldsymbol{x}_0[k,f]\boldsymbol{x}_0^{\dagger}[k,f]$$

Consider the response of the array to a plane wave arriving from θ. For the k^{th} data segment, we have

$$\boldsymbol{x}_0[k,f] \;=\; \alpha(f)\mathrm{r}_\theta(f)$$

where

$$\mathrm{r}_\theta(f) \;=\; \left(1,\; e^{j2\pi\frac{f\mathrm{d}}{\mathrm{c}}\sin(\theta)},\; e^{j2\pi\frac{2f\mathrm{d}}{\mathrm{c}}\sin(\theta)},\; \ldots\; e^{j2\pi\frac{(M-1)f\mathrm{d}}{\mathrm{c}}\sin(\theta)}\right)$$

For convenience let $\boldsymbol{x}_0[k, f = 1\text{Hz}] = \boldsymbol{x}_0[k]$. Suppose $\boldsymbol{x}_0[k]$ consists of the following statistically independent components

$$\boldsymbol{x}_0[k] = \mathbf{s}[k] + \mathbf{u}[k] + \mathbf{v}[k] + \boldsymbol{\eta}[k] \tag{6.1}$$

where $\mathbf{s}[k]$ represents the frequency coefficient vector of the desired the signal. Frequency coefficients of directional interferers are represented by a slowly time varying process $\mathbf{u}[k]$ and an impulsive component $\mathbf{v}[k]$. $\boldsymbol{\eta}[k]$ corresponds to frequency components of diffuse noise. Let $\mathbf{s}[k]$, $\mathbf{u}[k]$, and $\boldsymbol{\eta}[k]$ be modeled by wide sense stationary zero mean complex Gaussian random processes, such that

$$\mathcal{E}\left\{ \mathbf{s}[k]\mathbf{s}^{\dagger}[m] \right\} = \begin{cases} \mathsf{P}_{\mathbf{s}}\, \mathbf{1}_M \mathbf{1}_M^{\mathrm{T}}, & k = m \\ \mathbf{0}, & \text{otherwise} \end{cases} \tag{6.2}$$

$$\mathcal{E}\left\{ \mathbf{u}[k]\mathbf{u}^{\dagger}[m] \right\} = \begin{cases} \mathsf{P}_{\mathbf{u}}\, \mathbf{r}_{\theta_u}\mathbf{r}_{\theta_u}^{\dagger}, & k = m \\ \mathbf{0}, & \text{otherwise} \end{cases} \tag{6.3}$$

and

$$\mathcal{E}\left\{ \boldsymbol{\eta}[k]\boldsymbol{\eta}^{\dagger}[m] \right\} = \begin{cases} \mathsf{P}_{\mathbf{n}}\, \mathbf{I}_M, & k = m \\ \mathbf{0}, & \text{otherwise} \end{cases} \tag{6.4}$$

where $\mathbf{r}_{\theta_u} = \mathbf{r}_{\theta_u}(f = 1\text{Hz})$. We model the impulsive component $\mathbf{v}[k]$ by

$$\mathbf{v}[k] = \begin{cases} \sqrt{\mathsf{P}_{\mathbf{v}}}\, \mathbf{r}_{\theta_v}, & \text{for } T_1 \le k \le T_1 + t_{\mathrm{d}} - 1 \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

where $\mathsf{P}_{\mathbf{v}}$ is a constant. Note that t_{d}, the duration of the interferer, and T_1, the beginning point of the interferer, are now in units of snapshots.

In previous analysis, we computed beamformer weights from non-overlapping data blocks. An alternative to this approach is recursive adaptation. Since strong impulsive interference may lead to large eigenvalue spreads in CSDM estimates, adapting the first beamformer stage by least-mean-square (LMS) adaptation is precluded. We then turn to a recursive least squares (RLS) implementation, which is insensitive to the dynamic range of CSDM eigenvalues. In order for RLS adaptation to function correctly in an impulsive interference environment the scheme requires a minor

modification to operate as a Type I implementation. A RLS adaptation procedure is described in Appendix C.

6.2 Design Mismatch

To demonstrate the impact of design mismatch, consider a scenario consisting of a slowly time varying source and an impulsive event located at $\theta_u = -14^o$ and $\theta_v = -5^o$, respectively. From the previous discussion a signal component appears to the beamformer as a plane wave arriving from $\theta = 0^o$. Using this wavefield scenario set

$$\mathbf{w}_q = \frac{1}{\sqrt{12}}\mathbf{1}_{12}$$

and

$$\mathbf{C} = \frac{1}{\sqrt{12}}\mathbf{1}_{12}$$

Furthermore we set the duration of the impulsive event to $t_d = 1$ data segment. Components of the measured wavefield are scaled as follows

$$10\ \log_{10}\left(\frac{P_s}{P_n}\right) = 10\ \log_{10}(\overline{SNR}_{out}) = 15dB$$

$$10\ \log_{10}\left(\frac{P_u}{P_n}\right) = 15dB$$

and

$$10\ \log_{10}\left(\frac{P_v}{P_n}\right) = 30dB$$

The beamformer is described by

$$Q_1 = 1\ \text{DOF}$$

$$Q_2 = 10\ \text{DOF}$$

and

$$L_2 = 100\ \text{snapshots}$$

In the first part of our investigation we let the first stage signal blocking matrix $\overline{\mathbf{C}}_1$ equal

$$\overline{\mathbf{C}}_1 = \left(\mathbf{I}_{12} - \mathbf{C}\mathbf{C}^\dagger \right) \mathbf{r}_\theta \Big|_{\theta=36^\circ}$$

To see how $\overline{\mathbf{C}}_1$ responds to the different components of $\boldsymbol{x}_0[k]$ we plot the spatial magnitude response of the signal blocking matrix. The response of $\overline{\mathbf{C}}_1$ is given by

$$\overline{\mathbf{C}}_1^\dagger \mathbf{r}_\theta \mathbf{r}_\theta^\dagger \overline{\mathbf{C}}_1$$

from $-90 \leq \theta \leq 90$. Figure 6.1 shows a plot of the response. Notice first that the signal blocking matrix maintains the beamformer constraint by having zero response at the desired signal location (ie: $\theta = 0^\circ$). Observe as well that the slowly time varying interferer is located in a null of the response, ie:

$$\overline{\mathbf{C}}_1^\dagger \mathbf{r}_{\theta_u} \approx 0$$

Hence, $\mathbf{u}[k]$ appears effectively as a single component to the first stage. Since the impulsive interferer lies in a sidelobe of the signal blocking matrix response we say that $\overline{\mathbf{C}}_1$ is mismatched to $\mathbf{v}[k]$.

Assuming block adaptation, the optimal choice for L_1 follows from (5.39)

$$L_1 = \frac{3}{4(1 - Q_2/L_2)} Q_1 \left(1 + \sqrt{1 + \frac{8}{9Q_1} \left(\frac{L_2(1 + \overline{\mathrm{SNR}}_{\mathrm{out}})}{|\overline{\gamma}|^2} - l_{\mathrm{d}} \right) \left(1 - \frac{Q_2}{L_2} \right)} \right) \tag{6.5}$$

For the narrow-band beamformer we have

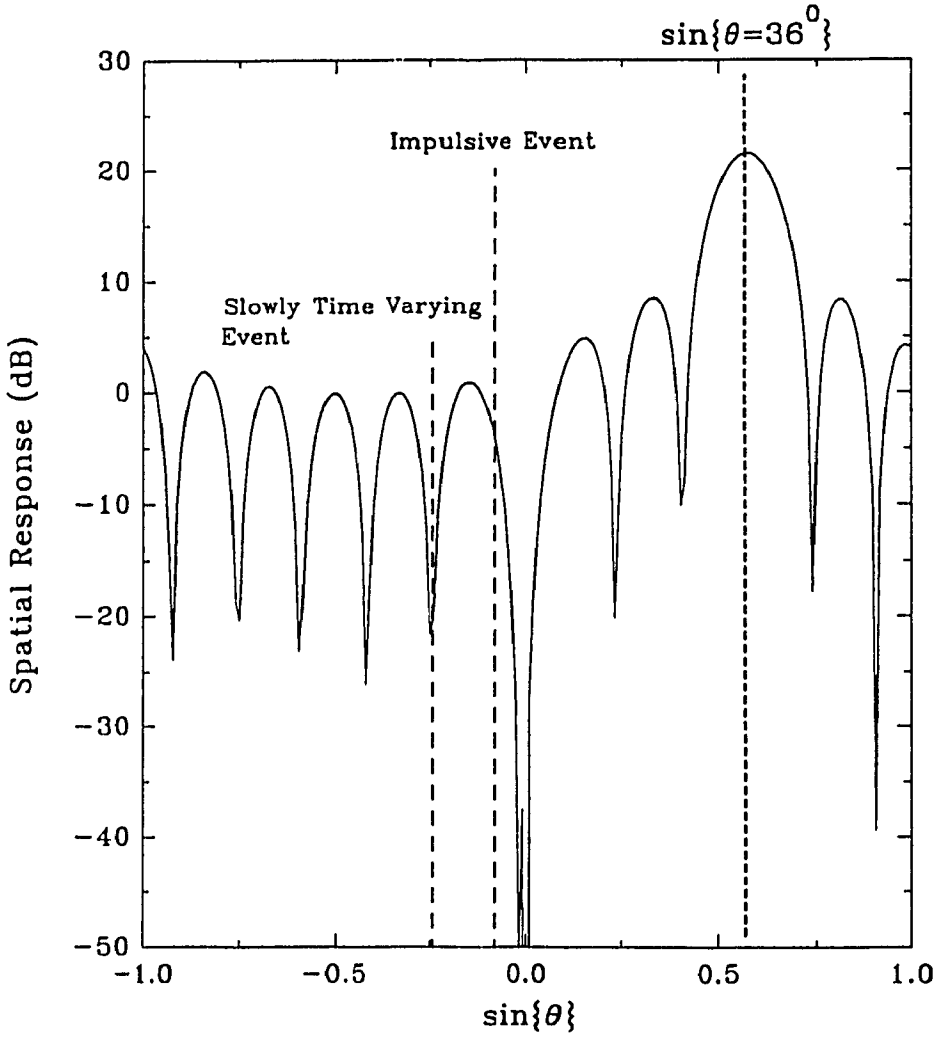
$$|\overline{\gamma}|^2 = \left| \mathrm{INR}_v \, \mathbf{w}_q^\dagger \mathbf{r}_{\theta_v} \mathbf{r}_{\theta_u}^\dagger \overline{\mathbf{C}}_1 \left(\mathrm{INR}_v \, \overline{\mathbf{C}}_1^\dagger \mathbf{r}_{\theta_u} \mathbf{r}_{\theta_u}^\dagger \overline{\mathbf{C}}_1 + 1 \right)^{-1} \right|^2$$

$$= \frac{\mathrm{INR}_v^2 \, \rho_v \overline{\rho}_v}{(\mathrm{INR}_v \overline{\rho}_v + 1)^2} \tag{6.6}$$

where

$$\rho_v = \mathbf{w}_\mathbf{q}^\dagger \mathbf{r}_{\theta_v} \mathbf{r}_{\theta_u}^\dagger \mathbf{w}_\mathbf{q}$$

$$\overline{\rho}_v = \overline{\mathbf{C}}_1^\dagger \mathbf{r}_{\theta_u} \mathbf{r}_{\theta_u}^\dagger \overline{\mathbf{C}}_1$$



Spatial Response: First Stage Signal Blocking Matrix

Figure 6.1

Beamformer	L_1
1	1
2	20
3	50

Table 6.1: First Stage Adaptation Time.

and

$$\mathtt{INR}_v \;=\; \frac{\mathtt{P_v}}{\mathtt{P_n}}$$

From (6.5)

$$L_1 \;\approx\; 20 \text{ snapshots}$$

Consider now the effect of varying L_1 over a range of values given in Table 6.1. For each entry in the table, we examine the evolution of the beamformer magnitude response and output noise power over a duration of 500 snapshots. To generate the magnitude response on a snapshot by snapshot basis we use a sliding rectangular window over the data input to each beamformer stage. Letting

$$\mathbf{w_a}[k] \;=\; \mathbf{w_q^\dagger}\left(\mathbf{I}_M - \overline{\mathbf{C}}_1\hat{\mathbf{B}}_1[k]\right)\left(\mathbf{I}_M - \overline{\mathbf{C}}_2\hat{\mathbf{B}}_2[k]\right)\mathbf{w_q} \tag{6.7}$$

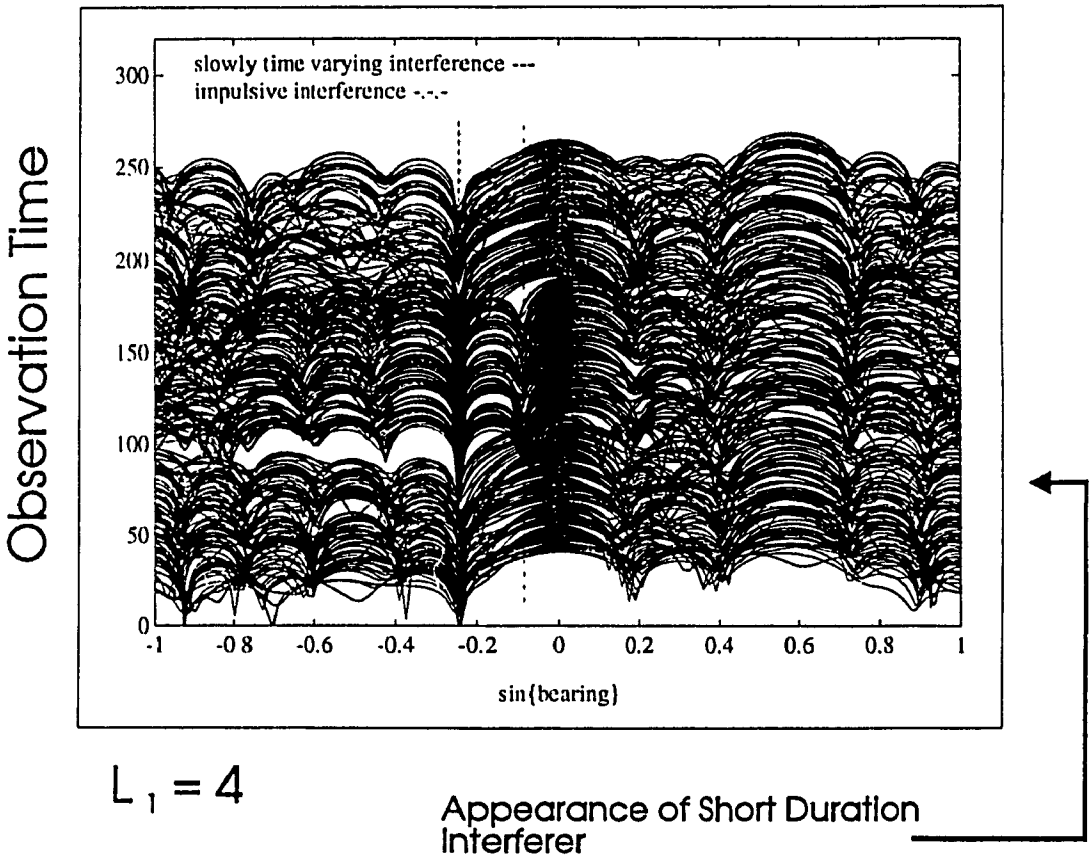
the magnitude response of the beamformer is

$$\mathtt{B}_\theta[k] \;=\; \mathbf{w_a^\dagger}[k]\mathbf{r}_\theta\mathbf{r}_\theta^\dagger\mathbf{w_a}[k]$$

Note that the maximum response of the beamformer is at $\theta = 0^o$. Output power due to diffuse noise at snapshot k follows from the inner product

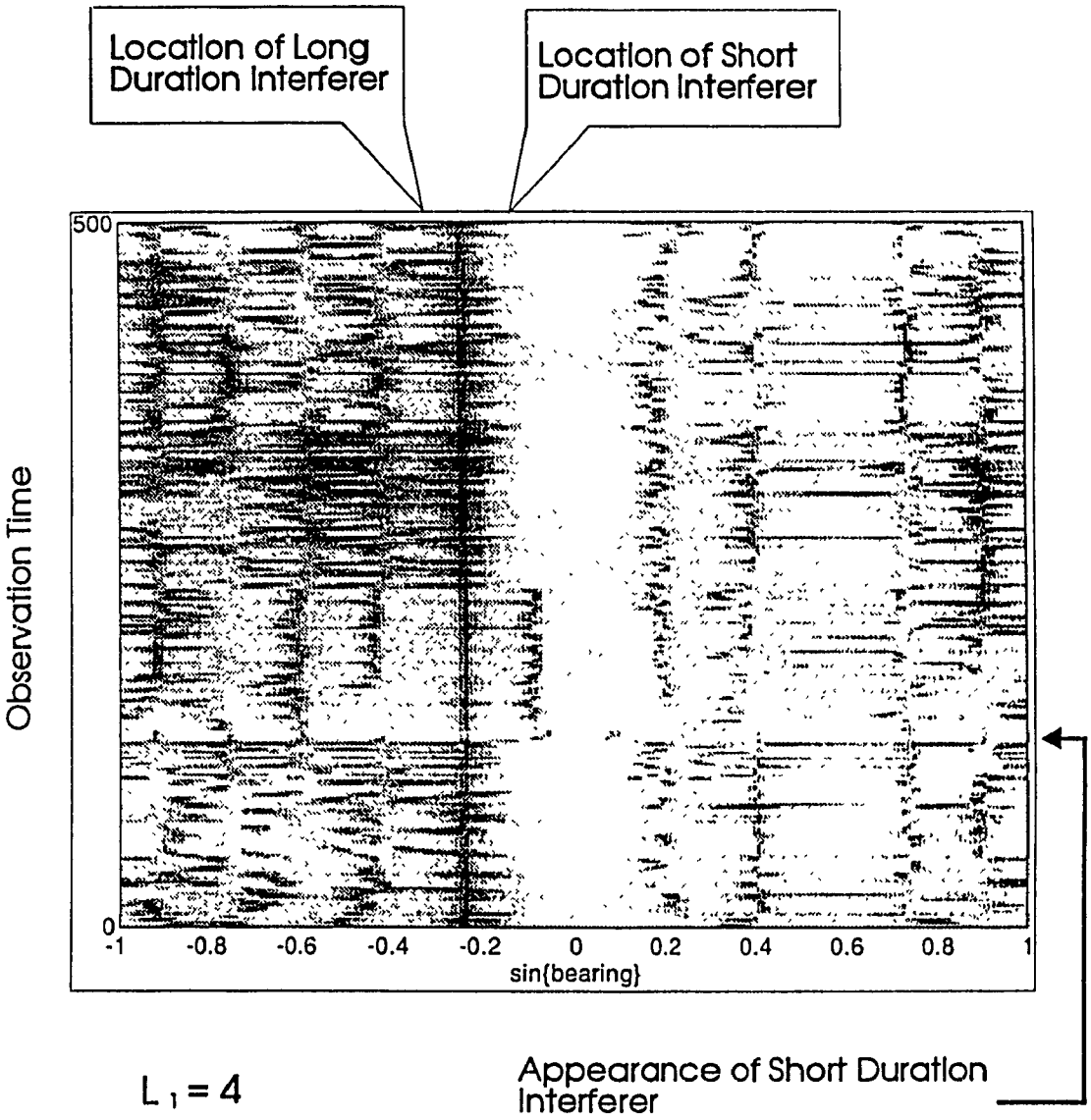
$$\mathtt{P_{out}}[k] \;=\; \mathtt{P_n}\mathbf{w_a^\dagger}[k]\mathbf{w_a}[k] \tag{6.8}$$

The evolution of the beamformer magnitude spatial response, under different adaptation times, is shown in the form of waterfall and grey level power maps in Figures 6.2 through 6.7. The grey level scale corresponds to Figures 1.6 and 1.7. The evolution of noise output power is summarized in Figure 6.8. From Figures 6.2, 6.3



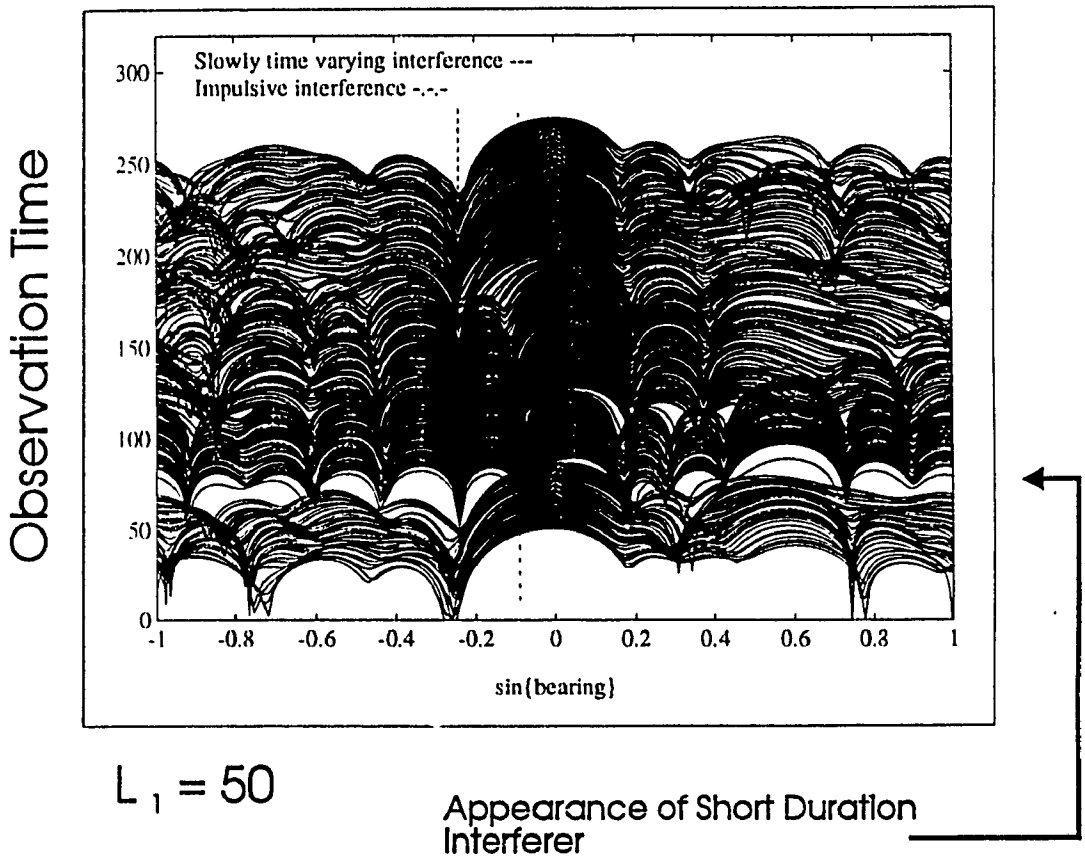
Magnitude Response (Waterfall Plot)

Figure 6.2



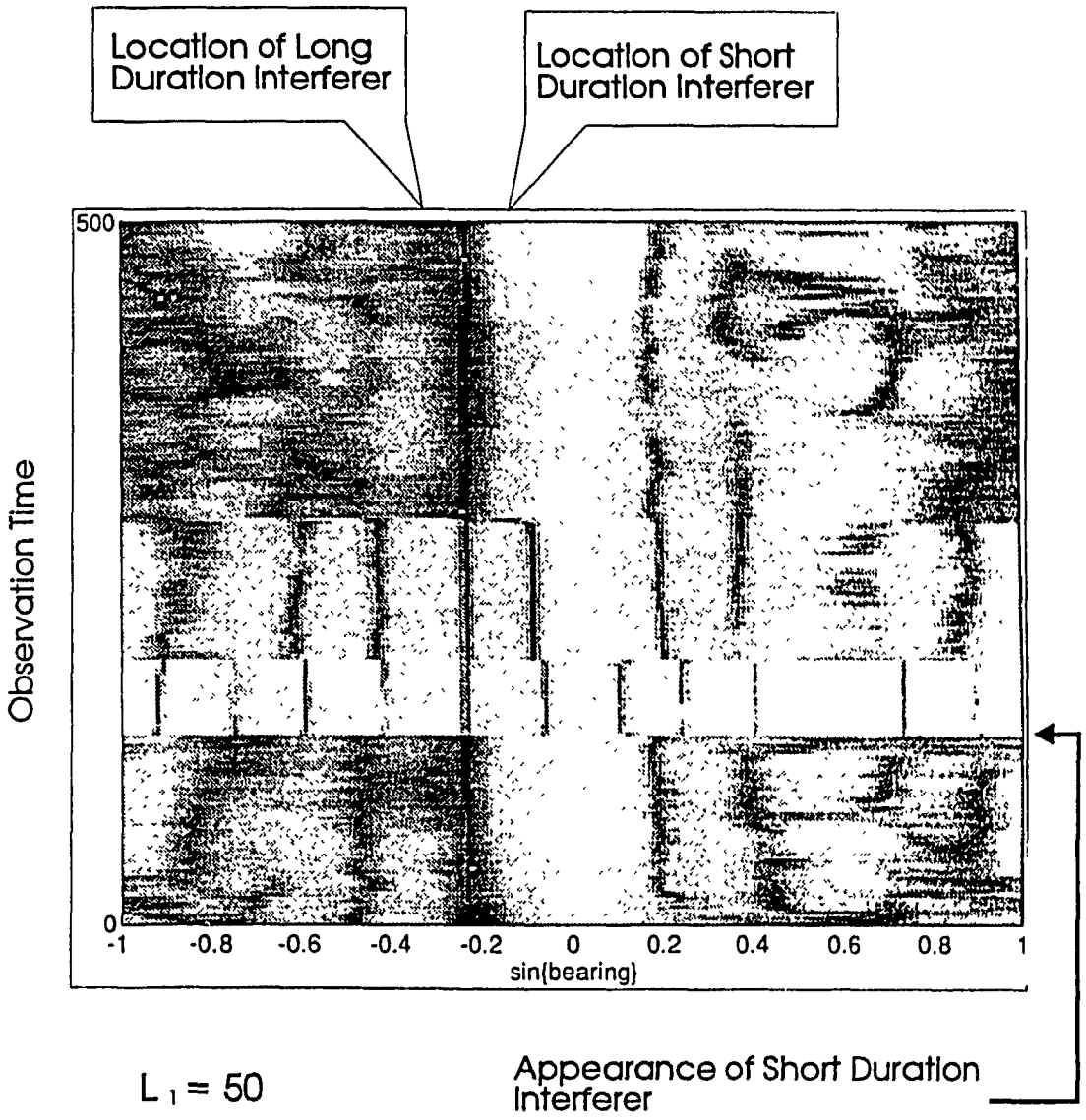
Magnitude Response - (Grey Level Map)

Figure 6.3



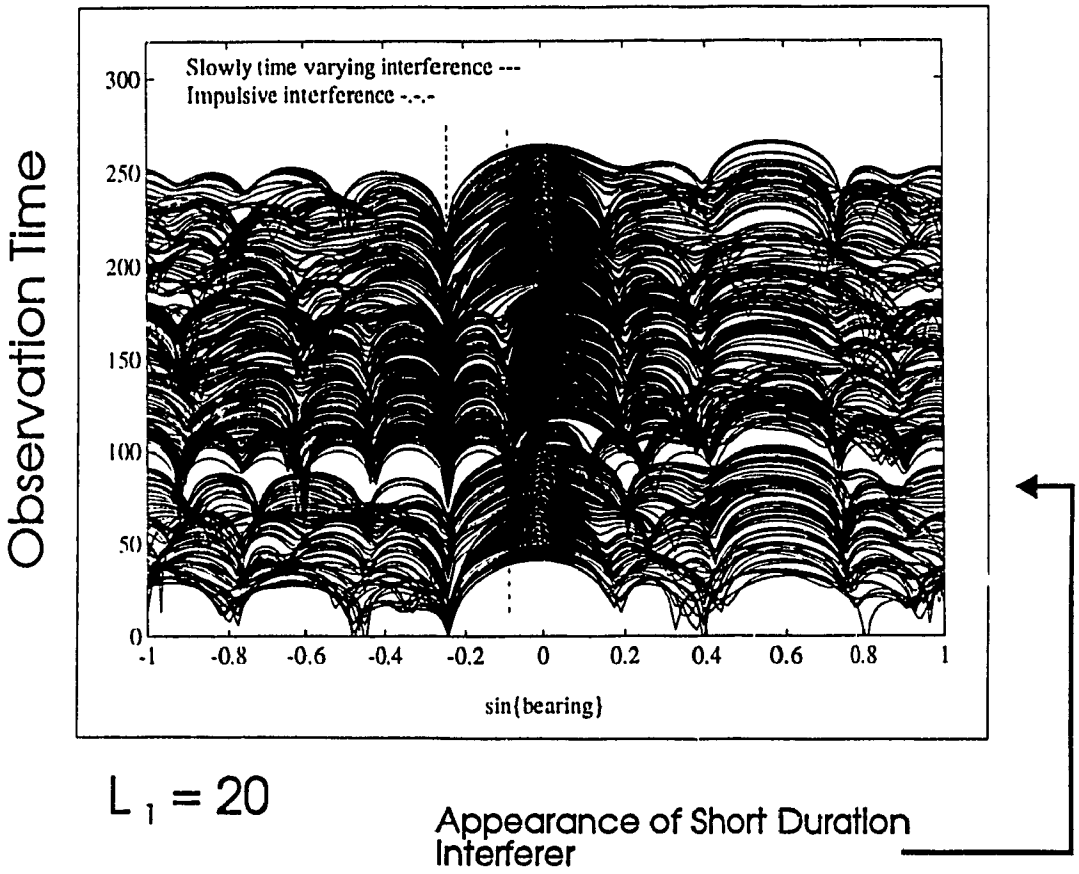
Magnitude Response (Waterfall Plot)

Figure 6.4



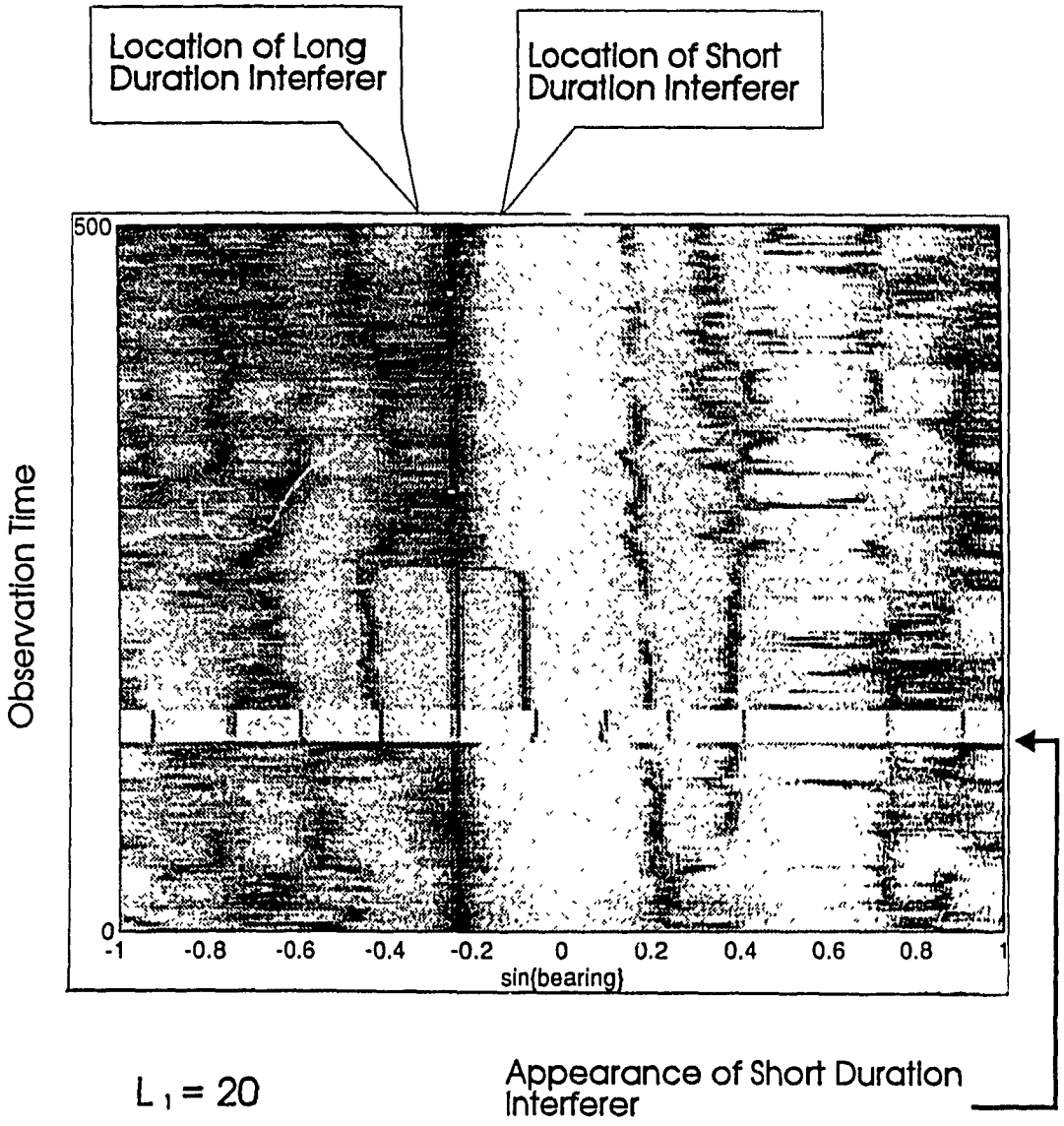
Magnitude Response - (Grey Level Map)

Figure 6.5



Magnitude Response (Waterfall Plot)

Figure 6.6



Magnitude Response - (Grey Level Map)

Figure 6.7

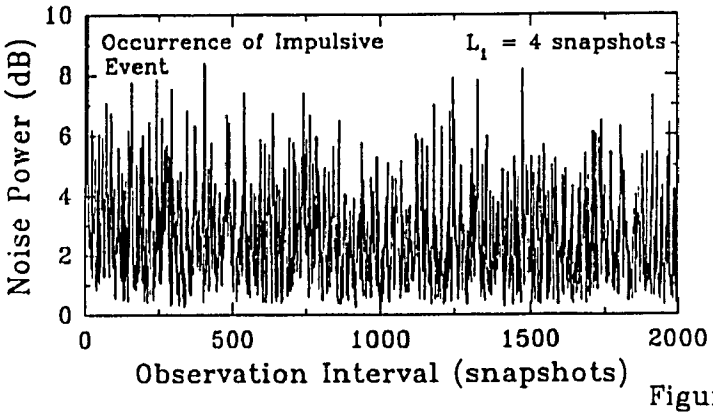


Figure 6.8a

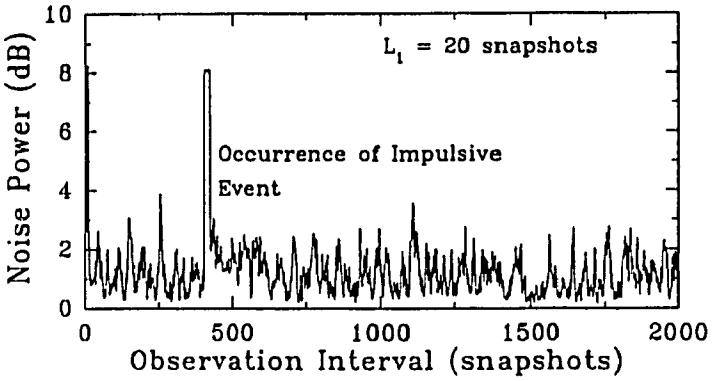


Figure 6.8b

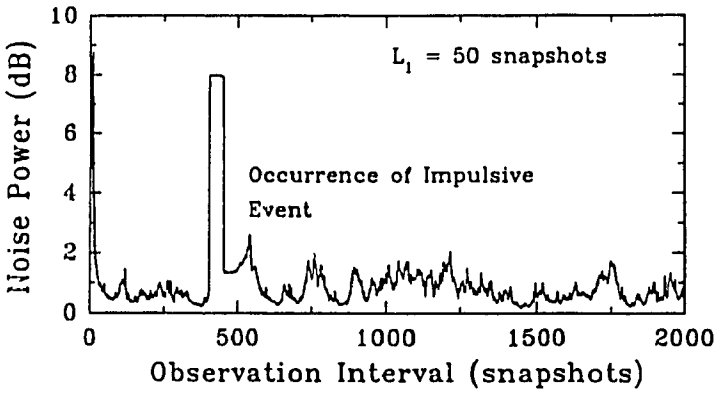


Figure 6.8c

Noise Output Power (Block Adaptation)
Figure 6.8

and 6.8a, we note the dominance of weight noise when $L_1 = 1$. Observe from Figure 6.3 the null in the magnitude response at θ_v. Since $\overline{\mathbf{C}}_1$ and $\mathbf{r}_{\theta_v}[n]$ are mismatched, the null appears after the event has vanished because of $\mathbf{v}[n]$ leaking into the second stage. As a consequence of interference leakage sidelobe levels remain unnecessarily elevated. However, the dominant feature here is weight noise variance. Increasing L_1 to 50 snapshots elevates sidelobe levels well past the duration of the impulsive interferer as indicated by Figures 6.4, 6.5 and 6.8c. Observe from these figures the reduction in sidelobe attenuation as indicated by the white regions on Figure 6.5 just following the appearance of the impulsive interferer. To optimally trade weight noise variance for detection performance in diffuse noise we set $L_1 = 20$. Results for this adaptation time are summarized in Figures 6.6, 6.7, and 6.8b. Our concern here is to minimize the degradation in detection performance after the impulsive interferer has vanished.

Repeating the experiments as outlined in Table 6.1, a summary of noise output power under RLS adaptation is shown in Figure 6.9. First, unlike the results obtained by block adaptation, weight noise is no longer a dominating factor. This is particularly evident in the case of $L_1 = 1$ snapshots. Second, we observe that output noise power following the impulsive event reflects the exponential weighting implicit in the RLS implementation outline in Appendix C. Although the effect of weight noise on output noise power is less than block adaptation, it is apparent from Figure 6.9 that the impulsive interferer illicits a longer response from the RLS algorithm than block adaptation. This results in an increase in output power over the duration of the response. Based on these observations the choice for L_1, as derived for block adaptation, may not provide an accurate tradeoff between weight noise variance and detection performance in diffuse noise in the case of RLS adaptation.

Our investigation so far has assumed a mismatch between $\overline{\mathbf{C}}_1$ and $\mathbf{v}[k]$. We now consider the case when the first stage blocking matrix is well matched to the

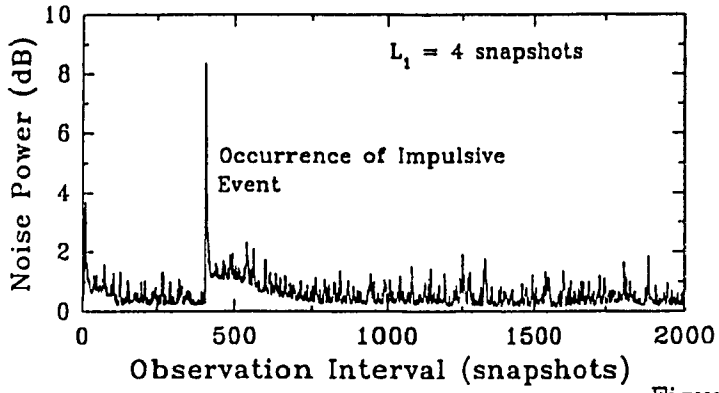


Figure 6.9a

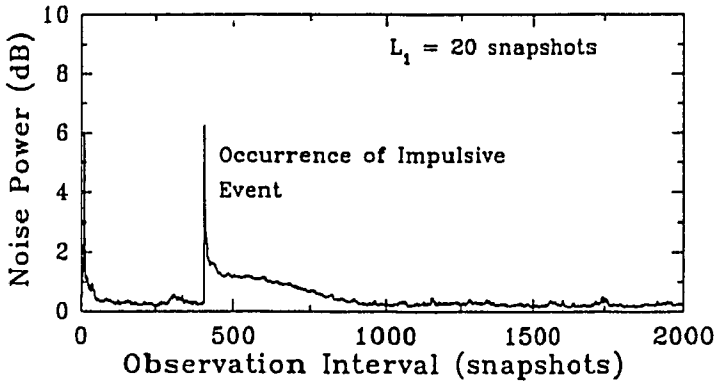


Figure 6.9b

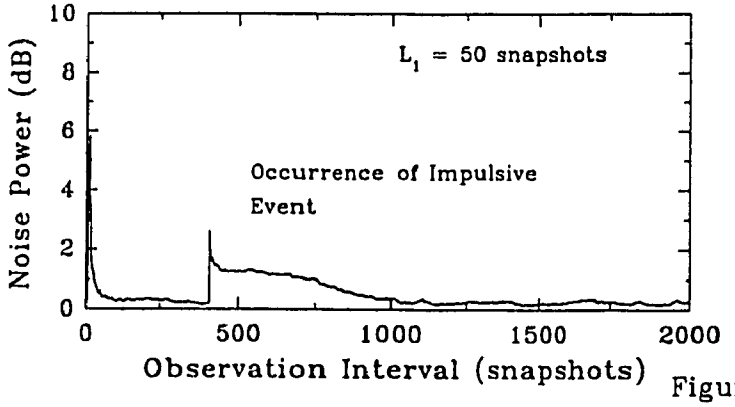


Figure 6.9c

Noise Output Power (RLS Adaptation)

Figure 6.9

impulsive interferer, ie:

$$\overline{\mathbf{C}}_1 \;=\; \left(\mathbf{I}_M \;-\; \mathbf{C}\mathbf{C}^{\dagger} \right) \mathbf{r}_{\theta} \Big|_{\theta=\theta_v}$$

The evolution of noise power, under block adaptation, for this choice of $\overline{\mathbf{C}}_1$ is shown in Figure 6.10, for values of L_1 given in Table 6.1. Evident from these results is the dominance of weight noise on performance for all values of L_1. From Figure 6.10 it appears that the best strategy here is to maximize adaptation time. Repeating our calculation of (6.5) we find that the optimal value for adaptation time approximately equals

$$L_1 \;\approx\; 100 \qquad \text{snapshots}$$

This suggests a strong coupling between the choice of signal blocking matrix and L_1. We investigate this result further by varying

$$\overline{\mathbf{C}}_1 \;=\; \left(\mathbf{I}_{12} \;-\; \mathbf{C}\mathbf{C}^{\dagger} \right) \mathbf{r}_{\theta} \tag{6.9}$$

from $-90^{o} \leq \theta \leq 90^{o}$. Figure 6.11 shows optimal values for L_1 for different choices of $\overline{\mathbf{C}}_1$ determined from (6.9). Of particular note is the increase in L_1 when $\overline{\mathbf{C}}_1$ is closely matched to the interference wavefront at $\theta = \theta_v$.

Set $L_1 = 100$ snapshots under the assumption that $\overline{\mathbf{C}}_1$ is optimally selected. We now continue our discussion by examining the impact an impulsive interferer has on the output power of the first stage after the interferer has vanished from sensor measurements. Immediately after the impulsive interferer has vanished from sensor measurements, we approximate output power of the first stage by

$$\mathsf{P}_{\mathsf{out}} \;=\; M\mathsf{P}_{\mathsf{s}} \;+\; \mathsf{P}_{\mathsf{n}} \left(1 \;+\; \left| \mathsf{P}_{\mathsf{v}} \mathbf{w}_{\mathbf{q}}^{\dagger} \mathbf{r}_{\theta_v} \mathbf{r}_{\theta_i}^{\dagger} \overline{\mathbf{C}}_1 \left(\mathsf{P}_{\mathsf{v}} \overline{\mathbf{C}}_1^{\dagger} \mathbf{r}_{\theta_v} \mathbf{r}_{\theta_v}^{\dagger} \overline{\mathbf{C}}_1 \;+\; \mathsf{P}_{\mathsf{n}} \right)^{-1} \right|^2 \right) \tag{6.9}$$

Given $\mathsf{C} = \frac{1_M}{\sqrt{12}}$ we rewrite $\overline{\mathbf{C}}_1^{\dagger} \mathbf{r}_{\theta_v}$ as

$$\begin{aligned} \overline{\mathbf{C}}_1^{\dagger} \mathbf{r}_{\theta_v} \;&=\; \mathbf{r}_{\theta_v}^{\dagger} \left(\mathbf{I}_M \;-\; \frac{1}{M} 1_M 1_M^{\mathsf{T}} \right) \mathbf{r}_{\theta_v} \\ &=\; M \;-\; \frac{1}{M} \mathbf{r}_{\theta_v}^{\dagger} \mathbf{1}_{12} \mathbf{1}_{12}^{\mathsf{T}} \mathbf{r}_{\theta_v} \\ &=\; M - M\cos^2(\xi) \;=\; M\sin^2(\xi) \end{aligned} \tag{6.10}$$

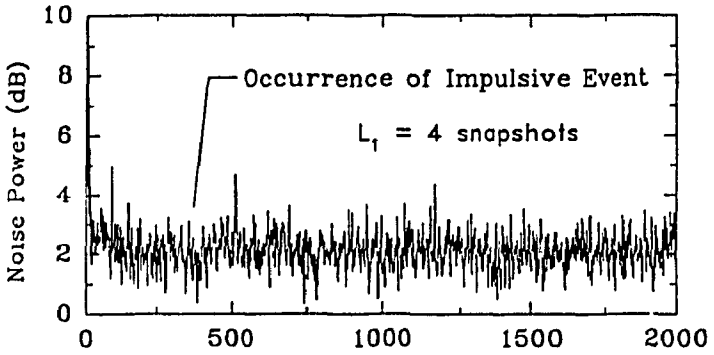


Figure 6.10a

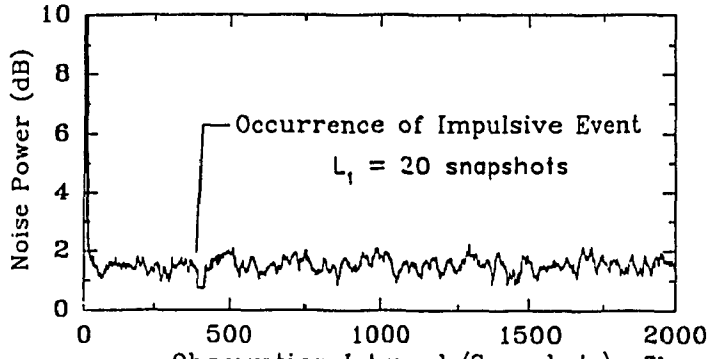


Figure 6.10b

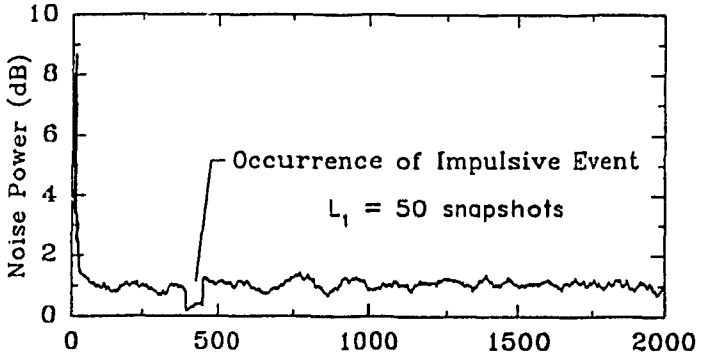
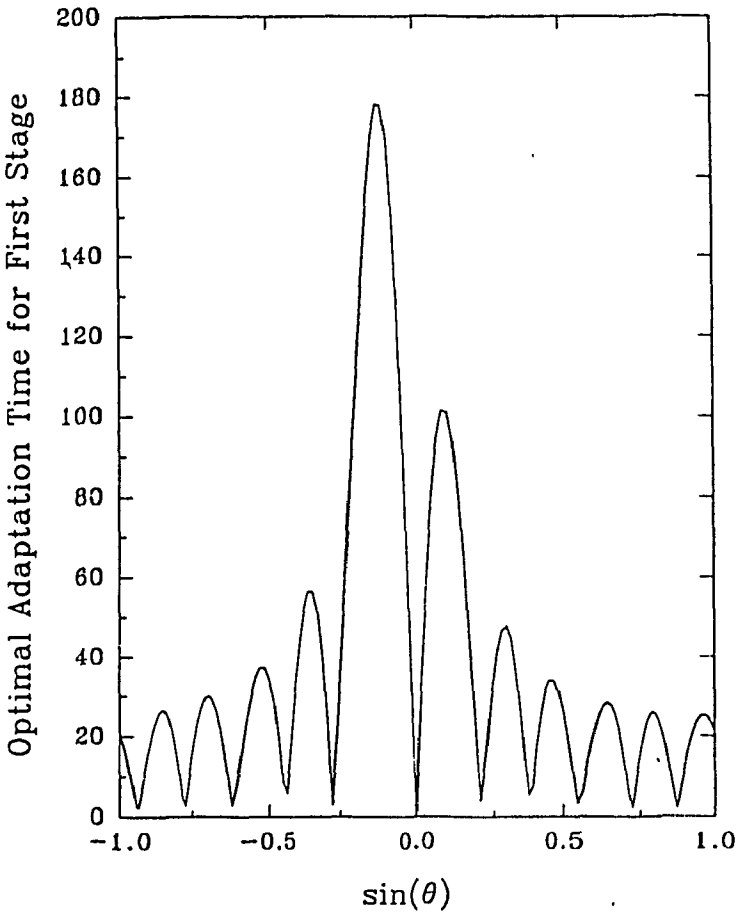


Figure 6.10c

Noise Output Power (Optimal Signal Blocking)

Figure 6.10



Relationship between Adaptation
Time of First Stage and the
Signal Blocking Matrix

Figure 6.11

Likewise

$$\begin{aligned} \mathbf{w}_q^\dagger \mathbf{r}_{\theta_v} \mathbf{r}_{\theta_v}^\dagger \mathbf{w}_q &= \frac{1}{M} \mathbf{1}_M^T \mathbf{r}_{\theta_v} \mathbf{r}_{\theta_v}^\dagger \mathbf{1}_M \\ &= M \cos^2(\xi) \end{aligned} \tag{6.11}$$

Substituting (6.10) and (6.11) into (6.9),

$$\mathsf{P}_{\mathsf{out}} = M\mathsf{P}_s + \mathsf{P}_n \left(1 + M \left(\frac{M\mathsf{INR}_v \sin^2(\xi)}{M\mathsf{INR}_v \sin^2(\xi) + 1} \right)^2 \cos^2(\xi) \right) \tag{6.12}$$

Since θ_v is in close proximity to $\theta = 0^o$ we have

$$\sin^2(\xi) \approx 0$$

and

$$\mathsf{P}_{\mathsf{out}} \approx M\mathsf{P}_s + \mathsf{P}_n$$

for the first stage assuming an optimal choice for signal blocking matrix. In correctly selecting $\overline{\mathbf{C}}_1$ the affect of $\mathbf{v}[n]$ on output power is shown to be negligible. Consequently the impact of $\mathbf{v}[k]$ on the selection of L_1 is minimal, making the minimization of weight noise variance the principle priority in optimizing adaptation time.

6.3 Spatial Power Estimates

An important feature of an adaptive beamformer is the ability to detect and localize signals under strong interference conditions. In this discussion we focus our simulation study on the spatial power estimates made by the M^{2}VDR beamformer and the single time-scale GSC when strong impulsive components are present in sensor measurements. Power directionality maps, relating a power estimate to a spatial location, are employed to provide a qualitative assessment of performance. We assume 2 directional sources in the wavefield. Table 6.2 summarizes three scenarios of interest. In scenario 1 we observe a weak, slowly time varying source and a strong

Scenario	*Source 1*			*Source 2*		
	Source Bearing	SNR (dB)	Duration (snapshots)	Source Bearing	SNR (dB)	Duration (snapshots)
1	-53^o	5	100	-43^o	30	100
2	-53^o	10	100	-57^o	30	100
3	-53^o	10	100	-57^o	30	1

Table 6.2: Wavefield Scenarios.

impulsive event. For the next two cases given in Table 6.2, resolution performance is evaluated by considering two strong sources placed in close proximity to each other.

We set the beamformer to

$$Q_1 = 1$$

$$Q_2 = 10$$

$$L_1 = 20$$

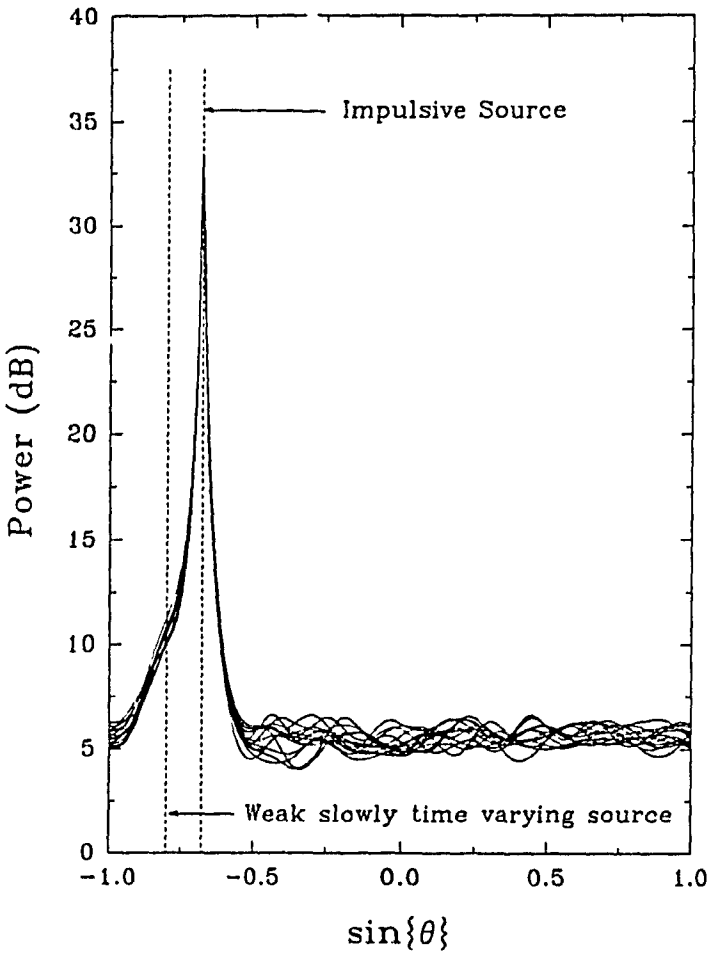
and

$$L_2 = 100$$

$\overline{\mathbf{C}}_1$ is selected from a eigenvector of (4.21), and the signal blocking matrix associated with the GSC is defined from (4.24). Adaptation time of the GSC equals $L = 100$.

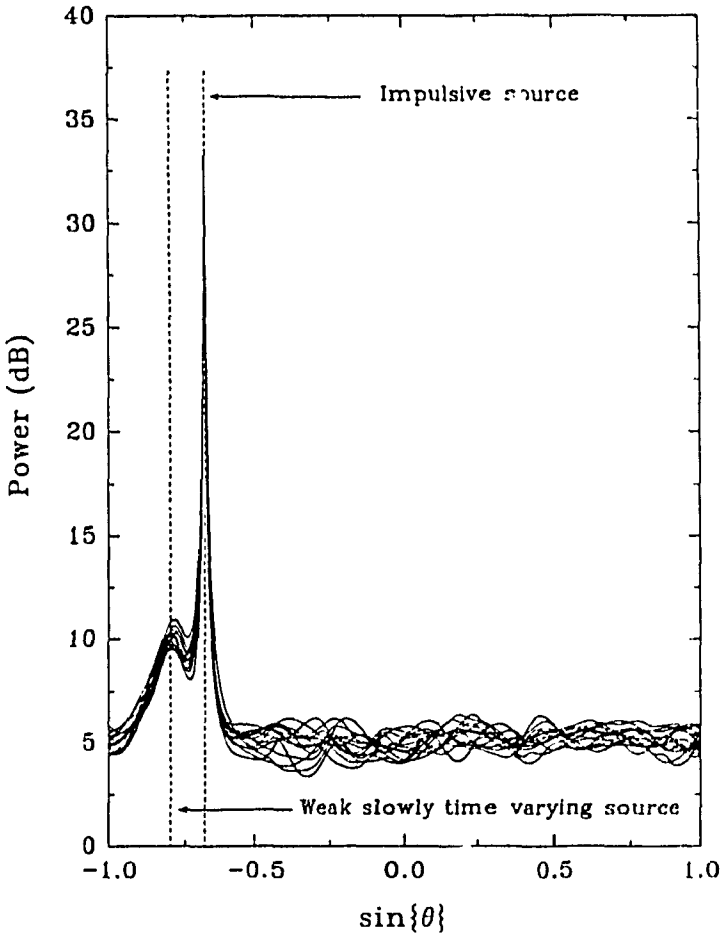
Power directionality maps for scenario 1 are shown in Figures 6.12 and 6.13, where results from 25 independent experiments have been overlaid. Observe from Figure 6.12 the inability of the GSC to detect the weak signal. We attributed this degradation in detection performance to increased noise output power.

Resolution performance describes the ability of the beamformer to discriminate closely spaced sources. Consider first scenario 2. In Figures 6.14 and 6.15 closely both the GSC and M^{2}VDR beamformer are unable to discern the presence of two slowly time varying sources. Replacing one source with an impulsive event, as indicated by scenario 3, observe from Figures 6.16 and 6.17 that the M^{2}VDR beamformer discriminates between the two sources in this case. To correctly interpret



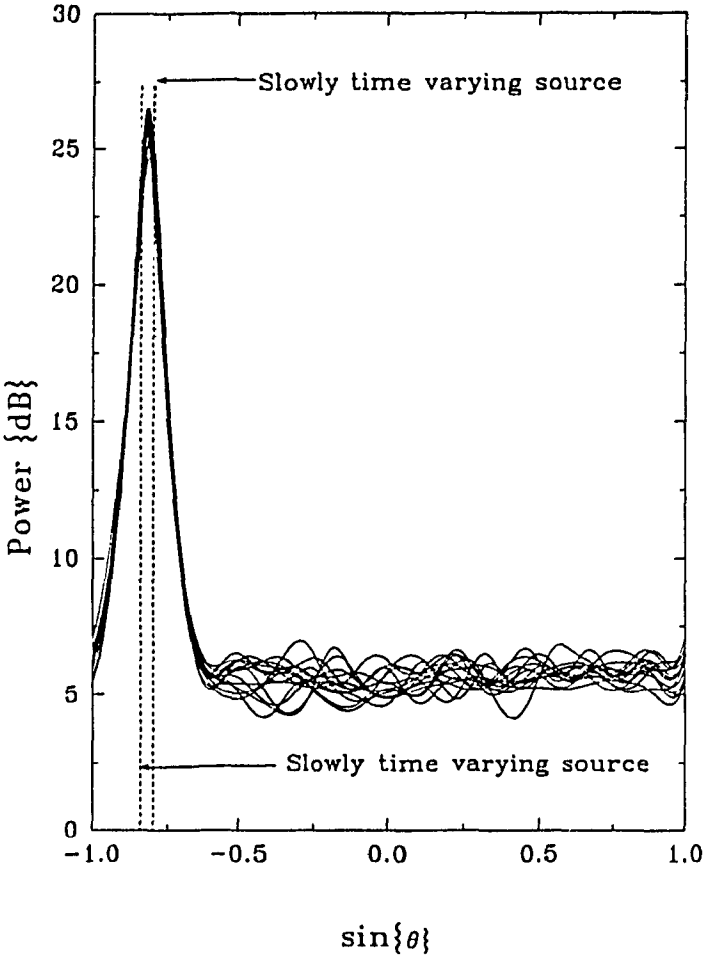
Detection: GSC

Figure 6.12



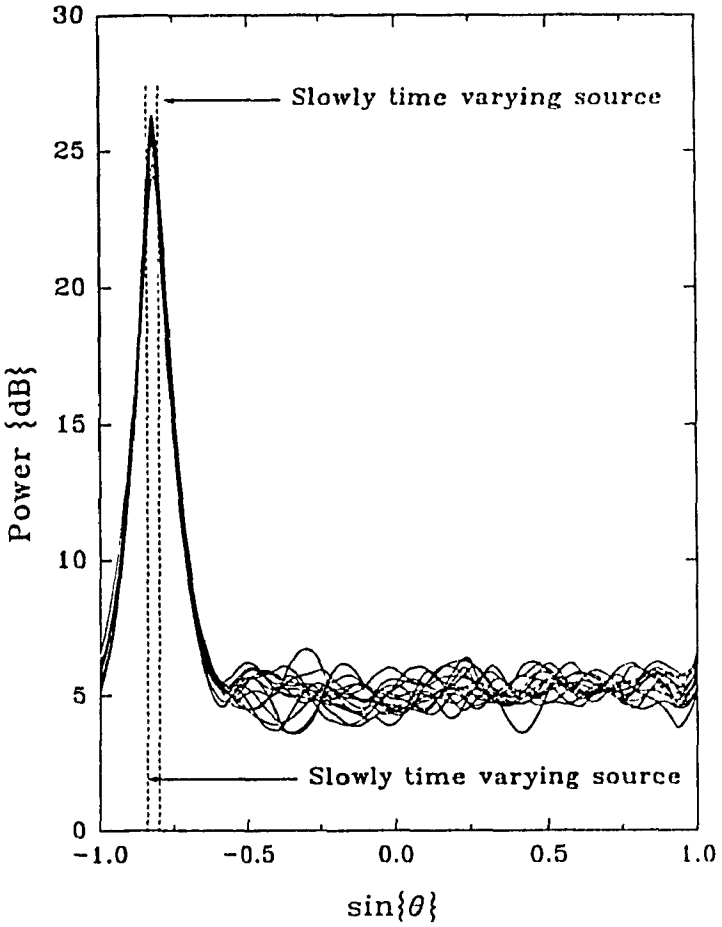
Detection: Two Stage M^2VDR Beamformer

Figure 6.13



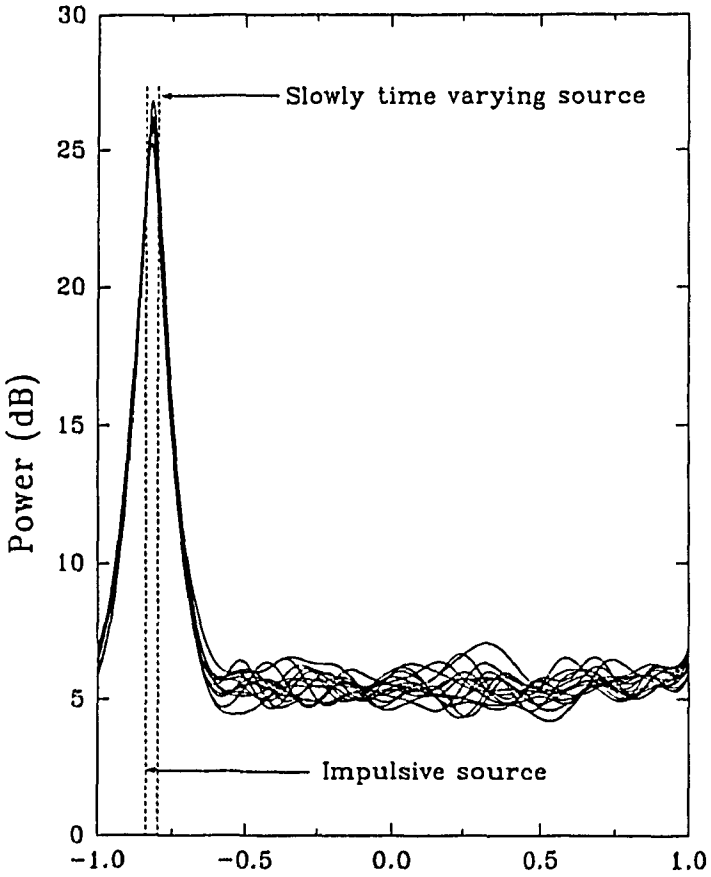
Resolution: GSC

Figure 6.14



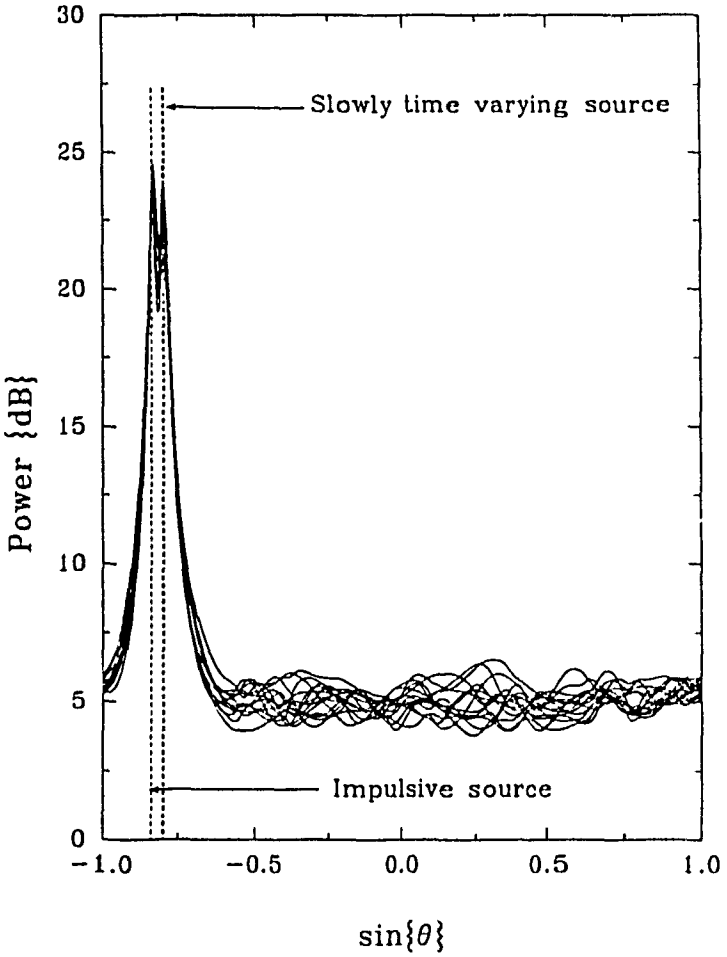
Resolution: Two Stage M^2VDR Beamformer

Figure 6.15



Resolution: GSC

Figure 6.16



Resolution: Two Stage M^2VDR Beamformer

Figure 6.17

the difference between Figures 6.16 and 6.17 recall that resolution performance of the MVDR beamformer is dependent on output SNR. In the case of the GSC, a null will be maintained in the direction of the impulsive source over $L \gg t_d$. Consequently, as we have previously seen, noise power at the output of the beamformer increases because of reduced attenuation in the response sidelobes. This has the effect of reducing output SNR and thus degrading resolution performance. Since noise output power is elevated for a far short period in the case of the M^{2}VDR beamformer, output SNR will be higher than for the GSC. We would then expect better resolution performance from the M^{2}VDR beamformer.

Chapter 7

Summary

The application of adaptive beamforming to the mixed time-scale interference environment has been addressed. The problem in such a setting arises when large differences in temporal extent between wavefield components exist. This presents a difficulty for MVDR beamformer realizations restricted to computing a fixed number of weights simultaneously on a single time-scale. When adaptation time is greater than the time-scale of an interferer an unnecessary degradation in detection performance may occur. Unfortunately reducing adaptation time does not remedy the problem. The reason for this is the increase in weight noise variance accompanying the change in adaptation time. The increase in variance may introduce a greater loss in performance than is gained through a reduction in adaptation time. For this reason a MVDR implementation limited to simultaneously computing all adaptive weights on one time-scale is not an ideal candidate for mixed time scale interference environments.

This thesis introduces a multiple stage beamformer that is capable of simultaneously adapting subsets of DOF on different time-scales. Because of the ability to adapt on different time-scales, the M^2VDR beamformer has an advantage in short duration interference environments over previous MVDR realizations. For instance, to cancel short duration interference, the beamformer computes a small number of

weights assigned to a stage on a time-scale approximating the duration of the interferer. With fewer weights, the beamformer stage introduces the possibility of larger reductions in adaptation time then could be possible for a GSC implementation that must adapt $MJ - K$ DOF. Furthermore, the M^{2}VDR realization does not curtail the ability to respond to multiple interferers since the beamformer applies the remaining DOF, retained in other stages, against long duration interference.

Strong impulsive interference has a particularly detrimental affect on detection performance. The principle reason for this is the short duration of the event, and the slow convergence rate of a single time-scale MVDR adapting a large number of DOF. To deal with impulsive type interference we propose a two stage M^{2}VDR beamformer. The first stage is assigned a rapid rate of adaptation and correspondingly a small number of DOF. The succeeding stage functions as a conventional MVDR beamformer in the sense that remaining DOF are adapted over a temporal duration matched to slowly time varying interferers.

Critical to the performance of the M^{2}VDR beamformer is the adaptation scheme used to compute weights, adaptation time, and signal blocking matrix. In this thesis we have discussed two adaptation schemes. Type I adaptation uses the same input data block to compute weights and update the output. Type II adaptation computes stage weights based on a past input data block. Correlation between the adaptive weights and the input data block characterizes Type I adaptation. This causes output power to be scaled. However, a comparison of SNR, generated by Type I and II implementations, indicates that both schemes yield equivalent output signal to noise ratio. In an impulsive interference environment Type I adaptation has an important advantage over Type II. Since the temporal duration of the impulsive interferer is short, weights computed by a Type II implementation may overlook the interferer since the weights are not synchronized with input data samples. Although Type II weights place a null in the location of the interferer, this will occur after the impulsive interference components have passed through the beamformer. This problem is

eliminated by synchronizing input data with the computation of beamformer weights (ie: Type I adaptation).

To ensure that each beamformer stage adapts on a different time-scale, stages must be arranged in increasing order of adaptation time. Hence, the adaptation time of a stage must be greater than or equal to adaptation times of preceding stages. Otherwise the adaptation of DOF will be coupled to the other stages. In addition, when stages are not correctly arranged, global minimum mean-squares error cannot be achieved.

In selecting an appropriate adaptation time for a rapidly adapting stage, two factors must be considered. First, when adaptation time is greater than the duration of an interferer, output power remains unnecessarily elevated after the event has vanished. Second, decreasing adaptation time causes weight noise variance to increase. In both these cases detection performance may degrade. To select adaptation time we propose an optimization criterion that trades weight noise variance against detection performance in diffuse noise. The approach selects an adaptation time that corresponds to the ROC representing the minimum PFA and maximum PD during a short duration interference event. Modeling beamformer output power by a Gaussian random variable, we derive a closed form solution for adaptation time. When a Gaussian assumption is not valid, the optimal choice for adaptation is derived numerically.

Another critical element for the M^{2}VDR beamformer is the design of the signal blocking matrix. In this design problem we recall two important characteristics of the M^{2}VDR beamformer. Even though the beamformer uses all DOF, $Q = MJ - K$, each stage operates as a partially adaptive beamformer, $Q_r < Q$. Second, a rapidly adapting stage can only use a few DOF to ensure that the effects of weight noise variance on detection performance are minimized. In designing a signal blocking matrix for a rapidly stage we must ensure that DOF assigned to the stage are efficiently used. Otherwise, interference suppression degrades. Along with reducing

interference suppression, a poorly designed signal blocking matrix may result in interference appearing in the adaptive weights of subsequent stages. This has the effect of prolonging the duration a response null is maintained in the direction of a short duration interferer. In this thesis we derive approximate expression relating AG to signal blocking matrix mismatch.

7.1 Future Extensions

A limiting factor affecting the performance of the M^{2}VDR beamformer is the match between the signal blocking matrix and the interference wavefield. Any future investigation of the beamformer must deal with the problem of partially adaptive design. Although extensive literature exists for the design of partially adaptive beamformers, the problem posed by the M^{2}VDR is unique. In addition to providing a efficient use of DOF at each stage, a partially adaptive design must account for any coupling between stages. Otherwise short duration interference components will leak into stages adapting over a longer time-scale. One possible approach would be to adaptively select the column elements of the first beamformer stage. In the case of a narrow-band beamformer this would simply imply a steering operation. However, adapting the signal blocking structure in a wide-band framework presents a far more challenging problem.

Recently multichannel parametric modeling has been applied to the array processing problem (ie: [63]-[65]). The primary motivation for applying multichannel parametric modeling is to decrease the statistical variability of wideband covariance matrix estimate through low order parametric models. As with previous work in adaptive beamforming, the approaches taken in [63]-[65] assume that the components of the measured wavefield are stationary over the same time period. An attempt at applying Kalman filtering to account for the time varying nature of the wavefield has been considered in [65]. Given the success of the multiple time-scale

beamformer in dealing with a mixed time-scale environment we are motivated to consider a multiple time-scale version of multichannel parametric techniques. Such an approach would partition the state space of the parametric model in such a way as to adapt subsets of state variables over different time-scales.

Throughout this thesis we have treated impulsive interference as a singular event. A discussion of the frequency of occurrence or other statistical properties of impulsive noise has not been included as part of designing the signal blocking matrix and adaptation time. Statistics of impulsive noise have been used (eg: [53, 66]) to formulate appropriate detectors under such noise conditions. To improve upon the design of the M^2VDR beamformer we suggest that an explicit model of impulsive noise be used (see [66]). In addition to providing an accurate model for signal blocking design, optimization of adaptation will be able to taken into account the temporal statistics of impulsive noise.

This thesis has viewed the M^2VDR beamformer as a vehicle for multiple time-scale adaptation. Alternatively, the decomposition can be used to design a dynamic adaptation scheme that permits a rapid rate of convergence without sensitivity to the input covariance matrix condition number. Self-orthogonalizing LMS strategies have been proposed as solution to this problem (eg:[62]). As suggested by our application of RLS in chapter 6, an alternative approach could be considered. Suppose we decompose the GSC into as many stages as there are adaptive DOF. By implementing RLS at each stage, adaptation becomes a scalar operation. This eliminates the issue of computational complexity. Furthermore, we decouple, or orthogonalize, each stage by selecting signal blocking matrices that are orthogonal to each other. Such a scheme allows for rapid convergence rates with immunity to the condition number of the input covariance matrix.

In conclusion this thesis has shown that a conventional MVDR beamformer is not the ideal beamformer structure to optimize detection performance when the time-scale of wavefield components differ substantially. Under such conditions, the

multiple time-scale M^2VDR beamformer introduces a greater degree of flexibility in optimizing detection performance.

Bibliography

[1] W. Vanderkulk, *"Optimal processing for acoustic arrays"*, I. R. E. , pp. 285-292, Oct. 1963.

[2] D. Middleton, J. H. Sampson, *"Detection of random acoustic signals by receivers with distributed elements"*, J. Acoust. Soc. America, pp. 727-737, Nov. 1965.

[3] W. S. Hodgkiss, L. W. Nolte, *"Bayes optimum and maximum-likelihood approaches in an array processing problem"*, IEEE Trans. Aerospace and Elec. Sys. , pp. 913-916, Sept. 1975.

[4] F. Bryn, *"Optimal signal processing of three-dimensional arrays operating on gaussian signals and noise"*, J. Acoust. Soc. America, pp. 289-297, March 1962.

[5] D. J. Edelblute, J. M. Fisk, and G. L. Kinnison, *"Criteria for optimal signal detection theory for arrays"*, J. Acoust. Soc. America, pp. 199-205, Jan. 1967.

[6] R. N. McDonough, *"A cononical form of the likelihood detector for gaussian random vectors"*, J. Acoust. Soc. America, pp. 402-406, Feb. 1971.

[7] W. J. Bangs, **Array Processing with Generalized Beamformers"**, Ph. D. Thesis, Yale University, 1971.

[8] V. H. MacDonald, P. M. Schultheiss, *"Optimal passive bearing estimation in a spatially incoherent noise environment"*, J. Acoust. Soc. America, pp. 37-43, July 1969.

[9] P. W. Howells, "*Intermediate frequency sidelobe canceller*", U. S. A. Patent 3202990, August 1966.

[10] S. P. Applebaum, **Adaptive Arrays**, Special Projects Lab. , Syracuse University, Res. Corp. , Rep. SPL TR 66-1, 1965.

[11] B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode, "*Adaptive antenna systems*", Proc. IEEE, pp. 2143-2159, Dec. 1967.

[12] O. L. Frost, "*An algorithm for linearly constrained adaptive array processing*", Proc. IEEE, pp. 926-935, August 1972.

[13] L. E. Brennan, I. S. Reed, "*Theory of adaptive radar*" IEEE Trans. Aero. Electron. Sys. , pp. 237-252, March 1973.

[14] R. A. Monzingo and T. W. Miller, **An Introduction to Adaptive Arrays**, Wiley, 1980.

[15] N. L. Owsley, in **Array Signal Processing**, S. Haykin (ed), Prentice Hall, 1985.

[16] J. Capon, "*High-resolution frequency-wavenumber spectrum analysis*", Proc. IEEE, pp. 1408-1418, 1969.

[17] R. O. Schmidt, "*Multiple emitter location and signal parameter estimation*", Proc. RADC Spectrum Estimation Workshop, Rome, New York, pp. 243-258, 1979.

[18] G. Bienvenu, L. Kopp, "*Optimality of high resolution array processing using the eigensystem approach*", IEEE Trans. ASSP, pp. 1235-1248, Oct. 1983.

[19] D. H. Johnson, S. Graaf, "*Improving the resolution performance of bearing in passive sonar arrays by eigenvalue analysis*", IEEE Trans. ASSP, pp. 638-647, Aug. 1982.

[20] D. H. Johnson, D. E. Dudgeon, **Array Signal Processing - Concepts and Techniques**, Prentice Hall, 1993.

[21] N. L. Owsley, "*Overview of adaptive array processing*", in NATO ASI **Adaptive Methods in Underwater Acoustics**, H. G. Urban (ed.), pp. 355-374, 1985.

[22] G. Bienvenu, "*Passive array processing: from conventional to high resolution concepts*" in NATO ASI **Underwater Acoustic Data Processing**, Y. T. Chan (ed.), pp. 329-343, 1989.

[23] H. Cox, " *Sensitivity considerations in adaptive beamforming*", in NATO ASI **Signal Processing**, J. Griffiths, P. Stocklin, and C. van Schooneveld (eds.), pp. 619-645, 1973.

[24] H. Cox, "*Resolving power and sensitivity of mismatch of optimum array processors*", J. Acoust. Soc. America, Vol. 54, pp. 771-785, 1973.

[25] B. Widrow, "*A review of adaptive antennas*", in NATO ASI **Underwater Acoustics and Signal Processing**, L. Bjorno (ed.), , pp. 287-306, 1981.

[26] B. Widrow, K. Duvall, R. Gooch, and W. Newman, "*Signal cancellation phenomena in adaptive antennas: Causes and cures*", IEEE Trans Ant. and Prop. , pp. 469-478, May 1982.

[27] M. J. Levin, **Maximum Likelihood Array Processing**", M. I. T. Lincoln Lab Tech Rep. DDC 455763, Dec. 1964.

[28] P. E. Green, E. J. Kelly, M. J. Levin, "*A comparison of seismic array processing methods*", Geophys. J. Roy. Astron. Soc. , Vol. 11, pp. 67-84, 1966.

[29] M. H. Er, A. Cantoni, "*Derivative constraints for broadband element space antenna array processors*", IEEE Trans. Acoustic Sig. Proc. , pp. 1378-1393, Dec. 1983.

[30] K. Buckley, "*Spatial/spectral filtering with linearly-constrained minimum variance beamformers*", IEEE Trans. ASSP, pp. 249-266, March 1987.

[31] H. Cox, R. Zeskind, M. Owen, "*Robust adaptive beamforming*", IEEE Trans. ASSP, pp. 1365-1375, Oct. 1987.

[32] B. D. Van Veen, "*Minimum variance beamforming*", in **Adaptive Radar Detection**, S. Haykin, A. Stienhardt (ed.), Wiley, New York, 1992.

[33] B. D. Van Veen, R. A. Roberts, "*Eigenstructure based partially adaptive array design*", IEEE Trans. Antenna Prop. , pp. 357-362, March 1988.

[34] B. D. Van Veen, R. A. Roberts "*Partially adaptive beamformer design via output power minimization*", IEEE Trans. Acoustic Sig. Proc. , pp. 1524-1532, Nov. 1987.

[35] B. D. Van Veen, **Optimal Partially Adaptive Beamforming**, Ph. D. Thesis, University of Colorado at Boulder, 1986.

[36] D. J. Chapman, "*Partial adaptivity for the large arrays*", IEEE Trans. Antenna Prop. , pp. 685-696, Sept. 1976.

[37] L. J. Griffiths, C. W. Jim, "*An alternative approach to linearly constrained adaptive beamforming*", IEEE Trans. Antennas Propog. , pp. 27-34, Jan. 1982.

[38] T. C. Liu, B. D. Van Veen, "*A modular structure for implementation of linearly constrained minimum variance beamformers*", IEEE Trans. Signal Proc. , pp. 2343-2346, Oct. 1991.

[39] B. D. Van Veen, K. M. Buckley, "*Beamforming: a versatile approach to spatial filtering*", IEEE ASSP Magazine, pp. 4-24, April 1988.

[40] R. N. Adams, L. L. Horowitz, K. D. Senne, "*Adaptive main-beam nulling for narrowband antenna arrays*", IEEE Trans. on AES", pp. I 509-516, July 1980.

[41] B. D. Van Veen, "*Adaptive convergence of linearly constrained beamformers based on the sample covariance matrix*", IEEE Trans. Signal Proc. , pp. 1470-1473, June 1991.

[42] J. L. Krolik, D. N. Swingler, "*On the mean-squared error performance of adaptive minimum variance beamformers based on the sample covariance matrix*", Accepted for publication IEEE Trans. Signal Proc. .

[43] G. H. Niezgoda, J. L. Krolik, " *A multiple time scale modularized adaptive beamformer*", Accepted for publication IEEE Trans. Signal Proc. .

[44] M. H. A. Davis and R. B. Vinter, **Stochastic Modelling and Control**, Chapman and Hall, 1985.

[45] N. R. Goodman, "*Statistical analysis based on a certain multivariate gaussian distribution*" Ann. Math. Stat. pp. 152-177, March 1963.

[46] J. Capon and N. R. Goodman, "*Probability distributions for estimators of the frequency-wavenumber spectrum*", Proc. IEEE, pp. 1785-1786, Oct. 1970.

[47] I. S. Reed, J. D. Mallet and L. E. Brennan, "*Rapid convergence rate in adaptive arrays*", IEEE Trans. Aerospace Elect. Sys. , pp. 853-863, Nov. 1971.

[48] A. O. Steinhardt, "*The pdf of adaptive beamforming weights*", IEEE Trans. Signal Proc. , pp. 1232-1234, May 1991.

[49] M. L. Eaton, **Multivariate Statistics - A Vector Space Approach**, Wiley, 1983.

[50] R. J. Muirhead, **Aspects of Multivariate Statistical Theory**, Wiley, 1982.

[51] L. L. Scharf, **Statistical Signal Processing - Detection, Estimation, and Time Series Analysis**, Addison Wesley, 1991.

[52] S. Haykin. **Adaptive Filter Theory**, Prentice Hall, 1986.

[53] R. F. Dwyer, "*A technique for improving detection and estimation of signals contaminated by under ice noise*", J. Acoust. Soc. America, pp. 124-130, 1983.

[54] D. Middleton, "*Channel modeling and threshold signals in underwater acoustics: An analytical overview*", IEEE J. Ocean. Eng. pp. 4-28, 1987.

[55] W. F. Gabriel, "*Using spectral estimation techniques in adaptive processing antenna systems*", IEEE Trans. Ant. Prop. , pp. 291-300, March 1986.

[56] A. M. Vural, "*A comparative performance study of adaptive array processors*", Proceedings IEEE ICASSP'77, pp. 665-700, 1977.

[57] E. Brookner, J. M. Howell, "*Adaptive-adaptive array processing*", Proc. IEEE, pp. 602-604, April 1986.

[58] B. D. Van Veen, "*An analysis of several partially adaptive beamformer designs*", IEEE Trans. Acoustic Sig. Proc. , pp. 192-203, Feb. 1989.

[59] M. Abramowitz, I. A. Stegan (eds.), **A Handbook of Mathematical Functions**, New York, Dover Press, 1972.

[60] R. D. DeGroat, R. A. Roberts, "*Efficient, numerically stablized rank-one eigenstructure updating*", IEEE Trans. Signal Proc. , pp. 301-316, Feb. 1990.

[61] Y. H. Chen, H. D. Fang, "*Frequency-domain implementation of Griffiths-Jim adaptive beamformer*", J. Acoust. Soc. America, pp. 3354-3366, June 1992.

[62] J. An, B. Champagne, "*GSC realizations using the two-dimensional transform-domain LMS algorithm*", Accepted for publication, IEE Proc. F.

[63] G. Niezgoda, J. Krolik, "*The Application of adaptive multichannel spectral estimation for broadband array processing*", ICASSP'91, Toronto, Canada, 1991.

[64] G. Niezgoda, J. Krolik, "*The application of laguerre partial series expansions for broadband array processing*", ICASSP'92, San Francisco, 1992

[65] G. Niezgoda, J. Krolik, E. Plotkin, "*Broadband source localization in a slowly time-varying environment by multichannel kalman filtering*", IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Victoria, British Columbia, Canada, 1991.

[66] H. Vincent Poor, " *Detection of stochastic signals in non-Gaussian noise* ", J. Acoust. Soc. America, pp. 2838-2850, Nov. 1993.

Appendix A

Derivations - Chapter 2

A.1 M^{2}VDR and GSC Equivalence

Define

$$\epsilon(\mathbf{B}) = \mathbf{w}_q^\dagger \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}\mathbf{B}\right)^\dagger \mathbf{R} \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}\mathbf{B}\right) \mathbf{w}_q \tag{A.1}$$

and

$$\epsilon(\mathbf{B}) = \mathrm{tr}\left(\left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}\mathbf{B}\right)^\dagger \mathbf{R} \left(\mathbf{I}_{MJ} - \overline{\mathbf{C}}\mathbf{B}\right)\right) \tag{A.2}$$

as the quadratic error for the GSC and single stage M^{2}VDR beamformer, respectively.

Setting the partial derivative of $\epsilon(\mathbf{B})$ with respect to $\mathbf{B}$ to zero,

$$\frac{\partial \epsilon(\mathbf{B})}{\partial \mathbf{B}} = -2\overline{\mathbf{C}}^\dagger \mathbf{R}\mathbf{w}_q\mathbf{w}_q^\dagger + 2\overline{\mathbf{C}}^\dagger \mathbf{R}\overline{\mathbf{C}}\mathbf{B}\mathbf{w}_q\mathbf{w}_q^\dagger = 0 \tag{A.3}$$

the minimum quadratic error associated with the GSC is determined from

$$\overline{\mathbf{C}}^\dagger \mathbf{R}\overline{\mathbf{C}}\mathbf{B}\mathbf{w}_q\mathbf{w}_q^\dagger = \overline{\mathbf{C}}^\dagger \mathbf{R}\mathbf{w}_q\mathbf{w}_q^\dagger \tag{A.4}$$

From (A.5) weight matrix $\mathbf{B}$ must then equal

$$\mathbf{B} = \left(\overline{\mathbf{C}}^\dagger \mathbf{R}\overline{\mathbf{C}}\right)^{-1} \overline{\mathbf{C}}^\dagger \mathbf{R} \tag{A.5}$$

Rewrite $\epsilon(\mathbf{B})$ in the form

$$\epsilon(\mathbf{B}) = \mathrm{tr}\left(-\mathbf{B}^\dagger \overline{\mathbf{C}}^\dagger \mathbf{R} - \mathbf{R}\overline{\mathbf{C}}\mathbf{B} + \mathbf{B}^\dagger \overline{\mathbf{C}}^\dagger \mathbf{R}\overline{\mathbf{C}}\mathbf{B} + \mathbf{R}\right) \tag{A.6}$$

Setting the partial derivative of (A.6) with respect to $\mathbf{B}$ to zero,

$$\frac{\partial e(\mathbf{B})}{\partial \mathbf{B}} = -2\overline{\mathbf{C}}^{\dagger}\mathbf{R} + 2\overline{\mathbf{C}}^{\dagger}\mathbf{R}\overline{\mathbf{C}}\mathbf{B} = 0 \tag{A.7}$$

From (A.7) the weight matrix $\mathbf{B}$ is determined from

$$\overline{\mathbf{C}}^{\dagger}\mathbf{R}\overline{\mathbf{C}}\mathbf{B} = \overline{\mathbf{C}}^{\dagger}\mathbf{R} \tag{A.8}$$

such that

$$\mathbf{B} = \left(\overline{\mathbf{C}}^{\dagger}\mathbf{R}\overline{\mathbf{C}}\right)^{-1}\overline{\mathbf{C}}^{\dagger}\mathbf{R} \tag{A.9}$$

thus establishing the equivalence between $e(\mathbf{B})$ and $\epsilon(\mathbf{B})$.

Appendix B

Derivations - Chapter 3

B.1 Proof: Proposition 3.1

Defining the $NQ \times NQ$ partitioned matrix by

$$\begin{aligned} \mathbf{A} &= \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{Y} \\ \mathbf{E}^{\dagger}\mathbf{Y} \end{pmatrix} \begin{pmatrix} \mathbf{Y}^{\dagger} & \mathbf{Y}^{\dagger}\mathbf{E} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{Y}\mathbf{Y}^{\dagger} & \mathbf{Y}\mathbf{Y}^{\dagger}\mathbf{E} \\ \mathbf{E}^{\dagger}\mathbf{Y}\mathbf{Y}^{\dagger} & \mathbf{E}^{\dagger}\mathbf{Y}\mathbf{Y}^{\dagger}\mathbf{E} \end{pmatrix} \end{aligned} \tag{B.1}$$

we note that

$$\mathbf{Y}\mathbf{Y}^{\dagger} \sim \mathrm{W}_{N}(L, \boldsymbol{\Sigma})$$

and

$$\mathbf{E}^{\dagger}\mathbf{Y}\mathbf{Y}^{\dagger}\mathbf{E} \sim \mathrm{W}_{Q}\left(L, \mathbf{E}^{\dagger}\boldsymbol{\Sigma}\mathbf{E}\right)$$

In addition by ([50], Corollary 3.2.6) the distribution of $\mathbf{A}$ is given by

$$\mathbf{A} \sim \mathrm{W}_{NQ}(L, \boldsymbol{\Sigma}_{\mathbf{A}}) \tag{B.2}$$

where

$$\boldsymbol{\Sigma}_{\mathbf{A}} = \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}\mathbf{E} \\ \mathbf{E}^{\dagger}\boldsymbol{\Sigma} & \mathbf{E}^{\dagger}\boldsymbol{\Sigma}\mathbf{E} \end{pmatrix}$$

Now letting

$$\begin{aligned} \boldsymbol{\Lambda} &= \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \\ &= \mathbf{Y}\mathbf{Y}^{\dagger} - \mathbf{Y}\mathbf{Y}^{\dagger}\mathbf{E}\left(\mathbf{E}^{\dagger}\mathbf{Y}\mathbf{Y}^{\dagger}\mathbf{E}\right)^{-1}\mathbf{E}^{\dagger}\mathbf{Y}\mathbf{Y}^{\dagger} \end{aligned} \tag{B.3}$$

we may applay ([50]. **Theorem** 3.2.10) and write the distribution of $\boldsymbol{\Lambda}$ as

$$\Lambda \sim \mathtt{W}_{\mathtt{V}}(L-Q, \boldsymbol{\Sigma}_{\Lambda}) \tag{B.4}$$

where

$$\boldsymbol{\Sigma}_{\Lambda} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{E}\left(\mathbf{E}^{\dagger}\boldsymbol{\Sigma}\mathbf{E}\right)^{-1}\mathbf{E}^{\dagger}\boldsymbol{\Sigma}$$

B.2 Proof: Proposition 3.2

Let

$$\begin{aligned} \mathcal{E}\left\{\mathbf{V}_2^{\dagger}\left(\mathbf{V}_2\mathbf{V}_2^{\dagger}\right)^{-1}\mathbf{V}_2\mathbf{V}_1^{\dagger}\right\} &= \mathcal{E}\left\{\mathcal{E}\left\{\mathbf{V}_2^{\dagger}\left(\mathbf{V}_2\mathbf{V}_2^{\dagger}\right)^{-1}\mathbf{V}_2\mathbf{V}_1^{\dagger} \mid \mathbf{V}_2\right\}\right\} \\ &= \mathcal{E}\left\{\mathbf{V}_2^{\dagger}\left(\mathbf{V}_2\mathbf{V}_2^{\dagger}\right)^{-1}\mathbf{V}_2\mathcal{E}\left\{\mathbf{V}_1^{\dagger} \mid \mathbf{V}_2\right\}\right\} \end{aligned} \tag{B.5}$$

Note that by ([50]. Theorem 1.2.11) the condition distribution of $\mathbf{V}_1^{\dagger}$ given $\mathbf{V}_2$ is

$$\mathbf{V}_1^{\dagger} \mid \mathbf{V}_2 \sim \mathtt{N}\left(\mathbf{V}_2^{\dagger}\boldsymbol{\Sigma}_{22}\boldsymbol{\Sigma}_{21}, \boldsymbol{\Sigma}_{11\,2}\right) \tag{B.6}$$

where

$$\boldsymbol{\Sigma}_{11\,2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

Since

$$\mathcal{E}\left\{\mathbf{V}_1^\dagger \mid \mathbf{V}_2\right\} = \mathbf{V}_2^\dagger \boldsymbol{\Sigma}_{22}\boldsymbol{\Sigma}_{21} \tag{B.7}$$

(B.5) is equal to

$$\begin{aligned}\mathcal{E}\left\{\mathbf{V}_2^\dagger\left(\mathbf{V}_2\mathbf{V}_2^\dagger\right)^{-1}\mathbf{V}_2\mathbf{V}_1^\dagger\right\} &= \mathcal{E}\left\{\mathbf{V}_2^\dagger\right\}\boldsymbol{\Sigma}_{22}\boldsymbol{\Sigma}_{21} \\ &= 0\end{aligned} \tag{B.8}$$

B.3 Proof: Proposition 3.3

Expressing $\mathbf{V}$ in the column partitioned form

$$\mathbf{V} = (\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_L) \tag{B.9}$$

where the $Q\times 1$ column vectors $\mathbf{v}_n$ are i.i.d., rewrite

$$\boldsymbol{\Gamma} = \mathbf{V}^\dagger\left(\mathbf{V}\mathbf{V}^\dagger\right)^{-1}\mathbf{V}$$

in the form

$$\boldsymbol{\Gamma} = (\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_L)^\dagger\left(\sum_{\ell=1}^{L}\mathbf{v}_\ell\mathbf{v}_\ell^\dagger\right)^{-1}(\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_L) \tag{B.10}$$

where

$$(\boldsymbol{\Gamma})_{mn} = \mathbf{v}_m^\dagger\left(\mathbf{v}_m\mathbf{v}_m^\dagger + \sum_{\substack{\ell=1\\ \ell\neq m}}^{L}\mathbf{v}_\ell\mathbf{v}_\ell^\dagger\right)^{-1}\mathbf{v}_n \tag{B.11}$$

Since the column vectors of $\mathbf{V}$ are symmetrically distributed about zero the expectation of $(\Gamma)_{mn}$ will not change if the random sample vector $\mathbf{v}_m$ is negated (ie: $-\mathbf{v}_m$). Notice that if $f(\mathbf{v}_m)$ is some function of the sample vector $\mathbf{v}_m$ then

$$\mathcal{E}\{f(\mathbf{v}_m)\} = \mathcal{E}\{f(-\mathbf{v}_m)\}$$

Hence from (B.11) the following relationship must hold

$$
\begin{aligned}
(\,\mathbf{\Gamma}\,)_{mn} &= \mathbf{v}_m^{\dagger}\left(\mathbf{v}_m\mathbf{v}_m^{\dagger} + \sum_{\substack{\ell=1,\\ \ell\neq m}}^{L} \mathbf{v}_\ell\mathbf{v}_\ell^{\dagger}\right)^{-1}\mathbf{v}_n \\
&= -\mathbf{v}_m^{\dagger}\left((-1)^2\mathbf{v}_m\mathbf{v}_m^{\dagger} + \sum_{\substack{\ell=1,\\ \ell\neq m}}^{L} \mathbf{v}_\ell\mathbf{v}_\ell^{\dagger}\right)^{-1}\mathbf{v}_n \qquad \text{(B.12)}
\end{aligned}
$$

implying that

$$\mathcal{E}\,\{(\Gamma)_{mn}\} = 0$$

for $m\neq n$.

Since $\mathcal{E}\,\{(\mathbf{\Gamma})_{mn}\}$ equals zero for $m\neq n$ we need only consider the matrix elements $\mathcal{E}\,\{(\mathbf{\Gamma})_{mm}\}$ in evaluating $\mathcal{E}\,\{\mathbf{\Gamma}\}$. Noting that the diagonal elements of $\mathcal{E}\,\{\mathbf{\Gamma}\}$ are identically distributed we can then write

$$\mathcal{E}\{\,\mathbf{\Gamma}\,\} = \mathcal{E}\,\{(\mathbf{\Gamma})_{\ell\ell}\}\,\mathbf{I}_N \qquad \text{(B.13)}$$

where

$$\mathcal{E}\{(\mathbf{\Gamma})_{\ell\ell}\} = \frac{1}{Q}\mathcal{E}\,\{\mathtt{tr}\,(\mathbf{\Gamma})\}$$

Evaluating $\mathtt{tr}\,(\mathbf{\Gamma})$,

$$
\begin{aligned}
\mathtt{tr}\,(\mathbf{\Gamma}) &= \mathtt{tr}\left(\mathbf{V}^{\dagger}\left(\mathbf{V}\mathbf{V}^{\dagger}\right)^{-1}\mathbf{V}\right) \\
&= \mathtt{tr}\left(\left(\mathbf{V}\mathbf{V}^{\dagger}\right)^{-1}\mathbf{V}\mathbf{V}^{\dagger}\right) \\
&= Q \qquad \text{(B.14)}
\end{aligned}
$$

substituting this result back into (B.13) we then have

$$\mathcal{E}\{\Gamma\} = \frac{Q}{L}\mathbf{I}_L \tag{B.15}$$

B.4 The Conditional Distribution of $\mathbf{Z}_{1,n}$

Expand $\mathbf{Z}_{1,n}$ in terms of $\mathbf{X}_{0,n}$ as

$$\mathbf{Z}_{1,n} = \mathbf{X}_{0,n} - \mathbf{X}_{0,n}\mathbf{X}_{0,n}^{\dagger}\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n}\mathbf{X}_{0,n}^{\dagger}\overline{\mathbf{C}}_1\right)^{-1}\overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n} \tag{B.16}$$

Applying ([50] Theorem 1.2.11) the distribution of $\mathbf{X}_{0,n}$ conditional on $\overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n}$ is defined by

$$\mathbf{X}_{0,n} \mid \overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n} \sim \mathrm{N}\left(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n},\ \boldsymbol{\Sigma}_{11\cdot 2}\right) \tag{B.17}$$

where

$$\boldsymbol{\Sigma}_{12} = \mathbf{R}_{0,n}\overline{\mathbf{C}}_1$$

$$\boldsymbol{\Sigma}_{22} = \overline{\mathbf{C}}_1^{\dagger}\mathbf{R}_{0,n}\overline{\mathbf{C}}_1$$

and

$$\boldsymbol{\Sigma}_{11\cdot 2} = \mathbf{R}_{0,n} - \mathbf{R}_{0,n}\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{R}_{0,n}\overline{\mathbf{C}}_1\right)^{-1}\overline{\mathbf{C}}_1^{\dagger}\mathbf{R}_{0,n}$$

From (B.17) $\mathbf{Z}_{1,n} \mid \overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n}$ is Gaussian distributed with a mean value equal to

$$\mathcal{E}\left\{\mathbf{Z}_{1,n} \mid \overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n}\right\} = \mathcal{E}\left\{\mathbf{X}_{0,n} \mid \overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n}\right\}$$

$$-\mathcal{E}\left\{\mathbf{X}_{0,n}\mathbf{X}_{0,n}^{\dagger}\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n}\mathbf{X}_{0,n}^{\dagger}\overline{\mathbf{C}}_1\right)^{-1}\overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n} \mid \overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n}\right\}$$

$$= \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n}$$

$$-\mathcal{E}\left\{\mathbf{X}_{0,n} \mid \overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n}\right\}\mathbf{X}_{0,n}^{\dagger}\overline{\mathbf{C}}_1\left(\overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n}\mathbf{X}_{0,n}^{\dagger}\overline{\mathbf{C}}_1\right)^{-1}\overline{\mathbf{C}}_1^{\dagger}\mathbf{X}_{0,n}$$

$$= \Sigma_{12}\Sigma_{22}^{-1}\overline{C}_1^{\dagger}\mathbf{X}_{0,n}$$

$$- \Sigma_{12}\Sigma_{22}^{-1}\overline{C}_1^{\dagger}\mathbf{X}_{0,n}\mathbf{X}_{0,n}^{\dagger}\overline{C}_1\left(\overline{C}_1^{\dagger}\mathbf{X}_{0,n}\mathbf{X}_{0,n}^{\dagger}\overline{C}_1\right)^{-1}\overline{C}_1^{\dagger}\mathbf{X}_{0,n}$$

$$= 0 \tag{B.18}$$

B.5 Proof: Proposition 3.4

Consider first the distribution of $\mathbf{CYY}^{\dagger}\mathbf{w}$ conditional on $\mathbf{C}^{\dagger}\mathbf{YY}^{\dagger}\mathbf{C}$. Define

$$\mathbf{A} \sim \mathbf{W}_{Q+1}(L,\ \Sigma_{\mathbf{A}}) \tag{B.19}$$

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{w}^{\dagger}\mathbf{Y} \\ \mathbf{C}^{\dagger}\mathbf{Y} \end{pmatrix}\begin{pmatrix} \mathbf{Y}^{\dagger}\mathbf{w} & \mathbf{Y}^{\dagger}\mathbf{C} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{w}^{\dagger}\mathbf{YY}^{\dagger}\mathbf{w} & \mathbf{w}^{\dagger}\mathbf{YY}^{\dagger}\mathbf{C} \\ \mathbf{C}^{\dagger}\mathbf{YY}^{\dagger}\mathbf{w} & \mathbf{C}^{\dagger}\mathbf{YY}^{\dagger}\mathbf{C} \end{pmatrix}$$

and

$$\Sigma_{\mathbf{A}} = \begin{pmatrix} \mathbf{w}^{\dagger}\Sigma\mathbf{w} & \mathbf{w}^{\dagger}\Sigma\mathbf{C} \\ \mathbf{C}^{\dagger}\Sigma\mathbf{w} & \mathbf{C}^{\dagger}\mathbf{YY}^{\dagger}\mathbf{C} \end{pmatrix}$$

From ([50], Theorem 3.2.10) observe that

$$\mathbf{CYY}^{\dagger}\mathbf{w} \mid \mathbf{C}^{\dagger}\mathbf{YY}^{\dagger}\mathbf{C} \sim \mathrm{N}\left(\mathbf{C}^{\dagger}\mathbf{YY}^{\dagger}\mathbf{CBw},\ \mathrm{P}\left(\mathbf{C}^{\dagger}\mathbf{YY}^{\dagger}\mathbf{C}\right)\right) \tag{B.20}$$

where

$$\mathbf{B} = \left(\mathbf{C}^{\dagger}\Sigma\mathbf{C}\right)^{-1}\mathbf{C}^{\dagger}\Sigma$$

$$\mathrm{P} = \mathbf{w}^{\dagger}\Sigma\mathbf{w} - \mathbf{w}^{\dagger}\Sigma\mathbf{C}\left(\mathbf{C}^{\dagger}\Sigma\mathbf{C}\right)^{-1}\mathbf{C}^{\dagger}\Sigma\mathbf{w}$$

It then follows from (B.20) that

$$\hat{\mathbf{B}}\mathbf{w} \mid \mathbf{C}^t\mathbf{Y}\mathbf{Y}^t\mathbf{C} \sim \mathrm{N}\left(\mathbf{B}\mathbf{w},\ \mathrm{P}\left(\mathbf{C}^t\mathbf{Y}\mathbf{Y}^t\mathbf{C}\right)^{-1}\right) \tag{B.21}$$

where

$$\mathcal{E}\left\{\mathbf{B}\mathbf{w} \mid \mathbf{C}^t\mathbf{Y}\mathbf{Y}^t\mathbf{C}\right\} = \mathbf{B}\mathbf{w}$$

and

$$\begin{aligned}
\mathrm{cov}\left\{\hat{\mathbf{B}}\mathbf{w} \mid \mathbf{C}^t\mathbf{Y}\mathbf{Y}^t\mathbf{C}\right\} &= \left(\mathbf{C}^t\mathbf{Y}\mathbf{Y}^t\mathbf{C}\right)^{-1} \mathrm{cov}\left\{\mathbf{C}^t\mathbf{Y}\mathbf{Y}^t\mathbf{w} \mid \mathbf{C}^t\mathbf{Y}\mathbf{Y}^t\mathbf{C}\right\}\left(\mathbf{C}^t\mathbf{Y}\mathbf{Y}^t\mathbf{C}\right)^{-1} \\
&= \left(\mathbf{C}^t\mathbf{Y}\mathbf{Y}^t\mathbf{C}\right)^{-1}\left(\mathbf{C}^t\mathbf{Y}\mathbf{Y}^t\mathbf{C}\right)\left(\mathbf{C}^t\mathbf{Y}\mathbf{Y}^t\mathbf{C}\right)^{-1}\mathrm{P} \\
&= \left(\mathbf{C}^t\mathbf{Y}\mathbf{Y}^t\mathbf{C}\right)^{-1}\mathrm{P}
\end{aligned}$$

Appendix C

Recursive Least Squares Adaptation

From [52], the RLS update of $\hat{\mathbf{B}}_1[n]$ proceeds from the error

$$\mathbf{e}_1[n] = \mathbf{x}[n] - \hat{\mathbf{B}}_1^{\dagger}[n-1]\overline{\mathbf{C}}_1^{\dagger}\mathbf{x}[n] \tag{C.1}$$

the gain estimate

$$\hat{\mathbf{k}}[n] = \frac{\tau^{-1}\hat{\mathbf{P}}[n-1]\overline{\mathbf{C}}_1^{\dagger}\mathbf{x}[n]}{1 + \tau^{-1}\mathbf{x}^{\dagger}[n]\overline{\mathbf{C}}_1\hat{\mathbf{P}}[n-1]\overline{\mathbf{C}}_1^{\dagger}\mathbf{x}[n]} \tag{C.2}$$

and the error covariance estimate

$$\hat{\mathbf{P}}[n] = \tau^{-1}\hat{\mathbf{P}}[n-1] - \tau^{-1}\hat{\mathbf{k}}[n]\mathbf{x}^{\dagger}[n]\overline{\mathbf{C}}_1\hat{\mathbf{P}}[n-1] \tag{C.3}$$

where the effective memory of the system is determined from

$$0 < \tau < 1$$

From the above, the weight matrix update equals

$$\hat{\mathbf{B}}_1[n] = \hat{\mathbf{B}}_1[n-1] + \mathbf{k}[n]\mathbf{e}_1^{\dagger}[n] \tag{C.4}$$

where

$$\hat{\mathbf{P}}[0] = \xi^{-1}\mathbf{I}_Q \quad \text{for } 0 < \xi << 1$$

and

$$\hat{\mathbf{B}}_1[0] = \mathbf{0}$$

This particular modification to RLS applies $\mathbf{B}_1[n]$ to the input data of the first stage at sample n instead of at $n+1$,

$$\mathbf{z}_1[n] = \mathbf{x}[n] - \hat{\mathbf{B}}_1^{\dagger}[n]\overline{\mathbf{C}}_1^{\dagger}\mathbf{x}[n] \tag{C.5}$$

Note from (C.2) and (C.3) that the computational complexity of the RLS algorithm is dictated by M and the number of adaptive DOF assigned to the stage. Hence, to realize a practical implementation of the beamformer it is necessary for Q_1 to be small. Adaptation of the second stage can proceed by either a LMS (eg: [15, 14]) scheme or by block adaptation [40]. Note that the memory of the RLS algorithm can be related to the adaptation period of a block implementation by

$$L_1 = \frac{1}{1-\tau} \text{ snapshots}$$