Partially Adaptive Array Signal Processing with Application to Airborne Radar

Iain Scott

A thesis submitted for the degree of Doctor of Philosophy.

The University of Edinburgh.

- February 1995 -

Abstract

An adaptive array is a signal processor used in conjunction with a set of antennae to provide a versatile form of spatial filtering. The processor combines spatial samples of a propagating field with a variable set of weights, typically chosen to reject interfering signals and noise. In radar, the spatial filtering capability of the array facilitates cancellation of hostile jamming signals and aids in the suppression of clutter.

In many applications, the practical usefulness of an adaptive array is limited by the complexity associated with computing the adaptive weights. In a partially adaptive beamformer only a subset of the available degrees of freedom are used adaptively, where adaptive degrees of freedom denotes the number of unconstrained or free weights that must be computed. The principal benefits associated with reducing the number of adaptive degrees of freedom are reduced computational burden and improved adaptive convergence rate. The computational cost of adaptive algorithms is generally either directly proportional to the number of adaptive weights or to the square or cube of the number of adaptive weights. In radar it is often mandatory that the number of adaptive weights be reduced with large antenna arrays because of the algorithms computational requirement. The number of data vectors needed for the adaptive weights to converge to their optimal values is also proportional to the number of adaptive weights. Thus, in some applications, adaptive response requirements dictate reductions in the number of adaptive weights. Both of these aspects are investigated in this thesis.

The primary disadvantage of reducing the number of adaptive weights is a degradation in the steady–state interference cancellation capability. This degradation is a function of which adaptive degrees of freedom are utilised and is the motivation for the partially adaptive design techniques detailed in this thesis. A new technique for selecting adaptive degrees of freedom is proposed. This algorithm sequentially selects adaptive weights based on an output mean square error criterion. It is demonstrated through simulation that for a given partially adaptive dimension this approach leads to improved steady–state performance, in mean square error terms, over popular eigenstructure approaches. Additionally, the adaptive structure which results from this design method is computationally efficient, yielding a reduction of around 80% in the number of both complex multiplications and additions.

When the adaptive weights are computed using a finite number of data vectors, the adapted response of the array may experience very noisy sidelobe fluctuations and main beam perturbations. This random behaviour occurs because finite sampling causes spurious cross–correlation effects, so that the background noise component differs greatly from the asymptotic value. Large sidelobe levels present a considerable problem in a radar application, as processing is typically performed in a non–concurrent manner, i.e. the weights are computed from a different set of data from that to which they are applied. High sidelobes can render the adaptive array very vulnerable to sidelobe clutter, sudden changes in the interference environment, or pulsed interference that can benefit from post-processing gain. A statistical analysis of the transient response of an adaptive array is presented. In particular, the transient sidelobe levels are examined, showing the error to be a function of the number of adaptive weights and the number of data vectors combined.

Declaration of originality

I hereby declare that this thesis and the work reported herein was composed and originated entirely by myself, in the Department of Electrical Engineering at the University of Edinburgh.

Iain Scott

Acknowledgements

I would like to thank the following people for their invaluable assistance during the course of this PhD:

- Bernard Mulgrew and Peter Grant, my supervisors, for their continuous support and guidance. Also for reading and checking this thesis.
- The other members of Signal Processing Group for their support throughout my PhD.
- GEC Marconi Avionics Ltd and the EPSRC for providing financial support.

Contents

List of Figures

List of Tables

Abbreviations

ACF	Adaptive cancellation factor
DF	Doppler filter
DOF	Degrees of freedom
DPCA	Displaced phase centred array
GSC	Generalised sidelobe canceller
LCMV	Linearly constrained minimum variance
MSE	Mean square error
MTI	Moving target indication
PRF	Pulse repetition frequency
SLAR	Sideways looking airborne radar
SMI	Sample matrix inversion
SNR	Signal to noise ratio
SVD	Singular value decomposition

List of principal symbols

d	Interelement spacing.
θ	Direction of arrival.
ω	Angular frequency.
k	Discrete time variable.
σ^2	Variance.
$r_d(\theta, \omega)$	Desired response of an arbitrary beamformer.
$d(\theta, \omega)$	Array steering vector for signal of frequency ω arriving from bearing θ.
$\rho^2(\theta, \omega)$	Power spatial/spectral density of a source received at an array.
λ_i	Eigenvalues of a given matrix.
w	General weighting vector.
w_{opt}	Optimised weighting vector.
$\boldsymbol{f}$	Constraint response vector in linearly constrained beamforming.
$\boldsymbol{h}$	Finite–impulse response filter tap weight vector.
w_q	Generalised sidelobe canceller quiescent weight vector.
$\boldsymbol{C}_n$	Generalised sidelobe canceller signal blocking matrix.
w_a	Generalised sidelobe canceller adaptive weight vector.
$\boldsymbol{T}_n$	Partially adaptive generalised sidelobe canceller transformation matrix.
$\mathbf{x}(k)$	Vector of elemental signals at instant k.
$\boldsymbol{x}(k)$	Stacked snapshot vector at instant k.
$\boldsymbol{X}(k)$	Matrix of stacked snapshot vectors.
$s(k)$	Stacked signal snapshot vector at instant k.
$\boldsymbol{n}(k)$	Stacked interference snapshot vector at instant k.
$y(k)$	Output signal at instant k.
$\boldsymbol{R}_s(\tau)$	Autocorrelation matrix for desired signal at lag τ seconds.
$\boldsymbol{R}_n(\tau)$	Autocorrelation matrix for interfering signals at lag τ seconds.
$\boldsymbol{R}_{sn}(\tau)$	Cross–correlation matrices of desired signal and interference at lag τ seconds.
$\boldsymbol{R}_x(\tau)$	Autocorrelation matrix for snapshot vector $\boldsymbol{x}(k)$ with lag τ seconds.

$\boldsymbol{R}_x$	Covariance matrix for snapshot vector $\boldsymbol{x}(k)$.
$\bar{\boldsymbol{R}}_n$	Ensemble averaged covariance matrix.
$\hat{\boldsymbol{R}}_n$	Estimated covariance matrix.
$\boldsymbol{I}$	Identity matrix.
μ	Step size parameter in steepest gradient and least mean squares algorithms.
$N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Normal multivariate distribution.
$W_m(n, \boldsymbol{\Sigma})$	Wishart distribution.
$E\{\cdot\}$	Expectation operator.
$tr()$	Matrix trace.
cov $\{\}$	Matrix covariance.
Re $\{.\}$	Real part of bracketed quantity.
Im $\{.\}$	Imaginary part of bracketed quantity.
O$\{.\}$	Of the order of bracketed quantity number of operations.
$(.)^T$	Simple transposition.
$(.)^*$	Simple conjugation.
$(.)^H$	Complex conjugate transpose.
$(.)^+$	Pseudo inverse.

Chapter 1
Introduction

1.1 Introduction

In an airborne radar desired signals often have to compete with strong ground clutter returns. These returns are usually strongest in the areas illuminated by the mainlobe of the transmit beam. When conventional (non–adaptive) signal processing techniques are applied, mainlobe clutter returns are typically translated to zero Doppler prior to rejection with low order digital filters. With an electronically scanned beam, clutter rejection is further complicated by the increase in beamwidth arising from scanning away from broadside. If the electronically scanned beam is to be formed adaptively the clutter filtering operation becomes of even greater importance because the mainlobe clutter coincides with the look direction. Great care must be taken to avoid the extreme sensitivity of the adaptive processor to mainlobe returns, which means that clutter rejection filters must attain significant rejection over the range of Doppler frequencies illuminated.

Within this thesis the problem of designing a partially adaptive beamformer which attains near fully adaptive performance is considered. The contribution of this work is the investigation of techniques for reducing the required adaptive dimension of an adaptive beamformer. A new technique is proposed which selects adaptive degrees of freedom on an output mean square error (MSE) criterion. The convergence performance of such a beamformer is also examined, and expressions for various parameters are derived, most notably the transient sidelobe levels. The purpose of this chapter is to introduce the work undertaken in this project. The chapter begins with a discussion of the principal motivations for the work and then gives an overview of conventional interference cancellation techniques. Following this, the principal areas of work within the thesis are summarised and other salient research reviewed. Finally the organisation of the thesis is described.

1.2 Motivation

Target recognition and identification in airborne radar typically consists of a reflector antenna or waveguide array with some fixed gain pattern which is mechanically steered over all look directions of interest. The disadvantages of such a system are many. The physical size of the antenna requires a scanning strategy which is continuous and usually periodic. In a hostile situation it may be preferable to follow a particular target whilst maintaining a conventional

scanning strategy. Use of an electronically steered phased array, in which the antenna stays fixed in space and the radar beam is moved by introducing a phase delay across the array face, can allow inertialess scanning and even random scanning strategies. Scanning in such a manner not only prevents the detection of a periodic scan, but allows continuous tracking of a target as a single task within a larger scheduling scheme.

With a fixed pattern antenna low sidelobes are relied upon to provide exclusion of undesired noise and jamming signals. Most airborne radars will operate in environments in which substantial jamming and interfering signals will be present. These interferences, whether intentional or not, will lead to degradation in the ability of the array to identify desired signals. By using an adaptive phased array "nulls" can be steered in the directions of jamming signals so cancelling their effect. This type of array is easily reconfigurable from within software, allowing many different modes of operation to be implemented. This flexibility allows performance which is superior to that of existing fixed pattern arrays.

Phased array antennae are, however, expensive and complex to implement. For a fully adaptive array, each element will require separate gain and phase control. Typical phased arrays may contain several thousand active elements, so much work has been focussed on reducing the complexity of such structures. Many different approaches have arisen. They may, though, be classified into two groups. In the first, auxiliary elements are selected according to some algorithm. In the latter, all the array elements are used, with elemental outputs being precombined in a fixed beamformer to form a reduced set of signals, thus reducing subsequent processing. These two approaches can be termed *element–space* and *beam–space* thinning respectively. There are many inherent problems with both of these regimes. This thesis attempts to combine the simplicity of the former techniques with the benefits of the latter. Despite the complexity of such structures it is widely agreed that addition of an adaptive capability will form the basis of future developments. Investigation of reduction networks will lead to interconnection regimes which are optimum (or near optimum) over all interference scenarios. The allowed interconnection and combinational rules of the elemental outputs can be incorporated in the design phase. In this way the the manner in which partial adaptivity is achieved may be seen as only another factor to be optimised in the adaptive processor.

1.3 Airborne radar

There are several motivations for carrying a radar on an aircraft, one of them being the raised position which enables the radar to look from above. This gives improved detection of low flying aircraft and vehicles in a hilly landscape. By doing so, one encounters two serious problems. Firstly, the clutter returns will be much larger in amplitude because of the steeper aspect angles. Secondly, the clutter returns will be Doppler shifted due to the aircraft motion. Whereas in ground–based radar, clutter suppression is a relatively simple filtering operation, suppression of clutter returns in an airborne radar becomes a far more complex task. Filters which operate in either space or time only will be sub–optimum, reducing the ability to detect a desired signal.

To overcome this, two dimensional sampling of the received field (space and time) has been employed. This is realised by using a phased array and coherent pulse trains. Such space–time processors as applied to suppression of airborne clutter returns are examined in this thesis. Partially adaptive processors, namely those that employ a subset of the available degrees of freedom are examined and their performance compared.

The type of radar considered here is termed a pulse–Doppler radar. In this mode of operation the radar can yield both range, bearing and Doppler characteristics of targets signals. This type of pulsed radar can be divided into three classes, depending on the pulse repetition frequency (PRF):

- *Low PRF* – the radar will be ambiguous in frequency, but unambiguous in range
- *Medium PRF* – the radar will be ambiguous in both frequency and range
- *High PRF* – the radar will be unambiguous in frequency, but ambiguous in range

In this thesis a medium–PRF pulse–Doppler radar is employed. However, the PRF is assumed sufficiently high to be greater than the Nyquist sampling rate for the return clutter field. This means that the clutter field will be sampled unambiguously in frequency. There will however be a set of ambiguous range rings, spread throughout the range profile. At the PRF chosen, the clutter returns from all but the first ambiguous range ring will be so small in magnitude, that they will not significantly affect the performance. In existing radar applications the term *narrowband* is used to refer to applications which utilise a bandwidth of up to 10% of the carrier frequency. In these terms, the problem considered here could be considered narrowband. However, because of the azimuth–Doppler structure which exists in the clutter returns, *wideband* steering delays are used in the receiver rather than simple phase-shifers, and the problem is treated as wideband, even though the clutter returns could be called narrowband by the above definition. For this reason, this thesis discusses clutter cancellation in terms of a wideband problem, despite the relatively small bandwidth of clutter returns.

The discussion above has indicated that the clutter field is two dimensional, so that a two dimensional filter is needed for effective suppression. These comments, though, have not given any indication of how many adaptive degrees of freedom are required, nor the best manner in which to employ these degrees of freedom. One technique for determining the required number of degrees of freedom is to examine the eigenstructure of the return clutter field. This examination leads to an estimate of the dimension of the clutter signals, and thus an idea of how many degrees of freedom are required. This technique has been employed by many investigators.

1.4 Existing cancellation techniques

By incorporating an array as the antenna in a radar system, the designer gains an additional spatial dimension, in which there are as many degrees of freedom as there are array elements. This is seen as the principal benefit of array antennae. Without an array only temporal filtering of the received field can be performed. A conventional moving target indicator (MTI) is ineffective in cancelling airborne clutter because it uses temporal degrees of freedom only, and can only cancel clutter at the look direction. It is ineffective at cancelling the remaining sidelobe clutter. It is evident that in order to cancel both mainlobe and sidelobe clutter, one must add some spatial degrees of freedom. The simplest system that combines spatial and temporal degrees of freedom is the displaced phase centred array (DPCA). In this system platform motion is compensated for by arranging two sub–apertures along the aircraft's axis and switching between them such that every two beams are spatially fixed. This is basically true motion compensation, also correcting for the sidelobe clutter. The problem with non–adaptive DPCA is that it is sensitive to antenna errors, and requires that the platform velocity is known well enough to adjust the interpulse period to ensure pulse to pulse cancellation. Additionally, DPCA needs some form of adaptive algorithm to compensate for tolerances and platform dynamics, so that it is better to use an adaptive clutter filter without the DPCA technique, which will in any case compensate for platform motion effects.

General adaptive clutter suppression has been discussed in several papers [1,3–8]. These adaptive processors are called space–time processors because they combine both spatial and temporal samples of the received field with an adaptive set of weights to effectively suppress all clutter. For an adaptive array with N elements, which processes L pulses the total number of adaptive weights is NL. The optimum processor will be practically useless because of this large dimension. The inversion of a $NL \times NL$ dimensional covariance matrix requires a computational cost of $\mathrm{O}\{(NL)^3\}$ per iteration which is considered too expensive for real time implementation. The extent to which the adaptive dimension may be reduced will ultimately be determined by what is considered an acceptable level of performance. However, as an initial estimate, the eigendecomposition techniques mentioned earlier can be applied to estimate the dimension of the clutter subspace. This was the approach taken in [1, 5–7]. In this clutter suppression was achieved by forming a single search beam and a set of auxiliary beams used for cancelling clutter echoes. More recently Su & Zhou [9] proposed a partially adaptive implementation of a clutter space–time filter which uses an on–line estimate of the clutter eigenstructure.

1.5 Linearly constrained beamforming

Linearly constrained beamformers are a class of optimum beamformer which allow general control over large regions of the angular and spectral response. As such they represent an important technique for overcoming the problems in adaptive beamforming. Most notably additional linear constraints can be used to control the sidelobe behaviour of an adapted array, reducing the high sidelobes which often occur, and thus reducing susceptibility to sudden changes in the

interference environment and sidelobe jamming. Another use of linear constraints is in broadening the mainlobe response of the beamformer, therefore making it less directive, and hence increasing the tolerance to beam steering errors. Many constraint regimes exist [10–21], but in most applications a combination of the different types of constraints is most effective. Each linear constraint uses one degree of freedom in the weight vector, so that with K constraints there are only $NL - K$ degrees of freedom available for minimising interference.

The basic idea behind linearly constrained beamforming is to constrain the response of the beamformer so signals from the direction of interest are passed with specified gain and phase. The weights are chosen to minimise output power subject to the response constraint. This has the effect of preserving the desired signal while minimising contributions to the output due to interfering signals and noise arriving from directions other than the direction of interest. The generalised sidelobe canceller (GSC) [20,22–26] implementation of the linearly constrained beamformer provides a useful structure from which to approach partial adaptivity. The GSC splits the weight vector into two portions, one non–adaptive part which satisfies the constraints, and another adaptive portion which is orthogonal to the constraints. The desired signal is effectively prevented from entering the latter path, meaning the adaptation takes place upon interfering signals only. Having decomposed the problem as such, it is possible to allocate any number of degrees of freedom up to $NL-K$ to the minimisation of interference. Many different techniques for selecting the reduced set of degrees of freedom exist, including that reported in this thesis [9,19,27–40].

1.6 Summary of the work

This thesis examines the problem of ground clutter suppression in an airborne array radar. More specifically, this thesis presents a new algorithm for weight selection in a partially adaptive beamformer, and investigates the convergence performance of this class of beamformer. The techniques developed and the ideas presented are applicable to a large class of partially adaptive beamforming structures.

The problem considered is the suppression of ground clutter, so naturally the first part of the study was concerned with developing a simple model for the clutter returns received at an airborne phased array. Eigendecompositions were performed on typical computed clutter returns as a means of providing an estimate of the dimension of the clutter subspace. The influence of various radar parameters upon the computed spectra were also examined.

The problem of choosing a suitable beamforming structure was considered, and the generalised sidelobe canceller was proposed as a solution. This structure answers many of the questions posed of adaptive processors for airborne environments. Most notably, the adaptive weights found are computed from signal free data, overcoming signal cancellation and the super–directive problems of maximised signal–to–noise beamformers. Additionally, the generalised sidelobe canceller offers the opportunity to adaptively select the number of adaptive

weights which are allocated to interference suppression, without altering the response of the beamformer to desired signals.

Many techniques are available for reducing the adaptive dimension of the generalised sidelobe canceller. In this thesis, a weight selection criteria is developed based upon an output mean square error cost function. This approach leads to a structure which is computationally efficient, yielding up to 80% reductions in the number of both complex multiplications and additions that must be performed. The cancellation performance (in mean square error terms) is found to exceed that of eigenstructure designed beamformers for a given partially adaptive dimension.

The adaptive weights that are selected by the algorithm presented are chosen according to a mean square error performance measure. This will yield a set of weights which are near optimum, and which will result in good signal–to–noise performance. However, no consideration is given to the adapted response of the beamformer. Typically, an adaptive array used in an airborne radar will operate in a non–concurrent mode, that is different data is used to compute the adaptive weights from that to which the weights are applied. High sidelobe levels can therefore present a considerable limitation to the performance in a non–concurrent mode. The maximum and average sidelobe levels of the generalised sidelobe canceller were examined through simulation when the adaptive weights were computed using a sample covariance matrix algorithm. The improvement in sidelobe performance obtained by employing diagonal loading of the sample covariance matrix was also examined, showing that significant improvements can be obtained for a small amount of diagonal loading, even when a small number of samples are used in estimating the data covariance.

1.7 Thesis organisation

After this brief introduction, the problem of adaptive beamformer design for airborne radar is considered. In chapter 2 much of the necessary background to beamforming is presented and subjects such as data independent array design, the terminology employed, optimal beamformer design when subject to linear constraints, and the generalised sidelobe canceller are discussed. The chapter then proceeds to consider the application of different sets of constraints to linearly constrained minimum variance (LCMV) beamformers, and how these can be used to improve system response.

In chapter 3 the ground clutter problem is examined in detail and a simple clutter model is described. The principal objective of this chapter is to examine the likely interference conditions under which the beamformer will operate, and to use these to get a feel for the complexity required in an adaptive processor. Typical clutter power spectra are computed and discussed, and then subsequently eigenspectra plots are presented for the clutter model. These results are used to give an estimate of the number of adaptive degrees of freedom that will need to be incorporated in an adaptive processor. The influence of various radar parameters on the computed eigenspectra is examined, and the implications discussed.

Chapter 4 proposes a new technique for selecting adaptive degrees of freedom in an generalised sidelobe cancelling structure. The algorithm is based upon the assumption that the limiting factor in applying adaptive algorithms is not in collecting the delayed samples, but in computing the adaptive coefficients. With this assumption, the algorithm described selects the best weights using a mean square error criterion. The adaptive algorithm uses a sub–optimum approach to sequentially select the adaptive weights which best minimise the output mean square error. The performance of this new algorithm is contrasted with that of several existing techniques and some conclusions are drawn.

Chapter 5 considers the convergence performance of linearly constrained beamformers, in particular partially adaptive generalised sidelobe cancellers. The output mean square error when the beamformer operates in concurrent and non–concurrent modes is examined. The transient sidelobe levels of the beamformer are also considered and are demonstrated through simulation to be significantly reduced by the addition of diagonal loading to the sample covariance matrix. In chapter 6 the conclusions of this work are summarised and areas in which further work may prove useful are suggested.

Chapter 2
Adaptive Beamforming

2.1 Introduction

A beamformer is a processor used in conjunction with an array of sensors to provide a versatile form of spatial filtering. The sensor array collects spatial samples of propagating waves, which are processed by the beamformer. An *adaptive* beamformer is one which adapts the sensor weights to the propagating wave field with the objective of identifying a desired signal in the presence of interference and noise. A beamformer can perform spatial filtering to separate signals which have overlapping frequency content but originate from differing locations. This chapter provides an introduction to beamforming from a signal processing perspective. Data independent, and statistically optimum linearly constrained beamformers are discussed.

The problem in adaptive processing is to obtain a set of weighting coefficients w which result in an output signal $y(k)$ having "better" characteristics than would be observed at the output of a conventional fixed weight processor. One commonly used weight selection criterion is linearly constrained adaptive beamforming [10] in which the weights are constrained such that any signal arriving from the desired look direction will appear at the beamformed output with prescribed temporal filtering.

The operation of an adaptive beamformer can be most easily visualised by considering the response in terms of the array sensitivity pattern. Interfering signal suppression is obtained by appropriately steering beam pattern nulls and reducing pattern sidelobes in the directions of interferences, while desired signal reception is maintained by preserving desirable mainlobe features. For the radar of interest, the adaptive algorithm therefore relies on the spatial and temporal characteristics of the interference to improve performance. Many weight adaptation algorithms have been developed over the past three decades with varying degrees of success. The reader is referred to [41] for a comprehensive bibliography. At the present time adaptive nulling is considered to be the principal benefit of the adaptive techniques employed by adaptive array systems, and automatic cancellation of sidelobe jamming provides a valuable electronic counter–countermeasures capability for radar systems.

2.2 Data independent beamformer design

Synthesis techniques for computing the elemental weights of an array beamformer which result in a desired response have existed for many years [42]. The majority of work has focussed on

designs which achieve reduced sidelobe levels whilst retaining prescribed mainlobe characteristics. Such beamforming techniques are independent of the data present at the array and are thus referred to as deterministic design approaches.

In this section we will consider beamformers which can be made to approximate an arbitrary desired response. This may prove useful in several instances. For example, if we may desire to receive a signal over a range of frequencies or directions, and would therefore like to ensure unity response over these regions. Alternatively we may have *a priori* knowledge of the location and frequency band of interfering or jamming signals and would consequently like to ensure zero response to these signals. Both these concepts are familiar in FIR filter design [43]; the former being bandpass filtering, whilst the latter is an example of bandstop filtering.

Consider matching a desired response $r_d(\theta,\omega)$ with the weight vector $\boldsymbol{w}$ at P points. The beamformer response to a sinusoidal source of frequency ω, incident from a bearing of θ radians, is

$$r(\theta,\omega) \;=\; \boldsymbol{w}^H \boldsymbol{d}(\theta,\omega). \tag{2.1}$$

where $\boldsymbol{d}(\theta,\omega)$ is the array steering vector, and H denotes complex conjugate transpose. For notation, we use boldface lowercase and uppercase symbols to denote vectors and matrices. For a general array, the elements of $\boldsymbol{d}(\theta,\omega)$ indicate the relative time delays of the received field samples within the beamforming structure. When the array has a periodic structure, e.g. a linear equi–spaced array, expressions may be derived for the elements of $\boldsymbol{d}(\theta,\omega)$. Consider the linear array of N elements, with L tap delays per element shown in Figure 2.1. For convenience the first element is taken as the phase reference. The steering vector takes the form

$$\boldsymbol{d}(\theta,\omega) \;=\; \left[1 \;\; e^{j\omega\tau_2(\theta)} \;\; e^{j\omega\tau_3(\theta)} \;\; \ldots \;\; e^{j\omega\tau_{NL}(\theta)}\right]^H. \tag{2.2}$$

The $\tau_k\,(\theta)$, $2 \leq k \leq NL$, represent the time delays due to propagation and any tap delays to the point at which the kth weight is applied. Suppose k represents the lth tap of the nth element, then

$$\tau_k\,(\theta) \;=\; \Delta_n\,(\theta) + l\,T_s, \tag{2.3}$$

in which T_s is the sampling interval. $\Delta_n\,(\theta)$ represents the time delay due to propagation from the first to the nth sensor, i.e.

$$\Delta_n \;=\; \frac{d}{c}\,(n-1)\cos\theta, \tag{2.4}$$

in which d is the element spacing in metres, and c is the speed of propagation in ms^{-1}. Matching the beamformer response $r(\theta,\omega)$ to the given response $r_d(\theta,\omega)$ in a least squares sense can be done as follows. First form the overdetermined least squares problem

$$\min_{\boldsymbol{w}} \left|\boldsymbol{A}^H \boldsymbol{w} - \boldsymbol{r}_d\right|^2, \tag{2.5}$$

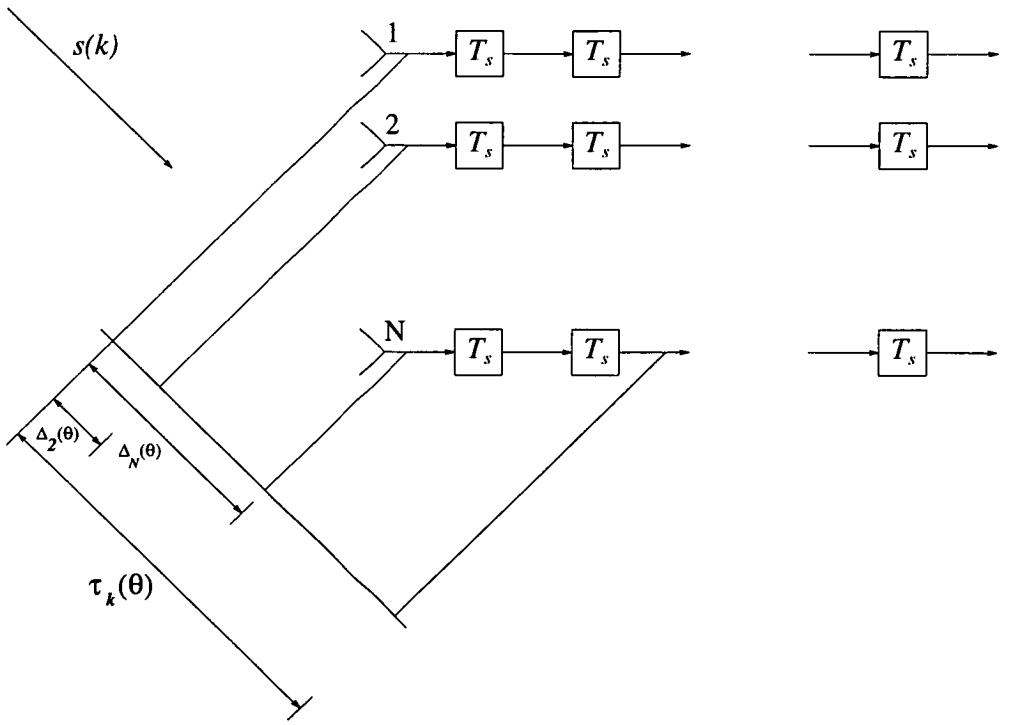

Figure 2.1: *Array with attached delay lines showing the sampling of a signal propagating in plane waves from a source located at θ radians.*

where

$$\begin{aligned} \boldsymbol{A} &= [\boldsymbol{d}(\theta_1,\omega_1)\ \ \boldsymbol{d}(\theta_2,\omega_2)\ \ \cdots\ \ \boldsymbol{d}(\theta_P,\omega_P)], \\ \boldsymbol{r}_d &= [r_d(\theta_1,\omega_1)\ \ r_d(\theta_2,\omega_2)\ \ \cdots\ \ r_d(\theta_P,\omega_P)]. \end{aligned}$$

Assuming that $\boldsymbol{A}\boldsymbol{A}^H$ is invertible (i.e., $\boldsymbol{A}$ is full rank), then the solution to (2.5) will be given by

$$w = \boldsymbol{A}^+ r_d, \tag{2.6}$$

where $\boldsymbol{A}^+ = (\boldsymbol{A}\boldsymbol{A}^H)^{-1}\boldsymbol{A}$ is the pseudo inverse of $\boldsymbol{A}$. This weight vector minimises the squared error between the actual and desired responses at the P points (θ_i,ω_i). The response of such a beamformer is shown in Figure 2.2. In this example the beamformer has been designed according to (2.5) with a desired response of unity over the normalised frequency interval $[0.4, 0.8]$ at 18 degrees. P was chosen equal to 200. The array is linear equi–spaced with 16 elements spaced at one half wavelength and 5 tap FIR filters used at each element.

Several points of caution should be made. Firstly, although this technique does provide good control over beamformer response within the design regions, nothing can be said about the response outwith these regions. The antenna pattern may display unacceptably large gain (i.e., large sidelobes) which may well be comparable in magnitude with the mainlobe response. Clearly this is not a desirable situation, and care must therefore be taken to ensure that this does not occur. Secondly, the weight vector found may lead to a large *white noise gain*, that

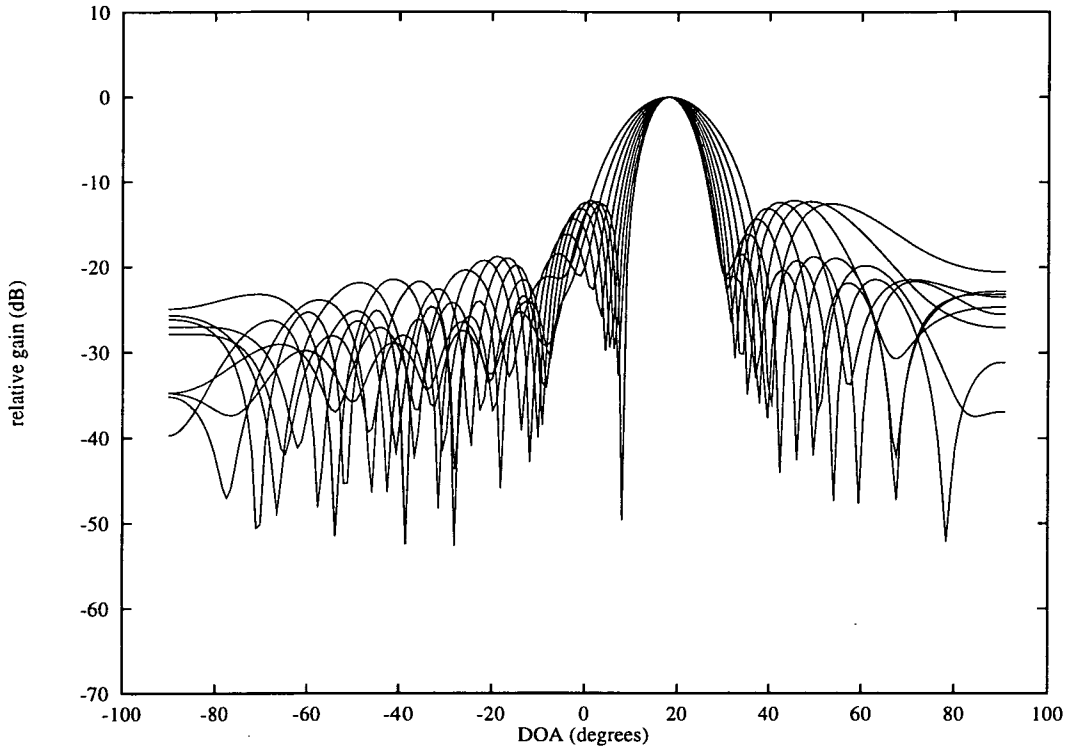

Figure 2.2: *Response of data independent beamformer at eight frequencies linearly spaced in the design band [0.4, 0.8]. The beamformer is designed to have unity gain at 18° over the normalised frequency band [0.4, 0.8].*

is the L_2 norm of the weight vector may be large. This may result in poor signal–to–noise performance because of the large gain experienced by white noise contributions. If $\boldsymbol{A}$ is ill–conditioned then the norm of $\boldsymbol{w}$ will be very large so that low rank estimates of $\boldsymbol{A}$ should be used whenever $\boldsymbol{A}$ is not full rank. A singular value decomposition can easily provide low rank approximations for $\boldsymbol{A}$ and $\boldsymbol{A}^+$.

The similarities between the synthesis techniques used in finite impulse response (FIR) filter design and linear array beamformer design are striking. A linear equi–spaced array can be seen as a spatial filter in which sampling occurs at multiples of the signal wavelength. The equivalent in FIR filter design is signal period. Many low sidelobe weighting functions have arisen for FIR filters. These may readily be applied in deterministic beamformer design. [43] provides a summary of many such shading functions. These deterministic design techniques can be used to emphasise certain directions or frequencies and to de–emphasise others. Continuing with the FIR filter analogy, a bandstop filter is equivalent to a spatially selective filter. The analogy fails when we consider either planar, or broadband arrays. In the former case we have discrimination in both azimuth and elevation, which has no time equivalence. In the latter case, we combine both spatial and temporal resolution, which has no FIR filter equivalent.

Since an array of sensors can be utilised to obtain some degree of spatial filtering or directional sensitivity much of the early literature on array processing was concerned with desirable beam patterns. These traditional array synthesis techniques used amplitude tapering of array ele-

ments to control the array response. During the 1960's interest moved toward arrays in which elements were not spaced on some periodic grid. Skolnik [44] and others [45, 46] investigated the use of statistically designed density tapers to control antenna pattern sidelobes. In the procedure described by Skolnik all of the array elements were excited with equal amplitude and the density of antenna elements was matched to the amplitude of the aperture illumination. Numazaki *et al.* [47] extended Skolnik's techniques to arrays in which element weights were quantised. Steinberg [46] studied various features of arrays in which the element positions were selected randomly. It was found that the ensemble averaged pattern, averaged over several random arrays would be equal to the Fourier transform of the probability distribution function from which element positions were drawn. This analysis applied only to the ensemble average of random arrays. However, the peak sidelobe of a single statistically designed array - an important measure of beamformer performance - can only be described in a statistical sense. An approximate expression was derived showing the peak sidelobe to be relatively independent of array size, beam steering angle, and taper function. The advantages of statistical thinning are many -

- Vastly reduced element numbers : up to 90% thinning without significant sidelobe degradation
- No grating lobes : no periodicity exists in element location
- Similar angular resolution to a filled array
- Reduced mutual coupling through greater interelement spacing
- Less tolerance required on element location and excitation
- Equal excitation means improved efficiency

These benefits are not penalty free. Array gain will be reduced considerably by the removal of array elements. In radar applications this may well preclude the use of these techniques. In addition the reduction in the number of elements will reduce the designer's control of radiation within the sidelobe region. Despite these problems, statistically designed arrays can provide solutions for many other applications where good angular resolution is required at minimal cost.

2.3 Data dependent beamformer design

The noise power received by a sensor array will be variable in space and time, so it follows that any optimum interference suppression can only be obtained through adaptive methods. A broadband adaptive processor combines both spatial and temporal samples of the received field in such a manner as to maximise (or minimise) some performance measure. Adaptive processors can be grouped into two distinct types, those that adapt on the interference field alone, and those that adapt on the received field with the desired signal included. The processors considered in this chapter are of both types, examples being the minimum power beamformer,

and the generalised sidelobe canceller. In all adaptive processors a set of weighting coefficients are derived which result in an output signal having better signal to noise characteristics than would be observed at the output of a conventional beamforming system.

Rather than try to attribute developments due to many different researchers in beamforming, the reader is referred to the following references: books - Hudson [48], Monzingo and Miller [4], Haykin ed. [49], Haykin [50]; special issues - *IEEE Transactions on Antennas and Propagation* [51,52]; tutorial - Gabriel [11]; and bibliography - Marr [41]. Papers devoted to beamforming are often found in the IEEE Transactions on: *Antennas and Propagation, Acoustics, Speech, and Signal Processing* (latterly *Signal Processing*), *Aerospace and Electronic Systems*, and in the IEE Proceedings on *Radar and Signal Processing*, (latterly *Radar, Sonar and Navigation*). There is a vast body of literature on various aspects of beamforming and only a subset will be referred to in this thesis. Much of the literature pertaining to adaptive filtering of time series is useful in beamforming discussions, since their histories are both parallel and overlapping.

Adaptive arrays are a radical departure from conventional thinking in antenna design, offering substantial improvements in performance over fixed pattern antennae in environments which include severe interference and jamming. They achieve this because of their ability to steer response nulls automatically in the direction of unknown or interfering signals and to generally alter their beampatterns in order to optimise performance. The class of beamformers which will be considered within this thesis are a subset of the general class of adaptive beamformer, termed linearly constrained beamformers. These select a set of adaptive weights which minimise the output power subject to a set of linear constraints. References to linearly constrained beamformers are contained in [9,10,12,13,15–25,27–30,32–40,53–60].

We will now review some of the data concepts and terminology for the remainder of the thesis, and then introduce the concept of linearly constrained adaptive beamforming.

2.4 Terminology

The interferences that will be considered in this thesis are two–dimensional in nature, that is they are distributed in both space and time. It is therefore necessary that the received field is sampled in both space and time. Evaluation of beamformer performance usually involves power or variance, so the second order statistics of the data play an important role. Throughout this thesis the received field will be assumed to be a wide–sense stationary discrete–time stochastic process. Suppose the field incident at a particular element i at time k is given by $x_i(k)$. The snapshot vector of all elemental signals at time k is

$$\mathbf{x}(k) \;=\; [\, x_1(k)\; x_2(k)\; \cdots\; x_N(k)\,]^T, \tag{2.7}$$

and the *stacked* snapshot vector containing the L previous snapshots up to and including the current snapshot is defined as

$$\boldsymbol{x}(k) \;=\; \left[\, \mathbf{x}^T(k)\; \mathbf{x}^T(k-T_s)\; \cdots\; \mathbf{x}^T(k-(L-1)T_s)\,\right]^T, \tag{2.8}$$

where T denotes transpose, and T_s is the sampling period. Let us define matrices

$$\boldsymbol{M}(\tau) \;=\; E\left\{\mathbf{x}(k)\,\mathbf{x}^H(k-\tau)\right\}, \tag{2.9}$$

which contain the spatial cross–correlation between two particular snapshots τ seconds apart of the data incident at an N element array. $E\{.\}$ denotes expectation. The complete $NL \times NL$ space–time autocorrelation matrix is then formed as

$$\boldsymbol{R}_x(\tau) \;=\; \begin{bmatrix} \boldsymbol{M}(\tau) & \cdots & \boldsymbol{M}(\tau + lT_s) & \cdots & \boldsymbol{M}(\tau+(L-1)T_s) \\ \vdots & \ddots & & & \vdots \\ \boldsymbol{M}^H(\tau+lT_s) & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \boldsymbol{M}^H(\tau+(L-1)T_s) & \cdots & \cdots & \cdots & \boldsymbol{M}(\tau) \end{bmatrix} \tag{2.10}$$

The spatial information is contained in the submatrices $\boldsymbol{M}(\tau)$ whereas the temporal information lies between them. In this way the structure of the input data is retained, allowing the beamformer to discriminate interferences in both bearing and frequency. The complete space–time autocorrelation matrix of (2.10) can be written more succinctly in terms of the stacked snapshot vector $\boldsymbol{x}(k)$ as follows

$$\boldsymbol{R}_x(\tau) \;=\; E\left\{\boldsymbol{x}(k)\,\boldsymbol{x}^H(k-\tau)\right\}. \tag{2.11}$$

The correlation matrices of principal concern in the analysis of adaptive array behaviour are those for which the time–delay variable τ is zero. Rather than write the argument explicitly as $\boldsymbol{R}_x(0)$, it is simpler to define $\boldsymbol{R}_x \triangleq \boldsymbol{R}_x(0)$. If the data is wide–sense stationary, then $\boldsymbol{R}_x$ will be independent of time. Covariance matrices are closely related to correlation matrices since the covariance between the vectors $\boldsymbol{x}(k)$ and $\boldsymbol{y}(k)$ is defined by

$$\mathrm{cov}\left[\boldsymbol{x}(k), \boldsymbol{y}(k)\right] \;=\; E\left\{\left(\boldsymbol{x}(k)-\bar{\boldsymbol{x}}\right)\left(\boldsymbol{y}(k)-\bar{\boldsymbol{y}}\right)^H\right\}, \tag{2.12}$$

where

$$\bar{\boldsymbol{x}} = E\left\{\boldsymbol{x}(k)\right\} \quad \text{and} \quad \bar{\boldsymbol{y}} = E\left\{\boldsymbol{y}(k)\right\}.$$

Consequently for zero–mean processes and at delay $\tau = 0$, correlation matrices and covariance matrices are identical, and the adaptive array literature frequently uses the two terms interchangeably. The covariance matrix has a key role in the statistical analysis and design of beamforming systems, so it is useful to understand its various properties and the implications thereof. In particular, using the definition of (2.11), the covariance matrix $\boldsymbol{R}_x$ has the following properties :

(i) $\boldsymbol{R}_x$ is Hermitian.

(ii) $\boldsymbol{R}_x$ is block Toeplitz.

(iii) $\boldsymbol{R}_x$ is non–negative and almost always positive definite.

This is not an exhaustive list, but these are the properties which are of particular relevance to the problem at hand. The Hermitian and Toeplitz nature of $\boldsymbol{R}_x$ is a direct consequence of the assumption that the stochastic process is wide–sense stationary. Indeed, we may even use these properties as a measure of the stationarity of the process. The inverse of $\boldsymbol{R}_x$ will also be Hermitian. Furthermore, because $\boldsymbol{R}_x$ is Hermitian and positive semidefinite the eigenvalues of $\boldsymbol{R}_x$ are real and positive, and all its eigenvectors are orthogonal. Therefore the eigenvectors of $\boldsymbol{R}_x$ can be used to form a basis for a subspace termed the signal subspace. Additionally, for any vector $\boldsymbol{w}$, we have

$$\boldsymbol{w}^H \boldsymbol{R}_x \boldsymbol{w} \geq 0. \tag{2.13}$$

This has a clear physical interpretation since $\boldsymbol{w}^H \boldsymbol{R}_x \boldsymbol{w}$ is the power output from the beamformer for a given weighting vector $\boldsymbol{w}$, and will therefore always be greater than or equal to zero. The Toeplitz form is desirable because the entire matrix can be reconstructed from the first row of submatrices, that is, $\boldsymbol{M}(0), \boldsymbol{M}(T_s), \cdots, \boldsymbol{M}((L-1)T_s)$. Consequently only a $N \times NL$–dimensional matrix is needed to be stored to have all the information contained in $\boldsymbol{R}_x$. To complete these definitions the crosscorrelation at lag τ between two signals $\boldsymbol{x}(k)$ and $\boldsymbol{y}(k)$ having stationary statistical properties is defined as

$$\boldsymbol{R}_{xy}(\tau) = E\left\{\boldsymbol{x}(k)\boldsymbol{y}^H(k-\tau)\right\}. \tag{2.14}$$

If two signals are totally uncorrelated then $\boldsymbol{R}_{xy}(\tau)$ will be zero, whereas if $\boldsymbol{x}(k)$ and $\boldsymbol{y}(k)$ are correlated, then the elements of $\boldsymbol{R}_{xy}(\tau)$ will reflect the correlation that does exist.

2.5 Linearly constrained broadband beamforming

The beamforming structure depicted in Figure 2.3 represents a basic broadband beamformer. The number of sensors is N, and the prefilters ρ_i are included to model demodulation and other filtering functions which may be present in the system. In the general case here, time–delay steering is *not necessarily* presumed to be present in the prefilters. The beamformer combines L successive samples from each prefilter through complex weights for a total of NL degrees of freedom. In a radar the time–delays will typically be matched to the pulse repetition interval. In vector notation, the beamformed output $y(k)$ is the vector inner product of the data vector and a NL–dimensional stacked weight vector $\boldsymbol{w}$,

$$y(k) = \boldsymbol{w}^H \boldsymbol{x}(k) = \sum_{n=1}^{N} \sum_{l=0}^{L-1} w_{nl}^* x_n(k-l). \tag{2.15}$$

Note that the form of this equation implies that the complex conjugates of the weights are actually applied to the data samples within the beamformer. The reason for using this formulation is that it leads to a compact notation which avoids the necessity of distinguishing between real and Hermitian transpose operators. A second advantage is that the form of the solutions for the tapped–delay line case are identical to those published widely for real tap weights, except for the replacement of real with Hermitian transpose operators. The beamforming weighting structure implied by (2.15) is termed the direct–form implementation.

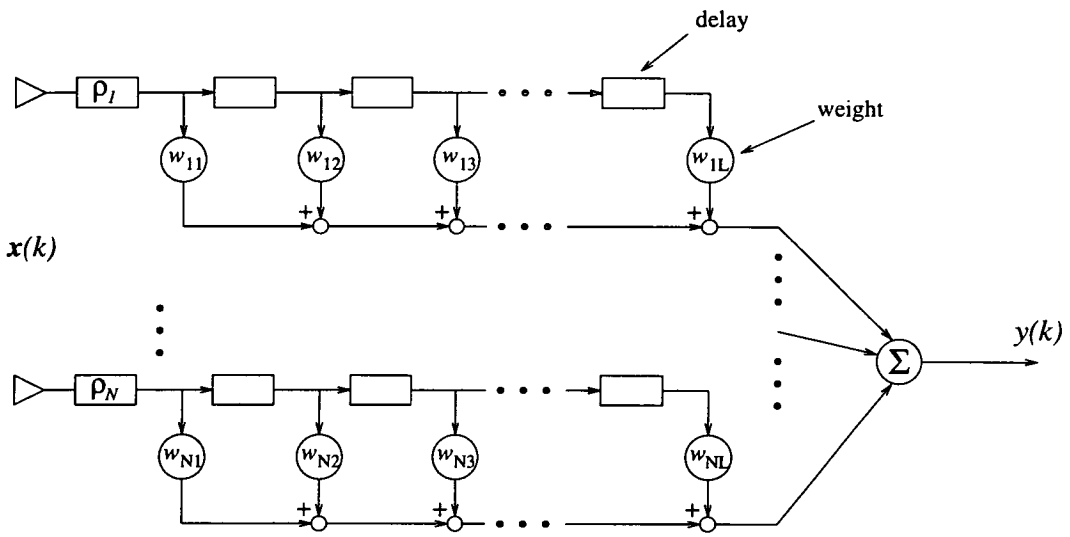

Figure 2.3: General form for digital broadband array beamforming system.

Data dependent weight synthesis methods are based upon the optimisation of a performance criterion for the array. For the case of narrow–band beamformers, virtually all known criteria result in the same weight vector response (within a scale factor) for a given signal, interference and noise environment. The weights obtained for broadband beamformers are, however, criteria dependent. During the past several years, much interest has focused on the use of linearly constrained minimum output power criterion [10]. The reasons for this interest stem largely from the fact that linear constraints can readily be used to control the mainlobe and sidelobe response of broadband arrays. This is achieved by fixing the gain of the array at selected response points corresponding to particular angles and frequencies of interest [12]; by forcing zero derivatives (with respect to bearing and/or frequency) in the beam pattern [18]; or, more recently, by using an orthogonal transformation based on eigenvector analysis [54]. In the last of these a somewhat surprising result was found, namely that the array response could be steered in arbitrary directions using only linear constraints, i.e., steering delays in the array elements are not required.

Each linear constraint on the array response is defined by a NL–dimensional vector $\boldsymbol{C}_k$. It is assumed that a total of K such linear constraints are used and that K is less than the total adaptive degrees of freedom available NL. The constraint equation is defined as $\boldsymbol{C}_k^H \boldsymbol{w} = f_k$,

where f_k is a scalar complex constant. The set of K linear constraint equations is given by

$$\boldsymbol{C}^H \boldsymbol{w} = \boldsymbol{f}, \tag{2.16}$$

where $\boldsymbol{C}$ is the $NL \times K$ constraint matrix, and $\boldsymbol{f}$ is the K-dimensional response vector containing the f_k. In the most simple case the columns of the matrix $\boldsymbol{C}$ are steering vectors for the array, and the elements of $\boldsymbol{f}$ specify the desired response in those steered directions. The linearly constrained mimimum variance (LCMV) beamformer is defined as the set of weights which satisfies (2.16) whilst simultaneously minimising the total output power, that is

$$\min_{\boldsymbol{w}} \boldsymbol{w}^H \boldsymbol{R}_x \boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{C}^H \boldsymbol{w} = \boldsymbol{f}. \tag{2.17}$$

The vector of direct–form beamformer coefficients that satisfies this can be found by the method of Lagrange multipliers as

$$\boldsymbol{w}_{opt} = \boldsymbol{R}_x^{-1} \boldsymbol{C} \left(\boldsymbol{C}^H \boldsymbol{R}_x^{-1} \boldsymbol{C} \right)^{-1} \boldsymbol{f}. \tag{2.18}$$

Constraining the array response with K linear constraints reduces the total adaptive degrees of freedom from NL to $NL-K$. It is these remaining degrees of freedom that are used to minimise the output power from the array. When the array operates in the presence of uncorrelated white noise, i.e., a benign environment, the response is termed the *quiescent* beamformer, given by

$$\boldsymbol{w}_q = \boldsymbol{C} \left(\boldsymbol{C}^H \boldsymbol{C} \right)^{-1} \boldsymbol{f}. \tag{2.19}$$

As the equation clearly demonstrates, the quiescent response of a linearly constrained beamformer is completely determined by the constraint equations.

A key feature of linearly constrained beamformers is the ability to perform a specified filtering operation upon signals incident from the steer direction. This is accomplished by employing steering–delays at the array elements [10]. The desired signal is identified by time–delay steering the sensor outputs so that any signal incident at the array from the direction of interest (look direction) appears as an identical replica at the outputs of the steering delays. All other signals which do not have this property are processed as noise or interference. As far as the desired signal is concerned the array then appears as a single tapped delay line, with tap weights given by the sum of the corresponding elemental tap weights for a particular tap. The pre-steering delays therefore allow control of the frequency response of the array in a prescribed look direction. Expressing this argument more formally, steer direction filtering is obtained by defining the constraint matrix $\boldsymbol{C}$ as

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{1}_N & \boldsymbol{0}_N & \cdots & \\ \boldsymbol{0}_N & \boldsymbol{1}_N & \cdots & \\ & & & \boldsymbol{0}_N \\ \vdots & \vdots & \vdots & \boldsymbol{1}_N \end{bmatrix}, \tag{2.20}$$

in which $\boldsymbol{1}_N$ and $\boldsymbol{0}_N$ are vectors of length N, with either all one or all zero entries. $\boldsymbol{C}$ has exactly

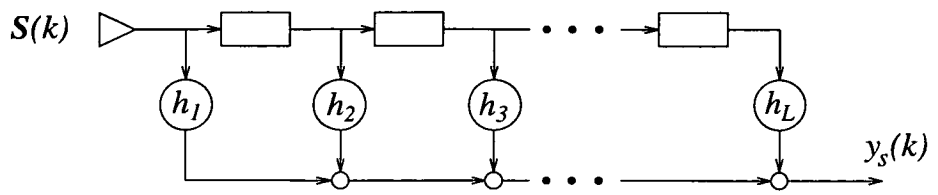

Figure 2.4: Equivalent structure for signal incident from the look direction.

L columns, one for each bank of taps in the beamformer. Frost [10] noted that the response of the beamformer to a desired signal $s(k)$, incident from the steered direction is equivalent to implementing a finite-impulse response (FIR) filter. This can be understood by noting that a NL-dimensional desired signal vector $s(n)$ can be expressed in terms of a L-dimensional time sampled vector of the desired signal

$$\boldsymbol{S}(k) \;=\; \boldsymbol{C}^H \boldsymbol{s}(k), \tag{2.21}$$

where $\boldsymbol{S}(k)$ contains the L delayed samples of the desired signal. The desired signal output component is therefore

$$y_s(k) \;=\; \boldsymbol{w}^H \boldsymbol{s}(k) \;=\; \boldsymbol{h}^H \boldsymbol{S}(k). \tag{2.22}$$

The L-dimensional vector $\boldsymbol{h} = [h_1 \;\cdots\; h_L]$, contains the equivalent FIR filter coefficients h_l, as shown in Figure 2.4. Controlling the look direction frequency response therefore simply becomes a case of selecting the desired number of FIR filter coefficients, and computing the values of the h_l. The corresponding values of f_l are h_l. The relationship between $\boldsymbol{w}$ and $\boldsymbol{h}$ is given explicitly as

$$\boldsymbol{h} \;=\; \left(\boldsymbol{C}^H \boldsymbol{C}\right)^{-1} \boldsymbol{C}^H \boldsymbol{w}. \tag{2.23}$$

Identification of desired signals can be seen as a two stage process; firstly, spatial filtering of undesired signals is performed by the steering delays on the sensor elements; secondly, temporal filtering of what remains is performed by the elements of the constraint vector $\boldsymbol{f}$. The constraints described by (2.20) are part of a general class of constraints known as directional or point constraints [12]. The example above is a particularly simple case - it provides a look direction gain constraint and some spectral filtering. It is possible, however, to constrain the response in other directions other than the steered direction, albeit with a non-sparse set of constraint vectors. Each constraint vector, along with the associated element of $\boldsymbol{f}$, will specify a relationship between all the elements of $\boldsymbol{w}$. This more general utilisation of the constraint equations may be motivated by a desire to remove the dependency upon steering delays, or to impose response conditions based upon some a *priori* knowledge of interferer locations and frequencies.

Figure 2.5 depicts graphically the operation of a linearly constrained beamformer. Contours of constant output power (cost) and the optimum constrained weight vector $\boldsymbol{w}_{opt}$ that minimises

the output power are shown. The constraint equations define a $NL-K$–dimensional hyperplane

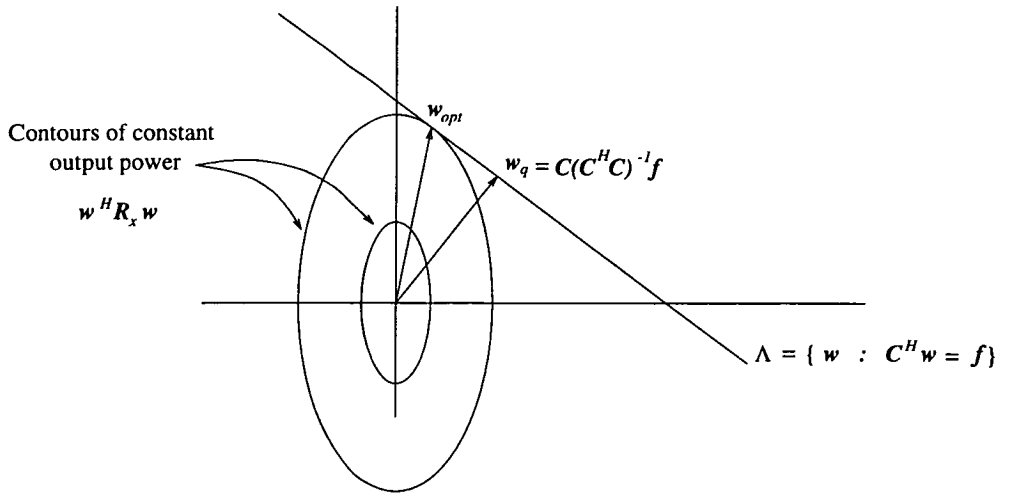

Figure 2.5: *Example showing contours of constant output power and the constrained weight vector that minimises output power,* $w_{opt} = \boldsymbol{R}_x^{-1}\boldsymbol{C}\left(\boldsymbol{C}^H\boldsymbol{R}_x^{-1}\boldsymbol{C}\right)^{-1}\boldsymbol{f}$.

Λ, in NL–dimensional space. This constraint plane is defined by

$$\Lambda \;=\; \left\{w \;:\; \boldsymbol{C}^H w = \boldsymbol{f}\right\}. \tag{2.24}$$

Additionally, it is possible to define a constraint subspace as the solution to the homogeneous set of equations

$$\Sigma \;=\; \left\{w \;:\; \boldsymbol{C}^H w = \boldsymbol{0}\right\}. \tag{2.25}$$

The optimum weight vector is that vector which terminates on the constraint plane and simultaneously minimises the output power. Vectors which point in a direction normal to the constraint plane (but not necessarily terminating on the plane) can be expressed as linear combinations of the constraint vectors $\boldsymbol{C}$. Thus, the vector $w_q = \boldsymbol{C}\left(\boldsymbol{C}^H\boldsymbol{C}\right)^{-1}\boldsymbol{f}$, points in a direction normal to the constraint plane, and terminates on the constraint plane, since $\boldsymbol{C}^H w_q = \boldsymbol{f}$. w_q is therefore the shortest vector terminating on the constraint plane, and forms the quiescent solution. Continuous adaptation algorithms have been developed for the linearly constrained beamformer, and can be easily interpreted in relation to Figure 2.5. In Frost's constrained least mean squares algorithm [10], the weight vector is initialised as $w(0) = \boldsymbol{C}\left(\boldsymbol{C}^H\boldsymbol{C}\right)^{-1}\boldsymbol{f}$. As each new snapshot becomes available a weight vector is computed. This new weight vector may not satisfy the constraints, so it is projected back onto the constraint subspace then returned to the constraint plane by adding w_q. The new weight vector $w\,(k+1)$ satisfies the constraints to within the accuracy of the arithmetic used in implementing the algorithm. Other references to continuous constrained adaptation algorithms can be found in [12, 22, 53].

Use of linear constraints is a very general approach that permits extensive control over the adapted response of the beamformer. However, the ability to control the response depends upon several variables, most notably the size of the constraint region, the number of constraints

employed and the manner in which the constraints are utilised. Following a process similar to that in section 2.2, we may get a qualitative feel for how well the beamformer can match a desired response. Suppose the desired response, defined in the region $\theta \in [\theta_a, \theta_b], \omega \in [\omega_a, \omega_b]$, is given by $g_1(\theta, \omega)$. A total of K linear constraints are assumed to be available and these are contained in the matrix $\boldsymbol{C}$. The response of the beamformer to a signal $s(\theta, \omega)$, lying within the constraint region is

$$r(\theta, \omega) = w^H \mathbf{s}(\theta, \omega), \tag{2.26}$$

where w will be computed according to (2.18), and the vector $\mathbf{s}(\theta, \omega)$ contains the temporal and spatial samples of $s(\theta, \omega)$. Denoting the partial response due to the kth column of $\boldsymbol{C}$ by $r_k(\theta, \omega)$, and $\boldsymbol{\Gamma} = \left(\boldsymbol{C}^H \boldsymbol{R}_x^{-1} \boldsymbol{C}\right)^{-1} \boldsymbol{f}$, then expanding yields

$$r_k(\theta, \omega) = \boldsymbol{\Gamma}^H \boldsymbol{C}_k^H \boldsymbol{R}_x^{-1} \mathbf{s}(\theta, \omega), \qquad 1 \leq k \leq K, \tag{2.27}$$

where $\boldsymbol{C}_k$ is the kth constraint vector in $\boldsymbol{C}$. (2.18) requires $r(\theta, \omega)$ to be expressed as a linear combination of the $r_k(\theta, \omega)$, $1 \leq k \leq K$, over the region $\theta \in [\theta_a, \theta_b], \omega \in [\omega_a, \omega_b]$. In general, this cannot be achieved so we conclude that perfect control of beamformer response cannot be accomplished. Several different philosophies can be employed for choosing the constraint matrix and response vector, a large number of which apply least squares approaches to the above problem. Constraint design techniques are discussed in section 2.7.

Before constraint design is discussed, we will describe an alternative implementation of the linearly constrained beamformer, called the generalised sidelobe canceller. This structure maps the constrained minimisation of the direct–form implementation to an unconstrained optimisation through a transformation which is orthogonal to the original constraint set $\boldsymbol{C}$. The resulting structure is flexible, and is readily applicable to the partially adaptive beamforming problem.

2.6 Generalised sidelobe canceller

The generalised sidelobe canceller (GSC) represents an alternative formulation of the LCMV problem, which provides new insight, is useful for analysis, and can simplify LCMV beamformer implementation. Essentially the GSC is a mechanism for changing a constrained minimisation problem into an unconstrained form. Griffiths and Jim [23] applied similar concepts to those found in some linear least squares problems to LCMV beamforming and coined the term GSC. Similar techniques were discussed by Applebaum and Chapman [51].

The linearly constrained adaptive processor defined by (2.17)–(2.18) can be implemented in one of two forms. In the first, the direct–form implementation, each coefficient in the beamformer is updated by the adaptive processor which computes new weights using an adaptive algorithm. An alternative implementation which yields equivalent steady–state performance can be derived from Frost's initial algorithm [22, 23]. This structure is termed the generalised

sidelobe canceller form and is depicted in Figure 2.7. The generalised sidelobe canceller is a useful vehicle for generalising the linearly constrained beamformer to include any arbitrary quiescent response. Additionally, the adaptive dimension of the adaptive processor can be modified without changing the quiescent response.

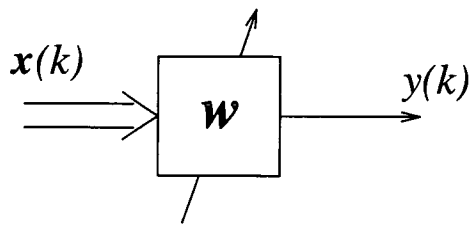

Figure 2.6: Direct implementation of linearly constrained adaptive processor.

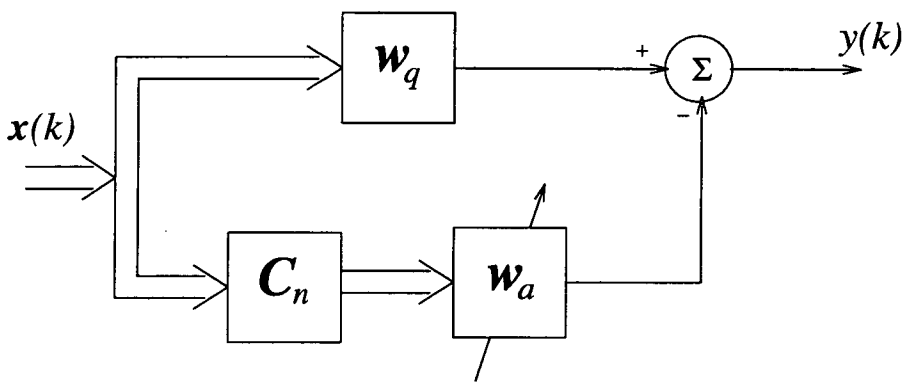

Figure 2.7: Generalised sidelobe canceller structure.

In a generalised sidelobe canceller the weight vector is decomposed into two orthogonal terms, one that lies in the constraint space spanned by the columns of $\boldsymbol{C}$, denoted by $\boldsymbol{w}_q$, and another that is orthogonal to the space spanned by the columns of $\boldsymbol{C}$, represented by $-\boldsymbol{C}_n\boldsymbol{w}_a$, so that $\boldsymbol{w} = \boldsymbol{w}_q - \boldsymbol{C}_n\boldsymbol{w}_a$. These two terms are implemented in the upper and lower processing paths. In the upper path a conventional non–adaptive beamformer combines the array tap data with the fixed weights $w_{q1}, w_{q2}, \ldots, w_{qNL}$, producing some non-adaptive signal $y_q(k)$

$$y_q(k) \;=\; \boldsymbol{w}_q^H \boldsymbol{x}(k), \tag{2.28}$$

where $\boldsymbol{w}_q = [\, w_{q1}\ w_{q2}\ \cdots\ w_{qNL}\,]^T$. This conventional beamsteering system is identical to that which would be used in a fixed non–adaptive array. In most applications the weights $\boldsymbol{w}_q$ are chosen to provide a trade–off between beamwidth and sidelobe level.

The lower path in Figure 2.7 is the sidelobe cancelling path. It consists of a matrix preprocessor $\boldsymbol{C}_n$ followed by an unconstrained set of adaptive weights $\boldsymbol{w}_a$. In it's most simple form, the purpose of the $NL \times J$ matrix $\boldsymbol{C}_n$ is to prevent the desired signal $s(k)$ from entering the lower path. The adaptive weights $\boldsymbol{w}_a$ are then allowed to adapt in an unconstrained manner. Therefore, the generalised sidelobe canceller can be seen as a mapping from the constrained minimisation problem to that of an unconstrained minimisation problem. The matrix $\boldsymbol{C}_n$ performs the mapping and required reduction in adaptive dimension. The generalised sidelobe

canceller and the Frost beamformer [10] will have equivalent steady–state performance if $J = NL - K$. If additionally $\boldsymbol{C}^H \boldsymbol{C}_n = \boldsymbol{0}$, then $\boldsymbol{w}_q$ will determine the quiescent response [17]. If $\boldsymbol{w}_q$ is chosen to specify the quiescent response then the optimum value for $\boldsymbol{w}_a$ under quiescent conditions is zero. This can be ensured if two conditions are met; firstly that the columns of $\boldsymbol{C}_n$ are linearly independent, and secondly that each column of $\boldsymbol{C}_n$ is orthogonal to all the columns of the constraint matrix $\boldsymbol{C}$. The generalised sidelobe canceller problem can thus be expressed as

$$\boldsymbol{C}^H \boldsymbol{w}_q = \boldsymbol{f} \; ; \; \boldsymbol{C}^H \boldsymbol{C}_n = \boldsymbol{0} \; ; \; rank[\, \boldsymbol{C} \mid \boldsymbol{C}_n \,] = NL, \tag{2.29}$$

and

$$\min_{\boldsymbol{w}_a} \; (\boldsymbol{w}_q - \boldsymbol{C}_n \boldsymbol{w}_a)^H \; \boldsymbol{R}_x \; (\boldsymbol{w}_q - \boldsymbol{C}_n \boldsymbol{w}_a) \, . \tag{2.30}$$

The solution of (2.30) is given by

$$\boldsymbol{w}_a \;\; = \;\; \left(\boldsymbol{C}_n^H \boldsymbol{R}_x \boldsymbol{C}_n\right)^{-1} \boldsymbol{C}_n^H \boldsymbol{R}_x \boldsymbol{w}_q . \tag{2.31}$$

The orthogonality of $\boldsymbol{C}$ and $\boldsymbol{C}_n$ implies that the constraints are satisfied independently of $\boldsymbol{w}_a$. $\boldsymbol{w}_a$ is thus unconstrained and represents the available degrees of freedom in $\boldsymbol{w}$. It is straightforward to show that the quiescent weight vector $\boldsymbol{w}_q = \boldsymbol{C}\left(\boldsymbol{C}^H \boldsymbol{C}\right)^{-1} \boldsymbol{f}$. Thus to form an equivalent implementation to the Frost beamformer the problem becomes a simple case of finding the $NL - K$ linearly independent columns of $\boldsymbol{C}_n$ such that

$$\boldsymbol{C}^H \boldsymbol{C}_{ni} \;\; = \;\; \boldsymbol{0} \; ; \quad i = 1, \; \cdots, \; NL - K \tag{2.32}$$

For an array with wideband steering delays this can be easily achieved, as will be discussed later.

The GSC is a useful structure from which to view linearly constrained beamforming. As an example, assume that the linear constraint defined by $\boldsymbol{d}^H(\theta, \omega)\, \boldsymbol{w} = g$ is imposed. $\boldsymbol{C}_n$ satisfies $\boldsymbol{d}^H(\theta, \omega)\, \boldsymbol{C}_n = \boldsymbol{0}$ so each column $[\boldsymbol{C}_n]_k$ can be viewed as a data independent beamformer with a null in the direction θ at frequency ω: $\boldsymbol{d}^H(\theta, \omega)\, [\boldsymbol{C}_n]_k = 0$. Thus, a signal of frequency ω and direction θ arriving at the array will be blocked or nulled by the matrix $\boldsymbol{C}_n$. In general if the constraints are designed to present a specified response to signals with set directions and frequencies, then the columns of $\boldsymbol{C}_n$ will block those directions and frequencies. These signals will only be processed by $\boldsymbol{w}_q$, and since $\boldsymbol{w}_q$ satisfies the constraints, they are presented with the desired response independent of $\boldsymbol{w}_a$. Signals incident from directions and frequencies not controlled by $\boldsymbol{C}$ will pass through both the upper and lower path; the upper channel forms a fixed response through $\boldsymbol{w}_q$, whilst the lower branch forms an estimate of the signal in the upper path with a linear combination of the tap data. This is familiar to traditional estimation problems, in which auxiliary sensors are combined linearly in order to estimate a primary channel output.

It may seem that if the generalised sidelobe canceller has equivalent steady-state performance to the direct form implementation of the linearly constrained beamformer, then the additional complexity involved in the GSC would make it foolish to use this implementation. This view, though, only considers the beamformer in a fully adaptive sense. The GSC structure becomes far more useful in the form of a partially adaptive processor. The desirable features of a partially adaptive beamformer are many. Primarily it should have a low adaptive dimension. Partial adaptivity is achieved by either deleting a portion of the sensor elements, or by combining all the elemental signals in a pre–beamformer. Of equal importance is that the structure should retain as near to fully adaptive performance as possible. This is the objective of the techniques discussed in chapter 4. Another concern, particularly in the radar community, is that reduction in adaptive dimension should not result in loss of gain to a desired signal. This effectively rules out thinned or aperiodic arrays, since a overriding concern in airborne radar is target observability. If a GSC like structure is used, desired signal gain can be maintained whilst simultaneously reducing the computational load. Full gain is ensured by the upper channel, and an arbitrary adaptive dimension can be selected in the lower path.

Linearly constrained beamforming, and latterly the GSC structure, have been an active topic of research almost since their first appearance. This is evidenced by the large number of research papers published each year. Many aspects of GSC design have been examined, not least of which is the applicability to partially adaptive processing. To summarise GSC literature: general and review papers [20,22–25]; constraint design and filtering properties [15,18,21,34,59,61]; partial adaptivity [9,19,27–40]; quiescent pattern control [17,33]; modified filtering structures [55,57, 58]; and coherent interference suppression [35,36].

2.7 Constraint design

A variety of different methods for constraining the beamformer response have arisen. The most obvious constraints are *point constraints* which constrain the beamformer response at particular points of spatial direction or temporal frequency. Each constraint will constrain the beamformer response at a particular point, so that controlling the beamformer response over large spatial or spectral regions would require an overly large number of point constraints. Er and Cantoni [13] extended point constraints to a more general class of *derivative constraints.* These constraints are used to broaden the beamformer response about the constraint point in order to make the beamformer less sensitive to steering mismatch. A typical situation in which these would be used is the case in which the direction of the desired signal is known only approximately. Latterly techniques have become available which offer a more general method of constraining the beamformer response over regions of space and bands of frequency. These constraints are termed *eigenvector constraints* [16,19].

2.7.1 Point constraints

Optimum beamforming with multiple linear constraints is now a well known technique in array processing. In the simplest case, a single linear constraint is imposed, namely unity gain response in the steer direction; the weight vectors are then calculated by minimising the beamformer mean output power subject to this linear constraint. For a narrowband beamformer this simple case is equivalent to the maximised signal–to–noise ratio beamformer.

The constraints used here constrain the array response in the direction of multiple desired signals, and are hence often called directional constraints [10,12,18]. Each directional constraint is formed by using *a priori* knowledge of the direction from which desired signals will impinge upon the array. For each desired signal direction, a single constraint vector and response value are specified. Suppose the beamformer response is required to be controlled at the multiple points $(\theta_1,\omega_1),(\theta_2,\omega_2),\cdots,(\theta_P,\omega_P)$, where P is less than the total number of degrees of freedom. The set of linear point constraint equations is then formulated as

$$\boldsymbol{C}^H \boldsymbol{w} \;=\; \boldsymbol{f},$$

where

$$\boldsymbol{C} \;=\; \left[\, \boldsymbol{d}^T(\theta_1,\omega_1)\;\; \boldsymbol{d}^T(\theta_2,\omega_2)\;\; \cdots\;\; \boldsymbol{d}^T(\theta_P,\omega_P) \right], \tag{2.33}$$

$$\boldsymbol{f} \;=\; [f_1\; f_2\; \cdots\; f_P]. \tag{2.34}$$

The f_i describe the response of the beamformer at the constraint points (θ_i,ω_i). Each linear constraint utilises one degree of freedom in the processor. A linear array of N elements with L tap FIR filters will have at most NL degrees of freedom, so that at most NL point constraints may be employed. If the number of desired signals is small, and their locations are known, then these constraints will represent an efficient use of beamformer degrees of freedom. In situations where direction of arrival, or frequency band information is inaccurate then undesirable signal cancellation may occur. A problem frequently encountered with point constraints is the inability to control response over large regions. Each constraint controls the response at a single point, so that control over regions of beamformer response can only be achieved by adding many closely spaced point constraints. For a general array configuration, the issue of which and how many point constraints should be selected to efficiently implement required control over a response region remains open.

2.7.2 Derivative constraints

In many cases of interest, for example in communication systems, the direction of arrival of the desired signal is known only within some angular tolerance. Any signal that is not exactly matched to one of the beam steer directions will be treated as an unwanted interference signal by the beamformer and will therefore tend to be suppressed. To overcome this problem it

is desirable to broaden the width of each adapted beam, while preserving the ability of the beamformer to reject unwanted signals from outwith the main beam. *Derivative* constraints [13, 15,18,21,60] are used to impose zero, first and second order derivative constraints with respect to bearing or frequency upon the beamformer response.

Consider an array of N elements each with L taps. It is desired to constrain the first order derivatives of the power response with respect to both θ and ω to be zero. The beamformer output is formed by applying a complex set of weights, subject to these constraints, to the data present in the array. The response of the beamformer to a signal of frequency ω, coming from direction θ is expressed as before as

$$b(\theta,\omega) \;=\; w^H V(\theta,\omega). \tag{2.35}$$

The complex vector $V(\theta,\omega)$ is a steering vector; its components specify the individual response of each element and the relative phase shifts among them. By definition, the power response F is given by

$$F \;=\; w^H V V^H w, \tag{2.36}$$

where $V(\theta,\omega)$ has been written V for convenience. The power response F and its derivatives are invariant to the spatial reference point since F is phase independent. Closed form expressions for the derivative constraints with respect to θ and ω have been derived in [13]. In general, these derivative constraints are nonlinear and are therefore difficult to implement. Linear constraints were used in their place, which resulted in an unnecessary additional constraint upon the phase response of the beamformer. For a periodic array, these derivative constraints with respect to phase are dependent upon the position of the array phase reference [15]. This phenomenon is undesirable, since the choice of spatial reference point should be nothing more than a notational convenience. In [60] an approach using double dimension variables to linearise the constraints was presented, which removed the unnecessary phase constraints. Taking the first derivative with respect to θ, one finds

$$\begin{aligned}\frac{\delta F}{\delta\theta} &= w^H \dot{V}_\theta V^H w \;+\; w^H V \dot{V}_\theta^H w \\ &= 2\,\mathrm{Re}\left\{w^H \dot{V}_\theta V^H w\right\},\end{aligned} \tag{2.37}$$

where $\mathrm{Re}\{.\}$ denotes the real part of the bracketed quantity, and $\dot{V}_\theta$ indicates the first derivative of V with respect to θ. Together with a unity gain constraint in the look direction, $w^H V = 1$, this yields

$$\frac{\delta F}{\delta\theta} \;=\; 2\,\mathrm{Re}\left\{w^H \dot{V}_\theta\right\}. \tag{2.38}$$

The first derivative of F with respect to ω has a similar form to this. We may therefore express the phase independent first derivative constraints (with an assumed gain constraint) as

$$\mathrm{Re}\left\{w^H \dot{V}_\theta\right\} = 0, \qquad \mathrm{Re}\left\{w^H \dot{V}_\omega\right\} = 0. \tag{2.39}$$

Notice, however, that these constraints are nonlinear. In order to form a set of linear constraints, (2.39) is rewritten in terms of a *real version* of w with double dimension defined as

$$\tilde{w} = \begin{bmatrix} \mathrm{Re}\{w\} \\ \mathrm{Im}\{w\} \end{bmatrix}. \tag{2.40}$$

With this definition $\mathrm{Re}\left\{w^H \dot{V}_\theta\right\}$ is written as

$$\mathrm{Re}\left\{w^H \dot{V}_\theta\right\} = [\mathrm{Re}\{w\}\ \mathrm{Im}\{w\}] \begin{bmatrix} \mathrm{Re}\left\{\dot{V}_\theta\right\} \\ \mathrm{Im}\left\{\dot{V}_\theta\right\} \end{bmatrix}. \tag{2.41}$$

By transforming the remainder of the variables, the problem readily becomes that of linearly constrained minimum variance beamforming in the real domain. The generalised sidelobe canceller structure can therefore be applied to implement a first–order case in an adaptive fashion. The modified constraint opinions are

$$\min_{\tilde{w}} \tilde{w}^H \tilde{R}_x \tilde{w} \quad \text{subject to} \quad \tilde{C}^H \tilde{w} = \tilde{f}, \tag{2.42}$$

where

$$\tilde{C} = \left[\tilde{V}\ \ \hat{V}\ \ \dot{\tilde{V}}_\theta\ \ \dot{\tilde{V}}_\omega\right], \tag{2.43}$$

$$\tilde{f} = [1\ 0\ 0\ 0]^T, \tag{2.44}$$

in which

$$\tilde{V} = \begin{bmatrix} \mathrm{Re}\{V\} \\ \mathrm{Im}\{V\} \end{bmatrix}, \qquad \hat{V} = \begin{bmatrix} -\mathrm{Im}\{V\} \\ \mathrm{Re}\{V\} \end{bmatrix}. \tag{2.45}$$

Higher order derivative constraints can readily be applied to the beamformer, although care must be taken to ensure that $\tilde{C}$ remains full rank. In periodic array structures the higher order derivatives often become linearly dependent. For example (as noted in [13]), implementing steer direction gain plus both first and second order derivative constraints to an arbitrary array geometry reduces the available degrees of freedom in the beamformer by 9. For the case of an equally spaced linear array the number of linearly independent constraint vectors is equal to 3. Redundant constraint vectors should be deleted from $\tilde{C}$ to prevent singularity occurring.

2.7.3 Eigenvector constraints

The linear constraints described in this section use beamformer degrees of freedom most efficiently in a 2nd–order statistical sense. These constraints are based on a low rank orthonormal representation of a composite broadband design region, and control response for spatial/spectral regions directly instead of at multiple points. In contrast to point and derivative constraints, these constraints are referred to as *eigenvector* constraints. When employing response sampling directly to control response over a region of source location and frequency, the response sample

points, and their number P, must be selected. For P too large, beamformer degrees of freedom will be ineffectually utilised and the constraint matrix will be ill conditioned. For P too small, desired response control will not be realised. The number of eigenvectors required to effectively control the constraint region can be found by determining the constraint region's effective time–bandwidth product within the array/beamformer structure. This time–bandwidth product has been discussed by many investigators, e.g. Buckley [16], Gabriel [11], and Van Veen [19,20]. It is also discussed in chapter 4. Consider again constraining the beamformer response at P points (θ_i, ω_i). In this case P is much greater than the available number of degrees of freedom, i.e. $P \gg NL$. We now have the overdetermined least squares problem

$$\min_{\boldsymbol{w}} \left| \boldsymbol{A}^H \boldsymbol{w} - \boldsymbol{r}_d \right|^2 \tag{2.46}$$

where

$$\begin{aligned} \boldsymbol{A} &= [\boldsymbol{d}(\theta_1,\omega_1)\ \boldsymbol{d}(\theta_2,\omega_2) \cdots \boldsymbol{d}(\theta_P,\omega_P)], \\ \boldsymbol{r}_d &= [r_d(\theta_1,\omega_1)\ r_d(\theta_2,\omega_2) \cdots r_d(\theta_P,\omega_P)], \end{aligned}$$

and

$$\boldsymbol{d}(\theta_i,\omega_i) = \left[1\ e^{j\omega_i\tau_2(\theta_i)}\ e^{j\omega_i\tau_3(\theta_i)}\ \ldots\ e^{j\omega_i\tau_{NL}(\theta_i)}\right].$$

P is chosen to be significantly greater than the time–bandwidth product for the broadband signal. The similarity between this formulation and the deterministic beamformer design example presented earlier in the chapter is obvious. In a deterministic design a singular value decomposition (SVD) has been suggested as a practical method of computing the beamformer weights. When designing eigenvector constraints a SVD is used explicitly as part of the design method. Consider constraining the beamformer response to a broadband signal incident from θ_0 over the band $[\omega_a, \omega_b]$. The $\boldsymbol{d}(\theta_0, \omega_i)$ oversample $[\omega_a, \omega_b]$ and $\boldsymbol{A}$ is therefore ill–conditioned. A rank D approximation of $\boldsymbol{A}$ can be obtained from it's singular value decomposition as

$$\boldsymbol{A}_D = \boldsymbol{V}\boldsymbol{\Sigma}_D\boldsymbol{U}^H, \tag{2.47}$$

where $\boldsymbol{\Sigma}_D$ is a $D \times D$ diagonal matrix containing the largest singular values of $\boldsymbol{A}$, and the D columns of $\boldsymbol{V}$ and $\boldsymbol{U}$ are respectively the left and right singular vectors corresponding to these singular values. Replacing $\boldsymbol{A}$ in (2.46) by its rank D approximate and bringing $\boldsymbol{U}\boldsymbol{\Sigma}_D$ to the right hand side (the pseudo inverse of $\boldsymbol{U}$ is $\boldsymbol{U}^H$), yields

$$\boldsymbol{V}^H\boldsymbol{w} = \boldsymbol{\Sigma}_D^{-1}\boldsymbol{U}^H\boldsymbol{r}_d. \tag{2.48}$$

Equation (2.48) has the same form as the constraint equation $\boldsymbol{C}^H\boldsymbol{w} = \boldsymbol{f}$, so by comparing terms we see that

$$\boldsymbol{C} = \boldsymbol{V}, \qquad \boldsymbol{f} = \boldsymbol{\Sigma}_D^{-1}\boldsymbol{U}^H\boldsymbol{r}_d.$$

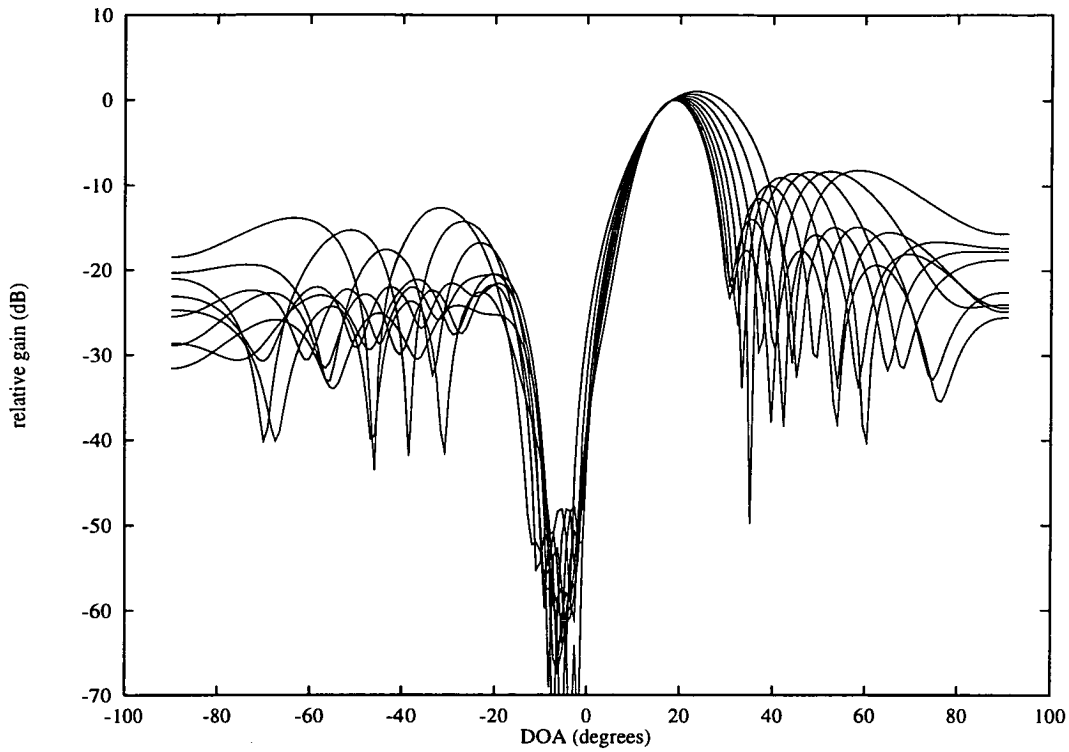

Figure 2.8: *Response of a LCMV beamformer at eight frequencies in the design band. The beamformer is designed to have unity gain at 18° over the normalised frequency band [0.4, 0.8]. A broadband interferer is incident over the angular region [-7.5°, -2.5°], and normalised frequency [0.4, 0.8]. The array is the same as that in Figure 2.2, and the frequencies plotted are identical.*

The columns of $\boldsymbol{V}$ correspond to the eigenvectors of the matrix $\boldsymbol{A}\boldsymbol{A}^H$, hence the name eigenvector constraints. It is these eigenvectors that form an orthogonal basis for the constraint subspace, the dimension of which is determined by the observation time–bandwidth product. Figure 2.8 depicts the response of a LCMV beamformer, designed with eigenvector constraints, when a broadband interferer is incident in the presence of white noise. The interferer has a flat spectrum on the normalised frequency band [0.4, 0.8], and extends over the spatial region $[-7.5°, -2.5°]$. Again the array has 16 elements, each with 5 taps, and is designed to present unity gain and linear phase to signals incident from 18° over the band [0.4, 0.8]. 7 eigenvector constraints are used. The effectiveness of the constraints is evident, since all the frequency curves pass through 0dB at 18 degrees. The response places nulls in the direction of the interference over the entire band of interference.

2.7.4 Matching a desired quiescent response

Under conditions when only white noise is incident at the beamformer the response pattern is termed the quiescent response. Often the quiescent response of the beamformer will have undesirably high sidelobes or large gain away from the mainlobe, particularly when a small number of snapshots are used to compute the weights [2, 17]. However, even with a large

number of snapshots, the response may exhibit large sidelobes because of the tendency of simple adaptive algorithms to revert to sinc type pattern responses. Additionally it may be desirable to modify the quiescent response to force nulls in the beamformer response in the expected direction of interference or jamming. Adjusting the linear constraints employed in order to match a desired response is the aim in this section.

Several investigators [16, 17, 33] have considered modifying the quiescent response of LCMV beamformers. The problem can be summarised as follows. A linearly constrained minimum variance beamformer defines a set of weight vectors which satisfy the constraints $\boldsymbol{C}^H \boldsymbol{w} = \boldsymbol{f}$. The resulting quiescent response of the beamformer will be given by $\boldsymbol{w}_q = \boldsymbol{C}(\boldsymbol{C}^H \boldsymbol{C})^{-1}\boldsymbol{f}$, which satisfies the constraints but may not be acceptable for the reasons explained above. Suppose the desired beamformer response is given by the weight vector $\boldsymbol{w}_0$. This alternative weight vector may not satisfy the existing constraints. It is therefore necessary to modify this new quiescent weight vector in some way, to form $\bar{\boldsymbol{w}}_0$ which will satisfy the original constraint set. Once $\bar{\boldsymbol{w}}_0$ has been computed the existing constraint set needs to be modified so that the quiescent response of the new beamformer is exactly equal to this modified weight vector. Matching a desired response is therefore a two step process.

Assume that the desired response weight vector $\boldsymbol{w}_0$ which does not satisfy the linear constraints, i.e. $\boldsymbol{C}^H \boldsymbol{w}_0 \neq \boldsymbol{f}$, is known. This quiescent response is modified by changing the portion of $\boldsymbol{w}_0$ which projects onto the subspace spanned by the columns of $\boldsymbol{C}$ and leaving the portion which projects onto the null space of $\boldsymbol{C}$ as follows

$$\min_{\bar{\boldsymbol{w}}_0} \left\{ (\bar{\boldsymbol{w}}_0 - \boldsymbol{w}_0)^H (\bar{\boldsymbol{w}}_0 - \boldsymbol{w}_0) \right\} \quad \text{subject to} \quad \boldsymbol{C}^H \bar{\boldsymbol{w}}_0 = \boldsymbol{f}. \tag{2.49}$$

The resulting vector $\bar{\boldsymbol{w}}_0$ meets the constraints and will be a valid quiescent vector. Equation (2.49) results in a new vector $\bar{\boldsymbol{w}}_0$ which matches the desired response in a mean square error sense. In fact (2.49) defines a mean square error projection of the desired response upon the constraint hyperplane. The solution of (2.49) is given by [16]

$$\begin{aligned} \bar{\boldsymbol{w}}_0 &= \left(\boldsymbol{I} - \boldsymbol{C} \left(\boldsymbol{C}^H \boldsymbol{C} \right)^{-1} \boldsymbol{C}^H \right) \boldsymbol{w}_0 + \boldsymbol{w}_q \\ &= \left(\boldsymbol{C}_n \left(\boldsymbol{C}_n^H \boldsymbol{C}_n \right)^{-1} \boldsymbol{C}_n^H \right) \boldsymbol{w}_0 + \boldsymbol{w}_q, \end{aligned} \tag{2.50}$$

where $\boldsymbol{w}_q = \boldsymbol{C} \left(\boldsymbol{C}^H \boldsymbol{C} \right)^{-1} \boldsymbol{f}$ is the existing quiescent response. The first form for $\bar{\boldsymbol{w}}_0$ is useful in the direct implementation of the LCMV beamformer, whilst the second is useful in a generalised sidelobe canceller implementation. The second stage in the process is to modify the original constraint set so that the quiescent response of the beamformer is identical to this new response. The new response vector $\bar{\boldsymbol{w}}_0$ is decomposed into two orthogonal components $\bar{\boldsymbol{w}}_c$ and $\bar{\boldsymbol{w}}_s$ (in a manner analogous to the GSC development), one which lies in the $\boldsymbol{C}$ subspace and one which lies in the $\boldsymbol{C}_n$ subspace. These are given by

$$\left.\begin{aligned} \bar{\boldsymbol{w}}_c &= \boldsymbol{C}\left(\boldsymbol{C}^H\boldsymbol{C}\right)^{-1}\boldsymbol{C}^H\bar{\boldsymbol{w}}_0, \\ \bar{\boldsymbol{w}}_s &= \boldsymbol{C}_n\left(\boldsymbol{C}_n^H\boldsymbol{C}_n\right)^{-1}\boldsymbol{C}_n^H\bar{\boldsymbol{w}}_0. \end{aligned}\right\} \tag{2.51}$$

If we had just substituted $\bar{\boldsymbol{w}}_0$ for $\boldsymbol{w}_q$ in the upper path of the GSC, then no change would have occurred in the quiescent response because the lower path would cancelled the desired change. We require to append the vector $\bar{\boldsymbol{w}}_s$ to the constraint set and to remove it from the column space of $\boldsymbol{C}_n$, thus preventing the lower path negating the effect upon the quiescent response. This will consequently reduce the dimension of $\boldsymbol{C}_n$ by one. Once this has been done the lower path can no longer cancel the $\bar{\boldsymbol{w}}_s$ component. The modified constraint opinions are

$$\left.\begin{aligned} \bar{\boldsymbol{C}} &= [\boldsymbol{C}, \bar{\boldsymbol{w}}_s], \\ \bar{\boldsymbol{f}}^H &= [\boldsymbol{f}^H\ ,\ \bar{\boldsymbol{w}}_s^H\bar{\boldsymbol{w}}_0]\ . \end{aligned}\right\} \tag{2.52}$$

Once these have been formed a new signal blocking matrix $\bar{\boldsymbol{C}}_n$ must be computed. This can be done by either removing the $\bar{\boldsymbol{w}}_s$ component from any $(K-1)$ columns of $\boldsymbol{C}_n$ or alternatively recomputing a new signal blocking using $\bar{\boldsymbol{C}}$. To confirm that these modified constraints do indeed give the required quiescent response consider

$$\begin{aligned} \bar{\boldsymbol{w}}_q &= \bar{\boldsymbol{C}}\left(\bar{\boldsymbol{C}}^H\bar{\boldsymbol{C}}\right)^{-1}\bar{\boldsymbol{f}} \\ &= [\boldsymbol{C}\ \ \bar{\boldsymbol{w}}_s]\left(\begin{bmatrix} \boldsymbol{C}^H \\ \bar{\boldsymbol{w}}_s^H \end{bmatrix}[\boldsymbol{C}\ \ \bar{\boldsymbol{w}}_s]\right)^{-1}\begin{bmatrix} \boldsymbol{C}^H \\ \bar{\boldsymbol{w}}_s^H \end{bmatrix}[\boldsymbol{f}\ \ \bar{\boldsymbol{w}}_s^H\bar{\boldsymbol{w}}_0] \\ &= \boldsymbol{C}\left(\boldsymbol{C}^H\boldsymbol{C}\right)^{-1}\boldsymbol{f} + \bar{\boldsymbol{w}}_s \\ &= \boldsymbol{w}_q + \bar{\boldsymbol{w}}_s. \end{aligned} \tag{2.53}$$

The vector $\bar{\boldsymbol{w}}_s$ is precisely the additional term needed to modify the response. Thus, the two–step procedure of modifying $\boldsymbol{w}_0$ to $\bar{\boldsymbol{w}}_0$ using (2.50), then augmenting $\boldsymbol{C}$ with the $\bar{\boldsymbol{w}}_s$ vector provides a new quiescent solution which matches the original constraint set. Figure 2.9 shows the influence of an additional quiescent pattern constraint upon the response of an adaptive beamformer. A directional constraint only results in the familiar sinc type power response, whereas the addition of a single quiescent pattern constraint, in this case a -30dB Chebychev weighting, clearly improves the response.

2.8 Conclusion

This chapter has reviewed the basic concepts of adaptive beamforming as they relate to the linearly constrained class of adaptive beamformer. Some of the terminology which will be used subsequently in the thesis has been defined, and expressions for the important second–order statistics of the signals concerned have been given.

The linearly constrained minimum variance beamformer has been defined formally, and the principal features have been outlined. The generalised sidelobe canceller has been introduced, and

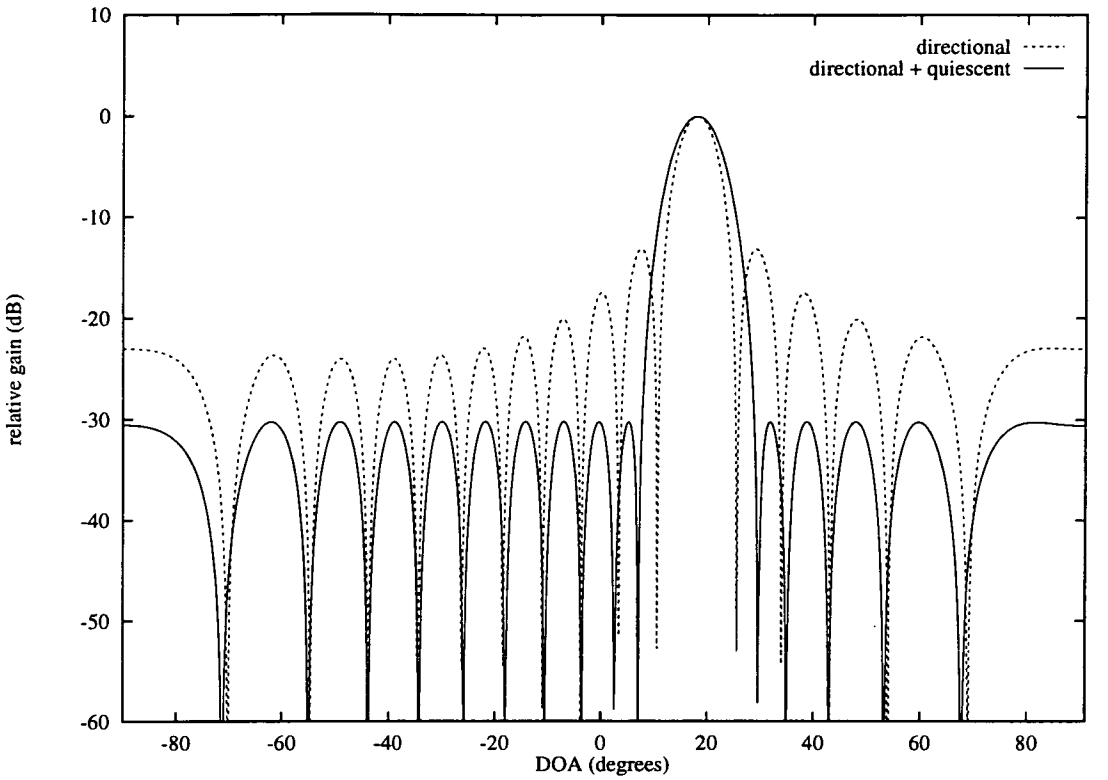

Figure 2.9: *Influence of quiescent pattern constraint upon adapted response of a linearly constrained beamformer.*

its application as a solution to linearly constrained beamforming has been described. Furthermore, various techniques for controlling aspects of beamformer response have been described.

Chapter 3

Airborne Radar

3.1 Introduction

This thesis is primarily concerned with developing efficient techniques for partially adaptive beamformers used in airborne radar. In particular the generalised sidelobe canceller structure is considered. To be able to clearly understand this work, and also to understand its position relative to other work, a knowledge of several disparate topics is required. This chapter will consider the airborne radar environment, and specifically the type of interference which an airborne look–down pulse–Doppler radar will experience. The chapter begins with an overview of the airborne radar scenario, and then considers the performance of traditional cancellation techniques.

Airborne surveillance radars will become increasing important in the near future because of their capability for detecting objects on or close to the ground. Such objects can hardly be detected by conventional ground–based surveillance radars because they are masked by the terrain, vegetation and other fixed objects. However, the target returns in such a radar have to compete with strong clutter returns which are broadband in nature. This has severe implications for conventional moving target indication (MTI) schemes, because these rely primarily on targets being either spatially or temporally separable. When the radar platform is in motion this may no longer be the case.

This chapter will discuss the problems inherent with airborne pulse–Doppler radar and describe several solutions, most notably a *Klemm* type radar. Typical computed eigenspectra for the airborne scenario will be presented and the implications thereof discussed.

3.2 Airborne pulse–Doppler radar

Airborne radar provides an important capability for reconnaissance and verification purposes. In conjunction with moving target indication it is well suited for detection of moving objects close to the ground (for example low flying aircraft or vehicles), because the degradation arising from terrain masking is avoided. The next generation of airborne radar will combine these advantages with an adaptive phased array. However, in realistic situations the strong clutter spectrum is Doppler broadened and time varying so that conventional MTI techniques perform poorly. The main problem is that moving target returns are submerged in the Doppler

broadened sidelobe clutter so that purely temporal filters cannot effectively suppress the sidelobe clutter which has the same Doppler as the moving target. However, because the sidelobe clutter comes from a different direction than that of the target, the moving target echo can be distinguished from the sidelobe clutter in the spatial domain and from mainlobe clutter in the time (frequency) domain. Optimum clutter suppression can therefore be performed by a filter which operates in both time and space. This space–time filter will require space–time sampling of the received field. This can be realised by a coherent pulse–Doppler radar with an array antenna. The sensor elements provide the spatial samples whilst the coherent pulse train realises the temporal sampling of the echo field.

Such coherent digital signal processing greatly alleviates the effects of clutter. A disadvantage of space–time processors is the relatively large amount of pulses (up to 10 or more) that must be transmitted at a stable frequency and pulse repetition frequency (PRF). A responsive jammer could measure the frequency of the first transmitted pulse and then centre the jammer to spot jam the following pulses. Also, the requirement for a stable PRF precludes the use of pulse–to–pulse jitter, which is one of the most effective techniques against deception and camouflage jammers, which rely on anticipating the radar transmitter's pulse.

Despite these limitations, it is assumed that the transmitted pulses are separated by a fixed period termed the *pulse repetition interval*. Because of the periodic spacing of the pulses there will be a greatest distant over which a pulse can travel and return as an echo without arriving after the next pulse is transmitted. This range is called the unambiguous range of the radar and is given by the expression

$$R_{unamb} = \frac{c}{2 f_r}, \tag{3.1}$$

where f_r is the pulse repetition frequency in Hz and c is the speed of light in ms^{-1}. Similarly there will exist an unambiguous Doppler frequency, again related to the PRF as follows

$$f_{max} = \frac{f_r}{2} = \frac{2 v_r f_c}{c}, \tag{3.2}$$

in which v_r is the relative velocity of the radar and target in ms^{-1}, and f_c is the carrier frequency. Studying these two expressions it can be seen that unambiguous range and unambiguous Doppler are competing ideals, i.e. an increase in unambiguous range will result in a corresponding decrease in unambiguous Doppler and an increase in unambiguous Doppler will cause a decrease in unambiguous range. As stated earlier, the radar considered in this thesis is operating in a medium PRF mode, that is both Doppler and range ambiguities will occur. However, it is assumed that the PRF is sufficiently high that the received clutter field is sampled unambiguously in Doppler. Furthermore, all processing is assumed to relate to a single range gate - this is a reasonable assumption with the PRF's considered.

The previous discussion has assumed some idealised receiver of which the output is zero when no echo pulse is received and one (or some other scale thereof) when a pulse is incident. This

would only be the case if the pulse had infinite length. When the pulse has finite length the output of the receiver will not be as described, but will have some sidelobe response - typically that of a weighted sinc ($\sin x/x$) function. Cook and Bernfeld [62] produced a valuable text on this subject. For the remainder of this thesis it is assumed that the transmitted pulse is a monotone of fixed duration, and that any problems arising from this finite duration may be ignored. Among the many good papers and books on airborne radar, the following are recommended: books [26,63,64], clutter modelling [65–67], clutter suppression [1,3,5–9,68,69].

3.3 Clutter model

The problem geometry is shown in Figure 3.1 The radar platform is taken to be at an altitude h above a planar earth. Points upon the ground may be measured relative to the radar by the range R, depression angle ϕ, and azimuth angle θ. Through simple geometric reasoning, the relationship between R and ϕ is found to be

$$\sin\phi \;=\; \frac{h}{R}\,. \tag{3.3}$$

Note that this expression is independent of the azimuth angle θ, so that a particular depression angle ϕ_0 will define a ring of range R_0 upon the planar earth, typically called a range gate. Range gating eliminates excess receiver noise from competing with the returned echoes and permits target tracking and range measurement. A radar signal returned from a point scatterer on the range ring has a two way Doppler shift of the amount

$$\omega \;=\; 2\pi\left(\frac{2\,v_r}{\lambda}\right)\cos\theta\cos\phi_0, \tag{3.4}$$

where λ is the carrier wavelength. This expression defines a relationship between the Doppler shift which will be observed at different points on a range ring due to the motion of the radar. The locus of all ground returns having the same Doppler shift, commonly called an isodop, is defined by

$$\cos\theta\cos\phi_0 \;=\; \gamma, \tag{3.5}$$

in which γ is a constant between 0 and 1 which specifies the particular isodop. With the above coordinate geometry, the clutter return can be expressed in terms of the problem geometry. The most important description of the radar process is the radar equation [64] which gives the range of a radar in terms of the radar characteristics. One form of the equation gives the received signal power P_r as

$$P_r \;=\; \frac{P_t\,G_t}{4\pi R^2}\times\frac{\sigma}{4\pi R^2}\times A_r. \tag{3.6}$$

The three factors in the right hand side convey the physical processes taking place. The first factor is the power density at a distance R metres from a radar that radiates P_t watts with an antenna of gain G_t. In the second factor σ is the target cross section in square metres, and the denominator represents the divergence on the return path with range and is exactly the same as

on the outward path. The product of the first two terms is the power per square metre returned to the radar, so that multiplying by the receiver aperture A_r gives the total power which the receiver collects. The power returned from a source illuminated by the radar therefore obeys an inverse R^4 law, and the maximum detectable range R_{max} is measured in terms of the minimum recoverable signal power S_{min} as

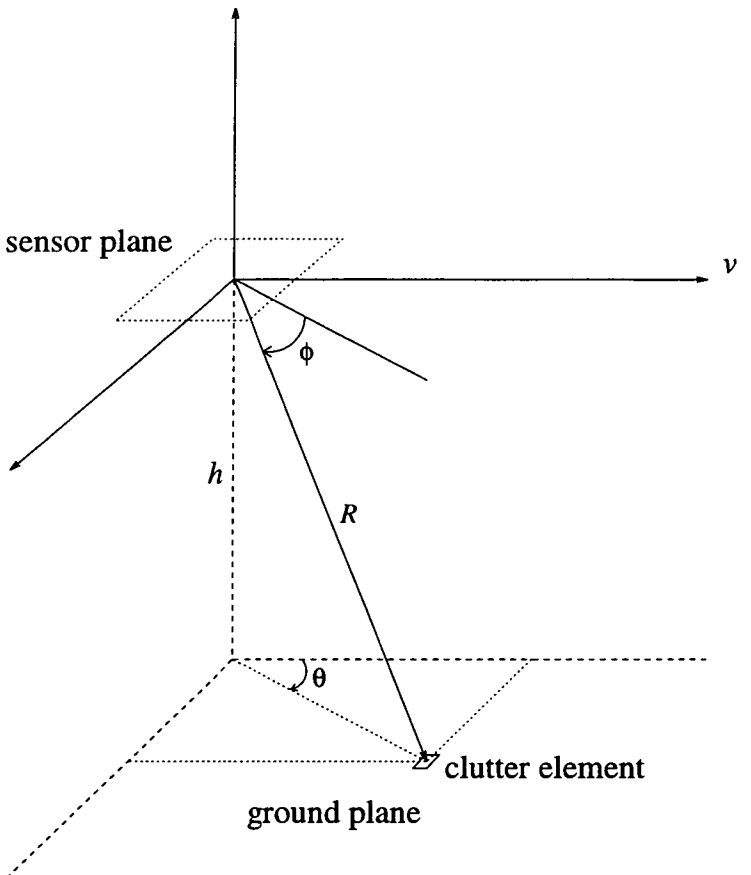

Figure 3.1: Airborne array geometry.

$$R_{max}{}^2 = \frac{P_t G_t A_r \sigma}{(4\pi)^2 S_{min}}. \tag{3.7}$$

Note that the important parameters are transmitting gain and receiver area. Antenna theory gives the relationship between antenna gain and effective area as

$$G_t = \frac{4\pi A_t}{\lambda^2}, \tag{3.8}$$

in which λ is the carrier wavelength. Assuming a common antenna is used for both transmission and reception, then $A_t = A_r = A_e$, and (3.6) can be rewritten as

$$P_r = \frac{P_t A_e^2 \sigma}{4\pi \lambda^2 R^4}. \tag{3.9}$$

This expression gives the power received by the radar for a target at range R for given antenna and target characteristics. These expressions have assumed that the transmit gain of the transmit antenna is independent of the relative location of the target, i.e. the antenna radiates and

receives power in an omni-directional manner. The transmitter will also experience a variation in gain with both bearing and frequency, so that G_t should be written $G_t(\theta,\omega)$, and the power returned becomes

$$\rho^2(\theta,\omega) = \frac{P_t\, G_t(\theta,\omega)\, A_r\, \sigma}{(4\pi)^2\, R^4}. \tag{3.10}$$

(3.10) can be interpreted as the power spatial/spectral density for a particular target σ which is returned to the receiver. The power distribution indicated by $\rho^2(\theta,\omega)$ forms the basis of many target recognition techniques. The points (θ,ω) which satisfy (3.4), define a region Ω over which the ground clutter returns within a range gate exist. Within this region the power spatial/spectral density can be computed from (3.10). Using the terminology defined in the previous chapter, the covariance matrix for the ground clutter returns in a particular range ring is given by

$$\boldsymbol{R}_x = \int_{\Omega} \rho^2(\theta,\omega)\, \boldsymbol{d}(\theta,\omega)\, \boldsymbol{d}^H(\theta,\omega)\, d\Omega, \tag{3.11}$$

where $\boldsymbol{d}(\theta,\omega)$ is the array steering vector. In practice, the integral of (3.11) will be approximated as a Riemann sum

$$\boldsymbol{R}_x \approx \sum_{p=1}^{P} \rho^2(\theta_p,\omega_p)\, \boldsymbol{d}(\theta_p,\omega_p)\, \boldsymbol{d}^H(\theta_p,\omega_p)\, \Delta\Omega, \tag{3.12}$$

where the (θ_p,ω_p) uniformly sample Ω. (3.12) may be conveniently rewritten in matrix form yielding

$$\boldsymbol{R}_x \approx \boldsymbol{A}\,\boldsymbol{\Gamma}^2\boldsymbol{A}^H, \tag{3.13}$$

where

$$\begin{aligned} \boldsymbol{A} &= [\,\boldsymbol{d}(\theta_1,\omega_1)\;\; \boldsymbol{d}(\theta_2,\omega_2)\;\; \cdots\;\; \boldsymbol{d}(\theta_P,\omega_P)\,], \\ \boldsymbol{\Gamma}^2 &= \mathrm{diag}\,\{\,\rho^2(\theta_1,\omega_1),\;\; \rho^2(\theta_2,\omega_2),\;\; \cdots,\;\; \rho^2(\theta_P,\omega_P)\,\}. \end{aligned}$$

This is the approach used when computing the clutter returns used in the simulations for this thesis. Each of the individual scatterers in the summation of (3.12) are modelled as normally distributed random scatterers with variance equal to the value of $\rho^2(\theta_p,\omega_p)$ for the particular scatterer's location. The variance in the estimate of the covariance matrix is reduced by averaging over a large number of individual covariance matrices computed according to (3.12).

3.4 Clutter spectra

The discussion presented here relates to an horizontal linear array lying along the axis of the aircraft. This need not be the case, and can be easily extended to more complex scenarios such as planar, spherical or conformal antennae. An horizontal linear array is sufficient to demonstrate the structure which exists within the clutter field and will make subsequent formulations less complex. This array orientation represents the sideways looking airborne radar

(SLAR) described by Klemm [1,5–7]. Once the clutter covariance matrix has been formed, a two–dimensional power spectrum can be calculated. For instance,

$$P(\theta,\omega) \;=\; \boldsymbol{d}^H(\theta,\omega)\;\boldsymbol{R}_x\,\boldsymbol{d}(\theta,\omega)\,, \tag{3.14}$$

is the familiar Fourier–type spectral estimator. The vector $\boldsymbol{d}(\theta,\omega)$ is a steering vector for the array. Typically the power spectrum defined by (3.14) has insufficient resolution and sidelobe behaviour to give an accurate spectral representation of the clutter distribution [70]. The minimum variance power estimator

$$P(\theta,\omega) \;=\; \frac{1}{\boldsymbol{d}^H(\theta,\omega)\,\boldsymbol{R}_x^{-1}\,\boldsymbol{d}(\theta,\omega)}\,, \tag{3.15}$$

gives a much more realistic impression of the clutter distribution in the $\theta-\omega$ plane.

3.5 Computed spectra

The following figures show actual computed two–dimensional power spectra using the clutter model outlined above. Table 3.1 summarises the radar parameters used during the simulations. The plots are all for arrays with a total adaptive dimension $NL = 128$, and examine the influence

Simulation Parameters

Platform height	1 km
Platform velocity	100 ms^{-1}
Range	$R = 2$ km
Look direction	$\theta = 0\,^\circ$
Wavelength	$\lambda = 0.1$ m
Pulse repetition frequency	$f_r = 4$ kHz
Sensor spacing	$d = \lambda/2$
Clutter–to–noise ratio (measured at the element)	20 dB
Element pattern : isotropic	
Transmit antenna : same as receiving antenna	
Clutter return : computed from 200 points distributed over range ring	
Number of elements	$N = 16$
Number of taps	$L = 8$
Total number of samples processed	$NL = 128$

Table 3.1: Radar parameters used during simulation.

of array geometry, sampling rate and transmit beamwidth. When computing the clutter returns, only half the range gate has been considered, $0 \leq \theta \leq \pi$. This is sufficient to demonstrate the structure of the clutter returns, and is the assumption made in several papers [1,5–8]. Figure 3.2 shows the case described by the parameters in the table. This is the geometry of the sideways looking airborne radar (SLAR) discussed by Klemm [1]. Clutter returns are spread along the diagonal of the azimuth–Doppler plane. The radar looks ahead ($\theta = 0\,^\circ$) so that the clutter

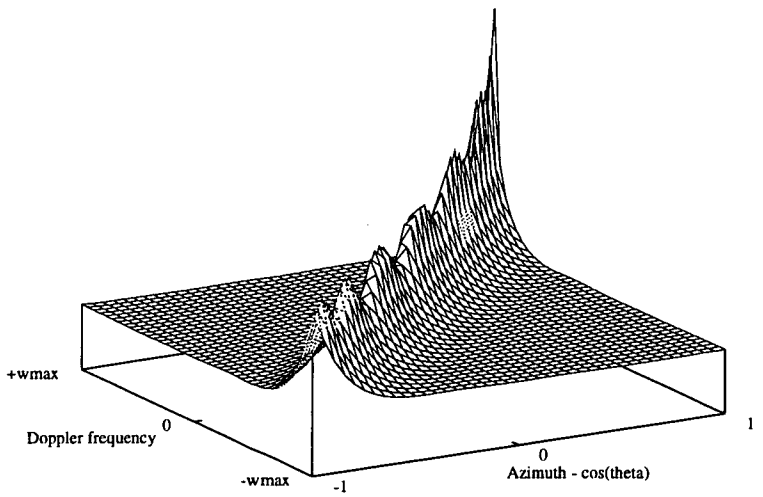

Figure 3.2: Airborne Clutter Spectrum: returns computed for parameters in Table 3.1.

maximum (mainlobe clutter) occurs at maximum Doppler $\omega = +\omega_{max}$. The sidelobe clutter is distributed over the whole of the azimuth range with the associated Doppler frequencies, resulting in this characteristic profile. Figure 3.3 shows the clutter returns for same beamformer as the previous figure, but with a reduced sampling rate. A reduction in the sampling rate (i.e. the PRF) will result in spectral aliasing of the clutter returns. This can be clearly seen in the figure. The Doppler spectrum of Figure 3.2 has been aliased several times in the Doppler plane, demonstrating the need for unambiguous sampling of received field. Figure 3.4 shows the same array as Figure 3.2 but with a different transmit pattern. In this case a low sidelobe equi–ripple response was used. The influence which all of these parameters have on the computed eigenspectra will be discussed later in the chapter.

3.6 Existing cancellation techniques

Desired targets are assumed to be illuminated by the mainlobe of the transmit pattern. They will have similar spatial characteristics as mainlobe clutter, and will therefore have to compete with large clutter returns. The most simple solution to the mainlobe clutter problem is to apply fixed digital filters in each elemental channel prior to spatial adaptive beamforming. These digital filters are configured to present a stop–band over the widest possible frequency extent of mainlobe clutter. This represents moving target indication (MTI) at its most basic. A potential enhancement to this fixed filtering is to employ adaptive filters in each spatial channel to adaptively match the stop band to the spread of mainlobe clutter. However, this suffers from the serious drawback that spatial correlation between jamming signals can be lost, meaning that subsequent adaptive spatial filtering will be ineffective. A substantial improvement can be obtained by performing adaptive processing in both the spatial and Doppler domains simultaneously. These space–time processors will provide the source of much interest in coming

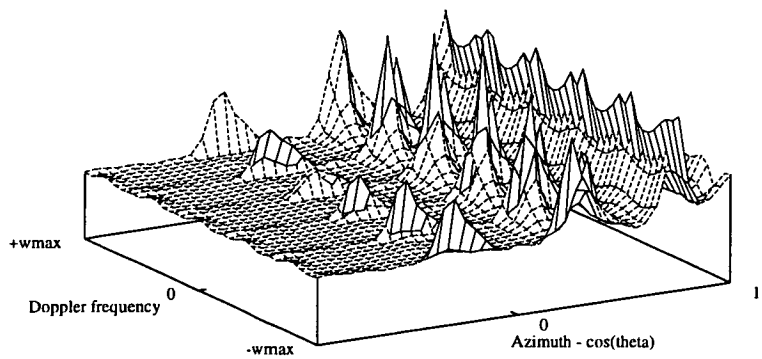

Figure 3.3: *Airborne Clutter Spectrum: sampling rate. The sampling rate has been reduced to 10% of the Nyquist sampling rate for the clutter field.*

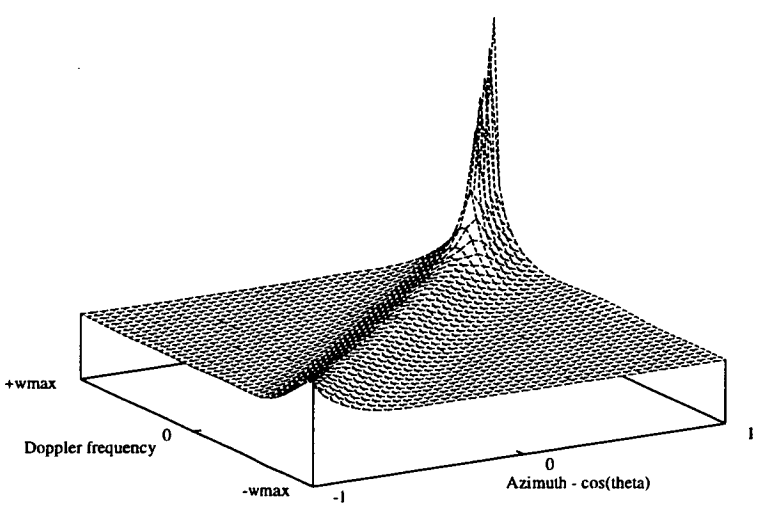

Figure 3.4: *Airborne Clutter Spectrum: transmit aperture. The transmit pattern has been modified to a low sidelobe equi-ripple pattern.*

years.

The basic idea behind conventional MTI is to place a notch in the filter response at the frequency corresponding to mainlobe clutter returns. The effective filtering operation is shown in Figure 3.5. It can be seen that this approach will provide good rejection of mainlobe clutter, but will be ineffective in cancelling sidelobe clutter. Target returns will therefore have to compete with sidelobe clutter returns, significantly reducing the ability to detect weak target signals. In "space–time" adaptive processing an adaptive tapped delay line processor is placed in each

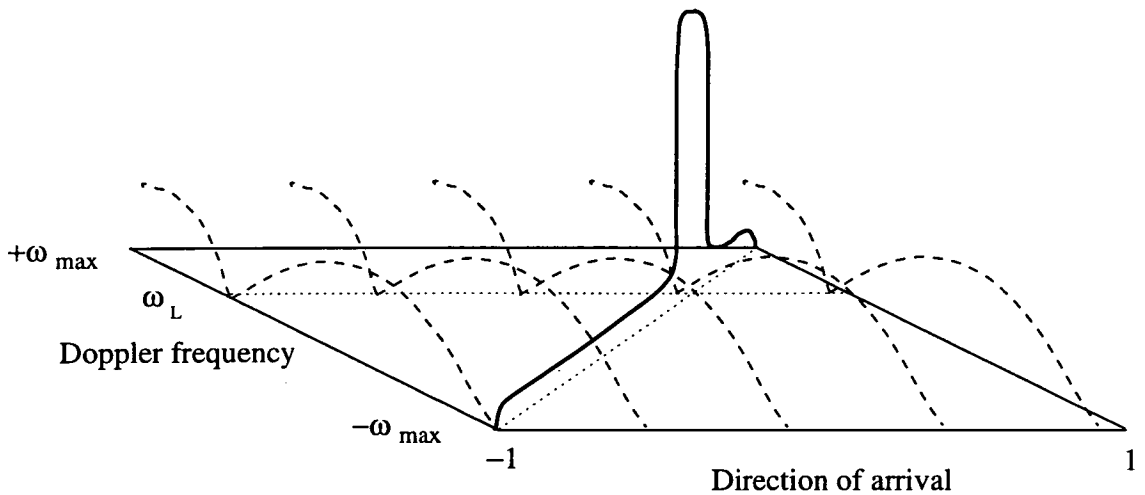

Figure 3.5: MTI filter response superposed on clutter returns.

spatial channel. The adaptive weights are computed for the entire array at once, overcoming the problems described above. In the radar geometries considered by Klemm, sensors are aligned along the direction of travel. As was noted by the author in [6], space–time processing with this type of array geometry results in a generalisation of the displaced phase centred array (DPCA) technique [68]. DPCA is the simplest system which combines spatial and temporal degrees of freedom. Platform motion is compensated for by varying the time delay between pulses, provided accurate estimates of the platform velocity exist. However, good estimates of platform velocity are not generally available so that an adaptive algorithm must be used to estimate the velocity. Since an adaptive algorithm is required for effective operation of a DPCA, many authors have highlighted adaptive space–time processing as a potential improvement. Adaptive space–time processors represent a powerful technique for suppressing both mainlobe and sidelobe clutter. Figure 3.6 demonstrates the filtering process achieved by combining spatial and temporal samples. The tap spacing needs to be matched to the pulse repetition frequency to ensure that each snapshot of data corresponds to the returns from a single range gate. Unlike fixed MTI schemes, no *a priori* knowledge is required of the clutter direction of arrival and Doppler. However, attenuation of the beam response at the known desired signal direction and unknown signal Doppler must be avoided. This can be achieved through use of linear constraints as described in chapter 2.

At this point it is important to examine the operation of the adaptive airborne MTI described

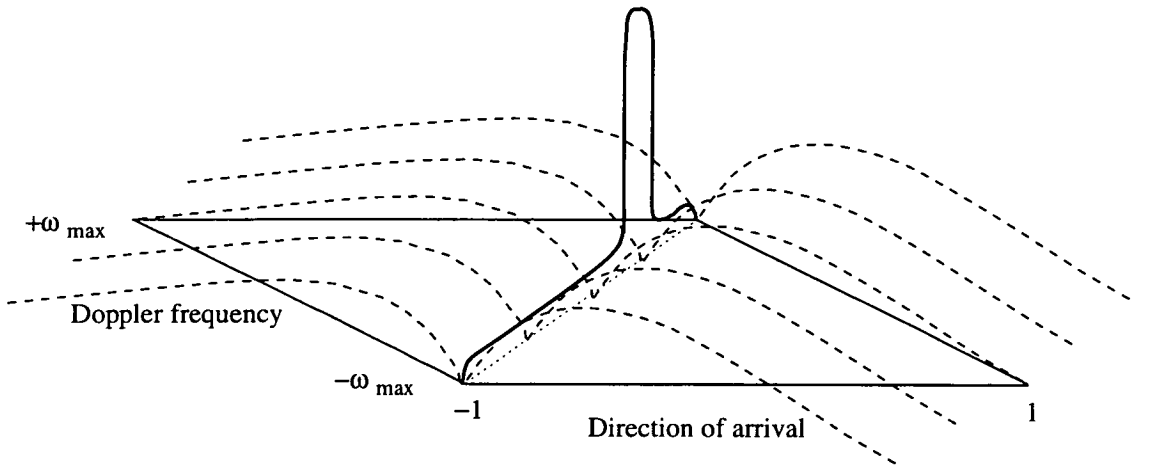

Figure 3.6: Space–time filter superposed on clutter returns.

by Klemm in [1]. This "auxiliary channel approach" is the most similar to that which will be presented later in this thesis, in that it uses a partially adaptive processor to cancel ground clutter. This approach is a generalisation of well known sidelobe canceller techniques, in which a total of $N + L + 1$ beams are formed, one of which is a search beam, the remaining beams being used for clutter cancellation. Each cancellation beam is subjected to Doppler processing such that the $N + L$ beams cover the diagonal of the clutter plane, effectively suppressing both mainlobe and sidelobe clutter. This receiver is shown in block diagram form in Fig. 3.7. On the

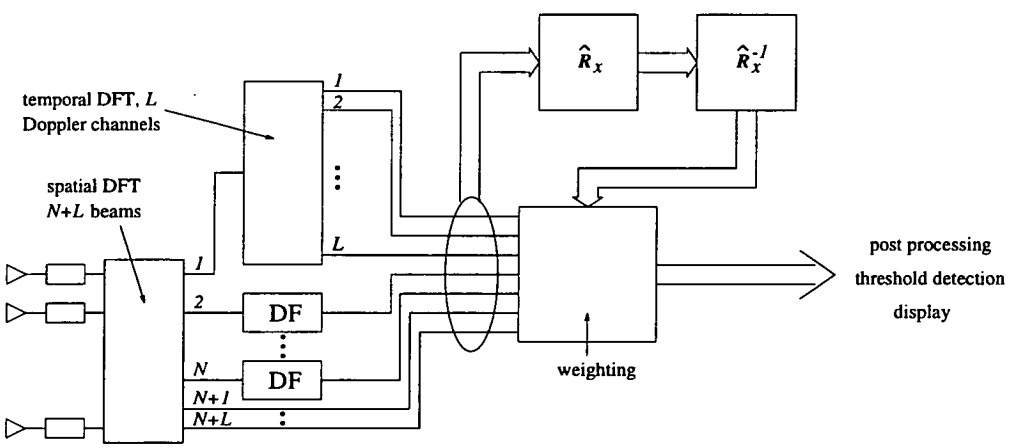

Figure 3.7: Block diagram of auxiliary channel receiver (after Klemm [1]).

left N elements with subsequent receiver channels including amplification, demodulation and I & Q sampling are shown. In the following network $N + L + 1$ beams are formed, one of which is the search beam. The beamformers number $2, \ldots, N$ are followed by Doppler filters (DF) matched to the clutter frequency of the corresponding beam. The search beam is connected to a Doppler filter bank matched to all possible target velocities. A set of weights are computed via an estimated covariance matrix $\hat{\boldsymbol{R}}_x$, after which there is some conventional detection and display circuitry. A total of $N + L$ clutter cancellation beams are formed because this was the dimension estimated for the clutter subspace. The analysis which provided this estimate is discussed in the following sections.

3.7 Relationship between space–time processing and DPCA

Several authors have commented upon the relationship between DPCA techniques and adaptive space–time processing. When considering the sideways looking radar geometry, adaptive space–time processing has been referred to as generalised DPCA [6,69], or adaptive DPCA [8]. However, DPCA and space–time processing differ fundamentally since the former is aimed at achieving complete clutter cancellation, whilst space–time processing attempts to maximise the signal to noise ratio. Therefore the solutions found in space–time processing cannot always be expected to replicate those of DPCA. In this section, the conditions under which adaptive space–time processing and DPCA have equivalent solutions are discussed.

The displaced phase centred array technique operates by matching the interval between transmission of two successive pulses to the distance travelled by two identical sensors (or subarrays) between the two pulses, that is

$$\tau = \frac{d}{2\,v_r}, \tag{3.16}$$

where τ is the pulse repetition interval, and d is the sensor spacing. When (3.16) is satisfied returns from stationary targets (i.e. the ground) can then be completely cancelled by subtracting the returns from two successive pulses. Assuming that all errors can be ignored, and that τ is exactly matched, then the effective filtering operation is as shown in Figure 3.6. Degradation in cancellation performance will occur through a variety of processes, mostly due to non–stationarity of the clutter returns and variability in the platform motion. Adaptive space–time processors have been proposed because of their ability to overcome these problems.

Figure 3.8 depicts the flow of clutter data within a space–time processor in which the tap spacing has been chosen to satisfy the DPCA condition. Subject to this condition, the clutter data lying at each point along the diagonals will be identical. These diagonals are indicated by dashed lines. The platform moves to the left and positions previously occupied by the sensors are indicated by 0, -1, -2, etc. These previous locations are termed "virtual" spatial sampling points, whilst the current N sensor locations are termed "real" spatial sampling points. Studying Figure 3.8, it can be seen that the clutter data present in the beamformer is collected from a set of N "real" spatial sampling points and from $L-1$ "virtual" spatial sampling points. Since the clutter data is derived from at most $N+L-1$ independent spatial locations, the clutter subspace will have dimension no greater than $N+L-1$. In the example shown, $N=5, L=4$, and the clutter subspace will have dimension $5+4-1=8$. This figure compares with that of $N+L=9$ derived empirically in [1].

This diagrammatic approach can also be used to gain insight into cases when the DPCA condition is or is not satisfied. When the pulse repetition interval is matched to the platform motion, then total clutter cancellation can be achieved by ensuring that the weights lying along any of the diagonals sum to zero. This is simply a restatement of the DPCA principal, that subsequent echoes should be combined such that the resultant output clutter signal is zero.

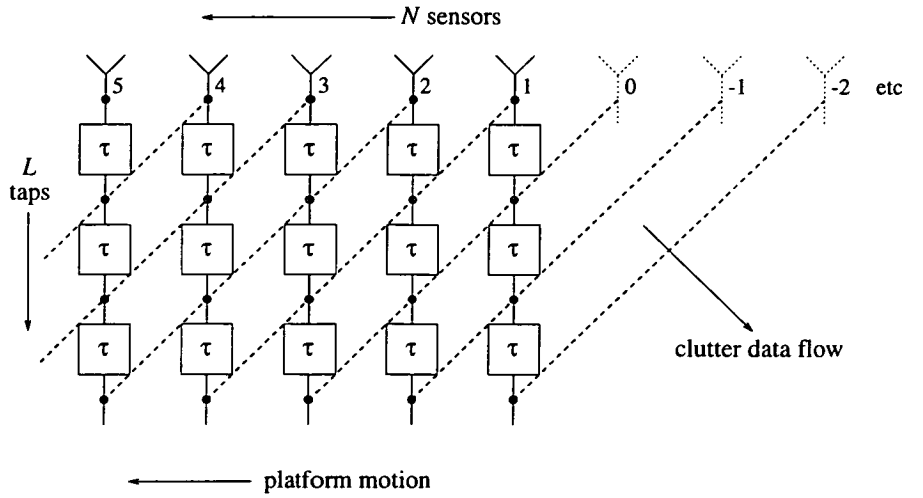

Figure 3.8: Space–time processor data flow diagram – DPCA condition.

However, when the DPCA condition in not satisfied, i.e. when the PRF is higher or lower than that indicated by (3.16), then the clutter data will be spread throughout the entire tap structure, causing the estimate for the clutter dimension to change. Both these situations are considered in simulations in the next section.

3.8 Eigenspectra

Power spectra like those seen in Section 3.5 are impressive to look at, but give little indication as to the complexity or design of a clutter suppression filter. An alternative representation of the clutter covariance matrix is it's eigenspectrum. This can give an indication of the rank of $\boldsymbol{R}_x$, and consequently the dimensionality required in the suppression filter. An eigenspectrum is simply a plot of the eigenvalues arranged in decreasing order of magnitude. Consider the eigendecomposition of $\boldsymbol{R}_x$

$$\boldsymbol{R}_x v_j = \lambda_j v_j, \qquad j = 1, 2, \cdots, NL, \tag{3.17}$$

where $\{\lambda_j \ ; \ \lambda_j \geq \lambda_{j+1}\}$ are the eigenvalues associated with the NL eigenvectors v_j. This is the eigenstructure of the interference covariance matrix $\boldsymbol{R}_x$. The eigenstructure, through $\boldsymbol{R}_x$, is a function of the frequency bands, the location ranges and the weighting function of the interference. The expansion described by

$$\hat{\boldsymbol{R}}_x = \sum_{j=1}^{D} \lambda_j v_j v_j^H, \tag{3.18}$$

where $D \leq NL$, is the discrete Karhunen–Loève expansion (KLE). With this expansion and a selected D, (3.18) is the most efficient rank D representation of $\boldsymbol{R}_x$ in a 2nd–order statistical sense. The eigenvalues λ_j represent the energy of a sample vector projected onto the corresponding basis vector. The representation dimension D is selected to obtain a required approximation error, and $\{v_1, \cdots, v_D\}$ span the broadband interference subspace. Such low

rank representations form the basis of many signal processing techniques. Any vector $\boldsymbol{x}(k)$ within the selected region Ω can be considered to be a sample vector of the "random process" represented by the covariance matrix. These vectors are therefore efficiently represented with the interference representation space described above.

Figure 3.9 shows a simplified eigenspectrum plot. This is often the most informative method for displaying the eigendecomposition. In this plot it can be seen that there are a total of NL eigenvalues. The first point to note is that a large number of the eigenvalues are of the same, small magnitude. Since these are small in magnitude we can make the reasonable assumption that they are related to the background noise within the input data, and can thus be classified as *noise* eigenvalues. The magnitude of the minimum eigenvalue is a function of the additive noise level, and will determine the obtainable gain in the adaptive processor. The remaining K eigenvalues are of a larger magnitude. The number of these "large" eigenvalues gives an approximate estimate of the degrees of freedom inherent within the process. Traditionally these

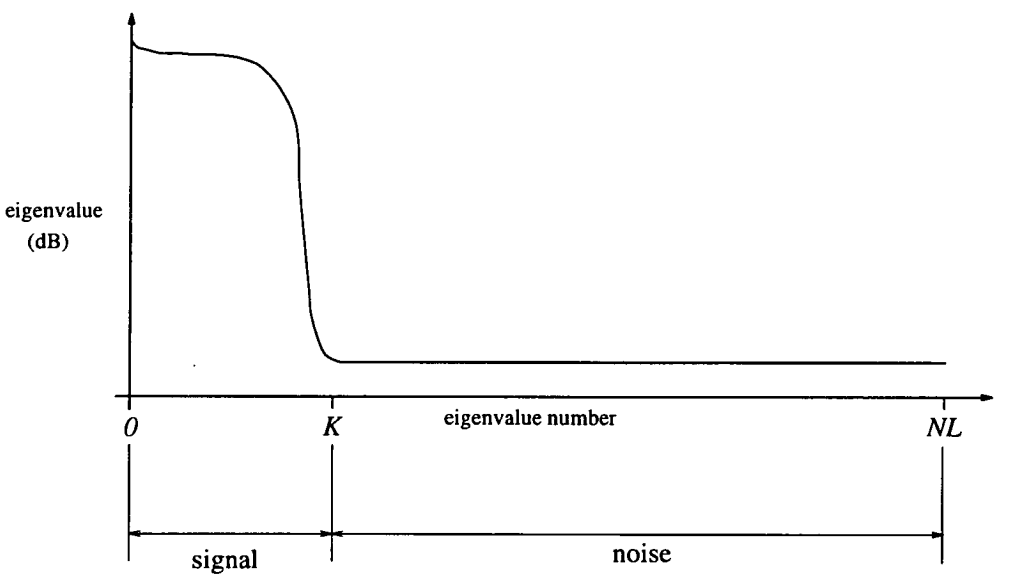

Figure 3.9: Simplified example eigenspectrum.

eigenvalues are termed *signal* eigenvalues, and their distribution will reflect the complexity of the generating process. For example, consider a simple case of a single narrowband interferer in the presence of uncorrelated white noise. In this scenario there will be a single large eigenvalue; the remaining eigenvalues will all have like magnitude, equal to the noise power. As the process becomes more complex, the number of non–noise eigenvalues will increase, not necessarily in relation to any measurable increase in the complexity of the generating process. Imagine our simple example was progressively extended by adding additional narrowband interferers. Initially the dimension K would increase as the number of sources, but not indefinitely. In the limit, when the number of interferers was much greater than the total number of samples processed, we may still observe a bounded eigenspectrum, similar to that depicted in Figure 3.9. The computed eigenspectra curves presented later demonstrate this behaviour.

Eigenspectra such as those presented later in the chapter, provide an estimate of the number of adaptive weights that might be required for effective suppression of interfering signals. For this reason, the first part of this study examined the influence various radar parameters have upon the eigenspectra curves. Figures 3.10 – 3.15 show the relative influence of the different parameters. The fixed radar parameters are those given in Table 3.1. Note that these values satisfy the DPCA condition. Until now, the discussion has related to a linear array of N elements lying along the axis of the aircraft. The eigenspectra presented here consider planar arrays of $N \times M$ elements lying parallel to the ground plane. In all plots the total number of samples processed (the fully adaptive dimension) is 128, but it can be seen by inspection that nearly all the scenarios demonstrate considerably less rank than this. The rank depends on various parameters, such as sampling rate, transmit beamwidth, the distribution of elements and the quiescent noise level. The number of non–noise eigenvalues are referred to as the clutter eigenvalues, since these eigenvalues relate to the underlying process which generates the clutter. The magnitude of the noise eigenvalues will depend upon the quiescent noise power, and will determine optimum processor gain, and target detectability.

Figures 3.10 - 3.13 show the eigenvalue spread as array geometry is varied. The three coefficients are N, M and L, and represent the number of elements parallel to the flight direction, the number of elements perpendicular to the flight direction, and the number of temporal snapshots combined in the processor. Consider first Figure 3.10. The most striking feature is the

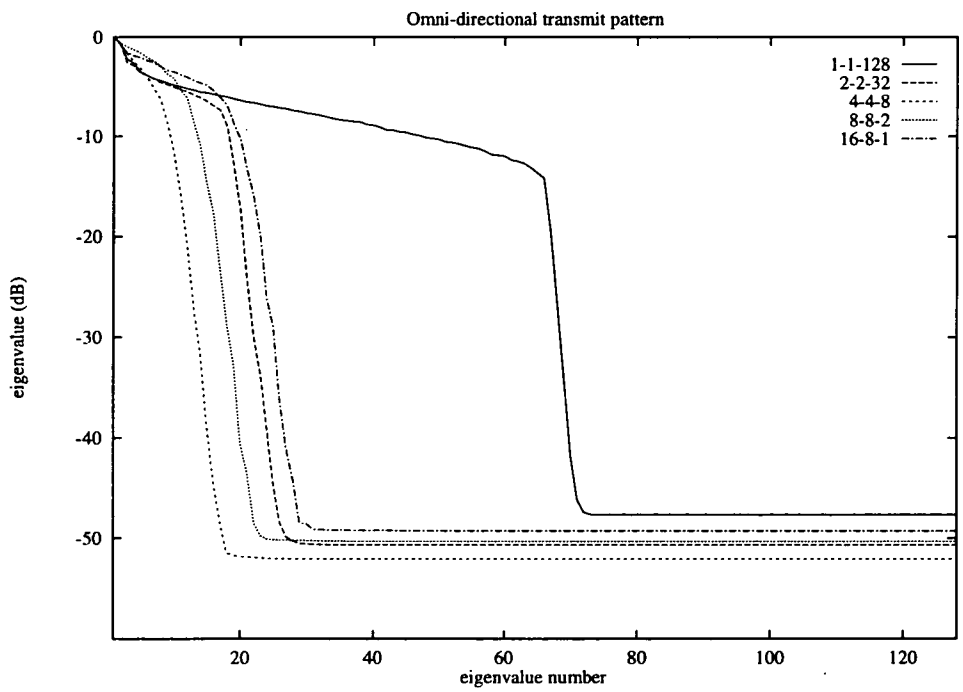

Figure 3.10: *Eigenspectrum: Distribution of elements - 2 dimensional array with omni–directional transmit pattern.*

rectangular nature of the curves. All the arrays display a stepped eigenspectra. This can be appreciated heuristically by noting that these result from an omni–directional transmit pattern. Such a uniform illumination will cause a rectangular Doppler spectrum, which could be seen to

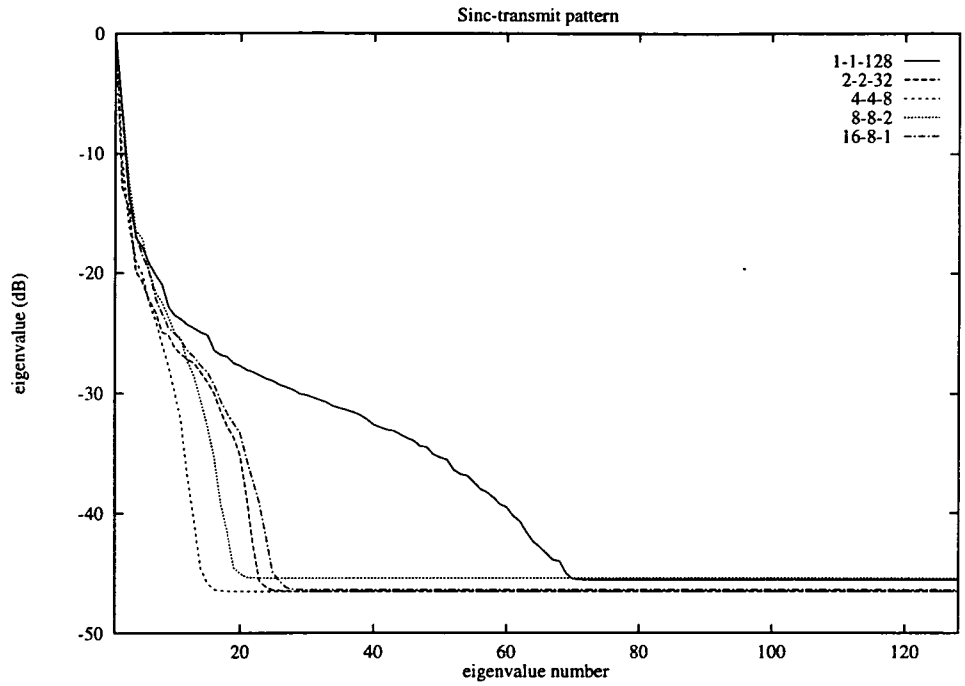

Figure 3.11: *Eigenspectrum: Distribution of elements - 2 dimensional array with* $\sin x/x$ *transmit pattern.*

result in a quasi-rectangular eigenspectrum. In Figure 3.10 it would appear that a large number of temporal samples (L large) leads to a larger clutter dimensionality. As L is reduced (and conversely $N \times M$ is increased) the clutter dimensionality decreases. Take for example curve 1–1–128. This represents the adaptive MTI type receiver depicted in Figure 3.5. Purely temporal filtering is relied upon to perform clutter suppression, which as was highlighted, will result in poor performance. This curve suggests that adaptive MTI filtering may additionally increase the number of significant degrees of freedom. The smallest clutter dimensionality can be seen to occur for case 4–4–8, the case in which the number of spatial degrees of freedom is nearest to the number of temporal degrees of freedom, i.e $N \times M \approx L$. In most situations an array with a combination of spatial and temporal resolution would prove to be the most versatile. Figure 3.11 shows the same array geometries, but in this case a $\sin x/x$ transmit pattern was used. The mainlobe beamwidth was 7.5 °. The clutter dimensionality remains more or less unchanged, but the distribution of clutter eigenvalues changes markedly. The ordered eigenvalues taper smoothly and do not exhibit the step–like behaviour of the omni–directional transmit pattern curves. Intuitively this would seem correct - a tapered Doppler spectrum will lead to a tapered eigenspectrum. This may be useful in a reduced-state processor because the performance would be expected to degrade more gracefully as the number of degrees of freedom in the processor are reduced.

Figures 3.12 and 3.13 show the eigenspectra for one dimensional arrays parallel and perpendicular to the flight direction, respectively. Both plots are similar in shape, but Figure 3.12 suggests that there may be some benefit to be derived from the array parallel to the flight direction. For all but the one element array the clutter dimensionality is approximately three less than for the

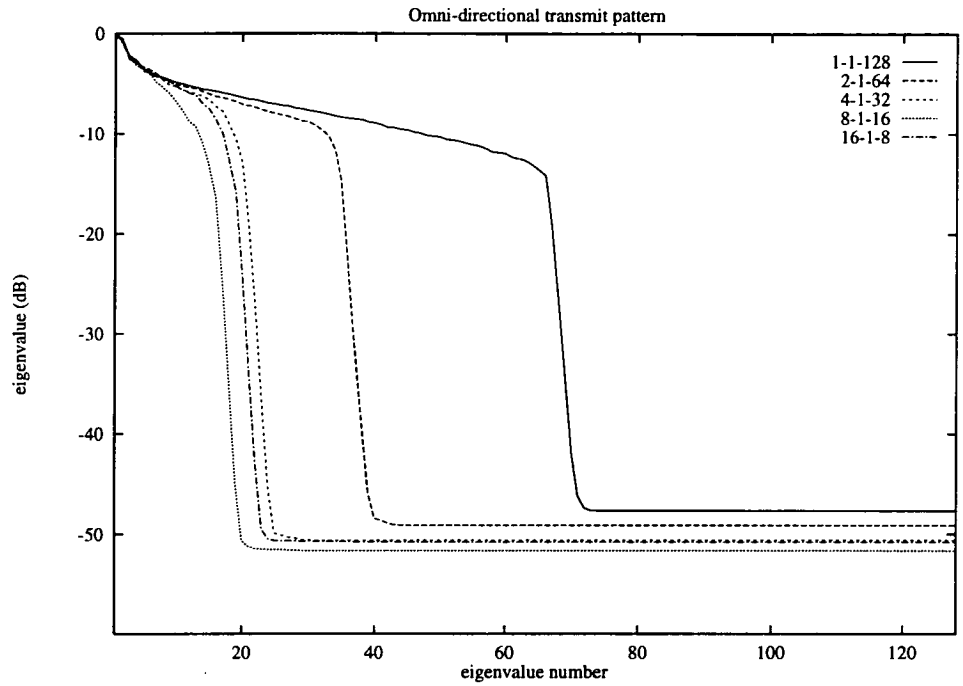

Figure 3.12: *Eigenspectrum: Distribution of elements - 1 dimensional array parallel to flight plane with omni–directional transmit pattern.*

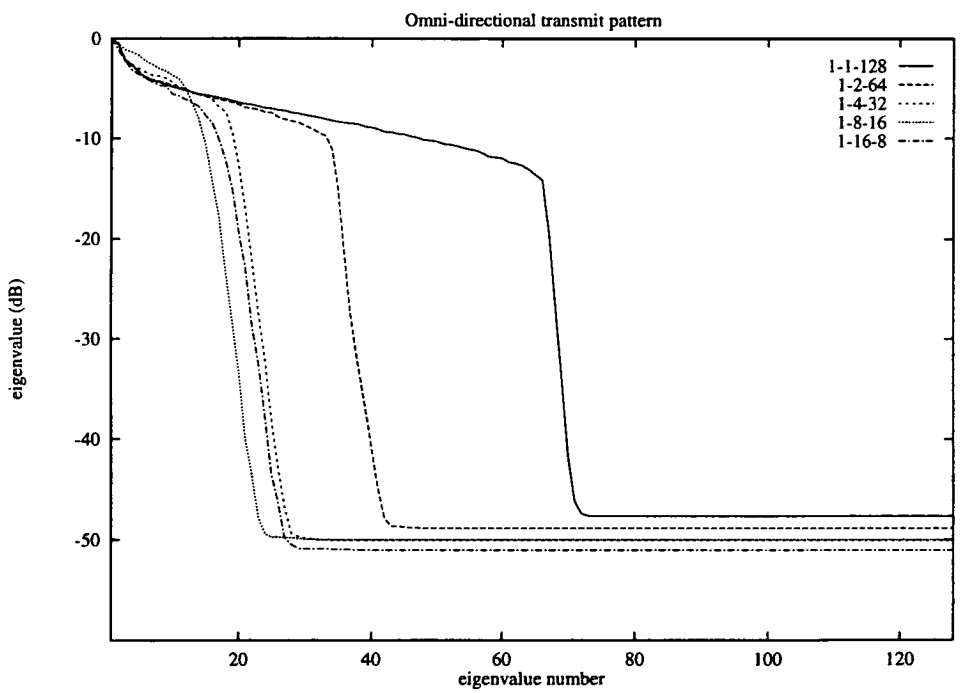

Figure 3.13: *Eigenspectrum: Distribution of elements - 1 dimensional array perpendicular to flight plane with omni–directional transmit pattern.*

equivalent perpendicular array. Klemm [1,5,6] suggested that, for a linear array, the clutter dimensionality was approximately equal to the sum of spatial and temporal degrees of freedom, i.e. $\approx N + L$, which will of course be a minimum when $N = L$. Su and Zhou [9] support this assertion, stating that the dimension will be bounded by $N + L$, providing the sampling of clutter is non-ambiguous and that no clutter fluctuations occur (i.e. the clutter statistics are stationary over the period during which the covariance matrix is formed). Subsequently Richardson [69] discussed this in relation to space–time processing which satisfies the DPCA criterion. The reasoning presented in section 3.7 suggested that the dimension of the clutter subspace equals $N + L - 1$. The plots presented here do not agree exactly, but can be seen to follow the trend. The difference may be attributed to different assumptions in the statistics of the clutter model, and effects due to finite sampling in the formation of the covariance matrices.

The next figure shows the influence of sampling rate upon the eigenspectrum. The Nyquist sampling rate for the received clutter field was $T_s = 0.25$ ms, which satsifies the DPCA condition. Undersampling ($T_s = 2.5$ ms, 25.0 ms) causes the eigenvalues to spread across the whole domain, whilst greatly oversampling ($T_s = 0.0025$ ms, 0.025 ms) reduces the clutter dimension. Again, these results seem intuitively correct. Undersampling of the received field leads to spectral aliasing of the clutter returns, effectively smearing the clutter returns over the whole azimuth–Doppler plane. By comparison with undersampling, a degree of oversampling will ensure that the clutter returns are localised upon the azimuth–Doppler plane.

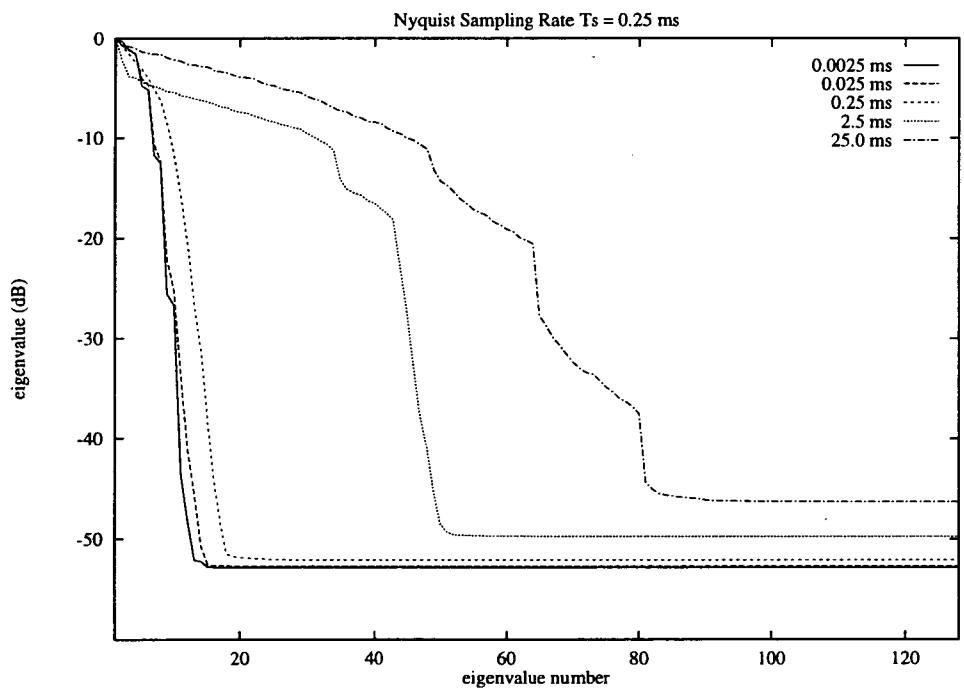

Figure 3.14: *Eigenspectrum: Influence of Sampling Rate. The Nyquist sampling rate for the clutter field was 0.25ms.*

The effect of differing transmit beamwidths upon the spectra is shown in Figure 3.15. All the curves demonstrate a similar clutter dimension. They show that a small transmit beamwidth

will tend to act as a narrow Doppler filter, selecting only a band of frequencies and thus resulting in a faster roll–off of the clutter eigenvalues. Alternatively, a large transmit beamwidth will illuminate a larger band of Doppler frequencies, and thus give rise to a slower roll–off of the clutter eigenvalues. This is an interesting point since it suggests that omni-directional transmit beams could be employed without increasing the required DOF of a subsequent processor. A similar effect was noted in [1].

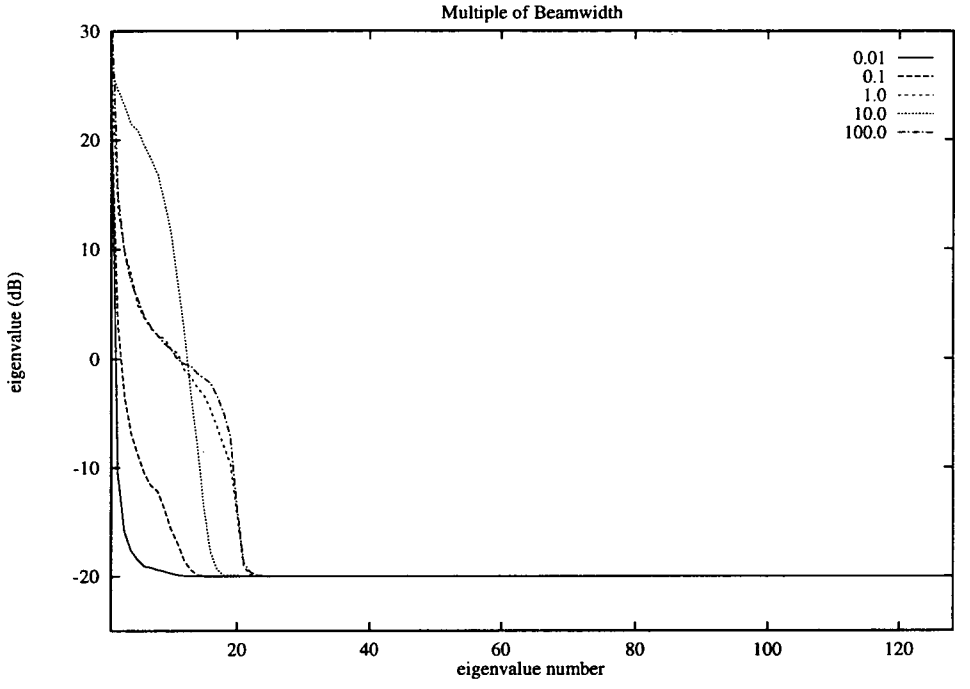

Figure 3.15: *Eigenspectrum: Influence of Transmit Beamwidth. The values indicate a multiple of the transmit beamwidth 7.5°. A larger multiple implies a smaller transmit aperture. The eigenvalues have been plotted normalised to the noise floor, as opposed to the largest eigenvalue.*

As was discussed in [5], eigendecompositions lack physical meaning. Any particular eigenvalue (or vector) cannot be related to some physical variable such as frequency or bearing, and as a result their usefulness is often overlooked. The number of clutter eigenvalues is a measure of the degrees of freedom of the process and tells the designer something about the obtainable gain and the number of degrees of freedom required for a clutter suppression filter. Many authors have considered this problem, suggesting eigen–decompositions of clutter space–time covariance matrices as techniques for partially adaptive processing. The approaches with most bearing upon this thesis are reviewed in the next chapter.

The plots that have been presented here indicate that the clutter dimensionality is significantly less than that of the data space. This is natural since the ground clutter returns within a single range gate represent a band limited process, that is, the clutter exists over a finite set of bearings, with a specific set of associated Doppler frequencies. Certain array geometries will lead to concise expressions for the dimension of the clutter subspace, e.g. the sideways looking space–time processor which satisfies the DPCA condition discussed in Section 3.7, whereas for

more general array geometries such expressions cannot be found. In these cases the clutter dimensionality will depend upon the assumptions made in the clutter model, specifically the statistical properties and number of the clutter scatterers, the depression angle of the particular range ring, and any sampling effects in forming the clutter covariance matrix.

3.9 Conclusion

This chapter has introduced pulse–Doppler radar, and examined the clutter suppression problem. A simple model for ground clutter returns has been given, and based upon this the implications of various radar and array parameters have been considered. This was achieved by computing an eigenspectrum for the simulated returns received at a moving platform. Through this it was shown that the dimension of the clutter subspace can be expected to increase as the number of samples processed is increased, and additionally as sampling rate is decreased, and transmit beamwidth is increased.

Chapter 4
An Iterative Algorithm

4.1 Introduction

The generalised sidelobe canceller was introduced in chapter 2 as an adaptive structure which is effective in the cancellation and suppression of two–dimensional noise and interference. Throughout the chapter the generalised sidelobe canceller was considered in a fully adaptive sense. No consideration was given to reducing the computational burden such structures place upon an adaptive processor. Within this chapter we will consider the problem of reducing the adaptive degrees of freedom which the processor is required to compute. A variety of techniques will be considered and their relative performance examined.

The computational requirements of each update in adaptive beamforming algorithms increases rapidly with the number of elements in the array. In many situations the beamformer will have an overly large number of degrees of freedom. The expression *degrees of freedom* denotes the number of unconstrained or "free" weights that must be computed. For example, an LCMV beamformer with L constraints upon N elements has $N - L$ degrees of freedom, the GSC implementation would separate these degrees of freedom into the unconstrained adaptive weight vector w_a. A fully adaptive beamformer uses all of these degrees of freedom whilst a *partially adaptive* beamformer will utilise only a subset of these degrees of freedom. When the system has too many degrees of freedom several undesirable results arise:

(i) the system will require many iterations before convergence; and

(ii) the computational burden per iteration will increase quickly as the number of weights.

It is therefore of great importance that we reduce the number of degrees of freedom available to the processor. One possible approach is to employ a linear (matrix) transformation to map the full dimension elemental data into a lower dimensional subspace, often called a beamspace, then to apply a signal processing algorithm to this new data set. However, the design of such transformations are generally guided by subjective criterion. This chapter will define the performance measures output mean square error (MSE) and signal–to–noise ratio (SNR) which are subsequently used to evaluate partially adaptive performance.

There is a performance penalty associated with partially adaptive beamforming. A partially adaptive beamformer will not converge to the same weight vector as the fully adaptive beamformer. Therefore the aim in partially adaptive beamformer design is to limit any degradation

in performance which occurs, whilst simultaneously reducing the adaptive dimension. This chapter will review several existing techniques for designing partially adaptive GSCs, and introduce an alternative approach - a technique based on iterative minimisation of the beamformer output mean squared error. It is shown that this approach not only leads to a sparse structure for the transformation matrix, but will also allow a reduction in the required partially adaptive dimension of the resultant beamformer. The relationship between this new approach and the existing techniques is described using a simple geometrical picture.

4.2 Partially adaptive beamformer design

Figure 4.1 depicts a generic sidelobe cancelling structure. N and M are respectively the numbers of elements in the array, and the number of adaptive weights that will be computed. Reduction in adaptive dimension is performed by the matrix $\boldsymbol{T}$. The fixed weights $w_{q1}, \ldots, w_{qN}$ are set to form a fixed beam with a peak in the direction of the desired signal, whilst the variable weights $w_{a1}, \ldots, w_{aM}$ are chosen so as to maximise some performance measure of the output $y(n)$. Several approaches to reducing degrees of freedom are based upon processing a subset of

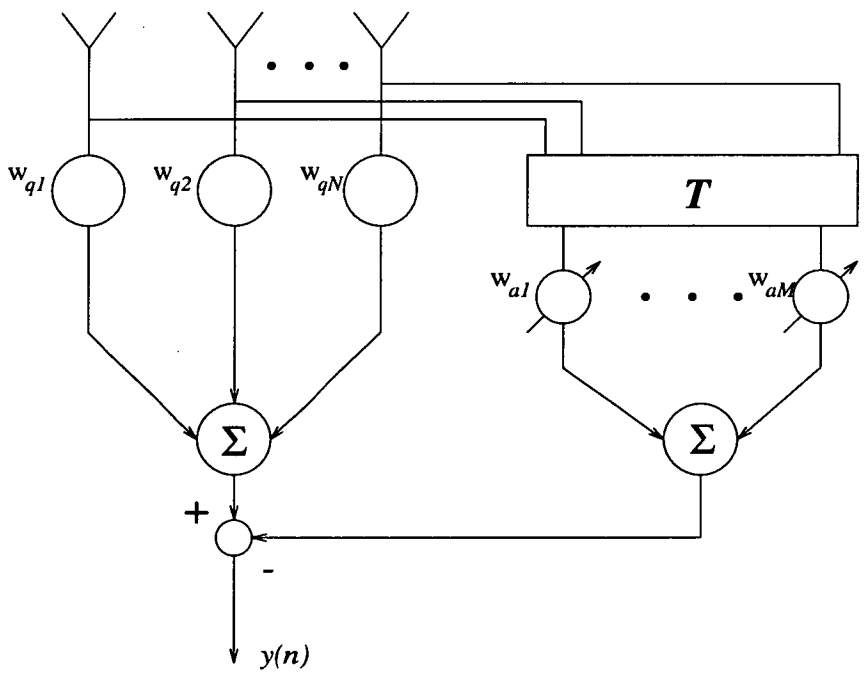

Figure 4.1: *Generic partially adaptive beamformer. N is the total number of elements, M the number of adaptive weights.*

the elemental outputs. This implies that the matrix $\boldsymbol{T}$ is a sparse matrix of zeros and ones. Morgan [71] evaluated partially adaptive beamformer performance for this multiple sidelobe canceller structure when the matrix $\boldsymbol{T}$ selected a subset of the elements to form auxiliary channel outputs. This configuration is termed an *element space* approach, since a subset of elemental outputs are utilised. An alternative approach has been described by several investigators, for example Adams [72] and Gabriel [73]. In this approach each column of $\boldsymbol{T}$ is used to form a beam. This technique is therefore termed a *beam space* approach. The columns of $\boldsymbol{T}$ are

designed as independent beamformers, each beam being steered in a different direction. The beams are matched to the locations of interfering signals. The objective is to direct a beam at each interfering source so that it can be subtracted from the fixed branch. Element space approaches are often preferred because of their simplicity. Improved cancellation performance can be obtained through beam space approaches, especially for interference due to either spatial distributed sources, or sources of appreciable temporal bandwidth. However, this improvement will be at the expense of implementing the required number of beams.

Chapman [74] considered selecting the columns of $\boldsymbol{T}$ to form subarrays, i.e. each column involves only a subset of the array elements. The weightings applied to each subarray (elements of $\boldsymbol{T}$) can be chosen in various ways, one of which is to use the subarray to form a beam. Cancellation performance depends upon the number of sensors in each subarray, the number of subarrays, and the weightings used to combine sensor outputs in the subarray. Note that each column of $\boldsymbol{T}$ will have zeros in the locations corresponding to elements that are excluded from that subarray, so that the overall $\boldsymbol{T}$ will be sparse in nature.

Takao [31] described a beam space partially adaptive antenna in which a subset of the active auxiliary beams were selected from the full set of potential auxiliary beams. The auxiliary beams were required to be orthogonal, such as those output from a Butler matrix. An eigen-decomposition similar to that in [73], was used to compute the optimum weight vector. This used a low rank approximation of the interference covariance matrix to form the optimum beam weightings.

The trade off between degrees of freedom and cancellation performance in partially adaptive beamforming merits some discussion. Consider a narrowband interferer of frequency ω_0, incident at the array from direction θ. The overall weight vector for the array is $\boldsymbol{w} = (\boldsymbol{w}_q - \boldsymbol{T}\boldsymbol{w}_a)$. The response of the fixed branch $\boldsymbol{w}_q$ to this source is $g_1 = \boldsymbol{w}_q^H \boldsymbol{d}(\theta,\omega_0)$, where $\boldsymbol{d}(\theta,\omega_0)$ is the steering vector. Perfect cancellation of this interferer will occur if $\boldsymbol{w}^H \boldsymbol{d}(\theta,\omega_0) = 0$. This implies

$$\boldsymbol{w}_a^H \boldsymbol{T}^H \boldsymbol{d}(\theta,\omega_0) = g_1, \tag{4.1}$$

that is the response of the adaptive branch will be equal to that of the fixed branch. Therefore perfect cancellation can be achieved by ensuring that the output of $\boldsymbol{T}$ is nonzero for this narrowband source. Now consider a broadband source of extent $\omega_a \leq \omega \leq \omega_b$ incident at the array. The response of the upper branch is given by $g_1(\omega) = \boldsymbol{w}_q^H \boldsymbol{d}(\theta,\omega)$. To achieve perfect cancellation $\boldsymbol{w}_a$ must satisfy

$$\boldsymbol{w}_a^H \boldsymbol{T}^H \boldsymbol{d}(\theta,\omega) = g_1(\omega). \qquad \omega_a \leq \omega \leq \omega_b \tag{4.2}$$

Define the response of each column of $\boldsymbol{T}$ as

$$f_i(\omega) = [\boldsymbol{T}]_i^H \boldsymbol{d}(\theta,\omega), \qquad 1 \leq i \leq M \tag{4.3}$$

where $[\boldsymbol{T}]_i$ denotes the ith column of $\boldsymbol{T}$. Equation (4.2) requires that $g_1(\omega)$ be expressed as a linear combination of $f_i(\omega)$, $1 \leq i \leq M$ over $\omega_a \leq \omega \leq \omega_b$. In general, this cannot be accomplished, so that we can conclude that total cancellation of a broadband interferer cannot be obtained. We can express the output power due to the broadband source as an integral over frequency of the magnitude squared of the difference between the fixed and adaptive branches responses, weighted by the interferer power spectrum. The degree of cancellation will vary greatly and will depend on spatial location, frequency extent and *critically* upon $\boldsymbol{T}$. For example, good cancellation could be obtained for $M = 1$, whereas poor cancellation might be achieved if M was large. These conclusions are equally valid for narrowband sources incident over a wide spatial region (spatial distributed interference). In the ground clutter suppression problem we will typically be faced with interference spread in both location and frequency.

4.3 The partially adaptive generalised sidelobe canceller

The generalised sidelobe canceller, because of its structure, is readily applicable to the design of partially adaptive beamformers. The basic GSC structure depicted in Figure 2.7 is an example of a beam space adaptive array. The signal blocking matrix forms a selection of beams, each beam having a null in the spatial location of the desired signal. In a partially adaptive GSC a transformation matrix $\boldsymbol{T}_n$ is inserted after the signal blocking matrix. We can therefore think of $\boldsymbol{T}$ as the combination of the signal blocking matrix $\boldsymbol{C}_n$, and the transformation $\boldsymbol{T}_n$, with the transformation acting as a beam selector/combiner. That is, partial adaptivity is achieved by either combining the beams output from $\boldsymbol{C}_n$ or else by selecting a subset of them.

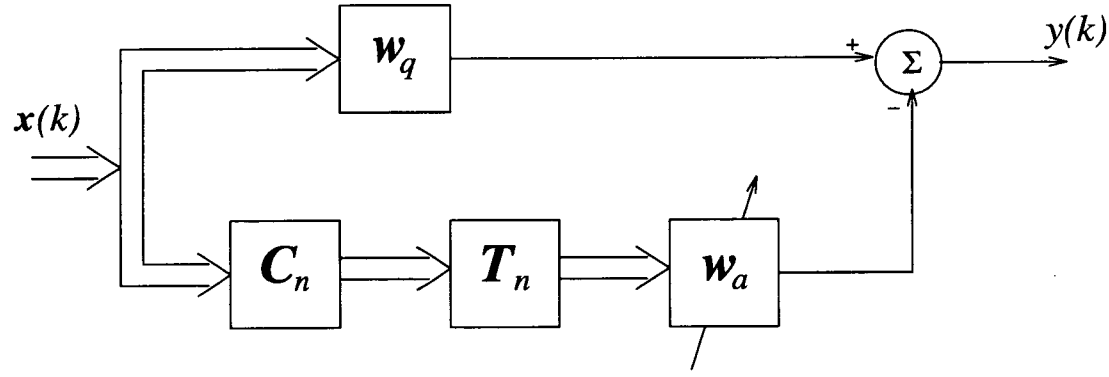

Figure 4.2: The partially adaptive generalised sidelobe canceller.

Recall that the fully adaptive GSC has dimension $NL - K$. The transformation $\boldsymbol{T}_n$ reduces this adaptive dimension from $NL - K$ to J. The partially adaptive GSC structure will allow an examination of the operation of a beam space partially adaptive array. Within this chapter we will be primarily concerned with designing a transformation matrix $\boldsymbol{T}_n$ which satisfies the twin goals of good cancellation performance and simplicity of implementation. This one statement summarises the whole problem in partially adaptive beamforming. Both objectives are of equal importance, yet each will have serious implications for the other. Good cancellation performance will be determined by how well the data output from the signal blocking matrix

is represented. Output from $\boldsymbol{C}_n$ will be a vector $\boldsymbol{C}_n^H \boldsymbol{n}(k)$, in which $\boldsymbol{n}(k)$ is a vector of noise and interference. This noise vector may have components arising from both correlated or uncorrelated interference. This data vector will lie within a space given by the expectation of the outer product of the vectors output from $\boldsymbol{C}_n$, i.e.

$$E\left\{\boldsymbol{C}_n^H \boldsymbol{n}(k)\boldsymbol{n}^H(k)\boldsymbol{C}_n\right\} = \boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n. \tag{4.4}$$

The square matrix $\boldsymbol{R}_n$ is the covariance matrix of interference and noise. How well the transformation $\boldsymbol{T}_n$ represents this space will determine how well the GSC operates as a partially adaptive beamformer.

Owsley [14] suggested a narrowband beamformer in which the columns of $\boldsymbol{T}_n$ were chosen as a basis for the space spanned by the fully adaptive weight vectors. The dimension of this space is given by the largest eigenvalues which represent the correlations in the input data sequence. Subsequently Van Veen [19] extended this to the broadband case. The dimension of the fully adaptive weight space can become particularly large since it is now given by the rank of the correlated part of the broadband covariance matrix. These approaches are capable of satisfactory performance when the interference is narrowband in nature, since each interferer will require only a single degree of freedom. However, they cannot be extended easily to the broadband environment whilst maintaining a small adaptive dimension. In the broadband case the cancellation beams must be matched over a range of frequencies at each interferer direction. As discussed in the previous section, in general it is not possible to perfectly cancel broadband interferences. Eqn. (4.2) requires control of the beamformer response over a continuous band of frequency, which can only be approximated. One example would be to use several banks of beams designed to span a range of directions, with each bank operating at a particular frequency. However, with this design the number of required beams will quickly become prohibitive.

Van Veen and Roberts [27,30] have considered a suboptimal sequential approach to this problem, in which each column of $\boldsymbol{T}_n$ is optimised in turn. The columns are designed to minimise the average interferer power over a range of likely interference environments. Each column of $\boldsymbol{T}_n$ is, however, optimised in an unconstrained manner, so that $\boldsymbol{T}_n$ will represent yet another beamformer. This impracticality is considered in the next section.

4.4 Practical realisation

The design of a partially adaptive beamformer cannot be guided solely by cancellation performance. Consideration must also be given to the practical problems of implementing the beamformer. Over complexity in the reduction network will negate any reduction in adaptive dimensionality obtained. A radar antenna designer would prefer to construct the majority of the beamforming structure at radio frequencies (RF) with demodulation and analogue–to–digital conversion occurring as far down the processing chain as possible. With this in mind, it is possible to begin formulating some guidelines for the design of the matrix $\boldsymbol{T}_n$. Firstly it would be preferable if only neighbouring elements were combined. Denote the matrix product

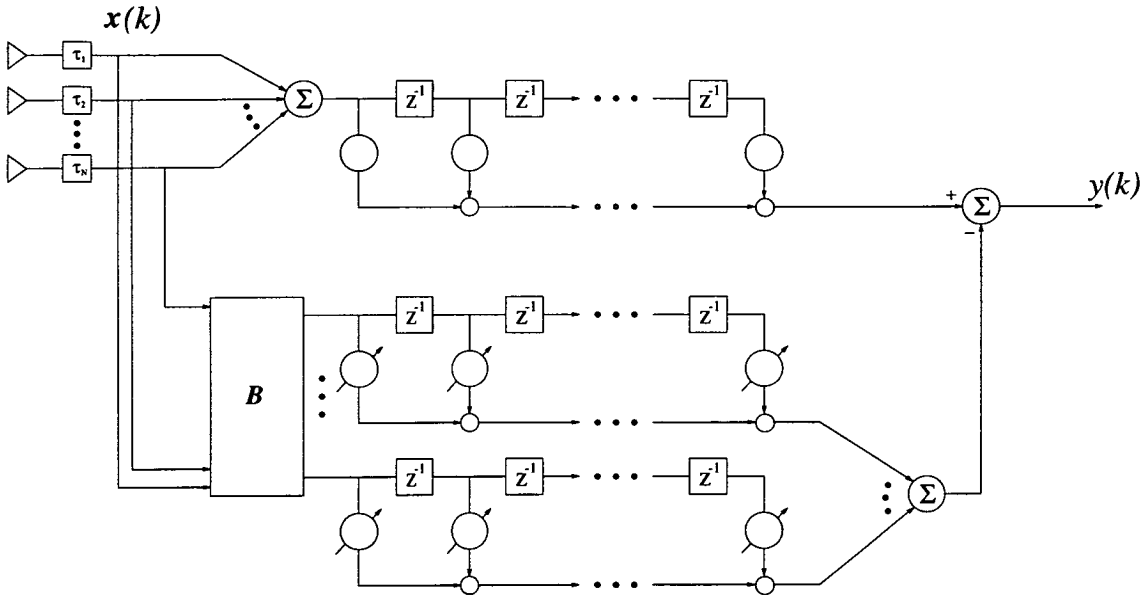

Figure 4.3: The generalised sidelobe canceller broadband beamformer.

$\boldsymbol{C}_n \boldsymbol{T}_n = \boldsymbol{T}$, and consider that $\boldsymbol{T}$ is some general complex matrix. To implement each column in hardware would require that a linear combination of all the elemental outputs is performed. This is clearly not practical. To ensure that only near neighbouring elements are combined, the first guideline is formed as (i) $\boldsymbol{T}$ should be broadly diagonal in structure (or some column–wise permutation). Secondly, it is preferable that $\boldsymbol{T}$ be a matrix of ones and zeros to obviate the use of amplitude weightings. Additionally $\boldsymbol{T}$ must still satisfy the signal blocking requirement imposed upon the lower branch. The second and third guidelines are therefore that (ii) $\boldsymbol{T}$'s only nonzero elements should be 1's, and (iii) $\boldsymbol{T}$ should satisfy $\boldsymbol{C}^H \boldsymbol{T} = \boldsymbol{0}$.

A key to the solution was provided by Frost in his early paper on LCMV beamforming [10]. The important feature of his beamformer was the use of *wideband* steering delays at each element. Figure 4.3 depicts of physical realisation of the basic GSC structure. The wideband steering delays τ_i steer the elemental outputs so that the desired signal appears identically at the input to w_q and $\boldsymbol{C}_n$. In the upper path, the outputs of the steering delays are summed linearly and the subsequent tap weights are used to identify the frequency of desired signals. The matrix $\boldsymbol{B}$ performs the signal blocking operation by simply differencing the outputs of the steering delays. Typically the $N \times (N-1)$ matrix $\boldsymbol{B}$ will take the form

$$\boldsymbol{B} = \begin{bmatrix} 1 & & & & & & \\ -1 & 1 & & & \mathbf{0} & & \\ & -1 & 1 & & & & \\ & & & \ddots & & & \\ & & & & -1 & 1 & \\ & & \mathbf{0} & & & -1 & 1 \\ & & & & & & -1 \end{bmatrix}. \tag{4.5}$$

This is a sparse bidiagonal matrix in which the only non-zero entries are either 1 or -1. The relationship between the signal blocking matrix $\boldsymbol{C}_n$ and the matrix $\boldsymbol{B}$ shown above is given explicitly as

$$\boldsymbol{C}_n = \boldsymbol{B} \bigotimes \boldsymbol{I}, \tag{4.6}$$

where $\boldsymbol{I}$ is an identity matrix of dimension equal to the number of taps within the tapped delay lines, and $\bigotimes$ indicates Hadamard product. Note that the signal blocking matrix will be full rank, and that each column of $\boldsymbol{C}_n$ represents a single adaptive weight in the lower path of the beamformer. The structures shown in Figures 4.2 and 4.3 will be equivalent provided $\boldsymbol{C}_n$ is as given in (4.6).

A signal blocking matrix whose elements consist entirely of ones and zeros has a significant implemental advantage, namely that the inputs to the adaptive processor are derived through a simple differencing of adjacent element outputs. For an array without wideband steering delays forming $\boldsymbol{C}_n$ is a more complex problem. In [16] a general approach for controlling beamformer spatial/spectral response is presented in which eigenvector constraints are derived using a low rank representation of the desired signal. The number of constraints required was found to be approximately equal to the time–bandwidth product of the signal of interest.

4.5 Beamformer performance measures

Design of the transformation matrix, and hence a partially adaptive beamformer requires that some suitable measure of beamformer performance exists. Measures relating to beamformer power patterns, e.g. beamwidth, peak and average sidelobe levels, null depths etc, have arisen because, historically, these were the measurable features of a beamformer. Since the advent of adaptive beamformers, especially adaptive array beamformers, these measures often do not provide a useful assessment of adaptive performance. It is often more meaningful to measure the performance not in terms of power patterns, but in terms of signal–to–noise characteristics, or cancellation performance. As way of an example, Figure 4.4 depicts the adapted response of a space–time processor when ground clutter returns are received. The sidelobe structure is complex which will make it difficult to measure the relative performance of different adaptive algorithms.

For the adaptive generalised sidelobe canceller a convenient measure is beamformer output mean squared error (MSE). This is defined as the mean squared difference between the output signal when only the desired signal is incident at the array and that which is output when both the desired signal and interference are incident. The data present in the array $\boldsymbol{x}(k)$ consists of a portion $\boldsymbol{s}(k)$ due to the desired signal and a portion $\boldsymbol{n}(k)$ due to the interference and noise, i.e. $\boldsymbol{x}(k) = \boldsymbol{s}(k) + \boldsymbol{n}(k)$. Thus the data covariance matrix will be given by four terms

$$\boldsymbol{R}_x = \boldsymbol{R}_s + \boldsymbol{R}_{ns} + \boldsymbol{R}_{sn} + \boldsymbol{R}_n, \tag{4.7}$$

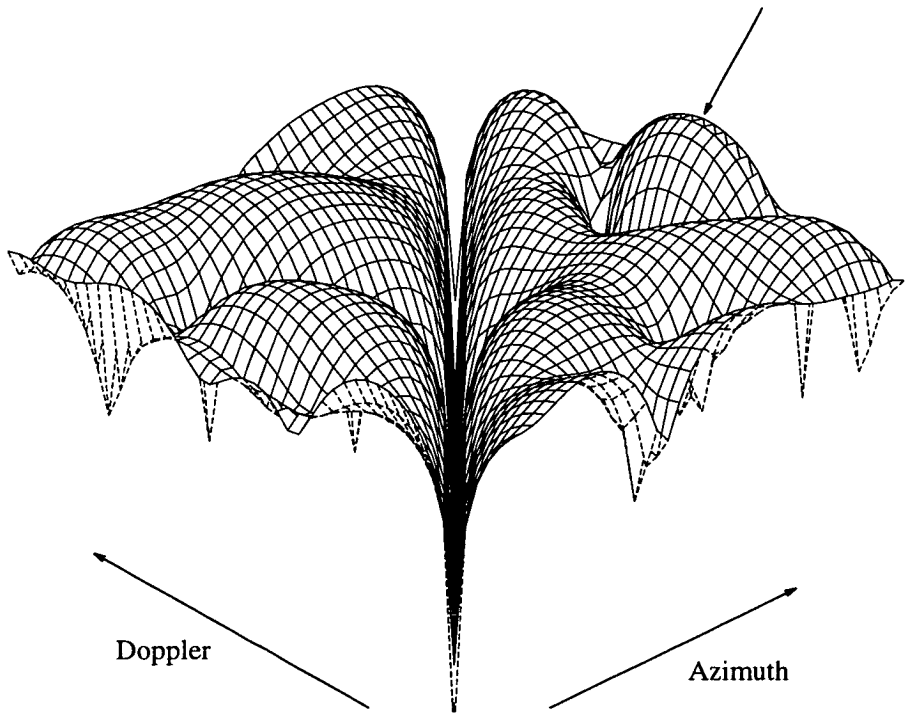

Figure 4.4: *Adapted response of a typical space–time processor, showing the complex sidelobe structure which often exists. The arrow indicates the direction from which the desired signal is incident.*

where $\boldsymbol{R}_n = E\left\{n(k)\,n(k)^H\right\}$, $\boldsymbol{R}_s = E\left\{s(k)\,s(k)^H\right\}$ and $\boldsymbol{R}_{ns} = \boldsymbol{R}_{sn}^H = E\left\{n(k)\,s(k)^H\right\}$. The cross–correlation terms $\boldsymbol{R}_{sn}$ and $\boldsymbol{R}_{ns}$ give a measure of the correlation between the desired signal and interferers. If no correlation exists then $\boldsymbol{R}_{sn} = \boldsymbol{R}_{ns} = \boldsymbol{0}$, but if the interferers are correlated with the desired signal (i.e. multipath propagation or intelligent jamming) then this will not be the case. Qian and Van Veen [35,36] have studied the use of partially adaptive generalised sidelobe cancellers in the suppression of coherent interferences. The LCMV adaptive beamformer will choose a weight vector which minimises the output power subject to a set of constraints upon the desired signal. This will be successful in the absence of coherent interferers. In the presence of coherent interferers though, the output power minimisation process acts not only in suppressing the interferers, but also to cause cancellation of the desired signal. This is demonstrated algebraically below. The generalised sidelobe canceller is an unconstrained implementation of the LCMV beamformer. Recall the unconstrained minimisation problem

$$\min_{\boldsymbol{w}_a} \left(\boldsymbol{w}_q - \boldsymbol{C}_n \boldsymbol{w}_a\right)^H \boldsymbol{R}_x \left(\boldsymbol{w}_q - \boldsymbol{C}_n \boldsymbol{w}_a\right), \tag{4.8}$$

the optimal weight vector being given by

$$\boldsymbol{w}_a \;=\; \left(\boldsymbol{C}_n^H \boldsymbol{R}_x \boldsymbol{C}_n\right)^{-1} \boldsymbol{C}_n^H \boldsymbol{R}_x \boldsymbol{w}_q. \tag{4.9}$$

The full rank signal blocking matrix will satisfy $\boldsymbol{C}^H \boldsymbol{C}_n = \boldsymbol{0}$, and will also satisfy $\boldsymbol{C}_n^H \boldsymbol{R}_{sn} = \boldsymbol{R}_{ns} \boldsymbol{C}_n = \boldsymbol{0}$ because the desired signal $s \in range(\boldsymbol{C})$. The weight vector $\boldsymbol{w}_a$ represents the available degrees of freedom in the beamformer. Substituting (4.7) into (4.9) and exploiting

the orthogonality between $s(k)$ and $\boldsymbol{C}_n$, we obtain

$$\boldsymbol{w}_a = \boldsymbol{w}_i + \boldsymbol{w}_s, \tag{4.10}$$

where

$$\begin{aligned} \boldsymbol{w}_i &= \left(\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n\right)^{-1} \boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{w}_q, \\ \boldsymbol{w}_s &= \left(\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n\right)^{-1} \boldsymbol{C}_n^H \boldsymbol{R}_{ns} \boldsymbol{w}_q. \end{aligned}$$

The first term on the RHS of (4.10) is the weight vector resulting from the noise and interference, whilst the second term is that which results from coherence between the desired signal and the interferers. Note that in the absence of coherent interferers, i.e. $\boldsymbol{R}_{ns} = \boldsymbol{0}$ the LCMV beamformer yields the best available weight set $\boldsymbol{w}_a = \boldsymbol{w}_i$ in the sense that it provides the maximum interference cancellation without causing any signal cancellation. In the presence of coherent interference, the additional term $\boldsymbol{w}_s$ will use the coherent interferers to cancel the desired signal.

Let $s(k)$ denote the desired signal at the beamformer output. We assume the constraints $\boldsymbol{C}^H \boldsymbol{w} = \boldsymbol{f}$ are chosen to ensure that $\boldsymbol{w}$ passes the desired signal with unit gain. Thus, $s(k) = \boldsymbol{w}_q^H \boldsymbol{s}(k) = \boldsymbol{w}^H \boldsymbol{s}(k)$. In other words, $s(k)$ is the beamformer output in the absence of noise and interference, $\boldsymbol{n}(k) = \boldsymbol{0}$. Now, the output mean squared error is defined as

$$\text{MSE} = E\left\{\left|s(k) - \boldsymbol{w}^H \boldsymbol{x}(k)\right|^2\right\}. \tag{4.11}$$

Substitution of $\boldsymbol{x}(k) = \boldsymbol{s}(k) + \boldsymbol{n}(k)$ and application of the constraint $s(k) = \boldsymbol{w}^H \boldsymbol{s}(k)$ yield

$$\begin{aligned} \text{MSE} &= E\left\{\left|s(k) - \boldsymbol{w}^H \boldsymbol{s}(k) - \boldsymbol{w}^H \boldsymbol{n}(k)\right|^2\right\} \\ &= E\left\{\left|\boldsymbol{w}^H \boldsymbol{n}(k)\right|^2\right\} \\ &= \boldsymbol{w}^H \boldsymbol{R}_n \boldsymbol{w}. \end{aligned} \tag{4.12}$$

Expanding this expression for output MSE, using (4.10) we get

$$\begin{aligned} \text{MSE} &= (\boldsymbol{w}_q - \boldsymbol{C}_n \boldsymbol{w}_a)^H \boldsymbol{R}_n (\boldsymbol{w}_q - \boldsymbol{C}_n \boldsymbol{w}_a) \\ &= \boldsymbol{w}_q^H \boldsymbol{R}_n \boldsymbol{w}_q - \boldsymbol{w}_a^H \boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{w}_q - \boldsymbol{w}_q^H \boldsymbol{R}_n \boldsymbol{C}_n \boldsymbol{w}_a + \boldsymbol{w}_a^H \boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n \boldsymbol{w}_a \\ &= \boldsymbol{w}_q^H \boldsymbol{R}_n \boldsymbol{w}_q - \boldsymbol{w}_q^H \boldsymbol{R}_n \boldsymbol{C}_n (\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n)^{-1} \boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{w}_q \\ &\qquad - \boldsymbol{w}_q^H \boldsymbol{R}_{sn} \boldsymbol{C}_n (\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n)^{-1} \boldsymbol{C}_n^H \boldsymbol{R}_{ns} \boldsymbol{w}_q. \end{aligned} \tag{4.13}$$

The output MSE therefore consists of three components which are related specifically to the upper and lower processing paths in the GSC. The first term on the RHS of (4.13) is the power output from the non-adaptive upper path due to interference and noise, whilst the second term is the portion which will be cancelled by the adaptive weights. The third is the portion of signal power cancelled due to the presence of $\boldsymbol{R}_{ns}$. This is as one would expect - the expression for the output MSE mirrors exactly the structure of the GSC. In a partially adaptive generalised

sidelobe canceller [19,27,30,33,35,36,56] a $NL-K$ by J transformation matrix $\boldsymbol{T}_n$, which maps the fully adaptive problem to a lower (J) dimensional space, is inserted after the signal blocking matrix. To simplify notation, the product $\boldsymbol{C}_n\boldsymbol{T}_n$ is denoted by $\boldsymbol{T}$ where convenient. For the partially adaptive GSC the above expression is modified to

$$\begin{aligned}\text{MSE} &= \boldsymbol{w}_q^H\boldsymbol{R}_n\boldsymbol{w}_q - \boldsymbol{w}_q^H\boldsymbol{R}_n\boldsymbol{T}(\boldsymbol{T}^H\boldsymbol{R}_n\boldsymbol{T})^{-1}\boldsymbol{T}^H\boldsymbol{R}_n\boldsymbol{w}_q \\ &\quad - \boldsymbol{w}_q^H\boldsymbol{R}_{sn}\boldsymbol{T}(\boldsymbol{T}^H\boldsymbol{R}_n\boldsymbol{T})^{-1}\boldsymbol{T}^H\boldsymbol{R}_{ns}\boldsymbol{w}_q. \end{aligned} \tag{4.14}$$

Our problem now is that of designing $\boldsymbol{T}_n$ so as to minimise any degradation in array performance. Within this chapter the primary performance measure employed will be output MSE. The task is to minimise this by appropriate design of $\boldsymbol{T}_n$. Existing techniques for designing $\boldsymbol{T}_n$ are discussed in following sections, followed by a new iterative approach which can lead to improved performance.

Measuring the beamformer performance in terms of output MSE will give a good impression of the ability to cancel noise and interference at the output of the beamformer. The expressions for output MSE derived above are measures which relate only to the noise power output from the beamformer. An alternative performance measure which incorporates the desired signal and also allows for an assessment of any signal cancellation effects is output signal–to–noise ratio (SNR). This is defined as the ratio of the desired signal power to interference and noise power at the output of the beamformer. Consider first the upper path of the beamformer. The power due to the desired signal will be given by

$$P_s = \boldsymbol{w}_q^H\boldsymbol{R}_s\boldsymbol{w}_q, \tag{4.15}$$

and the portion due to interference and noise

$$P_n = \boldsymbol{w}_q^H\boldsymbol{R}_n\boldsymbol{w}_q. \tag{4.16}$$

For a non–adaptive beamformer these powers would determine the output SNR. For the adaptive generalised sidelobe canceller there are cancellation signals generated by the lower adaptive path. The portion of signal power cancelled by a fully adaptive lower path due to coherent interferers is

$$P_{sc} = \boldsymbol{w}_q^H\boldsymbol{R}_{sn}\boldsymbol{C}_n(\boldsymbol{C}_n^H\boldsymbol{R}_n\boldsymbol{C}_n)^{-1}\boldsymbol{C}_n^H\boldsymbol{R}_{ns}\boldsymbol{w}_q, \tag{4.17}$$

and the noise and interference cancelled by the adaptive weights is given by

$$P_{nc} = \boldsymbol{w}_q^H\boldsymbol{R}_n\boldsymbol{C}_n(\boldsymbol{C}_n^H\boldsymbol{R}_n\boldsymbol{C}_n)^{-1}\boldsymbol{C}_n^H\boldsymbol{R}_n\boldsymbol{w}_q. \tag{4.18}$$

In the absence of correlation between the desired signal and the interferers ($\boldsymbol{R}_{ns}=\boldsymbol{0}$), the adaptive weights should ensure that the output signal–to–noise ratio is high. P_{sc} is the amount of the desired signal that is cancelled by coherent interference and represents the additional mean squared error due to the presence of coherent interference. If severe signal cancellation

occurs P_{sc} will approach P_s, the desired signal output power. The overall output signal–to–noise ratio for a generalised sidelobe canceller will be given by

$$\text{SNR} = \frac{P_s - P_{sc}}{P_n - P_{nc}}, \tag{4.19}$$

and the output mean squared error in the presence of coherent interference is

$$\text{MSE} = P_n - P_{nc} + P_{sc}. \tag{4.20}$$

For a partially adaptive beamformer equations (4.17) and (4.18) are modified as follows

$$P_{sc} = w_q^H \boldsymbol{R}_{sn} \boldsymbol{T} (\boldsymbol{T}^H \boldsymbol{R}_n \boldsymbol{T})^{-1} \boldsymbol{T}^H \boldsymbol{R}_{ns} w_q, \tag{4.21}$$

$$P_{nc} = w_q^H \boldsymbol{R}_n \boldsymbol{T} (\boldsymbol{T}^H \boldsymbol{R}_n \boldsymbol{T})^{-1} \boldsymbol{T}^H \boldsymbol{R}_n w_q. \tag{4.22}$$

The signal and noise powers P_s and P_n will be unchanged. An effective adaptive beamformer should have large P_{nc} to maintain reasonable interference cancellation whilst ensuring that P_{sc} is small to prevent signal cancellation. One natural criterion for designing $\boldsymbol{T}_n$ is therefore to minimise the output MSE. Note that P_n is independent of $\boldsymbol{T}_n$, so that minimising the MSE is equivalent to maximising $P_{nc} - P_{sc}$. For a fuller discussion on this see [35]. Throughout the remainder of this chapter coherent interferences are assumed not to be incident at the array. With this assumption the total power output from the array may be expressed as

$$\begin{aligned} P_{out} = {} & w_q^H \boldsymbol{R}_s w_q + w_q^H \boldsymbol{R}_n w_q \\ & - w_q^H \boldsymbol{R}_n \boldsymbol{T} \left(\boldsymbol{T}^H \boldsymbol{R}_n \boldsymbol{T} \right)^{-1} \boldsymbol{T}^H \boldsymbol{R}_n w_q. \end{aligned} \tag{4.23}$$

Similarly the interferer output power is expressed as

$$P_I = w_q^H \boldsymbol{R}_n w_q - w_q^H \boldsymbol{R}_n \boldsymbol{T} \left(\boldsymbol{T}^H \boldsymbol{R}_n \boldsymbol{T} \right)^{-1} \boldsymbol{T}^H \boldsymbol{R}_n w_q. \tag{4.24}$$

It can be seen that the output powers, mean squared error and signal–to–noise ratio depend upon $\boldsymbol{R}_n$ which will generally be unknown. However, in order to illustrate the potential of the algorithms studied, $\boldsymbol{R}_n$ is assumed to be known. This allows direct comparison of the optimal performance of different techniques in the suppression of unwanted interferences and noises incident at the array. In practice $\boldsymbol{R}_n$ will need to be estimated on line, typically using a block average incorporating a forgetting factor to allow a degree of tracking capability. Chapter 5 discusses the implications of such sampled covariance matrices.

4.6 Transformation matrix design

Now that some performance measures have been identified which give a realistic measure of the partially adaptive beamformer performance, we may now examine some techniques for designing the partially adaptive beamformer. Several techniques for transformation matrix design have been considered. The techniques employed to form the transformation are

- Eigenvector / eigenvalue approach
- Projection method
- Random method
- Iterative column selection (sparse approach)

In the first the eigenvectors corresponding to the largest eigenvalues of $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ are used to form the transformation matrix. The eigenvectors of $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ are mutually orthogonal and will form a basis for the interference subspace. Using the eigenvectors which correspond to the largest eigenvalues should therefore give good cancellation performance provided the reduced dimension J is greater than or equal to the dimension of the interference subspace. For the case when the array is fully adaptive and all the eigenvectors are used the array performance will be equivalent to the unmodified array because the transformation will simply be a mapping of $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ onto itself. This solution will lead to a non–sparse transformation matrix though, and therefore increased complexity in implementation.

The non–sparse nature of the transformation matrix found in the eigenvector approach may not be acceptable in a practical adaptive array. A transformation matrix which is mostly filled with zero elements and a few non–zero elements should reduce the implementational complexity of the adaptive processor. Such a transformation matrix will be equivalent to selecting a subset of the array elements, or equivalently a selection of the columns of $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$. A technique for selecting a numerically well conditioned set of columns was described by Nisbet *et al.* [75]. In this the J columns with the largest projections upon a set of axis vectors are selected and used to form an approximate representation of the signal subspace. This algorithm is summarised in section 4.6.2.

The next method considered uses a transformation matrix with the same sparse structure as that in the projection method, but in this case the columns are selected randomly. In some situations (dependent on the columns selected) this technique may lead to better performance than the projection method. This occurs because the columns selected in the projection method are chosen for their numerical properties rather than their ability to model the signal subspace. However, they will give a set of weighting coefficients which are numerically better conditioned than those that might occur in other approaches.

The final approach examined is a new algorithm which has been developed during the course of this project. Termed an *iterative* approach, it is based upon iterative minimisation of the beamformer output mean square error. This technique retains the sparse nature incorporated in the projection method, but because it operates in an iterative manner allows the selection of columns which would lead to better performance than the projection techniques. This algorithm is described fully in section 4.6.3.

4.6.1 Eigenstructure techniques

Eigenstructure techniques provide a method of designing partially adaptive beamformers which will have nearly fully adaptive performance under steady state conditions. The experimental analysis carried out in the previous chapter predicts that the required adaptive dimension is less than or equal to the rank of the spatially/temporally correlated portion of the interference covariance matrix for any arbitrary LCMV beamformer. To perform this prediction, knowledge of the eigenstructure of the correlated portion of the interference covariance matrix will be required. In order to avoid having to adaptively estimate this, eigenstructure techniques employ the eigenstructure of an "averaged" covariance matrix which spans all interference scenarios of interest. The adaptive dimension which is chosen is that which is given by the rank of this averaged covariance matrix.

Assuming that no coherent interference exists, $\boldsymbol{R}_x$ will decompose into two components, a portion due to the signal $\boldsymbol{R}_s$, and a portion due to interference and noise $\boldsymbol{R}_n$. Furthermore, assuming $\boldsymbol{C}_n^H \boldsymbol{R}_s = \boldsymbol{0}$, the adaptive weight vector w_a will depend only on $\boldsymbol{R}_n$,

$$w_a = (\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n)^{-1} \boldsymbol{C}_n^H \boldsymbol{R}_n w_q.$$

Assuming that the matrix $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ is symmetric and positive definite it can be factorised as

$$\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n = \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}^H + \sigma^2 \boldsymbol{I}, \tag{4.25}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix containing the J eigenvalues of the correlated portion

$$\boldsymbol{\Lambda} = \text{diag}[\, \lambda_1 \; \lambda_2 \; \ldots \; \lambda_J \,],$$

and $\boldsymbol{V}$ is the orthonormal matrix whose jth column is the eigenvector of the correlated portion of $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ associated with the jth eigenvalue. σ^2 represents the additive noise power. If the correlated portion of the interference covariance matrix lies in a J dimensional subspace, then correspondingly the adaptive weight vector used to cancel this interference will lie in a subspace that is at most J dimensional. Assuming that the eigenstructure of the interference is known, it is possible to reduce the adaptive dimensionality of the system from $NL - K$ to J without loss in cancellation performance. Most interference scenarios will have eigendecompositions which have a relatively small number of large magnitude eigenvalues and a larger number of small magnitude eigenvalues. We can make use of this fact by using the eigenvectors corresponding to the larger eigenvalues to fill the columns of the matrix $\boldsymbol{T}_n$. The extent to which $\boldsymbol{T}_n$ represents the space $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ will depend upon the structure of the eigenvalues. By plotting the eigenvalues in order of magnitude some quick conclusions can be drawn. If the eigenstructure contains relatively few large eigenvalues and a large number of much smaller eigenvalues then $\boldsymbol{T}_n$ should adequately represent the noise subspace. If, conversely there are a large number of large magnitude eigenvalues and a small number of much smaller eigenvalues then, for the same number of columns, $\boldsymbol{T}_n$ will poorly represent this subspace. If there exists no distinct

boundary or "cliff–edge" at which the eigenvalues suddenly fall in magnitude then the number of columns required to span the subspace will be difficult to define.

Putting these comments in a more mathematical sense, a measure for determining how well $\boldsymbol{T}_n$ represents the subspace $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ can be defined. Numerically determining the rank of a matrix is a difficult problem. In [53] an approach which gives a measure of the percentage error incurred by representing the space spanned by $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ with J eigenvectors was presented. Consider the ratio

$$\frac{\sum_{i=1}^{J} \lambda_i}{\sum_{i=1}^{NL-K} \lambda_i} > \alpha, \tag{4.26}$$

where λ_i are the ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{NL-K}$, of the correlated portion of $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$. The constant α is chosen to be less than or equal to one. The percentage $(1-\alpha)*100$ gives a measure of the loss incurred by representing $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ with J eigenvectors. Numerically computing the eigenstructure will not yield exactly zero eigenvalues, so that choosing α involves a compromise. Choosing α to be too large will unnecessarily increase the adaptive dimension of the processor, whilst choosing α too small will lead to poor performance. The only comment that can really be made is that the value of α chosen will depend upon the particular performance required for a particular beamformer. Van Veen [19] and Buckley [16] have both commented upon this.

The transformation matrix $\boldsymbol{T}_n$ can be thought of as forming eigen–beams. The adaptive processor will form J such beams which are then used for interference cancellation. Su and Zhou [9] considered an adaptive eigen–beamformer as a solution to clutter suppression. This method used a reduced set of Doppler–eigen beams which were updated on–line using an algorithm to adaptively estimate the space-time covariance matrix. Eigen beamformers are a specific example within a larger class of beamformer, those which are designed by subspace fitting/matching. The eigenvectors form a basis for the noise subspace, so that choosing the eigenvectors as columns of $\boldsymbol{T}_n$ will ensure that the beamformer is matched to the particular interference environment. These techniques are referred to as noise subspace techniques.

An alternative data independent subspace matching technique was proposed by Baraboski and Steinhardt [76]. A method called localised subspace projection (LSP) was used in which the subspace was determined solely by the desired mainbeam location and width and a priori knowledge of the array manifold. Therefore the subspace would be independent of any directional interferences. Designing the transformation with such a subspace allows enhanced sidelobe control whilst maintaining nulling performance. However, the performance was found to be very sensitive to errors in determining the array manifold. This class of beamformer is generally referred to as signal subspace beamformers.

The problem with filling the columns of $\boldsymbol{T}_n$ with eigenvectors is that, in general this gives rise to a non–sparse matrix $\boldsymbol{T}_n$ which will lead to significant complexity, especially in a large phased arrays. It would be preferable to have an auxiliary beamformer which either used only a few of the actual beams output from $\boldsymbol{C}_n$ or else combined them in some sparse manner to form a reduced number of beams. The sparse approaches described below have these features.

4.6.2 Projection methods

One technique for designing a transformation matrix that is sparse in nature was proposed in [75]. In this an efficient algorithm for selecting the columns that should be retained in a reduced state combiner was proposed. The solution found was chosen to be "close to" the minimum–norm solution, and would therefore retain its numerical properties. This technique acknowledges that in the vast majority of cases the matrix $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ will be numerically rank deficient, i.e. there will be several small eigenvalues. These small eigenvalues will lead to instability in the solution. As a result of the matrix inversion employed in forming the solution, the adaptive weight vector will be very sensitive to fluctuation in the magnitude of these small eigenvalues. Gabriel [73] discussed this problem at length. The solution adopted is to set the small eigenvalues to zero and to factorise the matrix as follows

$$\boldsymbol{C}_n^H \boldsymbol{R}_x \boldsymbol{C}_n \approx \boldsymbol{V} \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \boldsymbol{V}^H, \tag{4.27}$$

where the columns of $\boldsymbol{V}$ contain the $NL - K$ eigenvectors

$$\boldsymbol{V} = [\, v_1 \; v_2 \; \ldots \; v_{NL-K} \,],$$

and $\boldsymbol{\Lambda}$ contains the J non–zero eigenvalues

$$\boldsymbol{\Lambda} = \mathrm{diag}(\, \lambda_1 \; \lambda_2 \; \ldots \; \lambda_J \,),$$

of the matrix $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$. The minimum–norm solution for the generalised sidelobe canceller is found by using the pseudo–inverse of $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ and substituting in (4.9)

$$w_{min} = \boldsymbol{V} \begin{bmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \boldsymbol{V}^H \boldsymbol{C}_n^H \boldsymbol{R}_n w_q. \tag{4.28}$$

When the matrix $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ is rank deficient there will be many possible vectors w_a which will minimise the output mean square error. The minimum–norm solution w_{min} is unique in that it is the only weight vector that will simultaneously (i) minimise the mean square error and (ii) have the smallest Euclidian norm possible. A threshold must be chosen which reflects the precision of the arithmetic processor being used. Eigenvalues below this threshold are set to zero, the remainder are left as before. However, setting eigenvalues to zero is not the same as setting elements in the weight vector w_a to zero. In the general case the weight vector w_a will

have a full complement of elements. The rank deficiency highlighted earlier will mean that not all these elements are required. Using an analogy with the normal set of linear equations

$$\boldsymbol{A}\,\boldsymbol{x} \;=\; \boldsymbol{b}, \tag{4.29}$$

in which $\boldsymbol{x}$ represents the weight vector $\boldsymbol{w}_a$, and $\boldsymbol{A}$ the space $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$, we can say that the equations are undetermined in that the rank of $\boldsymbol{A}$ is less than the number of coefficients in $\boldsymbol{x}$. Thus there will be many solution vectors $\boldsymbol{x}$ which satisfy (4.29), one of which will be the minimum norm solution. The degeneracy which exists in (4.29) is an attractive feature, but it does not indicate which of the coefficients in $\boldsymbol{x}$ can be set to zero (discarded). There are obviously many possible subsets that could be chosen. The aim of the projection method is to find a subset which will be close to the minimum–norm solution. A solution which is close to the minimum–norm solution will have the same performance as many other possible solutions, but will also retain it's desirable numerical properties.

The objectives of the projection method are to find a solution which: (i) has a smaller number of coefficients than the minimum–norm solution, (ii) achieves the same MSE performance as the minimum–norm solution, (iii) has good numerical properties. Let $\boldsymbol{Z}$ denote the set of axis vectors, $\boldsymbol{Z} = [z_1\ z_2\ \ldots\ z_{NL-K}]$. A subset of axis vectors is selected by projecting each axis vector in $\boldsymbol{Z}$ in turn onto the signal subspace. The length of the projection is a measure of the proximity of the axis vector to the signal subspace. The axis vectors with the largest projections are chosen to form the transformation matrix. This will therefore ensure that the subset of coefficients (weights) chosen will have good numerical properties, and additionally will lie close to the minimum–norm solution. The algorithm can be summarised as follows -

(i). Calculate the eigenvalues and eigenvectors of $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ and hence estimate the rank T. The eigenvectors are collected into a matrix $\boldsymbol{V}$, which is partitioned as

$\boldsymbol{V} = [\ \boldsymbol{V}_s\ \boldsymbol{V}_n\ ]$,

in which $\boldsymbol{V}_s$ contains the T eigenvectors corresponding to the T largest eigenvalues.

(ii). Each axis vector is projected in turn onto the signal subspace. For the trivial case where the axis vectors are given by

$z_j = [\ z_{j1}\ z_{j2}\ \ldots\ z_{jj}\ \ldots\ z_{jNL}\ ]^T$,

where $z_{ji} = \delta_{ji}$, the Kronecker delta, these projections can be defined as a vector $\boldsymbol{d}$ such that

$\boldsymbol{d} = \mathrm{diag}(\ \boldsymbol{V}_s\ \boldsymbol{V}_s^H\ )$.

The T indices with the largest projections give the indices of the elements (columns) that will be retained. The others are discarded.

(iii). The matrix $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ is then reduced in size by removing the rows and columns corresponding to the discarded directions. This leaves a T by T matrix which is used in the adaptive processor to form the optimum weight vector.

4.6.3 An iterative approach

The projection technique described above uses a subset of the axis vectors to form the transformation matrix $\boldsymbol{T}_n$. This subset of vectors is selected at once, i.e. all the projections are computed then those axes with the largest projections are selected, the remainder being discarded. These vectors will retain the favourable properties of the minimum–norm solution whilst allowing a reduction in the adaptive dimension. Selecting the subset of vectors at once though, is not necessarily the best method of designing the transformation. The subset of vectors chosen are a selection of unit vectors taken from an identity matrix which will define the fully adaptive transformation matrix. Choosing these vectors according to the criteria described above will not allow the best MSE or SNR performance for a given partially adaptive dimension. It would seem that a better method would be to search for vectors which will minimise the output MSE. This is the idea behind iterative techniques. Designing the transformation matrix in this manner will not necessarily yield the optimum $\boldsymbol{T}_n$, but should provide one with better MSE performance than one which is designed in a single stage.

Analytical minimisation of the mean squared error over all interference scenarios of interest presents a formidable problem. Van Veen and others considered this in several papers [27,28,30], and suggested employing iterative techniques. A large array will have a transformation matrix which contains a large number of free elements. Global optimisation over each element is therefore an unrealistic proposition. The problem in a sparse design of $\boldsymbol{T}_n$, is that of choosing which degrees of freedom should be retained, and which should be discarded. In this section we present a sub–optimum iterative approach based upon minimisation of the output MSE of the partially adaptive array. Once the desired adaptive dimension is specified the algorithm iteratively searches for degrees of freedom which will best minimise the output MSE. A fully adaptive beamformer has a solution space of dimension $NL - K$. The sparse solution found means that a reduced number of dimensions, those that have most influence upon the adaptation, will be selected for the optimisation. This results in a transformation matrix that will be composed of a selection of unit vectors, the non–zero entries indicating the degrees of freedom selected.

The proposed algorithm is summarised in Table 4.1. We denote the transformation matrix of dimension $(NL-K)\times i$ with the superfix i as $\boldsymbol{T}_n^i$, and the set of allowed degrees of freedom as $\{z^j\}$, $1 \leq j \leq NL-K$. z^j is the jth column of an $(NL-K)$ dimensional identity matrix. The selected degrees of freedom are collected in $\{i^j\}$, $1 \leq j \leq J$. The selection procedure can now be summarised as follows. Initially the algorithm selects the first of the set of allowed vectors and forms $\boldsymbol{T}_n^1$ as a matrix with the single column z^1. The vector z^1 has a single entry 1 in the first position, and $(NL-K)-1$ zero elements. The output mean square error is computed for this transformation matrix and stored. The column z^1 is now replaced with the second vector in the allowed set, namely z^2, and the output mean square is evaluated once more. This procedure is repeated until all the vectors in the allowed set have been tried. The column to

□ At the first step, for $1 \leq j \leq NL-K$, compute

$$\boldsymbol{T}_n^1 = z_j$$

$$\text{MSE}^{(z_j)} = \boldsymbol{w}_q^H \boldsymbol{R}_n \boldsymbol{w}_q - \boldsymbol{w}_q^H \boldsymbol{R}_n \boldsymbol{C}_n \boldsymbol{T}_n^1 \left(\boldsymbol{T}_n^{1\,H} \boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n \boldsymbol{T}_n^1 \right)^{-1} \boldsymbol{T}_n^{1\,H} \boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{w}_q$$

$$\text{Find } \text{MSE}_1^{(i_1)} = \min\{\text{MSE}^{(z_j)},\ 1 \leq j \leq NL-K\}$$

then select $\boldsymbol{T}_n^1 = i_1$.

□ At the k^{th} step, $k \geq 2$, for $1 \leq j \leq NL-K$, $z_j \neq i_1, \ldots, i_{k-1}$, compute

$$\boldsymbol{T}_n^k = \left[\boldsymbol{T}_n^{k-1} : z_j \right]$$

$$\text{MSE}^{(z_j)} = \boldsymbol{w}_q^H \boldsymbol{R}_n \boldsymbol{w}_q - \boldsymbol{w}_q^H \boldsymbol{R}_n \boldsymbol{C}_n \boldsymbol{T}_n^k \left(\boldsymbol{T}_n^{k\,H} \boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n \boldsymbol{T}_n^k \right)^{-1} \boldsymbol{T}_n^{k\,H} \boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{w}_q$$

$$\text{Find } \text{MSE}_k^{(i_k)} = \min\{\text{MSE}^{(z_j)},\ 1 \leq j \leq NL-K,\ z_j \neq i_1, \ldots, i_{k-1}\}$$

then select $\boldsymbol{T}_n^k = \left[\boldsymbol{T}_n^{k-1} : i_k \right]$

□ The procedure is terminated at the Jth step when

$$\frac{\text{MSE}_{NL-K}}{\text{MSE}_J} > \rho$$

where $0 < \rho < 1$ is some performance measure. This gives a beamformer of adaptive dimension J.

Table 4.1: The selection algorithm.

be selected is chosen as the allowed vector which resulted in the smallest output MSE. This selected column i^1, is then deleted from the allowed set and the algorithm commences upon a search for additional vectors. At any stage k, the transformation matrix can be partitioned into two portions

$$\boldsymbol{T}_n^k = \left[\boldsymbol{T}_n^{k-1} : z^j \right], \tag{4.30}$$

where $\boldsymbol{T}_n^{k-1}$ contains the previously selected columns and the allowed vector z^j is the vector for which the output MSE is currently being evaluated. This iterative search for vectors which best minimise the output MSE is continued until all the allowed columns z^j have been added, or until the output MSE reaches an acceptable level. The simulations presented later show that often only a small number of the allowed columns are required, and that addition of further columns does little to further improve the output MSE performance.

In the fully adaptive case ($J = NL-K$), $\boldsymbol{T}_n$ will be an identity matrix (or some column–wise permutation), but in the partially adaptive case ($J < NL-K$) the columns of $\boldsymbol{T}_n$ will be those degrees of freedom that have most influence upon the output MSE. At each step the algorithm searches the remaining DOF for one that results in the greatest reduction in output MSE.

Figure 4.5 shows graphically how the selection procedure operates. For some step k in the

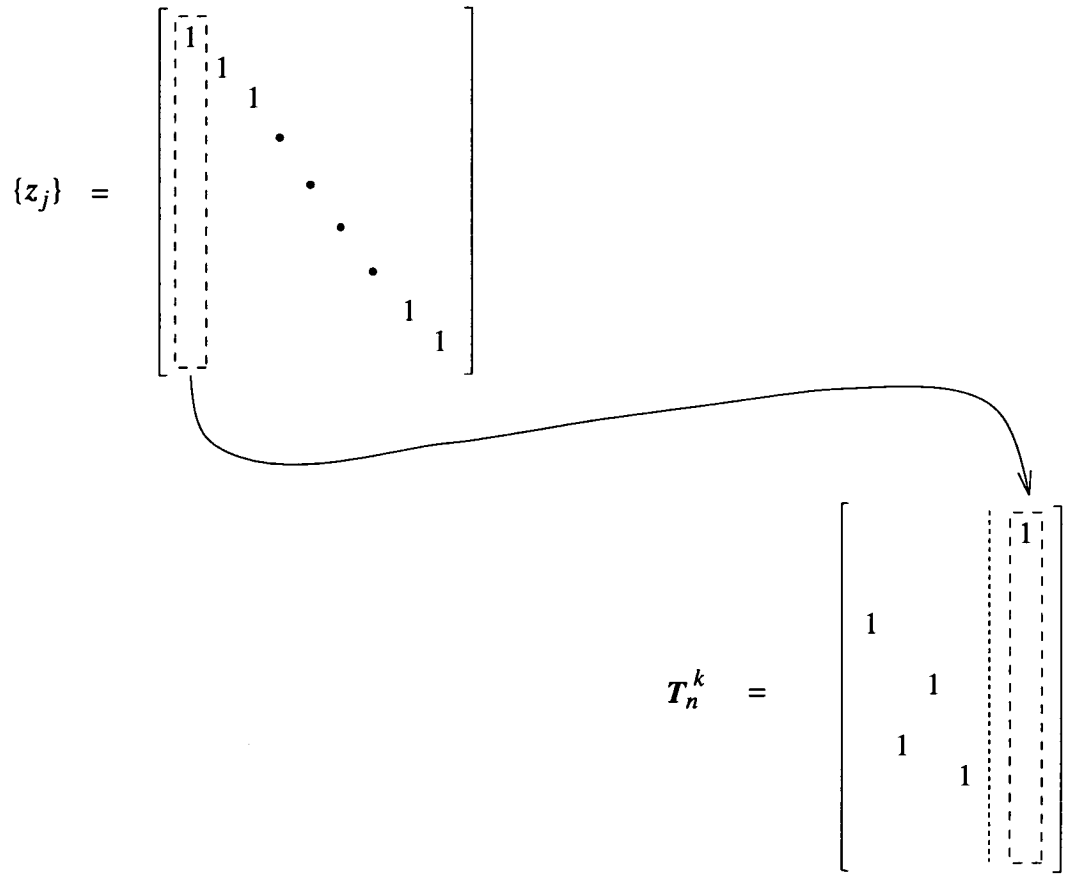

Figure 4.5: The selection procedure.

algorithm, $k-1$ columns of the matrix $\boldsymbol{T}_n$ will already have been selected. The algorithm begins a search for the kth column. This kth column is found by evaluating the MSE after appending each of the remaining columns in $\{z_j\}$ in turn to the current transformation matrix. This technique is similar to the orthogonal least squares algorithm described by Chen *et al.* [77]. In this radial basis function centres are chosen one by one so that each additional centre minimises the least squared error. After a centre is chosen the remaining basis vectors (columns) are made orthogonal to the chosen vector. This procedure is iterated until the output error is sufficiently small. The algorithm described above is simpler because of the sparse requirement upon the transformation matrix. By selecting a particular weight (axis vector) this iterative algorithm has effectively selected one of the columns of the matrix $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$. To follow the algorithm described in [77] the already selected columns would have to be orthogonalised with respect to this newly chosen column. The sparse nature of $\boldsymbol{T}_n$ precludes this operation. Orthogonality between the chosen columns is desirable, for several reasons. Firstly, an orthogonal set of columns will mean that stochastic gradient algorithms such as the least mean square (LMS) or recursive least squares (RLS) will have faster convergence to the optimum weight vector. Secondly, orthogonality between the columns should reduce the required adaptive dimension. The iterative nature of this algorithm will lead to a "good" set of vectors, but may not yield the optimum selection. Back–tracking may need to be added to the algorithm to allow an improved set to be chosen. Finding a globally optimum set would require a search of the order $(NL -$

$K)!/(J!(NL-K-J)!)$, whereas the sequential approach requires only $\sum_{j=1}^{J}(NL-K-j+1)$ steps.

4.7 Geometrical interpretation

The eigenvector, projection and iterative design approaches described above can be represented using a simple geometrical model. Using a simplified model will allow us a far clearer understanding of the operation of these algorithms.

Consider the three dimensional space depicted in Figure 4.6. In this it has been assumed that the fully adaptive weight has a dimension of three. This means that a fully adaptive weight vector will lie somewhere within this three dimensional space. Now consider the noise/interference subspace. Imagine, that after having completed some form of rank analysis upon the input data, i.e. an eigendecomposition, that this analysis had suggested that the input data had only 2 degrees of freedom, that is there would only be two non–zero eigenvalues, the third being identically equal to zero. This would mean that the input data would actually exist within a two–dimensional subspace (i.e. on a plane) within the three–dimensional space. This two–dimensional plane within the three–dimensional space is spanned by the eigenvectors computed in the eigendecomposition. The two–dimensional subspace in which the input data lies could be any general plane within the complex three–dimensional space. Figure 4.6 depicts one example.

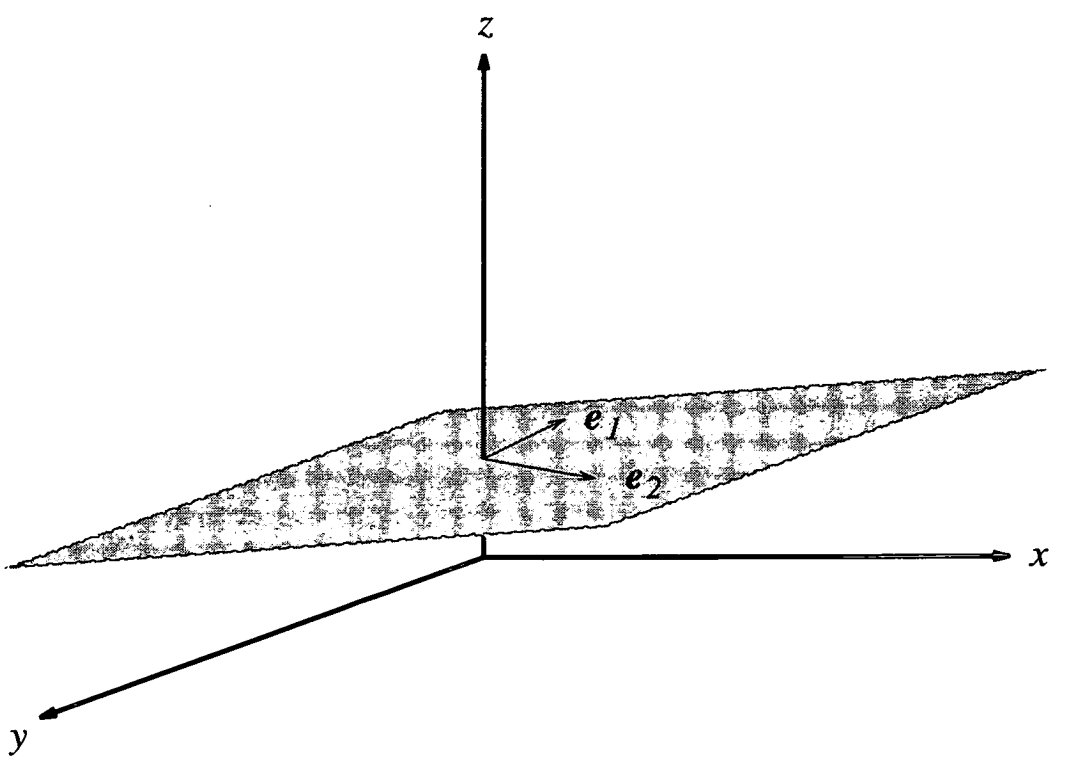

Figure 4.6: *Simplified subspace model showing a two–dimensional subspace (the plane) within a three–dimensional space. e_1 and e_2 are the two eigenvectors which span the interference subspace.*

In an eigenstructure design the transformation matrix is formed by choosing the eigenvectors associated with the largest eigenvalues of the matrix $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$, namely $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. Referring to the simplified model this transformation matrix can be interpreted as a mapping, one in which the two–dimensional subspace (plane) is mapped onto the "*nearest*" two–dimensional plane, in this case the plane indicated by the axes $x - y$. This holds true within our original definition of the plane - the two eigenvectors found will form a basis for this space so that the interference subspace is spanned by every possible linear combination of these eigenvectors. The mapping can be seen in this case to be an extremely efficient manner of representing the signal subspace. However, this is only true because the number of eigenvectors chosen was equal to the dimension of the space. If we were to only select one of the eigenvectors then the transformation matrix would very poorly represent the interference subspace. The subspace is two–dimensional, so that representing it with a single vector will cause drastic loss in performance. The simplified subspace model is obviously a particularly simple example. In most practical situations the eigenstructure will not be so simplified, but will consist of a certain number of near zero eigenvalues. This will mean that a precise definition of the signal dimensionality will not exist. If, for example, the third eigenvalue in our example had not been zero, but small in magnitude, then the interference subspace would not be a plane but would have a small but finite thickness. If this were the case then the remaining axis, the z axis would have a small but finite contribution in the representation of this subspace after the transformation.

The projection method described above attempts to select those axes which have the largest projections upon the signal subspace. The axes set in Figure. 4.6 are the simple set of unit vectors z_i. The projections upon the signal subspace are simply computed by evaluating the outer products of the eigenvectors and ordering them. In this simple example it is apparent that the axes which will have the largest projections are the x and y axes. The projection of the z axis upon the signal subspace will be significantly smaller than that of the x and y axes. In this example the signal subspace has been deliberately chosen so as to be nearly parallel to the $x - y$ plane. This allows us to demonstrate the projection technique. In reality there could be many axes which would have similar sized projections, so that there are many choices which would lead to similar performance. Imagine the plane in Figure 4.6 actually lay at 45° to the $x - y$ plane. In this case both the z and the x axes would have similar magnitude projections. Using the projection method would therefore require three degrees of freedom for adequate performance.

Now let us consider the iterative design approach. In this method it was stated that the transformation matrix would represent the important or significant dimensions within the space $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$. This can again be easily interpreted in the context of our simplified subspace model. Assuming the subspace model shown in Figure 4.6 it becomes clear that some dimensions will have a greater influence upon the adaptation than others. Following this argument, the iterative algorithm attempts to identify these dimensions that best approximate the interference

subspace. It can be seen that the dimensions given by the x and y axes will cover the majority of the interference subspace, whereas the dimension z will have little influence. These two dimensions (axes) happen to be the same as those chosen by the projection method. This occurs because of the simplified nature of this example subspace.

As one might expect, typical signal subspaces will be much more complex than the subspace depicted in the simplified model. We can expect to have an eigenstructure which has no cliff-edges and a large number of non-zero eigenvalues. In these situations the iterative design might be expected to perform better because it will select those degrees of freedom that best represent the interference subspace in terms of MSE performance. In simple interference environments both schemes will probably perform equally, but for the more complex environments, specifically those associated with the ground clutter problem encountered in airborne radar, the iterative approach offers a realistic hope of reducing the required partially adaptive dimension in the adaptive beamformer.

4.8 Training

An important part of these algorithms is the training data used in selecting the degrees of freedom to be retained. The covariance matrix of interference $\boldsymbol{R}_n$ should be formed as the "average" of a selection of scenarios. This was the approach taken in [19]. In this the transformation matrix was designed over all interferer scenarios of interest.. This can be done as follows. Suppose that a particular scenario can be described by the vector $\boldsymbol{\Theta}$. This vector will contain a group of parameters which define the interference – typically the angular and frequency extent, and the relative power levels. The averaged covariance matrix can thus be formed as

$$\bar{\boldsymbol{R}}_n = \int_{\boldsymbol{\Theta}_a}^{\boldsymbol{\Theta}_b} \boldsymbol{R}_n(\boldsymbol{\Theta})\ d\boldsymbol{\Theta}. \tag{4.31}$$

The space spanned by $\bar{\boldsymbol{R}}_n$ includes the space spanned by all scenarios $\boldsymbol{R}_n(\boldsymbol{\Theta})$ in the region $\boldsymbol{\Theta} \in [\boldsymbol{\Theta}_a, \boldsymbol{\Theta}_b]$. Once this averaged covariance matrix has been formed the transformation can be designed as described above. Thus $\boldsymbol{T}_n$ should operate effectively over all scenarios defined by $\boldsymbol{C}_n^H \bar{\boldsymbol{R}}_n \boldsymbol{C}_n$. The region $\boldsymbol{\Theta} \in [\boldsymbol{\Theta}_a, \boldsymbol{\Theta}_b]$ need not be continuous, but can be the union of several distinct regions. In practice the integral in (4.31) will be approximated by a discrete summation.

The choice of design region must reflect the interference scenarios that are liable to be encountered. On first inspection it may appear that utilising a large design region will give near optimum performance over a large range of scenarios. However, this neglects the influence the design region has upon the eigendecomposition of the matrix $\boldsymbol{C}_n^H \bar{\boldsymbol{R}}_n \boldsymbol{C}_n$. Increasing the scope of $[\boldsymbol{\Theta}_a, \boldsymbol{\Theta}_b]$ will result in an increase in the rank of $\boldsymbol{C}_n^H \bar{\boldsymbol{R}}_n \boldsymbol{C}_n$. Therefore, increasing the region over which the beamformer is expected to operate will cause a rise in the required adaptive dimension. Conversely, reducing the region $[\boldsymbol{\Theta}_a, \boldsymbol{\Theta}_b]$ over which the beamformer is de-

signed should lead to a reduction in rank, and thus a reduced adaptive dimension. Buckley [16] showed that the approximate rank of a covariance matrix for a broadband source of bandwidth B incident at an array from angle θ is

$$\lceil BT(\theta)/\pi \ + \ 1 \rceil, \tag{4.32}$$

where $T(\theta)$ represents the total temporal aperture of the array presented to a source at θ and $\lceil x \rceil$ indicates the next integer greater than x. This expression indicates that the rank of an interfering source will increase with both bandwidth and/or temporal aperture. Therefore we can expect the rank of $\boldsymbol{C}_n^H \bar{\boldsymbol{R}}_n \boldsymbol{C}_n$ to increase with increasing bandwidth or temporal aperture. A detailed analysis of the influence of varying design regions upon subsequent performance is beyond the present scope of this thesis, and would doubtless form a subjective study anyway.

4.9 Interference cancellation

The beamformer detailed in [19] used an eigenstructure technique to match the beamformer response over the required scenarios. The eigenstructure technique uses the decomposition in (4.25), by filling the columns of $\boldsymbol{T}_n$ with the most significant eigenvectors. This has the advantage that the resulting partially adaptive beamformer should have fully adaptive performance within the region described by $\boldsymbol{\Theta} \in [\boldsymbol{\Theta}_a, \boldsymbol{\Theta}_b]$. However, the required partially adaptive dimension is equal to the number of non–zero eigenvalues, which may be larger than that permitted by the adaptive processor. If this is the case the sparse solution will achieve better performance than the eigenstructure based approach for the following reasons. Eigenstructure techniques rely upon the assumption that there are a small number of significant eigenvalues and a collection of other much smaller, possibly zero, eigenvalues. In situations where this assumption is valid the eigenstructure technique will probably perform as well as any other scheme. An obvious example is the case of a single narrowband interferer in white noise. There will be only one large eigenvalue corresponding to the interferer, the remaining eigenvalues being equal to additive noise power. The interference will therefore be characterised by a single eigenvalue and eigenvector. When the interference eigenstructure is more complex, for example: when (a) the small eigenvalues are not all of the same magnitude; (b) the rank of $\boldsymbol{C}_n^H \bar{\boldsymbol{R}}_n \boldsymbol{C}_n$ is not easily defined, i.e. there is no "cliff–edge" at which the eigenvalues suddenly fall in magnitude; (c) most importantly, the allowable adaptive dimension is smaller than that demanded by the eigenstructure then there will exist better methods for forming the reduced–dimension processor. [32] considered this problem, showing that sequential power minimisation techniques can yield improved performance over eigenstructure designs. The sparse scheme described above is just one example. Additionally, it was found that cancellation is independent of eigenvalue size when the white noise power is small, meaning that a design based upon eigenvalue size alone may suffer severe loss of performance for particular bearing or frequency locations.

Another important point to note about eigenstructure techniques is that they make the implicit

assumption that all the information about the interference exists solely within the covariance matrix $\bar{\boldsymbol{R}}_n$. This does not allow the beamformer to take account of any other information which may exist about the interference structure, or for that matter the desired signal. Sequential algorithms, especially those using performance measures relating to the interference structure, make use of this additional information. Specifically, the eigenstructure design uses only the 2nd-order statistics of the input data, whereas sequential design techniques identify the most important dimensions, wherever this information is present. By successively searching the degrees of freedom available for one which best minimises the MSE the algorithm will necessarily have used any additional structure which exists.

The beamformer which results from the design procedure described above will be extremely simple to implement. Firstly, the upper path is only a summation. This is equivalent to a conventional unweighted beamformer. No complex weights or shading functions are required, since interference suppression will be performed by the lower adaptive branch of the beamformer. Secondly, the signal blocking operation and transformation $\boldsymbol{T}_n$ will consist purely of subtraction operations. The resultant signals can be passed straight to the adaptive portion of the beamformer. Contrasting this with the structure arising from an eigenstructure design the advantages become more clear. An eigenstructure design will give a transformation matrix which is full, i.e. a general complex matrix. For this matrix each input to the adaptive portion of the beamformer will be formed by combining all of the data samples present in the array. Such a computation negates the use of wideband steering delays at each array element - their insertion was to allow a sparse structure for $\boldsymbol{C}_n$ (and subsequently $\boldsymbol{T}_n$).

The discussion above provides a qualitative justification for the improved operation of an iterative beamformer. A quantitative description of the interference cancellation can be obtained by considering the following simple case which may be easily generalised. The following derivation is based on those variously presented in [16, 27, 30, 32]. Assume that the interference consists of a single broadband point interferer, uncorrelated with the desired signal, and spatially/temporally uncorrelated white noise. Letting $\boldsymbol{Q}(\theta)$ denote the covariance matrix for the interferer arriving from direction θ and σ_w^2 be the white noise level implies

$$\boldsymbol{R}_n = \boldsymbol{Q}(\theta) + \sigma_w^2 \boldsymbol{I}, \tag{4.33}$$

that is the interference covariance matrix consists of two independent terms. $\boldsymbol{Q}(\theta)$ is expressed in terms of the source power spectral density $\rho^2(\omega)$, the array response vector $\boldsymbol{d}(\theta, \omega)$, and the source frequency extent Ω as

$$\boldsymbol{Q}(\theta) = \int_{\Omega} \rho^2(\omega)\, \boldsymbol{d}(\theta, \omega)\, \boldsymbol{d}^H(\theta, \omega)\, d\omega. \tag{4.34}$$

An intuitive understanding of the interference cancellation is obtained by noting that P_{out} in (4.23) corresponds to $\boldsymbol{w}_q^H \boldsymbol{R}_s \boldsymbol{w}_q + (\boldsymbol{w}_q - \boldsymbol{C}_n \boldsymbol{T}_n \boldsymbol{w}_a)^H \boldsymbol{R}_n (\boldsymbol{w}_q - \boldsymbol{C}_n \boldsymbol{T}_n \boldsymbol{w}_a)$. Again writing

$\boldsymbol{T} = \boldsymbol{C}_n \boldsymbol{T}_n$ and substituting (4.34) into (4.23) yields

$$P_{out} = \boldsymbol{w}_q^H \boldsymbol{R}_s \boldsymbol{w}_q + \sigma_w^2 \left| \boldsymbol{w}_q - \boldsymbol{T}\boldsymbol{w}_a \right|^2 + \int_{\Omega} \rho^2(\omega) \left| \boldsymbol{w}_q^H \boldsymbol{d}(\theta,\omega) - \boldsymbol{w}_a^H \boldsymbol{T}^H \boldsymbol{d}(\theta,\omega) \right|^2 d\omega. \quad (4.35)$$

The output power in this simple case is therefore given by three terms. The first term is the signal output power, the second term is the white noise output power, and the last term is the interferer output power. The white noise power depends solely on the norm of $\boldsymbol{w}_q - \boldsymbol{T}\boldsymbol{w}_a$. The interferer output power is given by the average squared "error" between $\boldsymbol{w}_q^H \boldsymbol{d}(\theta,\omega)$ and $\boldsymbol{w}_a^H \boldsymbol{T}^H \boldsymbol{d}(\theta,\omega)$ over Ω weighted by the interferer power spectral density. $\boldsymbol{w}_q^H \boldsymbol{d}(\theta,\omega)$ and $\boldsymbol{w}_a^H \boldsymbol{T}^H \boldsymbol{d}(\theta,\omega)$ correspond to the responses of the fixed beamformer $\boldsymbol{w}_q$ and the adaptive branch $\boldsymbol{T}\boldsymbol{w}_a$, respectively, to a plane wave of frequency ω arriving from direction θ. Thus, the interferer output power is critically dependent on the degree to which the adaptive branch response matches the fixed beamformer response over the interferer frequency extent.

In [34] (reproduced in appendix A) an expression was derived giving the output power as a function of the adaptive weights as

$$P_{out} = \boldsymbol{w}_q^H \boldsymbol{R}_s \boldsymbol{w}_q + \boldsymbol{w}_q^H \boldsymbol{R}_n \boldsymbol{w}_q - \boldsymbol{g}^H \boldsymbol{g} \sum_{i=1}^{J} \frac{\sigma_i}{\sigma_i^2 + \sigma_w^2} \cos^2 \phi_i. \quad (4.36)$$

$\boldsymbol{w}_q^H \boldsymbol{R}_s \boldsymbol{w}_q$ represents the signal power at the output, $\boldsymbol{w}_q^H \boldsymbol{R}_n \boldsymbol{w}_q$ represents the interference plus white noise power at the fixed beamformer output, and the last term represents the reduction in output power resulting from the J adaptive weights. The σ_i^2 are the eigenvalues of the matrix $\boldsymbol{C}_n^H \boldsymbol{Q}(\theta) \boldsymbol{C}_n$. An adaptive cancellation factor (ACF) may be defined as

$$\text{ACF} = \sum_{i=1}^{J} \frac{\sigma_i}{\sigma_i^2 + \sigma_w^2} \cos^2 \phi_i, \quad (4.37)$$

which represents the relative cancellation performance of the beamformer. The adaptive cancellation is always bounded from above by one since $\sigma_i^2/(\sigma_i^2 + \sigma_w^2) < 1$ and $\sum_{i=1}^{J} \cos^2 \phi_i \leq 1$. It may seem plausible to select the columns of $\boldsymbol{T}_n$ corresponding only to the largest eigenvalues to minimise the squared error between the fully adaptive weight space and the space spanned by $\boldsymbol{T}_n$. However this can lead to dramatic performance breakdown. As was stated the problem with eigenstructure approaches is numerically determining the effective rank of $\boldsymbol{C}_n^H \boldsymbol{Q}(\theta) \boldsymbol{C}_n$. Choosing too few eigenvectors results in poor performance, while choosing too many increases the adaptive dimension unnecessarily. If the interferer direction is known, the eigenvalues of $\boldsymbol{C}_n^H \boldsymbol{Q}(\theta) \boldsymbol{C}_n$ are given by σ_i^2. Equation (4.36) shows that the cancellation associated with this mode depends upon the product $\sigma_i^2/(\sigma_i^2 + \sigma_w^2)$. When the white noise power is small the cancellation is therefore independent of eigenvalue size, thus designing $\boldsymbol{T}_n$ based in eigenvalue size is inappropriate.

Now let us consider how the above derivations can be interpreted for an iteratively designed beamformer. Recall that our objective is to design $\boldsymbol{T}_n$ to minimise the output interference

power over a range of possible scenarios. If we tighten this declaration to minimising the *average* interference power over a range of scenarios, we may express the problem as

$$\min_{\boldsymbol{T}_n} \int_{\Theta_a}^{\Theta_b} P_I(\boldsymbol{\Theta})\, d\Theta, \tag{4.38}$$

where $P_I(\boldsymbol{\Theta})$ is the interference power which will be a function of $\boldsymbol{T}_n$. The difficulty associated with analytical and numerical optimisation of (4.38) is such that sequential or iterative design techniques where $\boldsymbol{T}_n$ is designed one column at a time, with each new column depending upon the previously designed columns, provide the only practical solution. The interferer output power will critically dependent upon the degree to which the adaptive branch matches the response of the fixed weight branch over the interferer frequency extent. In one approach taken by Van Veen [27] each column was unconstrained, that is no restrictions were applied to the values of the elements in each column. This sequential technique attempted to match the partially adaptive weight vector to the fully adaptive weight vector over a given design region. It was shown that the error associated with this would increase as the size of the design region. If the region defined by $[\Theta_a, \Theta_b]$ is overly large then it should be broken up into a set of separate subregions, one for each column of $\boldsymbol{T}_n$, rather than design each column over the same large region. This recommendation arises directly from the discussion relating to (4.35). Increasing the size of the design region will cause a rise in adaptive dimension if cancellation performance is not be sacrificed. This was alluded to in section 4.8. The relationship between the error and region size implies that performance should improve as region size decreases. This characteristic was noted in [27].

Van Veen's approach, as with eigenstructure beamformers, leads to a full transformation matrix, with the associated computational penalties. The sparse solution found by the algorithm proposed above, attempts to retain the sequential nature of Van Veen, whilst satisfying the desire for computational efficiency. Columns are selected upon a MSE criterion, so satisfying the performance aspect of any partially adaptive design technique, but because of the nature of the columns selected, the goal of low computational cost is simultaneously met.

4.10 Computational expense

A key feature of the iterative technique is the saving in the number of operations that must be performed in computing the adaptive weight vector. Denoting the signal blocking and weight selection operations with the $NL \times J$ matrix $\boldsymbol{T} = \boldsymbol{C}_n \boldsymbol{T}_n$, the adaptive weight vector can be written as

$$\boldsymbol{w}_a = \left(\boldsymbol{T}^H \boldsymbol{R}_x \boldsymbol{T}\right)^{-1} \boldsymbol{T}^H \boldsymbol{R}_x \boldsymbol{w}_q. \tag{4.39}$$

In an eigenstructure beamformer $\boldsymbol{T}$ is a full complex matrix, whereas for the sparse beamformer it will contain at most $2J$ non–zero elements. For both beamformers, computing the weight vector consists of the following operations: evaluating the term inside the inverse, the inverse

itself, then the subsequent cross–correlation with the data in the upper path. The total number of operations for each beamformer are given by

$$\text{Eigenstructure} : \left.\begin{array}{ll} \text{Mult.} & J^3 + 2J(NL)^2 + 2J^2(NL) + J(NL), \\ \text{Add.} & J^3 + 2J(NL)^2 + 2J^2(NL) + J(NL). \end{array}\right\} \tag{4.40}$$

$$\text{Sparse} : \left.\begin{array}{ll} \text{Mult.} & J^3 + J(NL)^2 + J(NL), \\ \text{Add.} & J^3 + J(NL)^2 + 3J(NL) + J^2. \end{array}\right\} \tag{4.41}$$

These expression give the number of complex multiplication and addition/subtraction operations required, and allow an estimate of the relative expense of the differing approaches. The actual number of operations required will be computed later.

4.11 Simulation results

Having established the techniques which are available to us, we may now examine their relative performance. The results presented in this section relate to the two phases of beamformer operation, namely training and the subsequent operation. The performance of the new and existing algorithms is examined for the suppression of ground clutter received at an airborne phased array radar. The following computer simulations show the performance of the iterative algorithm as compared to existing design techniques. The clutter returns at a variety of GSC beamformers were computed. The results will consider the effects of both beamformer dimension and the differing radar parameters. Three radar scenarios are considered, for convenience these have been named scenarios 1 through to 3. Table 4.2 summarises the different radar parameters. All other parameters are as those in Table 3.1. Appendix D contains additional simulation results for the design techniques.

	Scenario 1	**Scenario 2**	**Scenario 3**
range	$2000m$	$1300m$	$2000m$
depression angle ϕ	$30°$	$50.3°$	$30°$
look direction θ	$18°$	$18°$	$90°$
target frequency band	$f_u = -761Hz$ $f_l = -1521Hz$	$f_u = -761Hz$ $f_l = -1521Hz$	$f_u = 1711.9Hz$ $f_l = 951Hz$

Table 4.2: Parameters for three training scenarios.

4.11.1 Training phase

The averaged transformation matrix used in the training phase was a covariance matrix that would result from an omnidirectional transmit pattern, i.e. one in which all Doppler frequencies

along the diagonal of the Doppler azimuth plane are illuminated equally. This is in fact the average of all covariance matrices $\boldsymbol{R}_n(\phi)$ for look directions θ within a particular range gate ϕ_0. This was the scenario employed by Klemm [1,5–7]. Figure 4.7 shows the ordered eigenspectrum for the first radar scenario. The step–like nature of the eigenspectra would suggest that the rank of the correlated portion of the interference covariance matrix is well defined. In this case the rank would appear to be approximately 33. For partially adaptive dimensions less than this number we might expect the iterative algorithm to have better performance. Another point to note is that there are a large number of small eigenvalues. This would suggest that our estimation of the rank may be inaccurate. We can expect the output MSE performance to mirror this.

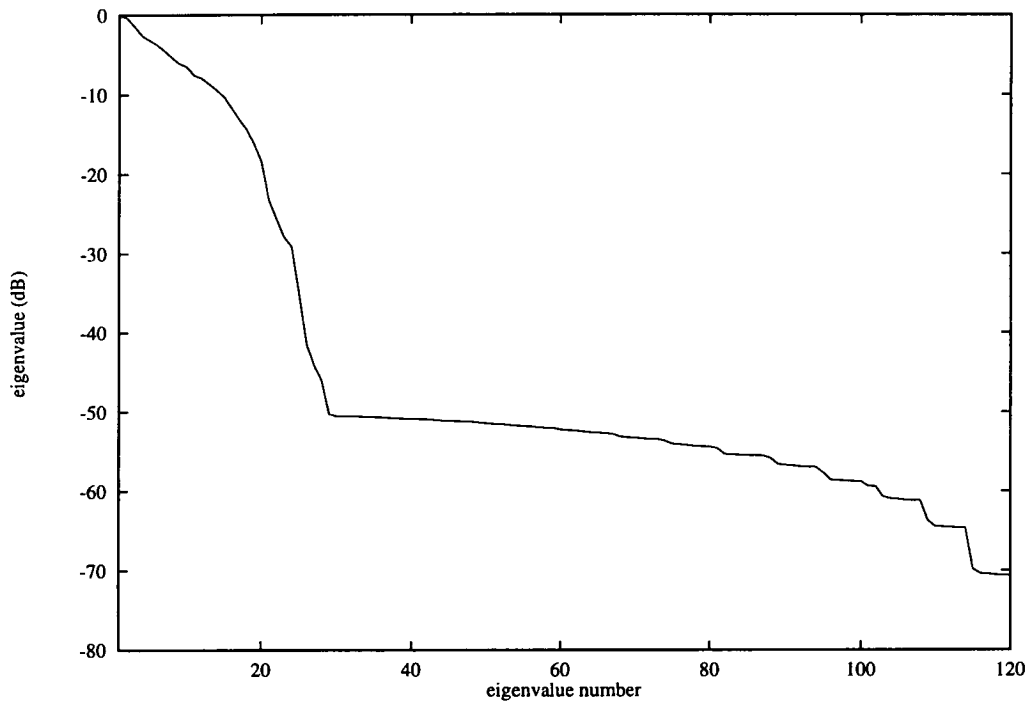

Figure 4.7: Eigenspectra for correlated portion of $\boldsymbol{C}_n^H \bar{\boldsymbol{R}}_n \boldsymbol{C}_n$ - Scenario 1.

Figure 4.8 compares the output MSE of the new iterative algorithm with that of the other techniques with the scenario 1 parameters. The projection and eigenvector approaches perform similarly, both exhibit a distinct step at 26 degrees of freedom. Subsequently both curves roll–off in almost linear fashion as additional degrees of freedom are added. At no point does either curve bottom out, meaning that no precise definition of the required adaptive dimension exists. The random design approach appears to have picked a particularly bad set of columns. The output MSE performance is by far the worst of the four techniques. The iterative algorithm performs significantly better. As can be seen, for low partially adaptive dimensions, the new algorithm has a lower output MSE than that of the other techniques. What is also interesting is that the new algorithm tends to the fully adaptive MSE much more quickly than the other approaches. In fact, for an output MSE of -35.5dB only 43 DOF are required in the sparse design, as opposed to 78 in the eigenvector case, representing an almost two–fold decrease in

adaptive dimension.

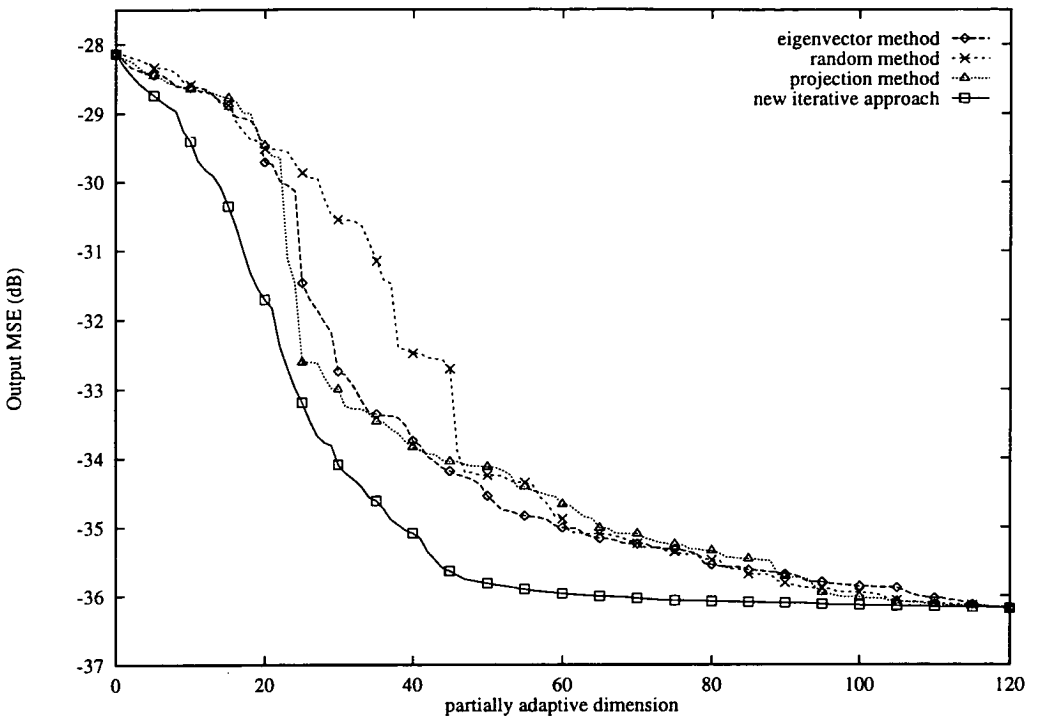

Figure 4.8: *Output mean squared error for new iterative design and existing techniques during training phase - Scenario 1.*

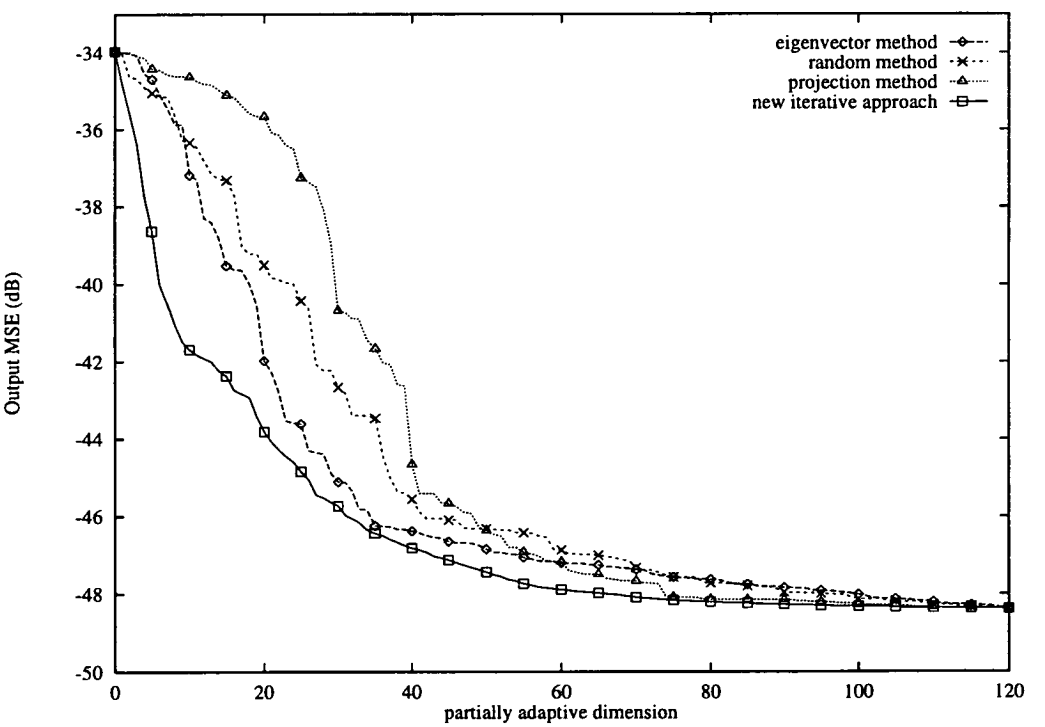

Figure 4.9: *Output mean squared error for new iterative design and existing techniques during training phase - Scenario 2.*

Figures 4.9 and 4.10 show the training curves of the same beamformer under the remaining

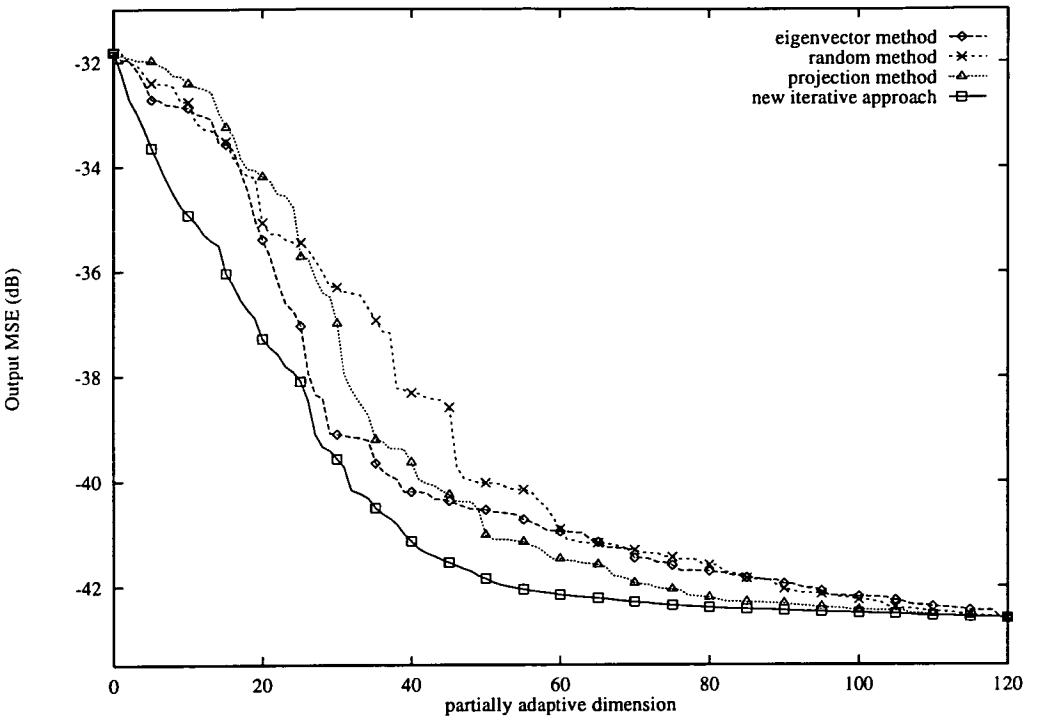

Figure 4.10: *Output mean squared error for new iterative design and existing techniques during training phase - Scenario 3.*

scenarios. As can be seen, the performance is similar for all three cases. The iterative design out–performs the other techniques, whilst the relative performance of the other techniques varies but never exceeds the performance of the new algorithm. The smooth nature of the MSE curves shown also suggests that the set chosen, if not the best, is certainly a good set. The output MSE may have a more stepped nature under other interference environments. It is interesting to note that the randomly designed beamformer of scenario 2 performs better than that designed using the projection approach.

Table 4.3 shows the cost of the two techniques in terms of multiplication and addition/subtraction operations. Three beamformer dimensions are compared; scenario 1 described above and two smaller cases. K is the number of constraints. For each example, the required partially adaptive dimension was found for the eigenstructure (J_e) and sparse (J_s) beamformers, as for Figure 4.8, then the operational costs were computed according to (4.40) and (4.41). The operational savings of the iterative approach are dramatic, typically of the order of 80% in the number of multiplications. Similar results can be seen in Appendix D.

4.11.2 Operational example

The next set of results consider the beamformer output signal–to–noise ratio (SNR) once the transformation matrix has been designed as outlined above. Two situations are considered - a narrowband desired signal and a broadband desired signal. In both cases the desired

Order of beamformer		Eigenstructure			Sparse			Saving %	
NL	K	J_e	Multiplies	Additions/ Subtract.	J_s	Multiplies	Additions/ Subtract.	Multiplies	Additions/ Subtract.
16	4	11	11011	11011	6	1848	3612	83	67
32	4	22	87384	87384	10	11560	22540	86	74
128	8	78	4597944	4597944	43	794003	1506892	82	67

Table 4.3: Operational expense of eigenstructure and iterative beamformers.

signal is identified by time delay steering of the elemental outputs and subsequent filtering according to the response vector $\boldsymbol{f}$. Suppression of clutter returns is now achieved through two processes. Non–look direction interference is cancelled by the adaptive portion as would be the case without broadband steering delays. Look direction interference (mainlobe clutter) is suppressed by designing the response vector $\boldsymbol{f}$ as a bandpass filter centred on the target signal frequency. In practice, where the target Doppler is unknown, there will be a selection of upper branches, one matched to each of the expected target Doppler frequencies. Both target signals are assumed to arrive from the look direction $(\theta_0, \phi_0) = (30^o, 18^o)$. Mainlobe clutter has Doppler $f_d = f_{max} \cos\theta_0$ where $f_{max} = (2V/\lambda)\cos\phi_0$ is the maximum Doppler of clutter returns. The narrowband signal has centre frequency $f_{\cdot} = -0.6 f_{max}$ and the broadband target has extent $[-0.8 f_{max}, -0.4 f_{max}]$. In both cases an equi–ripple bandpass filter with -30dB sidelobe level was used to isolate the desired signal. The elements of the response vector $\boldsymbol{f}$ are explicitly

$$\begin{aligned} \boldsymbol{f} \;=\; & [0.150 + j0.000,\; -0.006 + j0.100,\; -0.118 - j0.014,\; 0.024 - j0.127, \\ & 0.126 + j0.032,\; -0.036 + j0.114,\; -0.093 - j0.036,\; 0.062 - j0.136]^T . \end{aligned}$$

Figures 4.11 and 4.12 show the output SNR as degrees of freedom are successively added to the beamformer. The iteratively designed beamformer performs slightly better than the other designs, although are within 3dB of fully adaptive performance at a partially adaptive dimension of approximately 40. It should be remembered that the eigenstructure based design does this at a considerably greater implemental expense.

4.12 Limitations of the approach

The algorithm presented here operates by successively increasing the dimension of the adaptive weight vector $\boldsymbol{w}_a$. This increase in dimension arises from the addition of a column to the matrix $\boldsymbol{T}_n$, the added column being selected as described in Table 4.1. For a given dimension k (i.e. the kth iteration of the algorithm) the algorithm searches for the best vector to place in column k of $\boldsymbol{T}_n$. A question that must be answered is, is it possible that the algorithm could select a column which moves the MSE toward a local minima, rather than toward a global minima? It is important to note that when the algorithm selects the next column, this new column will be optimum in terms of the previously selected columns. For example, suppose that 9 columns had

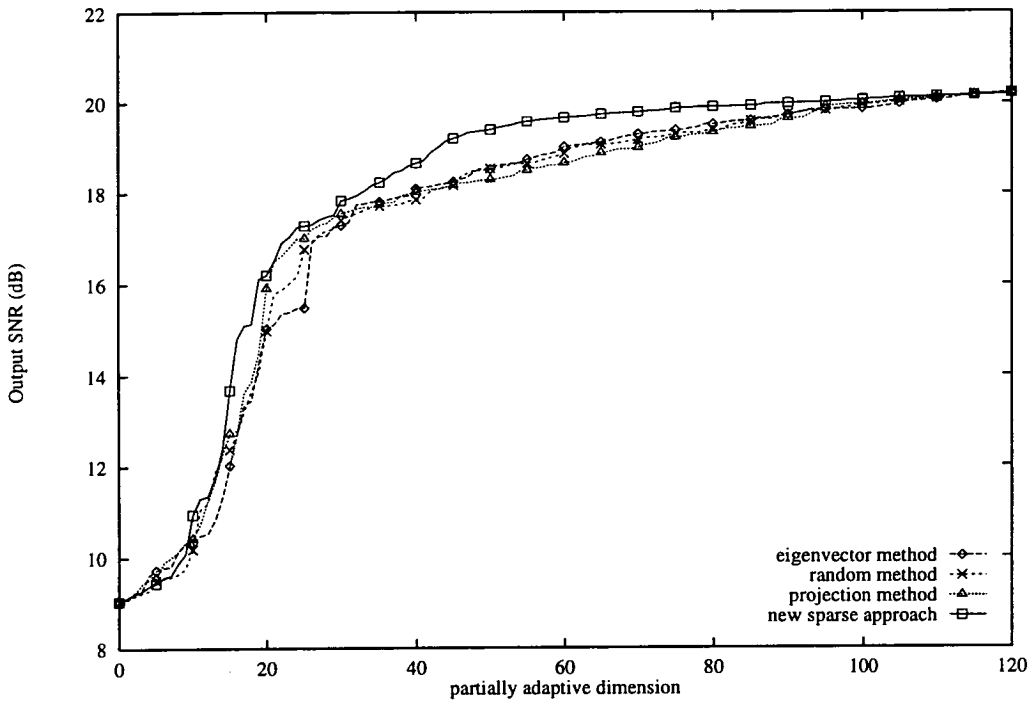

Figure 4.11: *Output signal–to–noise ratio for new iterative design and existing approaches with a narrowband target signal.*

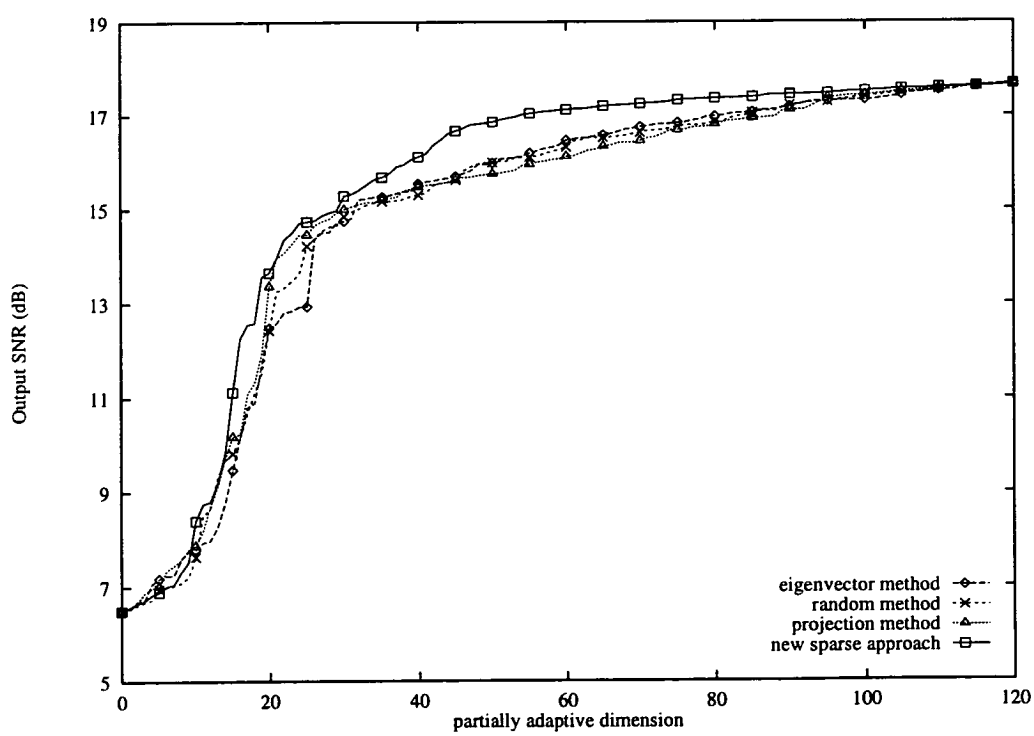

Figure 4.12: *Output signal–to–noise ratio for new iterative design and existing approaches with a broadband target signal.*

been selected, the 10th column is that which results in the greatest decrease in output MSE, so that this column will be the optimum choice (in a MSE sense) given the allowed set from which to choose, and the 9 previously selected columns. However, it may occur that these 10 columns are not the optimum 10 columns, and that an alternative 10 columns may result in improved MSE performance. This situation can occur because the algorithm does not look further ahead than a single column; at any instance it simply searches for the best column to add.

Incorporating an element of back–tracking has been proposed as a solution to this limitation of forward search techniques. In their paper Chng *et al* [78] contrasted the performance of a forward search algorithm with and without back–tracking. A simple example will indicate the advantage gained by back–tracking. Consider the linear system

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{e}, \tag{4.42}$$

where

$$\boldsymbol{X} = \begin{bmatrix} 3 & 0 & 3 \\ 0 & 0 & 0.1 \\ 0 & 3 & 3 \\ 0 & 0 & 0.1 \end{bmatrix}; \quad \boldsymbol{y} = \begin{bmatrix} 3 \\ 0 \\ 3 \\ 0 \end{bmatrix}, \tag{4.43}$$

and $\boldsymbol{e}$ is the error vector of approximating $\boldsymbol{y}$ with $\boldsymbol{X}\boldsymbol{h}$. The elements of $\boldsymbol{h}$ represent the columns of $\boldsymbol{X}$ that have been selected. Applying the algorithm described earlier to the least squares problem

$$\min_{\boldsymbol{h}} |\boldsymbol{y} - \boldsymbol{X}\boldsymbol{h}|^2, \tag{4.44}$$

will result in $\boldsymbol{h}$ being composed entirely of 1s or 0s. Suppose, initially, that only one column of $\boldsymbol{X}$ were to be selected, then the algorithm would select the third column of $\boldsymbol{X}$, i.e. $[3\ 0.1\ 3\ 0.1]^T$. Now consider adding another column, the algorithm will select one of the first or second columns to add to the third. Clearly this selection is poorer than having chosen both the first and second columns. This demonstrates that a sub–optimal subset of the columns could be selected without the addition of back–tracking. Back–tracking is implemented by measuring the drop in output MSE as the number of columns is increased. By studying how each column contributes to the output MSE, it is possible to determine whether this column should be added earlier. The idea of back–tracking is to introduce columns that provide better performance gain before those that provide a lesser performance gain. It can be seen that for back–tracking to provide a noticeable improvement, large drops in the MSE curves must occur. The simulations presented here have shown that, for the ground clutter problem considered, the MSE curves exhibit a smooth nature. This indicates that the set chosen, although maybe not the optimum, are certainly a good set of columns, and that back–tracking will not provide a significant improvement. The only method of finding the optimum set is to use a brute–force search. This is feasible for the example above, but for a practical array finding the best J columns out of a set of $NL - K$ requires the MSE

(equation (4.14)) to be evaluated

$$\frac{(NL-K)!}{J!(NL-K-J)!} \tag{4.45}$$

times. This is clearly impractical, even for small arrays. A sequentially designed beamformer will require only

$$\sum_{j=1}^{J}(NL-K-j) \;=\; \frac{NL-K-J}{2}(2J \;+\; (NL-K-J-1)) \tag{4.46}$$

independent evaluations of the MSE. Applying a brute–force search to the array considered in Figure 4.8 for which $NL = 128$, $K = 8$, $J_s = 43$, requires the MSE to be computed 7.627×10^{32} times, whereas the sequential design described requires only 4257 seperate evaluations! Forward selection algorithms will not necessarily provide the optimum selection of columns, but do represent a fair trade–off between performance and ease of calculation. Table 4.4 compares the beamformers highlighted earlier.

Order of beamformer			**Brute Force**	**Sequential**
NL	K	J_s	MSE computations	MSE computations
16	4	6	924	57
32	4	10	1.312×10^7	235
128	8	43	7.627×10^{32}	4257

Table 4.4: Cost of search techniques.

4.13 A reduced channel simplification

At this point it seems logical to examine the location of the weights which are chosen by the variety of methods. This will provide us with an insight into the distribution of weights and also where possible simplifications may be made. Figures 4.13–4.15 show the locations of weights chosen by the various algorithms described in the previous section. The figures are all for the case of an adaptive dimension of 48. Thus in each picture there are 48 points indicating the weights chosen. The GSC structure with $N = 16$ and $L = 8$ used previously is assumed, giving 15 tapped–delay–lines (channels) with 8 elemental weights in each channel. The reasons for choosing 48 are two–fold. Firstly, at this adaptive dimension the iterative beamformer has attained fully adaptive performance and secondly, 48 is a integer multiple of 8.

Studying the pictures of the weight distributions it becomes apparent that several channels have smaller numbers of weights than others. Take Figure 4.13 as an example. Channels 3, 5, 11, 13, 15 have one or less weights. Since our original aim was to produce a low–complexity beamformer, we might try to remove these channels (or any other whole channels) entirely from

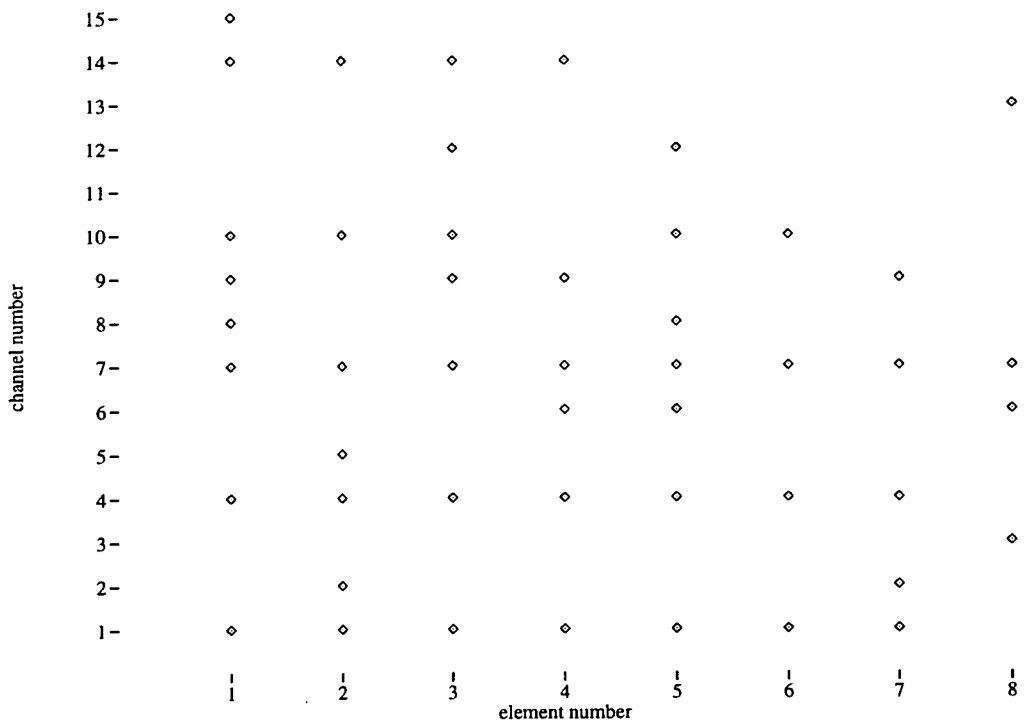

Figure 4.13: Actual elements chosen during training phase - iterative approach.

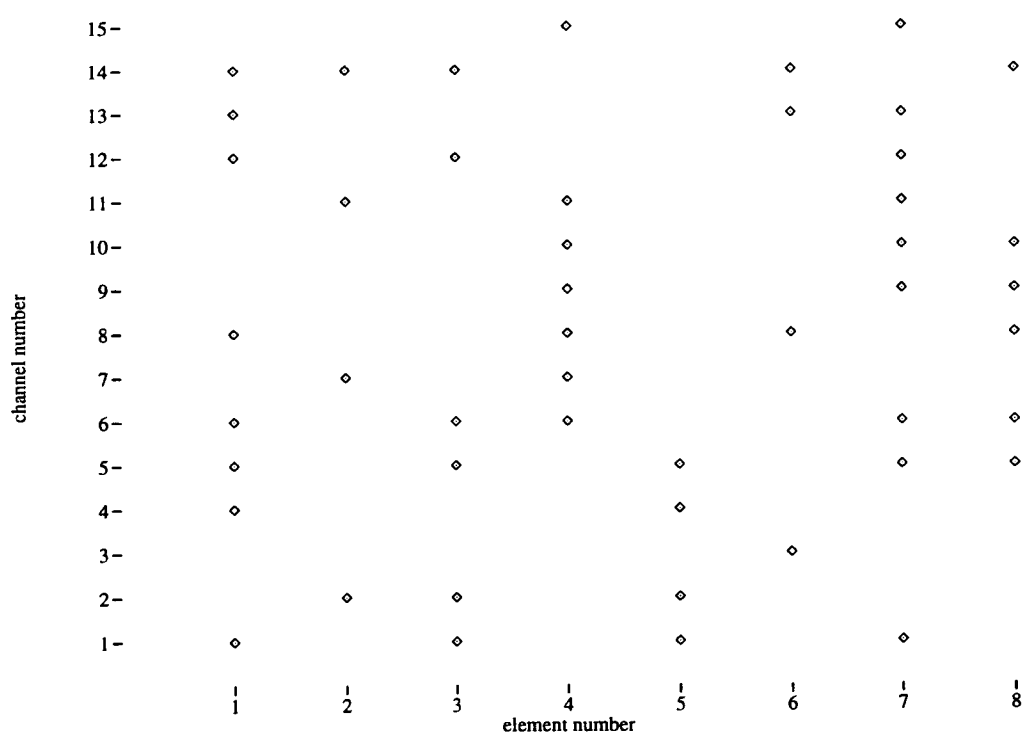

Figure 4.14: Actual elements chosen during training phase - projection method.

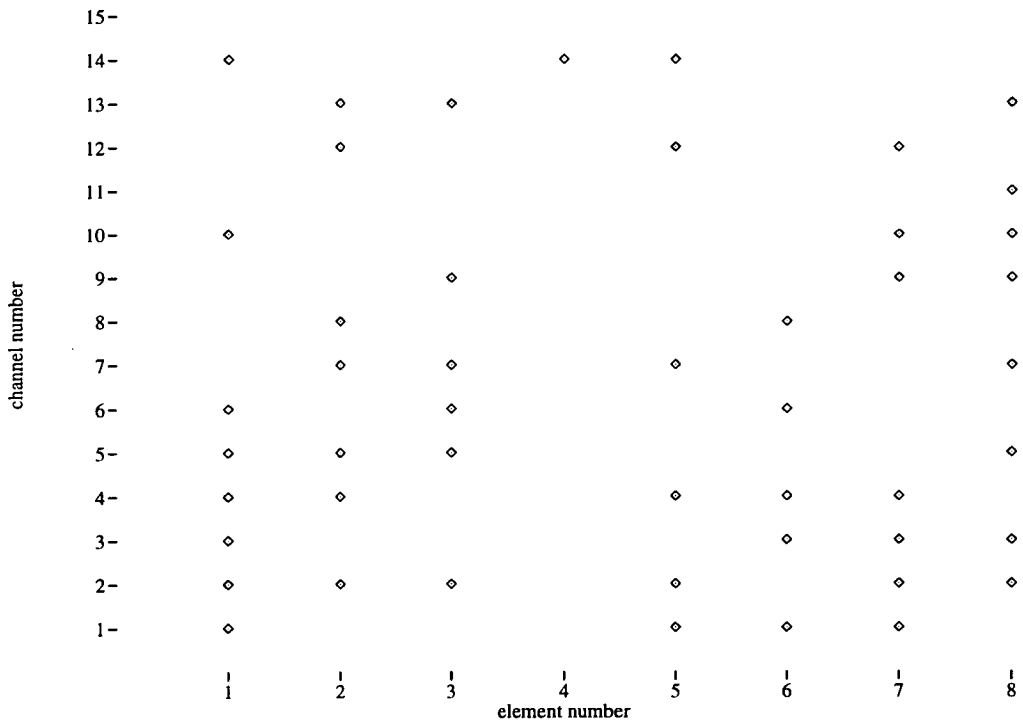

Figure 4.15: Actual elements chosen during training phase - random method.

the optimisation.

As an initial solution we can modify the iterative algorithm so that it adds a whole channel to the processor at each iteration. For the distributions shown in Figures 4.13–4.15, this will be represented by adding sets of 8 weights to the processor. Figure 4.16 depicts the channels chosen by this modified iterative algorithm, for an adaptive dimension of 56 (equivalent to 7 channels). The MSE performance of this modified algorithm is shown in Figure 4.17.

Now consider an improvement to this whole channel approach. Suppose that we had deleted channels 3, 5, 11, 13 and 15, and then allowed the iterative algorithm to run on this reduced set of channels. These deleted channels are those from which the original iterative algorithm selected one or less weights. These deletions mean that at each iteration the weights that may be selected cannot be chosen from any of these deleted channels. Curve *simulation 1* in Figure 4.17 contrasts the performance of an iterative beamformer using this reduced channel approach with that of a beamformer which uses all of the available channels. The first thing to note is that both beamformers have remarkably similar performance, the reduced channel beamformer performs only slightly worse. This slight degradation in performance is to be expected as obviously we have limited the number of columns which are available to the selection algorithm. That said, the degradation is small. This highlights the manner in which the benefits of using an iterative power minimisation technique are accrued. At any particular step in the selection process the improvement gained by choosing the optimum column as opposed to any other will be small. It is only when these small gains are compounded over a series of iterations that a significant benefit is observed. Thus, reducing the allowed set of channels from which we may choose

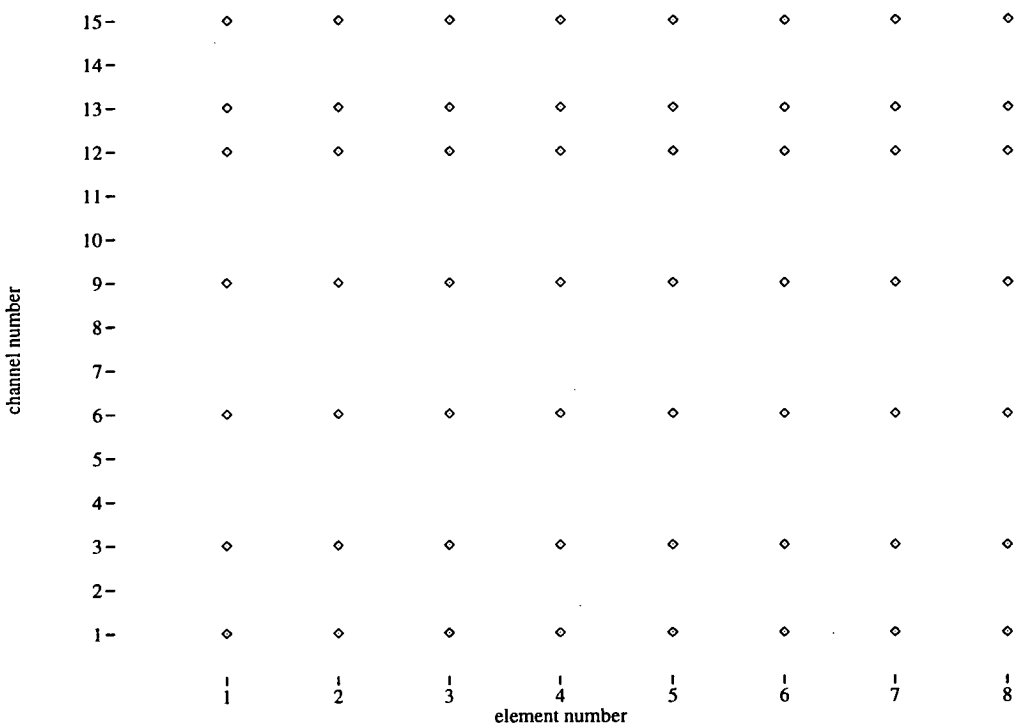

Figure 4.16: Actual elements chosen during training phase - whole channels approach.

degrees of freedom does not necessarily lead to a large loss in cancellation performance. It is interesting to observe that the reduced column technique performs marginally better around 40 degrees of freedom. This again demonstrates the operation of this simple iterative algorithm. At each step only the immediately next column to be added is considered. At no point is any consideration given to the implications this has for further iterations.

Now consider making further deletions. We now delete every channel with two or less weights. In addition to those deleted above, we remove channels 2, 8 and 12 from the optimisation. This will leave a fully adaptive dimension of $120 - 8 \times 8 = 56$. The MSE performance of this beamformer is shown in Figure 4.17, curve *simulation 2*. Figure 4.18 shows the elements which are chosen by the reduced channel processor of simulation 2 for an adaptive dimension of 56. Notably the channels 2, 3, 5, 8, 11, 12, 13, and 15 contain no weights, as was required by the design specification.

At this point we can truly see the advantage of using an iterative algorithm. Consider the two beamformer structures shown in Figures 4.16 and 4.18. Both these beamformers have identical adaptive dimension and implemental complexity, yet the difference in output MSE can be seen in Figure 4.17. The whole channel beamformer has an output MSE of -34.64dB, whilst the beamformer in simulation 2 has an output MSE of -35.81dB. When the dynamic range of the curves is 8dB, a gain of 1dB is significant.

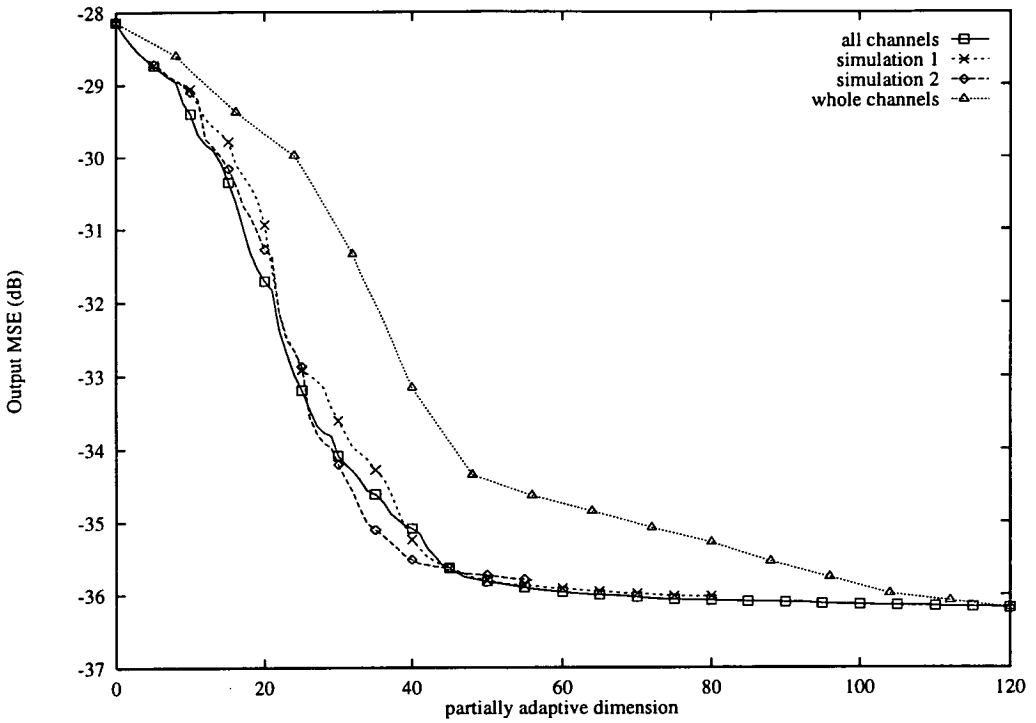

Figure 4.17: *Relative performance of an iteratively designed beamformer with all channels, and a reduced set of channels included in the design.*

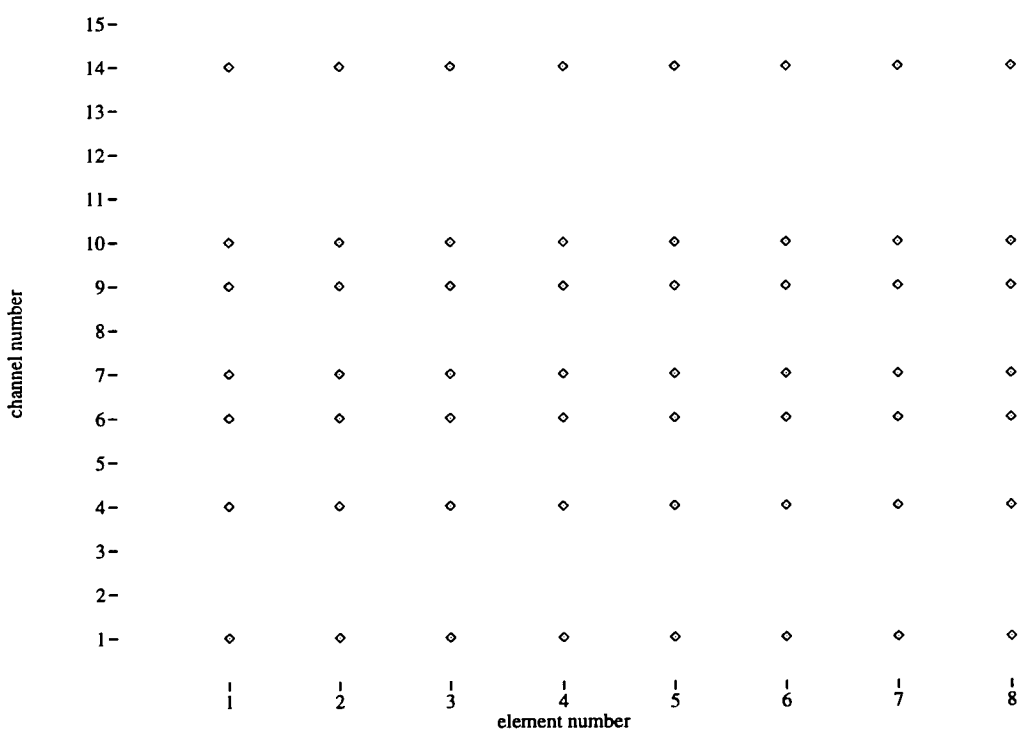

Figure 4.18: *Actual elements chosen during training phase - reduced channels approach. Channels 2, 3, 5, 8, 11, 12, 13, and 15 have been removed from the optimisation.*

4.14 Conclusions

Within this chapter various techniques for designing partially adaptive generalised sidelobe cancellers have been presented. The relative performance of each of these techniques has been established heuristically and by computer simulation. From these analyses it has been shown that iterative design techniques, particularly for complex interference environments, do offer significant performance advantages.

To justify the above conclusions it is necessary to compare the performances of each of the beamformer design techniques described. Considering firstly the eigenstructure approach, it was clear that for this beamformer to operate effectively, knowledge of the likely interference eigenstructure was required, or that it could be estimated accurately. If this is so, then good cancellation performance will be obtained. However, the complex eigenstructure typical of most interference scenarios will mean that this approach can require a prohibitively large partially adaptive dimension. Additionally, the complex nature of the transformation matrix found will lead to a significant computational load.

The projection technique outlined in section 4.6.2 used a low rank estimate of the interference subspace to form a solution which was "near" to the minimum–norm solution. This solution maintains the desirable numerical properties of the minimum–norm solution, but also reduces the number of adaptive weights that must be computed. In this approach, an eigendecomposition is performed from which an estimate of the interference subspace is constructed. Subsequently, coefficients (weights) are selected according to the projections upon the signal subspace of their corresponding axis vectors. The axis subset with the largest projections are chosen, ensuring that the weight vector will lie near to the minimum–norm solution. The low rank approximation of the signal subspace will, as before, require that a good estimate of the interference eigenstructure exists. A poor choice of threshold may radically alter the solution found.

In order to obviate the problems of the eigenstructure design, but to also further improve performance over that of the projection method, an iterative approach was presented. In this adaptive weights were selected in an iterative manner. This technique was demonstrated, through simulation, to attain a significant performance improvement over the conventional techniques. The operation of this algorithm, when viewed pictorially appears similar to that of the projection method. By relaxing the constraint upon finding a solution which is near to the minimum–norm solution, it is possible to achieve still further reductions in required adaptive dimension.

Comparing the MSE and SNR performance when applying adaptive algorithms, it can be seen that a substantial improvement is achieved. Furthermore, although it may appear that techniques such as the eigenstructure approach might better match the interference subspace, in fact it is the simpler structures designed to optimise certain performance measures that attains

the best performance.

Chapter 5

Convergence Performance

5.1 Introduction

The weights derived for the beamformers considered in this thesis are a function of the data covariance matrix $\boldsymbol{R}_x$. Thus far the true covariance has been assumed known, however in practice the data covariance is unknown and will have to be estimated from the available data. The purpose of an adaptive algorithm is to find a set of optimum weights using the tap data $\boldsymbol{x}(k)$ rather than the idealised covariance $\boldsymbol{R}_x$. Since adaptive algorithms determine the weights using the received data, the weights automatically adjust to changes in the environment in order to maintain interference suppression. This chapter considers the transient response of linearly constrained beamformers, and examines the convergence properties of various beamformer parameters.

In many applications the practical usefulness of an adaptive array is limited by it's convergence rate. The adaptively controlled weights must change at a rate equal to or greater than that of the external field. In a radar application this is further complicated by scanning of the array antenna. The convergence rate problem is greatest in systems with a large number of adaptive degrees of freedom. Adaptive algorithms which either directly or indirectly estimate the data covariance matrix are of particular interest to the radar community due to their convergence properties. Common gradient based schemes such a least mean squares have limited applicability because their convergence characteristics are strongly dependent upon the eigenvalue spread of the covariance matrix. A common estimate of $\boldsymbol{R}_x$ is the *sample covariance matrix*. This is formed by averaging the outer product of M data vectors, i.e.

$$\hat{\boldsymbol{R}}_x = \frac{1}{M}\sum_{k=1}^{M}\boldsymbol{x}(k)\boldsymbol{x}^H(k). \tag{5.1}$$

The sample covariance matrix represents the maximum likelihood estimate of $\boldsymbol{R}_x$ given no prior structural constraints [79], and can be used to obtain the maximum likelihood estimate of signals incident upon the array [80]. The adaptive weight vector is estimated by substituting the estimated covariance matrix in place of the true covariance in the expressions given earlier for the optimum weight vectors.

When the adaptive weights are computed via the sample covariance matrix the adapted response of the array can experience very "noisy" sidelobe fluctuations and main beam perturbations

even though the constraints are chosen to ensure a low sidelobe quiescent pattern. The random sidelobe behaviour occurs because finite sampling causes spurious cross-correlations in the covariance matrix, which causes the background noise component of the covariance matrix used in the adaptive weight determination to differ significantly from the asymptotic value. This was demonstrated in simulation results presented in [2, 73, 81]. Large sidelobe levels represent a considerable problem in radar applications, since processing is typical performed in a non–concurrent manner, i.e. the weights are computed from a different set of data from that to which they are applied. High sidelobes can render the adaptive array very vulnerable to sidelobe clutter, sudden changes in the interference environment, or pulsed interference that can benefit from post-processing gain. Therefore, an analysis of the transient sidelobe behaviour of sidelobe cancelling systems is of primary importance.

This chapter will derive expressions for the mean square error and transient sidelobe performance of a generalised sidelobe canceller beamformer, with an arbitrary number of adaptive degrees of freedom. Due to the nature of this beamforming structure, mathematical derivations are often quite complex so that theorems relating to multivariate statistical analysis will need to be used. A selection of these are given in appendix B. The key results presented are expressions for the concurrent and non–concurrent mean square error ((5.21) and (5.33)), and the transient sidelobe response ((5.39) and (5.40)). It is believed that the results relating to the transient sidelobe performance are completely new.

5.2 Transient weight vector

Adopting the notation and assumptions used commonly in the literature, let the columns of the $N \times M$ data matrix $\boldsymbol{X}$ represent M independently identically distributed (i.i.d.) zero–mean complex normally distributed random data vectors impinging at a sensor array which processes a total of N samples. Unless otherwise stated N represents the total number of spatial and temporal samples combined in the array. Assuming the generalised sidelobe canceller realisation of the LCMV beamformer, the beamformer weights are given by

$$\hat{w} \;=\; w_q \;-\; \boldsymbol{T}\left(\boldsymbol{T}^H\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{T}\right)^{-1}\boldsymbol{T}^H\boldsymbol{X}\boldsymbol{X}^H w_q, \tag{5.2}$$

where $\boldsymbol{T}$ denotes the $N \times J$ generalised signal blocking matrix, and J represents the number of adaptive degrees of freedom available to the beamformer. Let the input data matrix $\boldsymbol{X} = \boldsymbol{X}_s + \boldsymbol{X}_n$, where $\boldsymbol{X}_s$ and $\boldsymbol{X}_n$ are mutually uncorrelated signal and noise components, respectively. Assume the constraints are chosen so that the quiescent beamformer passes the signal without distortion $\boldsymbol{X}_s^H w_q = s$, where s denotes the M samples of the desired signal. Also, note that the signal blocking action of $\boldsymbol{T}$ implies that $\boldsymbol{T}^H\boldsymbol{X} = \boldsymbol{T}^H\boldsymbol{X}_n$. The estimated adaptive weight vector $\hat{w}_a$ is given by

$$\hat{w}_a \;=\; \left(\boldsymbol{T}^H\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{T}\right)^{-1}\boldsymbol{T}^H\boldsymbol{X}\boldsymbol{X}^H w_q. \tag{5.3}$$

This estimate of the adaptive weight vector can be decomposed into components arising due to interference alone, and due to the presence of the desired signal, as follows

$$\hat{w}_a = \hat{w}_{an} + \hat{w}_{as}, \tag{5.4}$$

where

$$\begin{aligned} \hat{w}_{an} &= \left(T^H X_n X_n^H T\right)^{-1} T^H X_n X_n^H w_q, \\ \hat{w}_{as} &= \left(T^H X_n X_n^H T\right)^{-1} T^H X_n X_s^H w_q, \end{aligned}$$

in which the fact that $T^H X_s = 0$ has been used. It is easy to show that $E\{\hat{w}_{an}\} = w_a$ and $E\{\hat{w}_{as}\} = 0$. This is done by defining

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} w_q^H X_n \\ T^H X_n \end{bmatrix}, \tag{5.5}$$

and noting that the $\hat{w}_{an}$ may be written in terms of conditional expectations as (theorem B.3)

$$\hat{w}_{an} = E\left\{\left(Z_2 Z_2^H\right)^{-1} Z_2 E\left\{Z_1^H | Z_2\right\}\right\}. \tag{5.6}$$

Note that the columns of Z are i.i.d. zero mean multinormal complex random vectors so that much of the multinormal distribution theory may be applied. Using theorem B.2 [82, Theorem 1.2.11], the conditional expectation $E\{Z_1^H | Z_2\}$ equals

$$E\left\{Z_1^H | Z_2\right\} = Z_2^H \left(T^H R_n T\right)^{-1} T^H R_n w_q, \tag{5.7}$$

where $R_n = M^{-1} E\{X_n X_n^H\}$ is the true noise covariance matrix. Substituting in (5.6) we see that

$$E\{\hat{w}_{an}\} = E\left\{\left(Z_2 Z_2^H\right)^{-1} Z_2 Z_2^H \left(T^H R_n T\right)^{-1} T^H R_n w_q\right\} = w_a, \tag{5.8}$$

which is the desired result. Now considering $\hat{w}_{as}$, note that $E\{\hat{w}_{as}\}$ may be expressed as

$$E\{\hat{w}_{as}\} = E\left\{\left(Z_2 Z_2^H\right)^{-1} Z_2 s\right\}, \tag{5.9}$$

which can be expressed in terms of conditional expectations as

$$E\{\hat{w}_{as}\} = E\left\{\left(Z_2 Z_2^H\right)^{-1} Z_2 E\{s | Z_2\}\right\}. \tag{5.10}$$

Since it is assumed that the signal blocking path is orthogonal to the constraints which passed the desired s then $E\{s|Z_2\} = 0$, and hence $E\{\hat{w}_{as}\} = 0$. Thus $E\{\hat{w}_a\} = w_a$, and we can conclude that the adaptive weights computed from the sample covariance matrix yield unbiased estimates of the steady–state weights. We conclude this section by defining the error in the adaptive weight vector as

$$\Delta w_a = \hat{w}_a - w_a, \tag{5.11}$$

and observe that Δw_a is a zero mean vector, and will consist of components arising from both the interference and the desired signal.

5.3 Transient mean square error

In an interesting recent correspondence by Van Veen [34], the expected output power and mean square error of the linearly constrained minimum variance beamformer were derived for an arbitrary number of adaptive degrees of freedom. The analysis in [34], which extended previous work of [4,79,83,84], was based upon the use of the sample covariance matrix as a covariance estimate and where the same input data was used to compute both the adaptive weights and the beamformer output. This mode of operation, which is referred to as block–mode or concurrent processing, suffers from the disadvantage that the mean output power is less than the corresponding "infinite–time" beamformer (i.e. one which employs the true covariance matrix rather than a finite–time estimate). The reduction in output power occurs because a portion of the desired signal is cancelled by the adaptive weights. This signal cancellation effect can be traced to the correlation which exists between the adaptive beamformer weights and the data to which they are applied. In a more recent communication [85], this cancellation phenomenon was examined more closely and it was found that the output signal component under finite–time conditions was, in fact, a biased version of the input signal, scaled by a factor less than or equal to one.

A disadvantage of concurrent processing is that the output data is available only after a delay corresponding to the data block length. An alternative implementation that relates to recursive processing schemes is to apply weights computed using previous data blocks to the current data. This mode of operation is commonly referred to as non-concurrent processing. In this mode the adaptive weights are uncorrelated from the input data, and hence signal cancellation is avoided. [85] considered this mode of beamformer operation, and showed interestingly that the concurrent beamformer could be made to perform identically to the non–concurrent beamformer if the concurrent output was corrected for the signal estimation bias. However, the earlier comments relating to high sidelobes and vulnerability to changing interference conditions apply particularly in the non–concurrent mode of operation. Because the weights are applied to data which was not used in their computation, the beamformer will be particularly susceptible to changing interference conditions.

5.3.1 Concurrent operation

Consider first the concurrent mode of beamformer operation. In this mode the adaptive weights are computed from the same block of data to that which they are applied. The sample covariance estimate will be given by $M^{-1}\boldsymbol{X}\boldsymbol{X}^H$, and the beamformer output is given by the $M \times 1$ vector

$y = \boldsymbol{X}^H w$. The sample mean square error $\hat{e}$ is defined as

$$\begin{aligned} \hat{e} &= \frac{1}{M} \left| s^H - y^H \right|^2 \\ &= \frac{1}{M} \left| s^H - w^H \boldsymbol{X} \right|^2 \\ &= \frac{1}{M} w^H \boldsymbol{X}_n \boldsymbol{X}_n^H w. \end{aligned} \tag{5.12}$$

Substituting (5.2) into (5.12) and simplifying, we obtain expressions for the mean square error due to the noise component alone $\hat{e}_n$, and the additional mean square error due to the presence of the desired signal $\hat{e}_s$ as

$$\hat{e}_n = \frac{1}{M} \left\{ w_q^H \boldsymbol{X}_n \boldsymbol{X}_n^H w_q - w_q^H \boldsymbol{X}_n \boldsymbol{X}_n^H \boldsymbol{T} \left(\boldsymbol{T}^H \boldsymbol{X}_n \boldsymbol{X}_n^H \boldsymbol{T} \right)^{-1} \boldsymbol{T}^H \boldsymbol{X}_n \boldsymbol{X}_n^H w_q \right\}, \tag{5.13}$$

$$\hat{e}_s = \frac{1}{M} \left\{ s^H \boldsymbol{X}_n^H \boldsymbol{T} \left(\boldsymbol{T}^H \boldsymbol{X}_n \boldsymbol{X}_n^H \boldsymbol{T} \right)^{-1} \boldsymbol{T}^H \boldsymbol{X}_n \, s \right\}. \tag{5.14}$$

At this point we make use of some basic multivariate statistical theory. Appendix B summarises most of the important theorems. The columns of $\boldsymbol{X}$ and $\boldsymbol{X}_n$ are assumed to be independent and identically distributed drawn from $N_N(\boldsymbol{0}, \boldsymbol{R}_x)$, and $N_N(\boldsymbol{0}, \boldsymbol{R}_n)$. Under these assumptions $\boldsymbol{X}\boldsymbol{X}^H$ and $\boldsymbol{X}_n \boldsymbol{X}_n^H$ are complex Wishart distributed ([79], and definition B.2), with distributions denoted by $W_N(M, \boldsymbol{R}_x)$ and $W_N(M, \boldsymbol{R}_n)$. Consider $\hat{e}_n$. Define the $J+1$ by $J+1$ partitioned matrices

$$\begin{aligned} \boldsymbol{A} &= [w_q \;\; \boldsymbol{T}]^H \boldsymbol{R}_n [w_q \;\; \boldsymbol{T}] \\ \hat{\boldsymbol{A}} &= [w_q \;\; \boldsymbol{T}]^H \boldsymbol{X}_n \boldsymbol{X}_n^H [w_q \;\; \boldsymbol{T}] \\ &= \begin{bmatrix} w_q^H \boldsymbol{X}_n \boldsymbol{X}_n^H w_q & w_q^H \boldsymbol{X}_n \boldsymbol{X}_n^H \boldsymbol{T} \\ \boldsymbol{T}^H \boldsymbol{X}_n \boldsymbol{X}_n^H w_q & \boldsymbol{T}^H \boldsymbol{X}_n \boldsymbol{X}_n^H \boldsymbol{T} \end{bmatrix}. \end{aligned} \tag{5.15}$$

Applying the identity for the inverse of a partitioned matrix, the first element in the first row of $\hat{\boldsymbol{A}}^{-1}$ is given by

$$\left[\hat{\boldsymbol{A}}^{-1} \right]_{1,1} = \left[w_q^H \boldsymbol{X}_n \boldsymbol{X}_n^H w_q - w_q^H \boldsymbol{X}_n \boldsymbol{X}_n^H \boldsymbol{T} \left(\boldsymbol{T}^H \boldsymbol{X}_n \boldsymbol{X}_n^H \boldsymbol{T} \right)^{-1} \boldsymbol{T}^H \boldsymbol{X}_n \boldsymbol{X}_n^H w_q \right]^{-1} \tag{5.16}$$

Studying (5.16) and (5.12), we see that the sample MSE may be written as

$$\begin{aligned} \hat{e}_n &= \frac{1}{M} \left[\left[\hat{\boldsymbol{A}}^{-1} \right]_{1,1} \right]^{-1} \\ &= \frac{1}{M} \left[u_1^H \hat{\boldsymbol{A}}^{-1} u_1 \right]^{-1}, \end{aligned} \tag{5.17}$$

where $u_1 = [1 \; 0 \; 0 \; \cdots \; 0]^H$. Application of theorem B.6 tells us that $\hat{\boldsymbol{A}}$ is complex Wishart distributed, with distribution denoted by $\hat{\boldsymbol{A}} \sim W_{NL}(J+1, \boldsymbol{A})$. Now applying theorem B.8, we see that $M\hat{e}_n$ is distributed as $W_1\left(M - J, \left[u_1^H \boldsymbol{A}^{-1} u_1\right]^{-1}\right)$. Note, however, that $\left[u_1^H \boldsymbol{A}^{-1} u_1\right]^{-1}$ is the steady state mean square error, i.e. the MSE with an infinite number of snapshots (see equation (4.14)). Now, $M\hat{e}_n$ is one half a chi squared random variable with $(M - J)$ complex degrees of freedom, i.e. it has $2(M - J)$ real degrees of freedom. Thus, the mean value of $\hat{e}_n$ is

given by [34]

$$E\{\hat{e}_n\} = \frac{M-J}{M} e_n. \tag{5.18}$$

The ratio $(M-J)/M$ determines the adaptive convergence of the mean when viewed as a function of the number of snapshots M. Equation (5.18) shows that the expected value of the excess MSE will rise towards the steady–state MSE with increasing M and will be within 3dB of the optimum after $M = 2J$ data vectors.

Now, consider the distribution of $\hat{e}_s$. The conditional distribution of $\hat{e}_s$ given s is obtained by defining $\boldsymbol{T}^H \boldsymbol{X}_n = \boldsymbol{V}$ as follows

$$\begin{aligned} \hat{e}_s &= \frac{1}{M}\left\{ s^H \boldsymbol{V}^H \left(\boldsymbol{V}\boldsymbol{V}^H\right)^{-1} \boldsymbol{V}\, s \right\} \\ &= \frac{a^2}{M}\rho, \end{aligned} \tag{5.19}$$

where a is a real variable which defines the magnitude of the desired signal, i.e.. $s = a s_0$, where $s_0^H s_0 = 1$. By forming a Cholesky factorisation of $\boldsymbol{V}\boldsymbol{V}^H$, Van Veen [34] showed that ρ is a beta distributed random variable, independent of s, with mean J. Thus the mean value of $\hat{e}_s$ is

$$\begin{aligned} E\{\hat{e}_s\} &= M^{-1} E\{a^2\}\, E\{\rho\} \\ &= \sigma_s^2 \frac{J}{M}, \end{aligned} \tag{5.20}$$

where σ_s^2 is the variance or power of the desired signal. Equation (5.20) shows that the average MSE associated with the signal presence is directly proportional to the signal power and the number of adaptive degrees of freedom, and inversely proportional to the number of data vectors. Thus, it can be seen that the infinite–time MSE associated with the desired signal will be zero, and that the presence of a strong desired signal can be expected to result in large transient MSE. Recalling $\hat{e} = \hat{e}_n + \hat{e}_s$, the total sample MSE for the concurrent beamformer configuration is given by

$$\mathrm{MSE}_c = \left(1 - \frac{J}{M}\right) P_n + \left(\frac{J}{M}\right) P_s, \tag{5.21}$$

where P_n and P_s are the steady–state interference and signal powers, respectively, and the subscript c indicates that the beamformer operates in a concurrent manner. As the number of data vectors increases equation (5.21) indicates that the sample MSE will tend towards the steady–state MSE defined by equation (4.14).

Figures 5.1 and 5.2 depict the MSE due to the noise component and the excess MSE arising due to the presence of a strong desired signal. A narrowband GSC beamformer with a fully adaptive dimension of 14 was employed. The quiescent response was designed to match a -30dB Chebychev pattern using the technique outlined in [2]. A single narrowband interferer and uncorrelated noise were included in the simulation. The desired signal and the interferer were statistically independent, and the output MSE was averaged over 100 Monte Carlo simulations

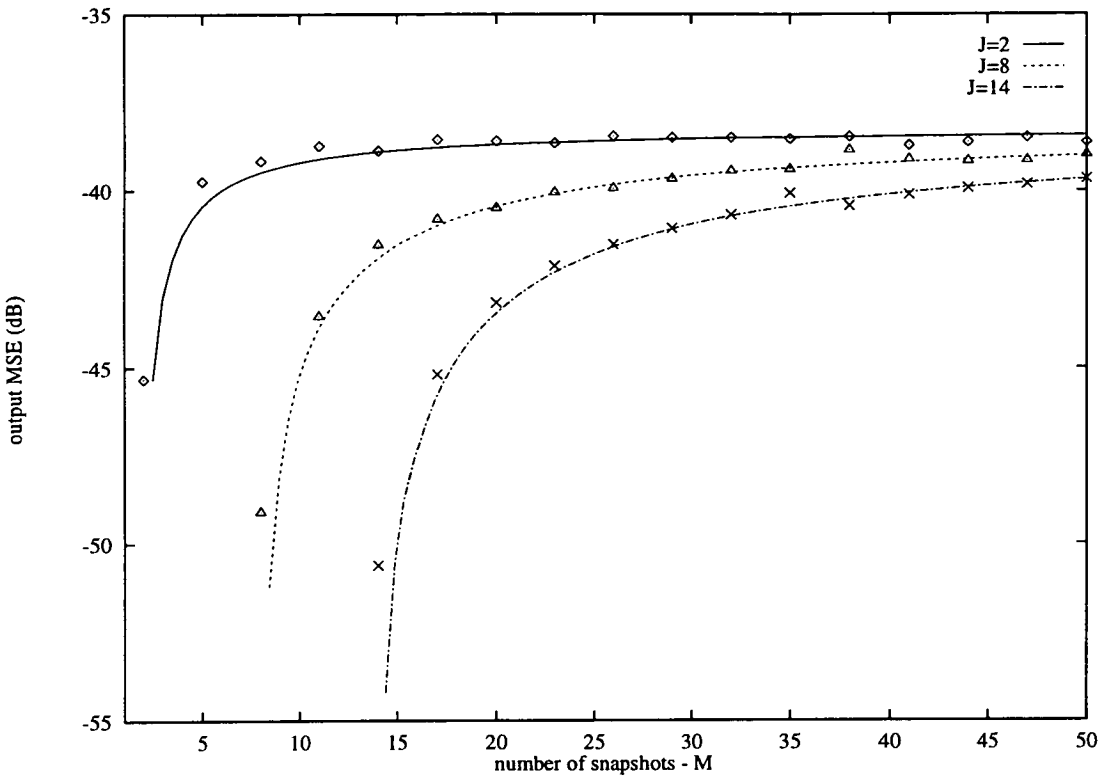

Figure 5.1: *Output sample mean square error due to noise and interference alone versus data matrix size - concurrent processing. Each point was computed from 100 Monte Carlo simulations. The curves indicate the theoretical values.*

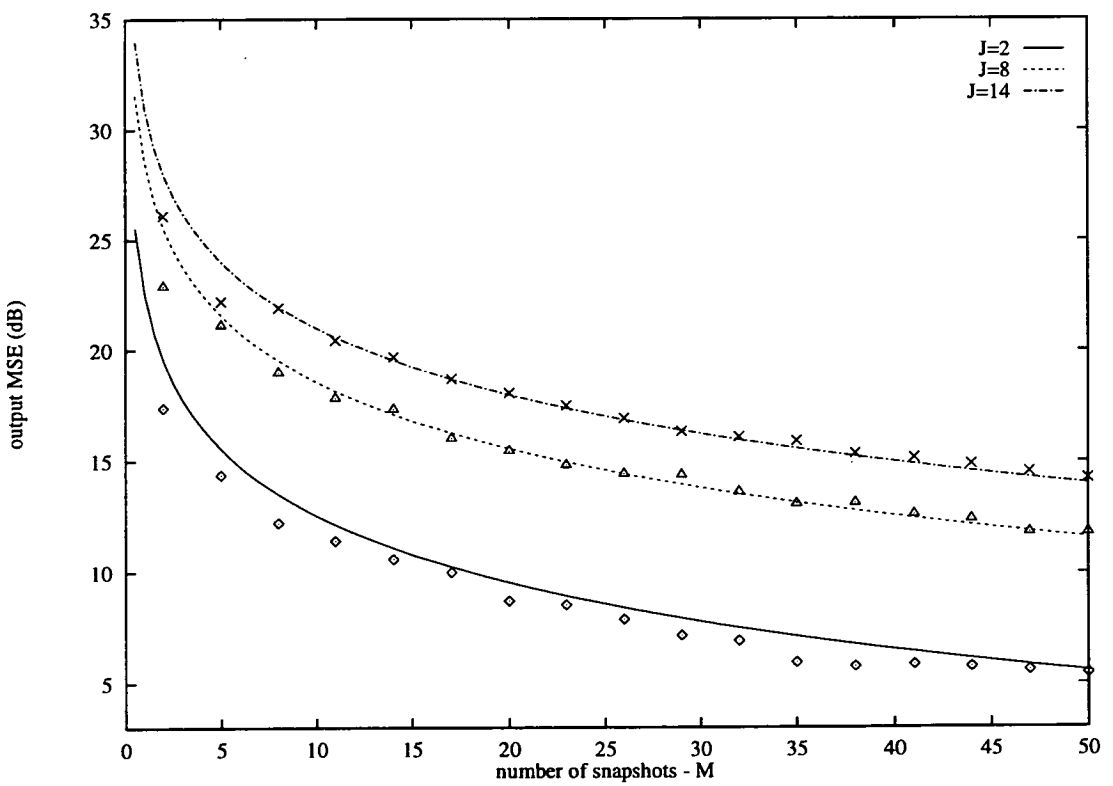

Figure 5.2: *Output sample mean square error due to presence of a 20dB desired signal versus data matrix size - concurrent processing. Each point was computed from 100 Monte Carlo simulations. The curves indicate the theoretical values.*

for the three adaptive dimensions $J = 2, 8$ and 14. Theoretical curves have been derived using (5.18) and (5.20), and as can be seen the simulation results follow the theoretical curves closely.

5.3.2 Non–concurrent operation

In the non–concurrent LCMV beamformer [85], weights computed from a previous block of data are applied to the current input vector. Mathematically this differs from concurrent processors in that now the input data may be assumed uncorrelated from the adaptive weights. Letting $\boldsymbol{u} = \boldsymbol{u}_s + \boldsymbol{u}_n$ denote the current input vector signal and noise components respectively, the corresponding beamformer output sample is $v = \boldsymbol{u}^H \boldsymbol{w}$, where $\boldsymbol{w}$ is given by (5.2). Note that $\boldsymbol{u}$ is assumed identically distributed, but uncorrelated with the columns of $\boldsymbol{X}$, and hence the weight vector $\boldsymbol{w}$. Having made this assumption, both $\boldsymbol{X}$ and $\boldsymbol{u}$ have the same covariance matrix, i.e.

$$\boldsymbol{R}_x = E\{\boldsymbol{u}\boldsymbol{u}^H\} = M^{-1} E\{\boldsymbol{X}\boldsymbol{X}^H\}. \tag{5.22}$$

Now, since there is no coupling of the current input signal through the adaptive weights, the mean (or infinite–time) adaptive weight vector may be written as $\boldsymbol{w}_0 = E\{\boldsymbol{w}\}$. The mean square error of the non–concurrent beamformer can be derived by writing $\text{MSE} = E\left\{|v - s|^2\right\}$, and expanding as

$$\text{MSE} = E\left\{|v|^2\right\} + E\left\{|s|^2\right\} - 2Re\{E\{s^* v\}\}, \tag{5.23}$$

where $Re\{x\}$ denotes the real part of x and superscript $*$ indicates complex conjugate. Applying theorems B.3 and B.4, we can write

$$\begin{aligned} E\{s^* v\} &= E\{E\{s^* v | \boldsymbol{u}_s\}\} \\ &= E\{s^* E\{v|\boldsymbol{u}_s\}\} \\ &= E\{|s|^2\} = P_s. \end{aligned} \tag{5.24}$$

In [85] it was stated that $E\{|v|^2\}$ could be written as

$$E\{|v|^2\} = \boldsymbol{w}_0^H \boldsymbol{R}_x \boldsymbol{w}_0 + tr(\boldsymbol{R}_x \boldsymbol{\Gamma}), \tag{5.25}$$

where $\boldsymbol{\Gamma}$ is the covariance matrix associated with $\boldsymbol{w}$, i.e. $\text{cov}\{\boldsymbol{w}\}$, and $tr()$ denote the matrix trace operation. Using these manipulations, the non–concurrent mean square error can be written as

$$\text{MSE} = \boldsymbol{w}_0^H \boldsymbol{R}_x \boldsymbol{w}_0 + tr(\boldsymbol{R}_x \boldsymbol{\Gamma}) - P_s. \tag{5.26}$$

Now, recalling that $\boldsymbol{T}$ is orthogonal to $\boldsymbol{w}_q$, and that $\boldsymbol{w}_q$ is deterministic, then $\boldsymbol{\Gamma}$ is given by

$$\boldsymbol{\Gamma} = \boldsymbol{T}\boldsymbol{\Upsilon}\boldsymbol{T}^H, \tag{5.27}$$

where $\Upsilon = \text{cov}\left\{\left(\boldsymbol{T}^H \boldsymbol{X}\boldsymbol{X}^H \boldsymbol{T}\right)^{-1} \boldsymbol{T}^H \boldsymbol{X}\boldsymbol{X}^H w_q\right\} = \text{cov}\left\{\left(\boldsymbol{Z}_2\boldsymbol{Z}_2^H\right)^{-1} \boldsymbol{Z}_2\boldsymbol{Z}_1^H\right\}$. Now applying theorem B.3, Υ can be written as

$$\begin{aligned}\Upsilon &= E\left\{\text{cov}\left\{\left(\boldsymbol{Z}_2\boldsymbol{Z}_2^H\right)^{-1} \boldsymbol{Z}_2\boldsymbol{Z}_1^H | \boldsymbol{Z}_2\boldsymbol{Z}_2^H\right\}\right\} \\ &= E\left\{\left(\boldsymbol{Z}_2\boldsymbol{Z}_2^H\right)^{-1} \text{cov}\left\{\boldsymbol{Z}_2\boldsymbol{Z}_1^H | \boldsymbol{Z}_2\boldsymbol{Z}_2^H\right\}\left(\boldsymbol{Z}_2\boldsymbol{Z}_2^H\right)^{-1}\right\}. \end{aligned} \tag{5.28}$$

Application of [82, theorem 3.2.10], shows that $\text{cov}\left\{\boldsymbol{Z}_2\boldsymbol{Z}_1^H | \boldsymbol{Z}_2\boldsymbol{Z}_2^H\right\} = P \cdot \left(\boldsymbol{Z}_2\boldsymbol{Z}_2^H\right)$, which yields

$$\Upsilon = P \cdot E\left\{\left(\boldsymbol{Z}_2\boldsymbol{Z}_2^H\right)^{-1}\right\}, \tag{5.29}$$

where P is the total power output. Further, application of theorem B.3 shows that $E\left\{\left(\boldsymbol{Z}_2\boldsymbol{Z}_2^H\right)^{-1}\right\} = \frac{1}{(M-J)}\left(\boldsymbol{T}^H \boldsymbol{R}_x \boldsymbol{T}\right)^{-1}$, and hence

$$\Upsilon = \frac{P}{(M-J)}\left(\boldsymbol{T}^H \boldsymbol{R}_x \boldsymbol{T}\right)^{-1}. \tag{5.30}$$

Substitution in the expression for Γ yields

$$\begin{aligned}\Gamma &= \frac{P}{(M-J)}\, tr\left(\left(\boldsymbol{T}^H \boldsymbol{R}_x \boldsymbol{T}\right)^{-1}\left(\boldsymbol{T}^H \boldsymbol{R}_x \boldsymbol{T}\right)\right) \\ &= \frac{J\,P}{(M-J)}, \end{aligned} \tag{5.31}$$

which allows the mean square error to be expressed as

$$\text{MSE} = w_0^H \boldsymbol{R}_x w_0 + \frac{J\,P}{(M-J)} - P_s. \tag{5.32}$$

Expanding for $w_0 = w_q + \boldsymbol{T}\left(\boldsymbol{T}^H \boldsymbol{R}_x \boldsymbol{T}\right)^{-1} \boldsymbol{T}^H \boldsymbol{R}_x w_q$, and writing $P = P_s + P_n$, yields the non–concurrent mean square error, after some manipulation, as

$$\text{MSE}_n = \left(1 - \frac{J}{M}\right)^{-1} P_n + \left(\frac{J}{M}\right)\left(1 - \frac{J}{M}\right)^{-1} P_s, \tag{5.33}$$

where now the subscript n indicates that non–concurrent processing is employed. This compares with the expression for the concurrent processor (5.21). Again, the presence of a strong desired signal will lead to a high transient MSE. Figures 5.3 and 5.4 depict the MSE due to the noise component and the excess MSE arising due to the presence of a 20dB desired signal (relative to jamming) for the non–concurrent mode of beamformer operation. The same beamformer was used as that in the concurrent mode of operation. The results of the Monte Carlo simulations generally follow the theoretical analysis.

5.4 Transient response

The previous sections have considered the transient response of the generalised sidelobe canceller in terms of the output mean square error. However, another important aspect of the transient behaviour is the sidelobe response of the adapted pattern. High sidelobe levels present a serious limitation to an airborne radar when the beamformer operates in a non–concurrent mode. The remainder of this chapter will examine the adapted response of the generalised sidelobe canceller

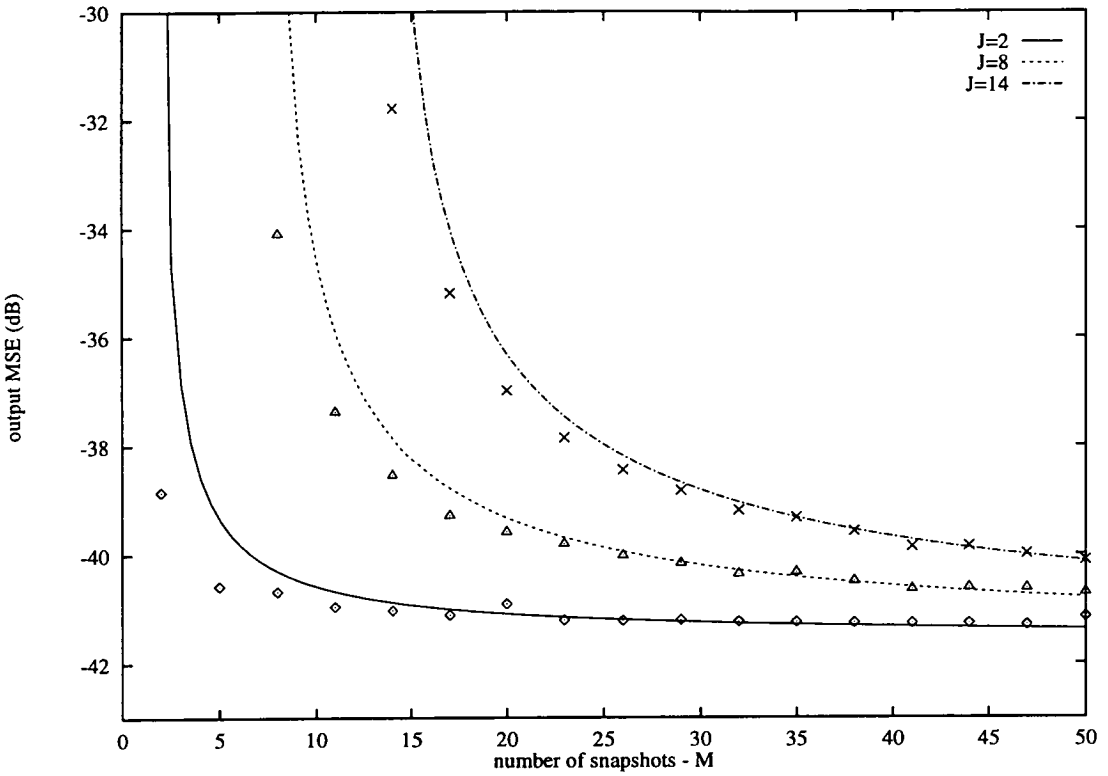

Figure 5.3: *Output sample mean square error due to noise and interference alone versus data matrix size - non-concurrent processing. Each point was computed from 100 Monte Carlo simulations. The curves indicate the theoretical values.*

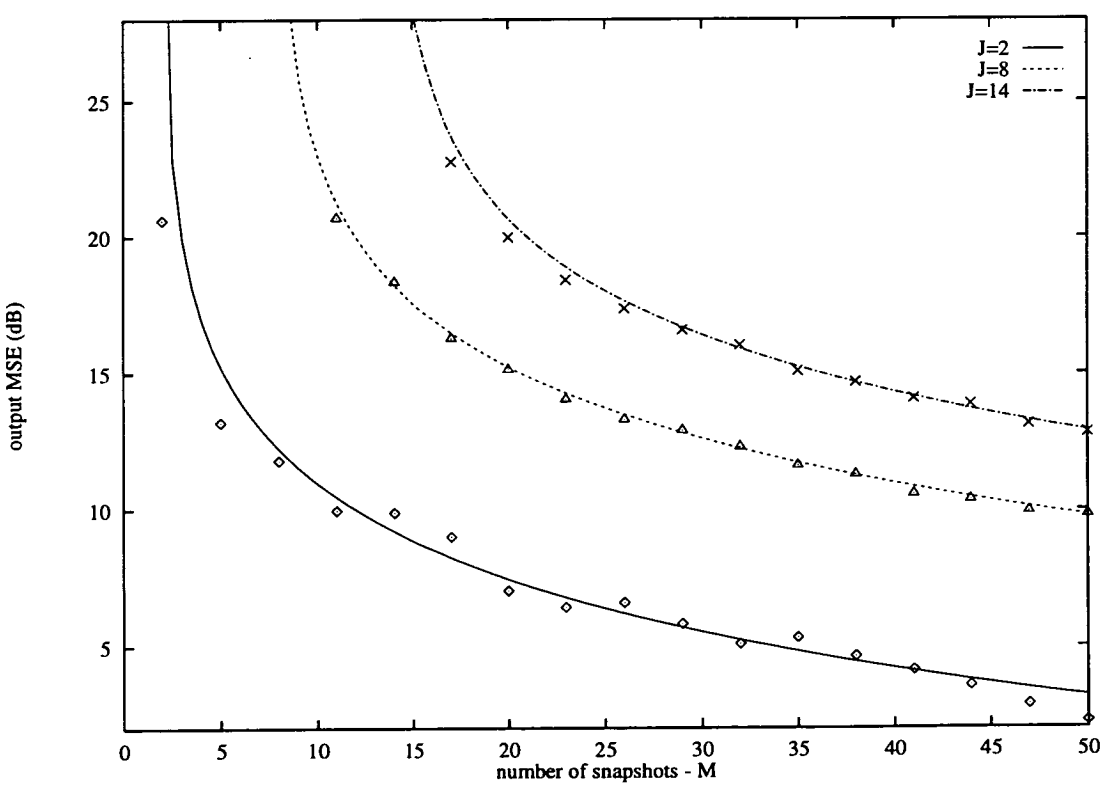

Figure 5.4: *Output sample mean square error when a 20dB desired signal is present versus data matrix size - non-concurrent processing. Each point was computed from 100 Monte Carlo simulations. The curves indicate the theoretical values.*

as a function of the number of snapshots used in estimating the sample covariance matrix, and also as a function of the number of adaptive degrees of freedom present in the beamformer.

Figures 5.5 and 5.6 demonstrate the high sidelobe levels which can be expected due to the sensitivity of the adaptive weights to fluctuations in the background noise field. In all examples a 16 element linear array has been used. The quiescent response of a generalised sidelobe canceller implementation was designed to match a -30dB Chebychev weighting using a technique outlined in [2], which leads to a fully adaptive dimension of 14. The beam was steered to 18°

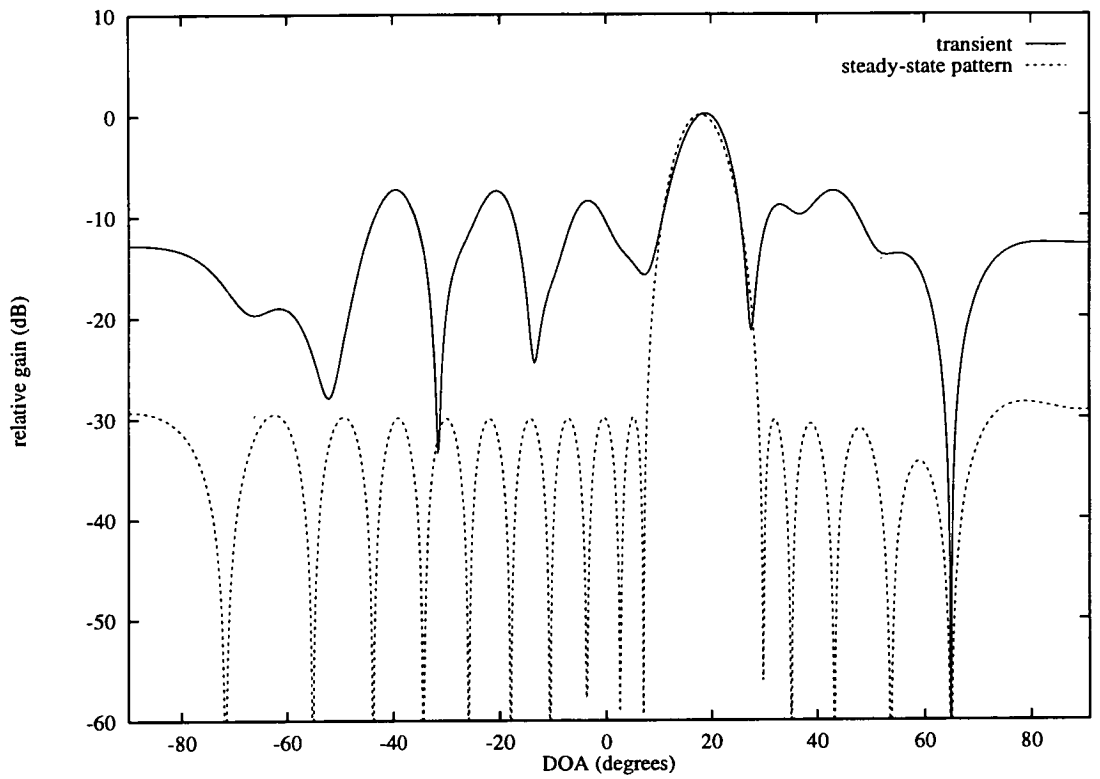

Figure 5.5: *Transient response of a narrowband GSC beamformer after 32 snapshots. The quiescent response was designed to match a -30dB Chebychev weighting using the technique outlined in [2]. A single jamming source is incident from 65°.*

from broadside, and a simulated interference source was placed at 65° with a power level of 30dB relative to the background noise level. Figures 5.5 and 5.6 show adapted beampatterns when 32 ($2N$) and 1024 ($64N$) samples are used, respectively. Virtually perfect nulling is achieved in both cases, with little main beam distortion. However, the average sidelobe levels for the two cases differ considerably (-15.1 and -30.1, respectively). This level of sidelobe distortion may be considered unacceptably high in many applications even though the beam shape is optimum from a signal-to-noise standpoint. An expression for the deviation in the adapted pattern can be derived by writing the adaptive weight vector in terms of the eigendecomposition of the estimated covariance matrix as follows. We begin by recalling the expression for the asymptotic value of the adaptive weight vector (i.e. that computed when $\hat{\boldsymbol{R}}_x = \boldsymbol{R}_x$)

$$\boldsymbol{w}_a = \left(\boldsymbol{T}^H \boldsymbol{R}_x \boldsymbol{T}\right)^{-1} \boldsymbol{T}^H \boldsymbol{R}_x \boldsymbol{w}_q. \tag{5.34}$$

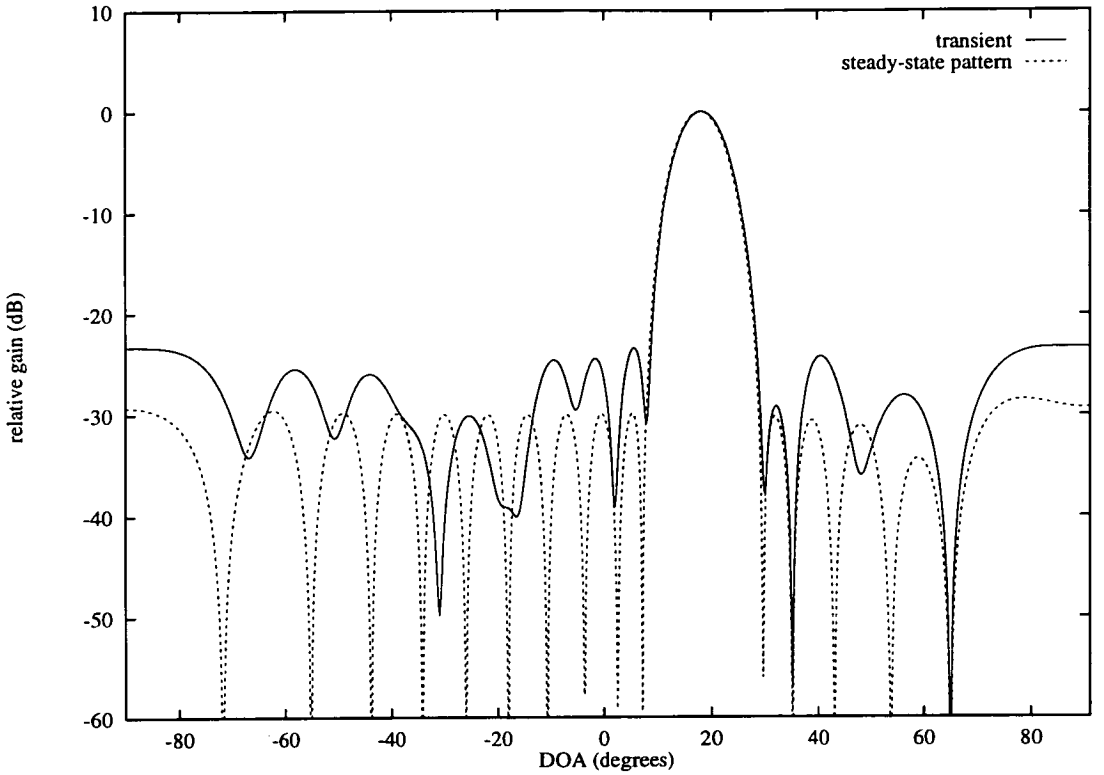

Figure 5.6: *Transient response of a narrowband GSC beamformer after 1024 snapshots. The array is the same as that it Figure 5.5.*

Suppose the interference is characterised by P narrowband interferers and uncorrelated noise, then we may represent the true covariance matrix as the sum of a rank P term and an uncorrelated term corresponding to the background noise

$$\boldsymbol{R}_x = \boldsymbol{E}\boldsymbol{S}\boldsymbol{E}^H + \sigma^2\boldsymbol{I}, \tag{5.35}$$

where $\boldsymbol{E}$ is an N by P matrix of eigenvectors and $\boldsymbol{S}$ is a diagonal matrix containing the P eigenvalues. Substituting (5.35) into (5.34) yields an expression for w_a in terms of the eigenvectors $\boldsymbol{E}$. Beginning with the term in the inverse

$$\boldsymbol{T}^H\boldsymbol{R}_x\boldsymbol{T} = \tilde{\boldsymbol{E}}\boldsymbol{S}\tilde{\boldsymbol{E}}^H + \sigma^2\boldsymbol{I}, \tag{5.36}$$

where $\tilde{\boldsymbol{E}} = \boldsymbol{T}^H\boldsymbol{E}$ and we have assumed $\boldsymbol{T}^H\boldsymbol{T} = \boldsymbol{I}$. This does not cause a loss of generality since it is the space which $\boldsymbol{T}$ spans which is of interest, not the individual elements. The inverse of (5.36) can be expanded as [48]

$$\left(\boldsymbol{T}^H\boldsymbol{R}_x\boldsymbol{T}\right)^{-1} = \sigma^{-2}\boldsymbol{I} - \sigma^{-4}\tilde{\boldsymbol{E}}\left(\sigma^{-2}\tilde{\boldsymbol{E}}^H\tilde{\boldsymbol{E}} + \boldsymbol{S}^{-1}\right)^{-1}\tilde{\boldsymbol{E}}^H. \tag{5.37}$$

$\boldsymbol{T}^H\boldsymbol{R}_x\boldsymbol{w}_q$ becomes

$$\begin{aligned}\boldsymbol{T}^H\boldsymbol{R}_x\boldsymbol{w}_q &= \tilde{\boldsymbol{E}}\boldsymbol{G} + \sigma^2\boldsymbol{T}^H\boldsymbol{w}_q \\ &= \tilde{\boldsymbol{E}}\boldsymbol{G},\end{aligned} \tag{5.38}$$

where $\boldsymbol{G} = \boldsymbol{S}\boldsymbol{E}^H \boldsymbol{w}_q$, and $\boldsymbol{T}^H \boldsymbol{w}_q = \boldsymbol{0}$. Combining (5.37) and (5.38) yields

$$\boldsymbol{w}_a = \boldsymbol{T}^H \boldsymbol{E}\, \boldsymbol{H}, \tag{5.39}$$

with the P dimensional vector $\boldsymbol{H}$ as

$$\boldsymbol{H} = \left[\sigma^{-2}\boldsymbol{I} - \sigma^{-4}\tilde{\boldsymbol{E}}\left(\sigma^{-2}\tilde{\boldsymbol{E}}^H\tilde{\boldsymbol{E}} + \boldsymbol{S}^{-1}\right)^{-1}\tilde{\boldsymbol{E}}^H\tilde{\boldsymbol{E}}\right]\boldsymbol{G}. \tag{5.40}$$

Equation (5.39) gives the desired result. The adaptive weight vector is formed from the weighted sum of the P eigenvectors. The weightings applied to the eigenvectors are given by the individual elements of $\boldsymbol{H}$, which in turn depend upon the eigenvalues $\boldsymbol{S}$. If only uncorrelated interference is incident upon the array, then $\boldsymbol{S}$ is identically zero, and hence from (5.40), $\boldsymbol{w}_a$ is zero. Recall that the overall weight vector for the array is $\boldsymbol{w} = \boldsymbol{w}_q - \boldsymbol{T}\boldsymbol{w}_a$, so that when P sources are incident, the adapted pattern is formed as a quiescent pattern minus the weighted sum of the eigenbeams corresponding to the P sources. An eigenbeam is taken to mean the array response when the individual elements of any particular eigenvector are used as the beamformer weights. This is a clear expression of the fundamental principle of pattern subtraction which applies in adaptive array analysis. The reader is referred to [11] for a more extensive discussion.

If the sample covariance matrix is used in place of the true covariance in (5.34) then the adaptive weight vector is given by $\hat{\boldsymbol{w}}_a = \boldsymbol{T}^H \hat{\boldsymbol{E}}\, \hat{\boldsymbol{H}}$, where the hats indicated estimated values. When a finite number of samples are used in estimating the covariance the eigenstructure is perturbed so that the correlated portion of $\hat{\boldsymbol{R}}_x$ increases in dimension from P to N. Associated with the $N-P$ additional eigenvalues are $N-P$ "noise" eigenvectors; the term noise eigenvector means those eigenvectors which correspond to the small eigenvalues generated by the background noise contained in the finite $\hat{\boldsymbol{R}}_x$ estimate. The adaptive weight vector is formed from a combination of the P eigenvectors relating to the interference, and the $N-P$ noise eigenvectors. The eigenvectors associated with the interferers are generally rather robust [73, 81] and tend to remain relatively stable from one estimate to the next, whereas the noise eigenvectors tend to fluctuate considerably because of the inherent random behaviour of noise. Thus, we expect that the sidelobe undulations of Figure 5.5 are associated primarily with the noise eigenvectors. Figures 5.7 and 5.8 show the noise eigenvalue spread (the ratio of the biggest to smallest noise eigenvalue in dB) and maximum and average sidelobe level as a function of the number of snapshots used in estimating the covariance matrix. The convergence rate of the noise eigenvalues is slow after a large initial improvements. Each point represents a single simulation run, and the curves represent least squares fits to these points.

5.5 Diagonal loading

If the noise eigenvalue spread can be minimised, then the effects of randomly shaped noise eigenbeams will be reduced and the adapted response will approach the ideal response. Gabriel [73] suggested that the estimated covariance matrix be modified to accomplish this result. This modification takes the form of augmenting the leading diagonal of the covariance matrix

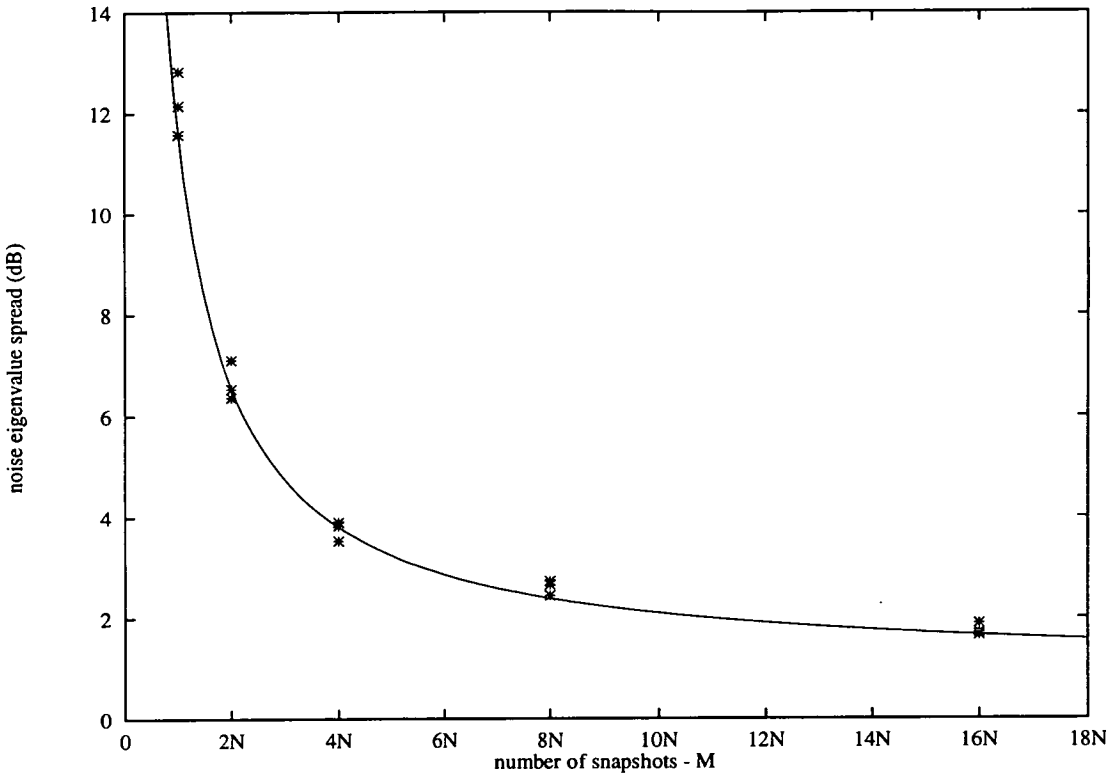

Figure 5.7: *Noise eigenvalue spread as a function of the number of snapshots. Each point represents a single realisation.*

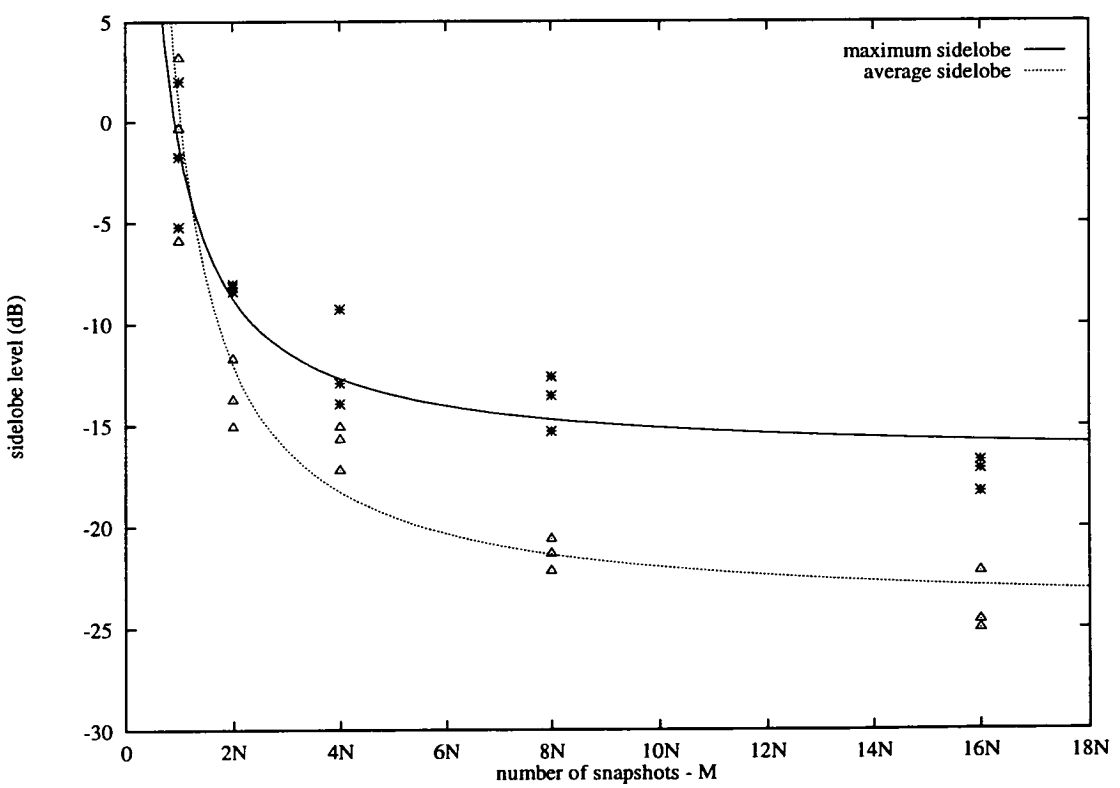

Figure 5.8: *Maximum and average sidelobe level as a function of the number of snapshots. Each point represents a single realisation.*

with a fixed (positive) term as follows

$$\hat{\boldsymbol{R}}_x = \frac{1}{M}\sum_{k=1}^{M}\boldsymbol{x}(k)\boldsymbol{x}^H(k) + F\boldsymbol{I}. \tag{5.41}$$

The effect of this modification is that the eigenvalues of the loaded matrix are individually increased by an amount equal to F. Large interference eigenvalues are unaffected by this small change, but eigenvalues well below the loading level are increased to and compressed at the level F. The corresponding eigenvectors remain unchanged by diagonal loading. Figure 5.9 depicts the adapted response of the beamformer of Figure 5.5 with diagonal loading equal to 12dB over the background noise level added. This example is identical to the conditions in Figure 5.5 with the exception of loading. The response shows that cancellation of the source at 65° is minimally affected, whilst the random sidelobe behaviour is considerably reduced, even with such a small number of snapshots. The weights computed using (5.41) deviate from the

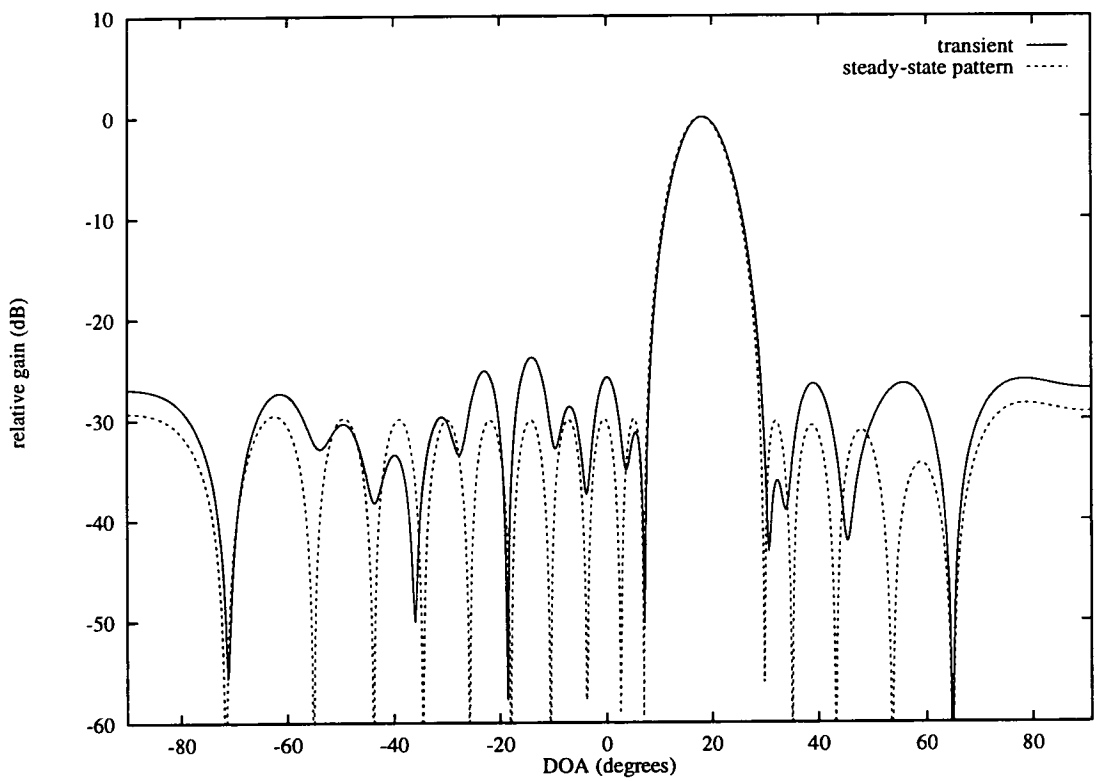

Figure 5.9: *Transient response of a narrowband GSC beamformer after 32 snapshots. The array is the same as that it Figure 5.5, but diagonal loading of 12dB above noise level has been added.*

optimum weights and will result in a slightly larger output residue, but the cost is negligible compared to the remarkably stable results achieved by this relatively simple approach.

Figure 5.10 depicts the reduction in noise eigenvalue spread achieved by diagonal loading of the covariance matrix. The three curves represent least squares fits to the results of 100 Monte Carlo simulations for the cases when $1N$, $3N$ and $6N$ snapshots are used in forming the covariance estimate. The physical implications of the reductions in eigenvalue spread can be seen in

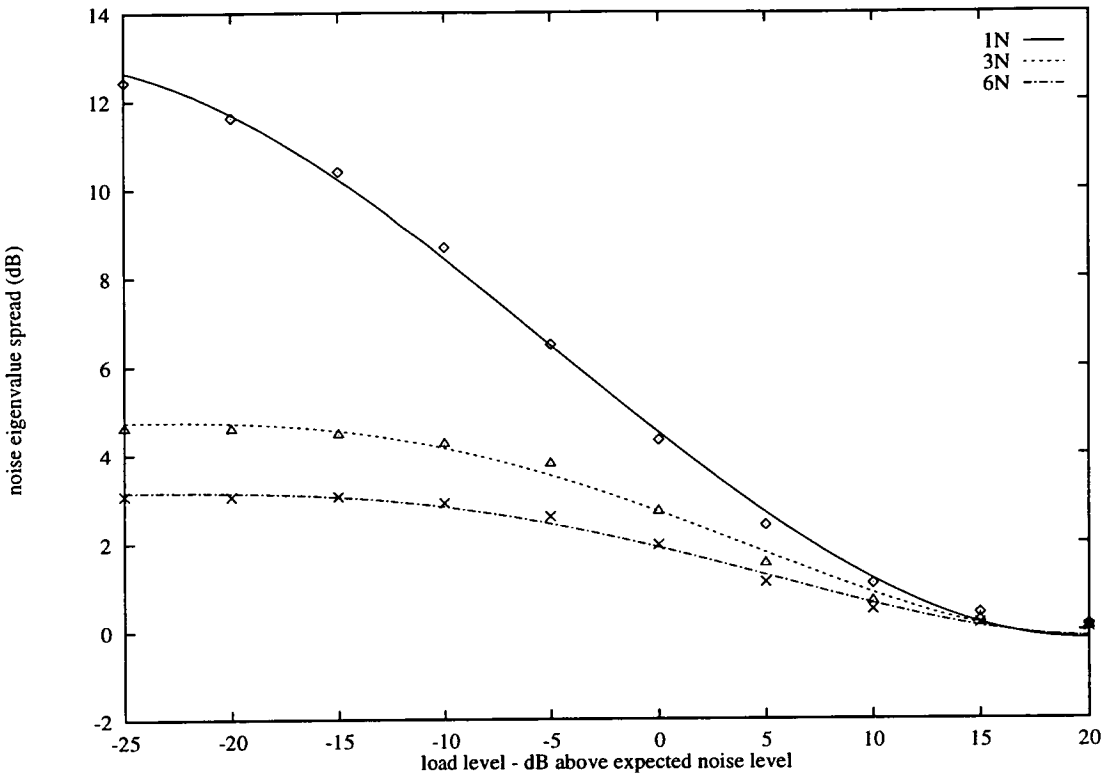

Figure 5.10: *Noise eigenvalue spread as a function of diagonal loading level for $1N$, $3N$ and $6N$ snapshots. Each point was computed from 100 Monte Carlo simulations.*

Figures 5.11 and 5.12, which show the array maximum and average sidelobe behaviour with respect to loading level. Loading the diagonal of the covariance estimate as in (5.41) to compress the noise eigenvalues is equivalent to desensitising the system by reducing its adaptive capability to small interference sources. This process can be thought of as artificially injecting a small noise component at each array element. Whilst this desensitivity has minimal effect upon cancellation of large interfering sources, as can be seen from (5.39), it may reduce the ability to counter interfering signals with small eigenvalues, such as occur from small jamming signals, residual jammer energy and dispersive paths.

The expressions derived earlier for the mean square error showed explicitly the link between MSE and adaptive dimension. One would also expect the random sidelobe fluctuations to be dependent upon adaptive dimension. Studying (5.39) it can be seen that the effect of the linear transformation $\boldsymbol{T}$ is to linearly combine the eigenbeams, prior to weighting. The number of these combined eigenbeams that are weighted (i.e. the dimension of $\boldsymbol{H}$) is now determined by the adaptive dimension, so that by reducing the adaptive dimension, random sidelobe fluctuations should be minimised. Figure 5.13 and 5.14 show the maximum and average sidelobe level as a function of load level for the three adaptive dimensions $J = 2, 8$ and 14. The conditions are the same as those for the earlier figures, namely a single interfering source and uncorrelated noise. $3N$ samples were used in forming the covariance matrix estimate. Best fit lines have be drawn to illustrate the trend. Clearly, for this simple example, a reduced adaptive dimension leads to lower transient sidelobes, both in peak and mean. This is as might be expected, a single

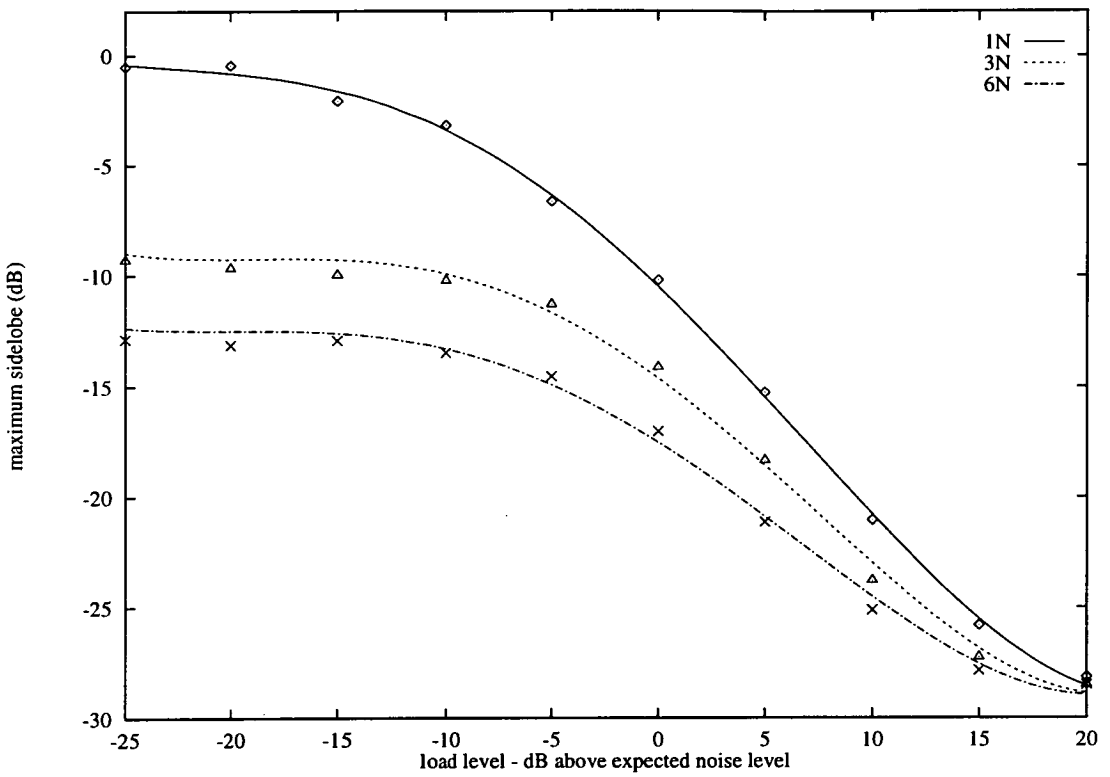

Figure 5.11: *Maximum sidelobe level as a function of diagonal loading level for $1N$, $3N$ and $6N$ snapshots. Each point was computed from 100 Monte Carlo simulations.*

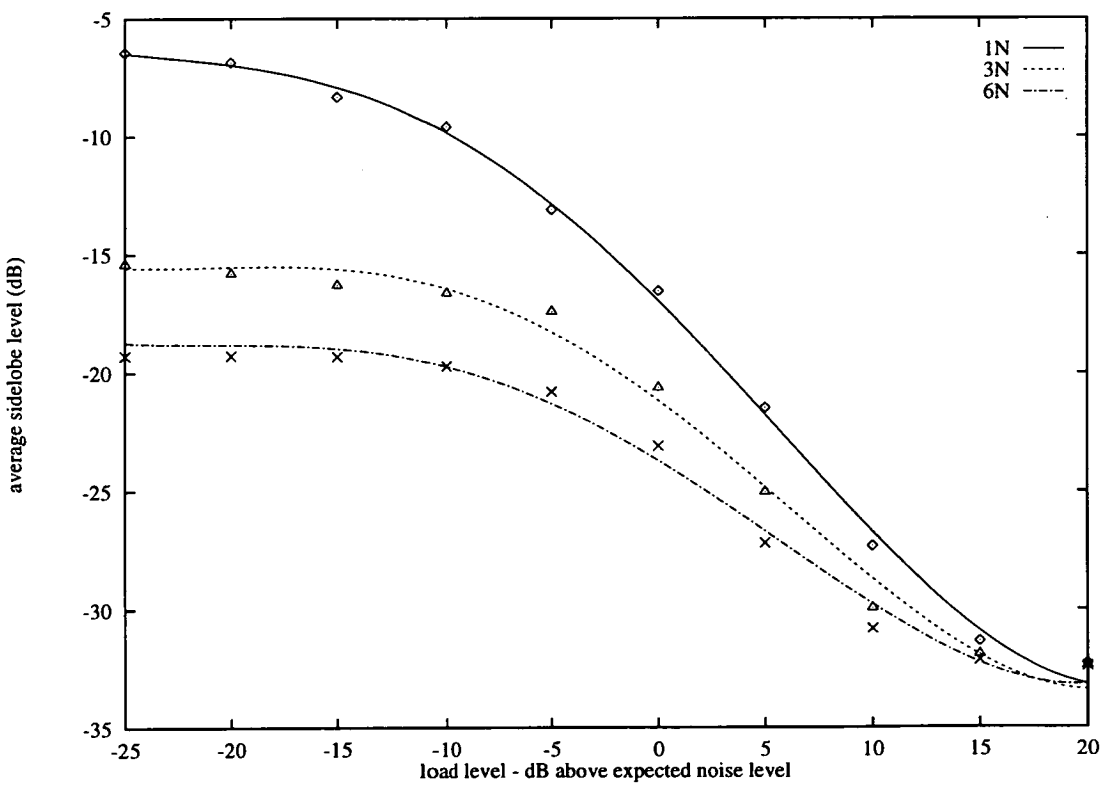

Figure 5.12: *Average sidelobe level as a function of diagonal loading level for $1N$, $3N$ and $6N$ snapshots. Each point was computed from 100 Monte Carlo simulations.*

jamming source will require a single adaptive degree of freedom, so that any additional degrees of freedom will adapt upon the cross–correlations in the background noise field due to finite sampling.

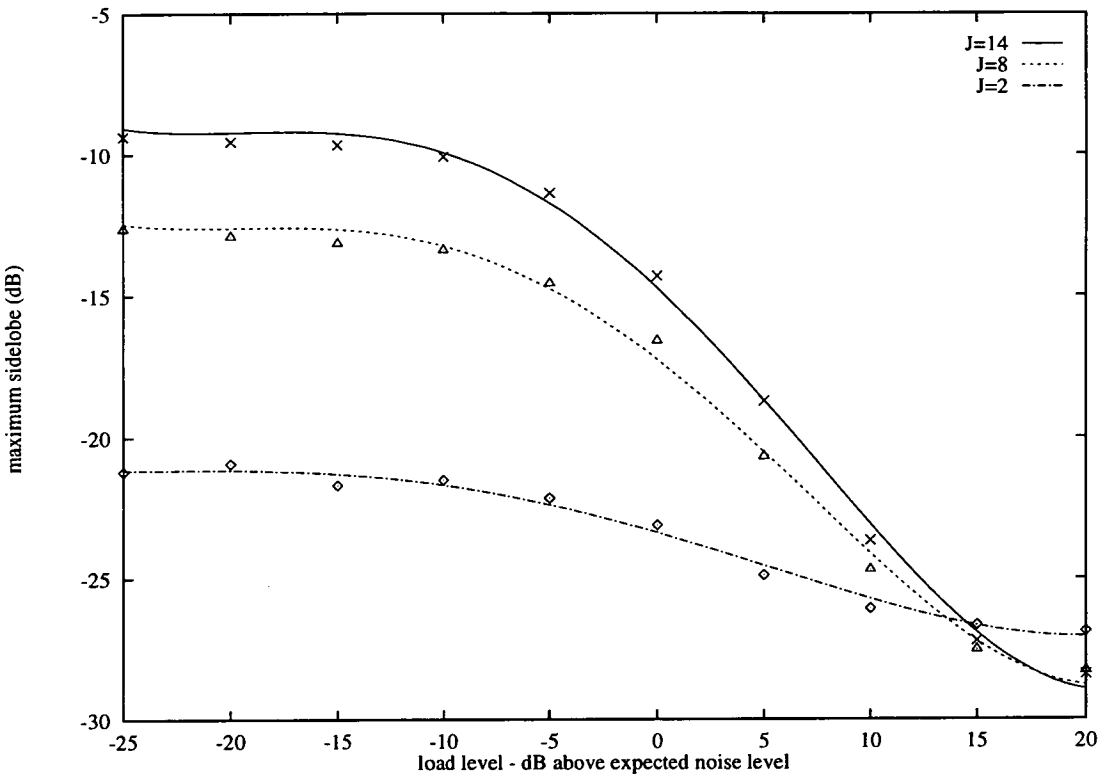

Figure 5.13: *Maximum sidelobe level after 3N snapshots as a function of diagonal loading level for $J = 2, 8$, and 14. Each point was computed from 100 Monte Carlo simulations.*

The interference scenario used to compute the results presented in this section is relatively simple. Noise and jammer eigenvalues are separated by a large margin (30dB) so that estimation of the number of jamming signals is a simple task. However, if the interference received at the array is derived from dispersive paths, or from a collection of closely spaced interfering sources the jammer eigenvalues may not necessarily be so distinct. In the latter case of several closely grouped jamming signals, the eigendecomposition will typically yield a single large eigenvalue and a selection of smaller eigenvalues. The correlation in spatial location between the sources is such that they do not yield distinct eigenvalues. The effects of a dispersive channel can be interpreted in a similar way. If secondary eigenvalues are similar in magnitude to the loading level, then reduced cancellation can be expected and residual interference power will be present after adaptation. However, the residual interference power may be acceptable since the sources are weak by definition and the sidelobes of the adaptive pattern provide additional attenuation if the interference is spatially separated from the steer direction. It is apparent from Figures 5.11 and 5.12 that a small amount of diagonal loading can considerably improve many aspects of adaptive performance when the number of snapshots used in forming the covariance estimate is small.

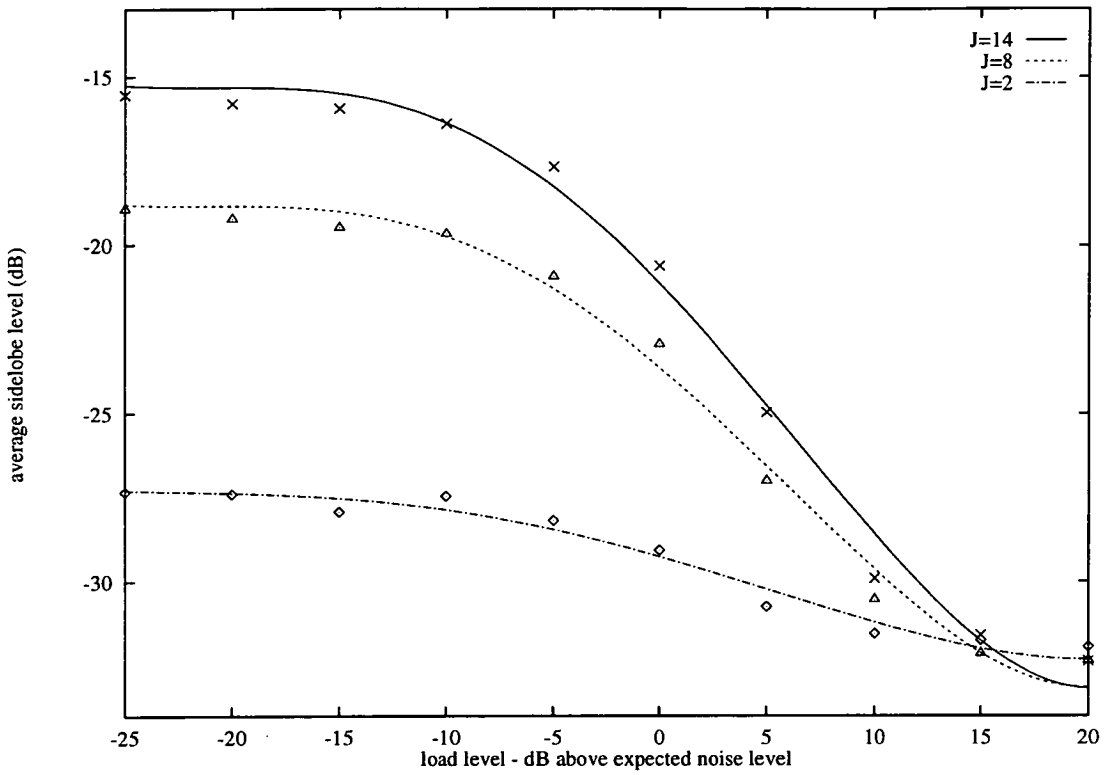

Figure 5.14: *Average sidelobe level after 3N snapshots as a function of diagonal loading level for $J = 2, 8$, and 14. Each point was computed from 100 Monte Carlo simulations.*

5.6 Conclusions

This chapter has considered the adaptive convergence of the generalised sidelobe cancelling beamformer structure. Expressions have been derived for a selection of beamformer characteristics. In particular, expressions for the concurrent and non–concurrent mean square error and transient sidelobe behaviour have been derived. Furthermore, these expressions have been examined and verified through Monte Carlo simulation.

If a finite number of data snapshots are used to compute the covariance matrix estimate the output mean square error will be different from that which would be observed if the true data covariance matrix was used. The level of excess mean square error depends upon whether the beamformer operates in a concurrent or non–concurrent mode. Expressions for the concurrent and non–concurrent mean square error were derived and simulation results showed close agreement.

When the covariance estimate is formed from a finite set of data vectors, the adapted response of the array was shown to experience noisy sidelobe fluctuations which can render the array susceptible to interfering signals and sidelobe clutter. As a means of characterising the random sidelobe fluctuations the eigendecomposition of the estimated covariance matrix was examined. It was found that for small numbers of snapshots a significant spread existed in the eigenvalues associated with the background noise. As more information was combined in the estimate the

noise eigenvalue spread decreased, which led to a corresponding decrease in the maximum and average sidelobe levels. A technique for reducing the noise eigenvalue spread, called diagonal loading, was examined. It was found that with a small level of loading, the random sidelobe fluctuations could be significantly reduced, thus reducing the vulnerability to interfering signals. No simulation results have been included for the space–time processors discussed in chapter 4. This is mainly due to the simulation time involved in running a large number of Monte Carlo simulations, but also because of the problem in accurately identifying the peak sidelobes in a two–dimensional response.

Chapter 6
Conclusions

6.1 Introduction

The work described in this thesis is primarily concerned with the analysis of techniques for designing partially adaptive beamformers suitable for use with a sideways looking airborne radar. In particular, techniques for reducing the adaptive dimension of a generalised sidelobe canceller array, when the radar operates in a look–down mode. A sparse algorithm for specifying the adaptive degrees of freedom to be selected has been described and applied to the suppression of ground clutter returns received at an airborne platform. The convergence performance of the resultant partially adaptive beamformer has been analysed and expressions have been derived for a variety of transient parameters. Within this chapter the main conclusions of the work are highlighted. Additionally, limitations associated with the algorithms and experimental techniques are discussed. The chapter concludes with some pointers towards future work.

6.2 Achievements of the work

Adaptive beamforming represents a very powerful technique for suppression and cancellation of interfering signals, and improving the detection performance of an airborne radar. However, adaptive beamforming represents a considerable expense, both in computation and in implementation. For these reasons, much interest has focussed on techniques for reducing the cost of adaptive algorithms, whilst maintaining near optimum performance. A feature of many of the common partially adaptive beamforming structures is what appears to be an assumption of access to large computing resources. This is an impractical assumption, so the approach taken in this thesis has been to produce a partially adaptive beamforming structure which attains near fully adaptive performance whilst minimising the computational cost.

Many of the simpler adaptive algorithms suffer performance degradation if desired signals are present in the received data. This occurs because of estimation errors in forming the data covariance matrices. For example, the maximised signal to noise ratio adaptive array computed the adaptive weights by forming a covariance matrix for the data received at the array, inverting it and then multiplying it by a steering vector. If any signal component is present in the covariance matrix from the steered direction, the effect of spurious correlations is to place a deep null in this direction. At no time can an airborne array make the assumption that there will not be signal components incident from the look direction during the period over which

the covariance matrix is formed. The generalised sidelobe cancelling beamformer structure was proposed as the adaptive structure to be used. This structure separates the beamformer weights into two components, a fixed beamformer, and an adaptive path. A matrix transformation is placed at the start of the adaptive path, which effectively blocks signal components from the steered direction, so that the adaptive weights are computed using data which is free of look direction signals. Thus, use of this two path structure prevents the signal cancellation phenomena associated with simple maximised signal to noise ratio beamformers.

The generalised sidelobe canceller is also useful as a partially adaptive beamforming structure. Full gain is retained for desired signals, whilst an arbitrary number of adaptive degrees of freedom may be assigned to interference cancellation. The number of adaptive degrees of freedom assigned to cancellation will be determined by the level of suppression which is deemed acceptable, and by the number of weights which can be computed. With computational efficiency in mind, a sparse approach for specifying the adaptive degrees of freedom in a partially adaptive beamformer was presented in chapter 4. The algorithm selects adaptive degrees of freedom based upon a mean square error performance criterion. The adaptive degrees of freedom which result in the greatest reduction in output mean square error are selected, the remaining adaptive weights being set to zero. The performance of this new algorithm was contrasted with several existing techniques for designing partially adaptive beamformers, most notably the eigenstructure based design. The computational expense of the sparse and eigenstructure beamformers was also considered, and it was shown that for a variety of beamformer dimensions the sparse beamformer could yield around 80% savings in the numbers of operations required to compute the adaptive weights.

Chapter 5 considered the convergence performance of the generalised sidelobe cancelling structure. A statistical analysis of the output mean square error when the beamformer operates in a concurrent and non–concurrent mode were presented. Simulations results for both modes of operation were seen to be in close agreement with theoretical analysis. The random sidelobe behaviour which exists when the adaptive weights are computed from a finite set of samples was also considered. An expression for the adaptive weight vector in terms of the eigendecomposition of the sample covariance matrix was derived. This expression showed that the transient sidelobes of the adapted pattern were a function of the spread of the noise eigenvalues of the sample covariance matrix. Monte Carlo simulation results were presented showing the maximum and average sidelobe levels as a function of the noise eigenvalue spread. A technique for overcoming the dependency upon the fluctuations in the small eigenvalues was discussed. It was seen through simulation that a small level of diagonal loading of the sample covariance matrix could yield significant improvements in both maximum and average sidelobe levels, even when a small number of samples were used in forming the sample covariance matrix.

6.3 Limitations of the experimental techniques

In order to achieve the primary object of the work, which was to develop techniques for specifying the best adaptive degrees of freedom to retain in a partially adaptive beamformer, a number of necessary simplifying assumptions were made. These can be broadly separated into two groups, (i) simplifications associated with the radar operation, (ii) and simplifications relating to the beamforming aspect of the work.

The most obvious simplifications relating to the radar are the flight characteristics, namely a assumption of an horizontal flight plane at constant velocity. These assumptions, coupled with the assumption that no dispersion or multipath propagation occurs, allow a very simple model for the operation of a range gated radar. The clutter returns in any particular range gate will be derived solely from scatterers (targets or the ground) at multiples of the gate range. In computing the clutter returns, the range gates are divided into a ring of individual uncorrelated scatterers, so that calculation of idealised covariance matrices is replaced by a discrete summation over the individual scatterers. Arguments can be formed in a similar manner to those relating to the choice of point constraints (section 2.7.1) in linearly constrained beamforming, to express how well (in a least squares sense) the sum matches the ideal.

Simplifications relating to the array itself are chosen more for mathematical convenience than computational ease. Each sensor is assumed to be followed with an ideal wideband steering delay, such that any signal, regardless of frequency, incident from the look direction appears identically at the output of the steering delays. Such perfect matching is unrealistic, and can only be achieved over small bands of frequency. Mismatch in array elements will destroy the correlation of a source within the beamforming structure, making separation of desired and interfering signals a considerably more complex problem.

6.4 Limitations of the work

The algorithm presented in chapter 4 is based upon the principal that the over–riding cost of performing adaptive beamforming is the computation of the adaptive weights, rather than in forming the data samples. Thus the algorithm is allowed to select which ever tap sample best minimises the output mean square error, regardless of the position which the weight occupies in the adaptive structure. The nature of the transformation matrix associated with this method of design is sparse, so that the required number of operations is considerably less than the popular subspace techniques. The comments and simulations presented later in chapter 4 were aimed at limiting the implemental cost of the iterative algorithm. The algorithm is nonetheless computationally intensive, which will most probably mean that it will be only used at a design stage, based upon likely interference data. The choice of training data will determine to a large extent how well the beamformer will operate once in flight. The question of how large to choose the range of scenarios the beamformer is designed to operate over remains open.

Adaptive array performance is generally measured in signal to noise or cancellation ability, and the iterative algorithm is based upon a mean square error performance measure. However, these measures do not consider the adapted response of the beamformer, which can often exhibit undesirably high sidelobe levels. The analysis of chapter 5 shows that the adaptive sidelobe level, whilst being a function of the noise scenario and desired signal strength, is directly proportional to the number of adaptive degrees of freedom present in the beamformer. Therefore the reduced adaptive dimension facilitated by the sparse approach should help to minimise the average transient sidelobe response.

The interference problem considered in this thesis was the cancellation of ground clutter returns in an airborne radar. Whilst the two–dimensional nature of this interference provides a stern test of a broadband adaptive beamformer, the analysis of section 3.7 demonstrated the inherent rank deficiency of this problem. The cancellation problem, under certain conditions, requires considerably fewer degrees of freedom than is provided by the tap structure. There may exist other interference problems which do not have this nature, and the performance enhancements achieved may not be as significant.

6.5 Areas for future work

To conclude the thesis, we provide some pointers to further areas of development, and provide suggestions for alternative applications.

The radar geometry considered in this thesis was that of a sideways looking airborne radar (SLAR). This is a common mode of operation for surveillance radars, but in the future attention will move to forward looking geometries. These are popular in fighter aircraft because the radar antenna is typically mounted in the aircraft's nose. The diagonal azimuth–Doppler characteristics of the ground clutter associated with sideways looking geometries are not mirrored in the forward looking geometry. When the radar is in a forward looking configuration clutter returns are localised around an ellipse in the azimuth–Doppler plane. This has major repercussions for the design of clutter cancellation filters. The simple DPCA cancellation system cannot be applied because sub–apertures no longer exist along the aircraft's axis. For this reason, an analysis of the forward looking geometry, and of typical clutter spectra will form the basis for the design of forward looking radars. From this analysis, conclusions may be drawn as to the beamforming requirements for the forward looking geometry.

The selection algorithm presented in chapter 4 selects the adaptive degrees of freedom that have most influence upon the output mean square error. This algorithm was applied to the weight selection for cancellation of ground clutter returns. However, this approach may be taken in a variety of interference cancellation problems in which a reduced adaptive dimension is useful. In fact a simple analogy can be drawn between the nature of this algorithm and tap selection in the RAKE type receivers common in spread spectrum communications. A RAKE receiver combines delayed samples of a received field to provide cancellation of interfering signals and

enhancement of desired signals in the presence of multipath propagation. The weights applied to the various multipaths are simply the tap weights of the filter structure. If the algorithm described in this thesis were to be applied to this problem it would simply result in the selection of the weights corresponding to the largest multipath components.

In connection with the work of chapter 5 a statistical analysis of the peak and mean sidelobe behaviour of the generalised sidelobe canceller may be performed. Previous work exists giving a measure of the peak and average sidelobes of a randomly spaced array [46]. This analysis, coupled with multivariate statistical theory could provide expressions similar in nature to the concurrent and non–concurrent mean square error expressions of chapter 5. These expressions would allow confidence intervals to be established for the adapted response of the beamformer.

References

[1] R. Klemm, "Adaptive airborne MTI: an auxiliary channel approach", *IEE Proceedings Part F*, vol. 134, no. 3, pp. 269–276, June 1987.

[2] P.G. Richardson, "A quiescent pattern control strategy for adaptive arrays", in Proceedings *EUSIPCO-94*, pages 1301–1304, September 1994.

[3] L.E. Brennan, J.D. Mallett, and I.S. Reed, "Adaptive arrays in airborne MTI radar", *IEEE Trans. Antennas and Propagation*, vol. AP-24, no. 5, pp. 607–615, September 1976.

[4] R.A. Monzingo and T.W. Miller, *Introduction to Adaptive Arrays*. John Wiley and Sons, New York, 1980.

[5] R. Klemm, "Adaptive clutter suppression for airborne phased array radars", *IEE Proceedings Part F*, vol. 130, no. 1, pp. 125–132, February 1983.

[6] R. Klemm, "Adaptive airborne MTI with two–dimensional motion compensation", *IEE Proceedings Part F*, vol. 138, no. 6, pp. 551–558, December 1991.

[7] R. Klemm, "Antenna design for adaptive airborne MTI", in Proceedings *IEE Conf. Radar 92*, pages 296–299, October 1992.

[8] E.C. Barille, R.L. Fante, and J.A. Torres, "Some limitations on the effectiveness of airborne adaptive radar", *IEEE Trans. Aerospace and Electronic Systems*, vol. 28, no. 4, pp. 1015–1032, October 1992.

[9] J. Su and Y.Q. Zhou, "Adaptive clutter suppression for airborne array radars using clutter subspace approximation", in Proceedings *IEE Conf. Pub. Radar 92*, pages 155–158, October 1992.

[10] O.L. Frost III, "An algorithm for linearly constrained adaptive array processing", *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, August 1972.

[11] W.F. Gabriel, "Adaptive arrays - an introduction", *Proc. IEEE*, vol. 64, pp. 239–272, Feb 1976.

[12] K. Takao, M. Fujita, and T. Nishi, "An adaptive array under directional constraint", *IEEE Trans. Antennas and Propagation*, vol. AP-24, no. 5, pp. 662–669, September 1976.

[13] M.H. Er and A. Cantoni, "Derivative constraints for broadband element space antenna array processors", *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-31, no. 6, pp. 1378–1393, December 1983.

[14] N.L. Owsley, in *Array Signal Processing*. Prentice–Hall, Englewood Cliffs, New Jersey, 1985.

[15] K.M. Buckley and L.J. Griffiths, "An adaptive generalized sidelobe canceller with derivative constraints", *IEEE Trans. Antennas and Propagation*, vol. AP-34, no. 3, pp. 311–319, March 1986.

[16] K.M. Buckley, "Spatial/spectral filtering with linearly constrained minimum variance beamformers", *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-35, no. 3, pp. 249–266, March 1987.

[17] L.J. Griffiths and K.M. Buckley, "Quiescent pattern control in linearly constrained adaptive arrays", *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-35, no. 7, pp. 917–926, July 1987.

[18] M.H. Er, "Adaptive antenna array under directional and spatial derivative constraints", *IEE Proceedings Part H*, vol. 135, no. 6, pp. 414–419, December 1988.

[19] B.D. Van Veen, "Eigenstructure based partially adaptive array design", *IEEE Trans. Antennas and Propagation*, vol. AP-36, no. 3, pp. 357–362, March 1988.

[20] B.D. Van Veen and K.M. Buckley, "Beamforming: A versatile approach to spatial filtering", *IEEE Acoustics, Speech, and Signal Processing Magazine*, pp. 4–24, April 1988.

[21] I. Thng, A. Cantoni, and Y.H. Leung, "Derivative constrained optimum broadband antenna arrays", *IEEE Trans. Signal Processing*, vol. 41, no. 7, pp. 2376–2388, July 1993.

[22] C. Jim, "A comparison of two LMS constrained optimal array structures", *Proceedings of IEEE*, vol. 65, pp. 1730–1731, December 1977.

[23] L.J. Griffiths and C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming", *IEEE Trans. Antennas and Propagation*, vol. AP-30, no. 1, pp. 27–34, January 1982.

[24] N.K. Jablon, "Steady state analysis of the generalized sidelobe canceller by adaptive noise cancelling techniques", *IEEE Trans. Antennas and Propagation*, vol. AP-34, no. 3, pp. 330–337, March 1986.

[25] K.M. Buckley, "Broadband beamforming and the generalized sidelobe canceller", *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no. 5, pp. 1322–1323, October 1986.

[26] A. Farina, *Antenna-Based Signal Processing Techniques for Radar Systems*. Artech House, London, 1992.

[27] B.D. Van Veen and R.A. Roberts, "Partially adaptive beamformer design via output power minimization", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-35, no. 11, pp. 1524–1532, November 1987.

[28] B.D. Van Veen and R.A. Roberts, "Analytic design of broadband partially adaptive beamformers", in Proceedings *ICASSP-87*, pages 54.4.1–54.4.4, April 1987.

[29] T.T. Ma and L.J. Griffiths, "A solution space approach to achieving partially adaptive arrays", in Proceedings *ICASSP-88*, pages 2869–2872, April 1988.

[30] B.D. Van Veen, "Improved power minimization based partially adaptive beamformer design", in Proceedings *ICASSP-88*, pages 2793–2796, April 1988.

[31] K. Takao and K. Uchida, "Beamspace partially adaptive antenna", *IEE Proceedings Part F*, vol. 136, no. 6, pp. 439–444, December 1989.

[32] B.D. Van Veen, "An analysis of several partially adaptive beamformer designs", *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, no. 2, pp. 192–203, February 1989.

[33] B.D. Van Veen, "Optimization of quiescent response in partially adaptive beamformers", *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 38, no. 3, pp. 471–477, March 1990.

[34] B.D. Van Veen, "Adaptive convergence of linearly constrained beamformers based on the sample covariance matrix", *IEEE Trans. Signal Processing*, vol. 39, no. 6, pp. 1470–1473, June 1991.

[35] F. Qian and B.D. Van Veen, "Coherent interference suppression via partially adaptive beamforming", in Proceedings *ICASSP-92*, pages 441–444, April 1992.

[36] F. Qian and B.D. Van Veen, "Quadratically constrained adaptive beamforming in coherent interference environments", in Proceedings *ICASSP-93*, pages 528–531, April 1993.

[37] I. Scott and B. Mulgrew, "A sparse approach in partially adaptive linearly constrained arrays", in Proceedings *ICASSP-94*, pages IV541–IV544, April 1994.

[38] I. Scott and B. Mulgrew. "Sparse LCMV beamformer design for suppresion of ground clutter in airborne radar". *IEEE Trans. Signal Processing.* Submitted for review.

[39] I. Scott and B. Mulgrew, "A sparse approach to partially adaptive airborne radar", in Proceedings *EUSIPCO-94*, pages 351–354, September 1994.

[40] J.S. Goldstein, M.A. Ingram, E.J. Holder, and R.N. Smith, "Adaptive subspace selection using subband decompositions for sensor array processing", in Proceedings *ICASSP-94*, pages IV–281–IV–284, April 1994.

[41] J. Marr, "A selected bibliography on adaptive antenna arrays", *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-22, pp. 781–796, November 1986.

[42] C.L. Dolph, "A current distribution for broadside arrays which optimizes the relationship between beam width and sidelobe level", *Proceeding of IRE*, vol. 34, no. 6, pp. 335–348, June 1946.

[43] F.J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform", *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, January 1978.

[44] M.I. Skolnik, "Statistically designed density-tapered arrays", *IEEE Trans. Antennas and Propagation*, vol. AP-12, no. 5, pp. 408–417, July 1964.

[45] Y.T. Lo, "A mathematical theory of antenna arrays with randomly spaced elements", *IEEE Trans. Antennas and Propagation*, vol. AP-12, no. 5, pp. 257–268, May 1964.

[46] B.D. Steinberg, "The peak sidelobe of the phased array having randomly located elements", *IEEE Trans. Antennas and Propagation*, vol. AP-20, no. 2, pp. 129–136, March 1972.

[47] T. Numazaki, S. Mano, T. Katagi, and M. Mizusawa, "An improved thinning method for density tapering of planar array antennas", *IEEE Trans. Antennas and Propagation*, vol. AP-35, no. 9, pp. 1066–1070, September 1987.

[48] J.E. Hudson, *Adaptive Array Principles.* Peter Peregrinus Ltd, London, 1981.

[49] S. Haykin ed., *Array Signal Processing.* Prentice Hall, Englewood Cliffs, New Jersey, 1985.

[50] S. Haykin, *Adaptive Filter Theory.* Prentice Hall, Englewood Cliffs, New Jersey, 1991.

[51] Special issue on adaptive beamforming. *IEEE Trans. Antennas and Propagation*, September 1976. vol. AP-24.

[52] Special issue on adaptive beamforming. *IEEE Trans. Antennas and Propagation*, September 1986. vol. AP-34.

[53] M.H. Er and A. Cantoni, "A new set of linear constraints for broadband time domain element space processors", in Proceedings *ICASSP-85*, pages 46.7.1–46.7.4, April 1985.

[54] K.M. Buckley and L.J. Griffiths, "Spatial filtering with broadband minimum variance beamformers", in Proceedings *Proc. Int. IEEE A/P-S Symp.*, pages 575–578, Philadelphia, PA, June 1986.

[55] T. Liu and B.D. Van Veen, "A modular structure for implementation of linearly constrained minimum variance beamformers", *IEEE Trans. Signal Processing*, vol. 39, no. 10, pp. 2343–2346, October 1991.

[56] B.D. Van Veen, "Minimum variance beamforming with soft response constraints", *IEEE Trans. Signal Processing*, vol. 39, no. 9, pp. 1964–1972, September 1991.

[57] Y. Chen and C. Chen, "A new structure for adaptive broadband beamforming", *IEEE Trans. Antennas and Propagation*, vol. 39, no. 4, pp. 551–555, April 1991.

[58] K. Huarng and C. Yeh, "Adaptive beamforming with conjugate symmetric weights", *IEEE Trans. Antennas and Propagation*, vol. 39, no. 7, pp. 926–932, July 1991.

[59] K. Huarng and C. Yeh, "Performance analysis of derivative constraint adaptive arrays with pointing errors", *IEEE Trans. Antennas and Propagation*, vol. 30, no. 8, pp. 975–981, August 1992.

[60] C.Y. Tseng, "Minimum variance beamforming with phase-independent derivative constraints", *IEEE Trans. Antennas and Propagation*, vol. 40, no. 3, pp. 285–294, March 1992.

[61] I. Claesson and S. Nordholm, "A spatial filtering approach to robust adaptive beaming", *IEEE Trans. Antennas and Propagation*, vol. 40, no. 9, pp. 1093–1096, September 1992.

[62] C.E. Cook and M. Bernfeld, *Radar Signals - An Introduction to Theory and Application*. Academic Press, 1967.

[63] M.I. Skolnik, *Introduction to Radar Systems*. McGraw–Hill, 1962.

[64] M.I. Skolnik ed., *Radar Handbook*. McGraw–Hill, 1970.

[65] J.L. Farrell and R.L. Taylor, "Doppler radar clutter", *IEEE Trans. Aerospace and Navigational Electronics*, vol. ANE-11, no. 3, pp. 162–172, September 1964.

[66] A.L. Friedlander and L.J. Greenstein, "A generalized clutter computation procedure for airborne pulsed Doppler radars", *IEEE Trans. Aerospace and Electronics Systems*, vol. AES-6, no. 1, pp. 51–61, January 1970.

[67] D.K. Barton ed., *CW and Doppler Radar*, vol. 7 of *Radars*. Artech House, 1978.

[68] G.A. Andrews, "Radar pattern design for platform motion compensation", *IEEE Trans. Antennas and Propagation*, vol. AP-26, no. 4, pp. 566–571, July 1978.

[69] P.G. Richardson, "Analysis of the adaptive space time processor technique for airborne radar", *IEE Proceedings part F*, vol. 141, no. 4, pp. 187–195, August 1994.

[70] R. Klemm, "Use of generalised resolution methods to locate sources in random dispersive media", *IEE Proceedings Part F*, vol. 127, no. 1, pp. 34–40, February 1980.

[71] D.R. Morgan, "Partially adaptive array techniques", *IEEE Trans. Antennas and Propagation*, vol. AP-26, no. 11, pp. 823–833, November 1978.

[72] R.N. Adams, L.L. Horowitz, and K.D. Senne, "Adaptive mainbeam nulling for narrowband antenna arrays", *IEEE Trans. Aerospace and Electronic Systems*, vol. 16, pp. 509–516, July 1980.

[73] W.F. Gabriel, "Using spectral estimation techniques in adaptive processing antenna systems", *IEEE Trans. Antennas and Propagation*, vol. AP-34, no. 3, pp. 291–300, March 1986.

[74] D.J. Chapman, "Partial adaptivity for the large array", *IEEE Trans. Antennas and Propagation*, vol. AP-24, no. 9, pp. 685–696, September 1976.

[75] K.C. Nisbet, B. Mulgrew, and S. McLaughlin. Reduced state methods in nonlinear prediction. submitted to: EURASIP Signal Processing Journal, January 1993.

[76] E.J. Baranoski and A. Steinhardt, "Localized subspace projection", in Proceedings *ICASSP-93*, pages IV-333 - IV-347, April 1993.

[77] S. Chen, C.F.N. Cowan, and P.M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks", *IEEE Trans. Neural Networks*, vol. NN-2, no. 2, pp. 302–309, March 1991.

[78] E.C. Chng, B. Mulgrew, and S. Chen, "Backtracking orthogonal least squares algorithm for model selection", in Proceedings *IEE Colloquium on Mathematical Aspects of Digital Signal Processing*, Digest No. 1994/034, Bristol, February 1994.

[79] N.R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution", *Ann. Math. Stat.*, vol. 34, pp. 152–177, March 1963.

[80] E.J. Kelly, "An adaptive detection algorithm", *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-22, no. 1, pp. 115–127, March 1986.

[81] B.D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays", *IEEE Trans. Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 397–401, July 1988.

[82] R.J. Muirhead, *Aspects of Multivariate Statistical Theory*. Wiley, New York, 1982.

[83] J. Capon and N.R. Goodman, "Probablility distributions for estimators of the frequency–wavenumber spectrum", *Proc. IEEE*, vol. 58, pp. 1785–1786, October 1970.

[84] I.S. Reed, J.D. Mallett, and L.E. Brennan, "Rapid convergence rate in adaptive arrays", *IEEE Trans. Aerospace and Electronic Systems*, vol. AES10, pp. 853–863, November 1974.

[85] J.L. Krolik and D.N. Swingler, "On the mean–square error performance of adaptive minimum variance beamformers based on the sample covariance matrix", *IEEE Trans. Signal Processing*, vol. 42, no. 2, pp. 445–448, February 1994.

[86] K.V. Mardia, J.T. Kent, and J.M. Bibby, *Multivariate Analysis*. Academic Press, London, 1979.

[87] M.L. Eaton, *Multivariate Statistics*. Wiley, New York, 1983.

[88] M. Siotani, T. Hayakawa, and Y. Fujikoshi, *Modern multivariate statistical analysis: A graduate course and handbook*. American Sciences Press, Inc., Columbus, Ohio, 1985.

[89] C. Chatfield and A.J. Collins, *Introduction to Multivariate Analysis*. Chapman and Hall, London, 1980.

[90] P.J. Bickel and K.A. Doksum, *Mathematical Statistics*. Holden–Day, Inc., San Francisco, 1977.

[91] J. Wishart, "The generalized product moment distribution in samples from a normal multivariate population", *Biometrika*, vol. 20A, pp. 32–43, 1928.

[92] R. Nitzberg, "Effects of errors in adaptive weights", *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-12, no. 3, pp. 369–373, May 1976.

[93] L.L. Horowitz, H. Blatt, W.G. Brodsky, and K.D. Senne, "Controlling adaptive antenna arrays with the sample matrix inversion algorithm", *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-15, no. 6, pp. 840–848, November 1979.

[94] L.L. Horowitz, "Convergence rate of the extended SMI algorithm for narrowband adaptive arrays", *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-16, no. 5, pp. 738–740, September 1980.

[95] R. Nitzberg, "Computational precision requirements for optimal weights in adaptive processing", *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-16, no. 4, pp. 418–425, July 1980.

[96] R. Nitzberg, "Detection loss of the sample matrix inversion technique", *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-20, no. 6, pp. 824–827, November 1984.

[97] K. Gerlach and F.F. Kretschmer, "Convergence properties of Gram-Schmidt and SMI algorithms", *IEEE Trans. Aerospace and Electronic Systems*, vol. 26, no. 6, pp. 44–56, November 1990.

[98] M.W. Ganz, R.L. Moses, and S.L. Wilson, "Convergence of the SMI and the diagonally loaded SMI algorithms with weak interference", *IEEE Trans. Antennas and Propagation*, vol. 38, no. 3, pp. 394–399, March 1990.

[99] K. Gerlach, "Implementation and convergence considerations of a linearly constrained adaptive array", *IEEE Trans. Aerospace and Electronic Systems*, vol. 26, no. 2, pp. 263–272, March 1990.

[100] K. Gerlach, "Adaptive array transient sidelobe levels and remedies", *IEEE Trans. Aerospace and Electronic Systems*, vol. 26, no. 3, pp. 560–568, May 1990.

[101] K. Gerlach and F.F. Kretschmer, "Convergence properties of Gram-Schmidt and SMI algorithms: part II", *IEEE Trans. Aerospace and Electronic Systems*, vol. 27, no. 1, pp. 83–90, January 1991.

[102] S.M. Yuen, "Comments on "Convergence properties of Gram-Schmidt and SMI algorithms"", *IEEE Trans. Aerospace and Electronic Systems*, vol. 27, no. 6, pp. 897–901, November 1991.

[103] K. Gerlach, "Convergence bounds of an SMI/Gram–Schmidt canceler in coloured noise", *IEEE Trans. Aerospace and Electronic Systems*, vol. 27, no. 4, pp. 655–666, July 1991.

[104] L. Chang and C. Yeh, "Performance of DMI and eigenspace–based beamformers", *IEEE Trans. Antennas and Propagation*, vol. 40, no. 11, pp. 1336–1347, November 1992.

Appendix A

Interference cancellation

In this appendix an expression for the adaptive cancellation associated with using J eigenvectors of the averaged interferer covariance matrix is formed. The following derivation is based on those variously presented in [16,27,30,32]. Assume that the interference consists of a single broadband point interferer, uncorrelated with the desired signal, and spatially/temporally uncorrelated white noise. Letting $\boldsymbol{Q}(\theta)$ denote the covariance matrix for the interferer arriving from direction θ and σ_w^2 be the white noise level implies

$$\boldsymbol{R}_n = \boldsymbol{Q}(\theta) + \sigma_w^2 \boldsymbol{I}. \tag{A.1}$$

$\boldsymbol{Q}(\theta)$ is expressed in terms of the source power spectral density $\rho^2(\omega)$, the array response vector $\boldsymbol{d}(\theta,\omega)$, and the source frequency extent Ω as

$$\boldsymbol{Q}(\theta) = \int_{\Omega} \rho^2(\omega)\, \boldsymbol{d}(\theta,\omega)\, \boldsymbol{d}^H(\omega)\, d\omega. \tag{A.2}$$

Writing $\boldsymbol{T} = \boldsymbol{C}_n \boldsymbol{T}_n$, the power output is given by

$$\begin{aligned} P_{out} &= \boldsymbol{w}_q^H \boldsymbol{R}_s \boldsymbol{w}_q + \sigma_w^2 \left| \boldsymbol{w}_q - \boldsymbol{T}\boldsymbol{w}_a \right|^2 \\ &\quad + \int_{\Omega} \rho^2(\omega) \left| \boldsymbol{w}_q^H \boldsymbol{d}(\theta,\omega) - \boldsymbol{w}_a^H \boldsymbol{T}^H \boldsymbol{d}(\omega) \right|^2 d\omega. \end{aligned} \tag{A.3}$$

The first term is the signal output power, the second term is the white noise output power, and the last term is the interferer output power. A more quantitative description of interference cancellation is obtained by approximating (A.2) as a Riemann sum

$$\boldsymbol{Q}(\theta) \approx \sum_{m=1}^{M} \rho^2(\omega_m)\, \boldsymbol{d}(\theta,\omega_m)\, \boldsymbol{d}^H(\theta,\omega_m)\, \Delta\omega, \tag{A.4}$$

where ω_m uniformly sample Ω, so $\Delta\omega = \omega_{m+1} - \omega_m$. Rewriting (A.4) in matrix form yields

$$\boldsymbol{Q}(\theta) \approx \boldsymbol{A}_\theta \boldsymbol{\Gamma}^2 \boldsymbol{A}_\theta^H, \tag{A.5}$$

where

$$\begin{aligned} \boldsymbol{A}_\theta &= [\, \boldsymbol{d}(\theta,\omega_1)\ \boldsymbol{d}(\theta,\omega_2)\ \cdots\ \boldsymbol{d}(\theta,\omega_M)\,], \\ \boldsymbol{\Gamma}^2 &= \operatorname{diag}\left\{ \rho^2(\omega_1),\ \rho^2(\omega_2),\ \cdots,\ \rho^2(\omega_M) \right\} \Delta\omega. \end{aligned}$$

The approximation error in (A.4) and (A.5) approaches 0 as $M \rightarrow \infty$ ($\Delta\omega \rightarrow 0$). Here M is assumed to be sufficiently large. We assume for the latter part of this discussion that $\boldsymbol{\Gamma}$ is the positive square root of $\boldsymbol{\Gamma}^2$, $\boldsymbol{T}^H \boldsymbol{T} = \boldsymbol{I}$, and $\boldsymbol{T}^H \boldsymbol{w}_q = 0$. Assuming $\boldsymbol{T}^H \boldsymbol{T} = \boldsymbol{I}$ does not reduce generality since it is the space spanned by the columns of $\boldsymbol{T}$ that is important, not the elements of $\boldsymbol{T}$; this is evident by noting that the output power is invariant to nonsingular transformations applied to $\boldsymbol{T}$ on the right. The assumption $\boldsymbol{T}^H \boldsymbol{w}_q = 0$ implies that $\boldsymbol{w}_q$ has been chosen to determine the quiescent response of the beamformer [17]. Substituting (A.5) and (A.1) into the expression for the optimum weight vector $\boldsymbol{w}_a$ and simplifying results in

$$\boldsymbol{w}_a = \left[\boldsymbol{T}^H \boldsymbol{A}_\theta \boldsymbol{\Gamma}\boldsymbol{\Gamma} \boldsymbol{A}_\theta^H \boldsymbol{T} + \sigma_w^2 \boldsymbol{I}\right]^{-1} \boldsymbol{T}^H \boldsymbol{A}_\theta \boldsymbol{\Gamma}\boldsymbol{\Gamma} \boldsymbol{A}_\theta^H \boldsymbol{w}_q. \tag{A.6}$$

Now, we may approximate the M by J matrix $\boldsymbol{\Gamma} \boldsymbol{A}_\theta^H \boldsymbol{T}$ by the rank d singular value decomposition

$$\boldsymbol{\Gamma} \boldsymbol{A}_\theta^H \boldsymbol{T} = \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^H. \tag{A.7}$$

Here $\boldsymbol{U}$ and $\boldsymbol{V}$ are composed of the d left and right singular vectors corresponding to the d largest singular values and $\boldsymbol{\Sigma}$ is a d by d diagonal matrix containing the largest singular values. The approximate low rank nature of $\boldsymbol{Q}(\theta)$ (and thus $\boldsymbol{\Gamma} \boldsymbol{A}_\theta^H \boldsymbol{T}$) is discussed by Buckley [16].

Consider the physical meaning of $\boldsymbol{V}, \boldsymbol{\Sigma}$, and $\boldsymbol{U}$. $\boldsymbol{V}$ represents an orthonormal basis for the rows of $\boldsymbol{\Gamma} \boldsymbol{A}_\theta^H \boldsymbol{T}$ and is also the set of eigenvectors of $\boldsymbol{T}^H \boldsymbol{Q}(\theta) \boldsymbol{T}$ corresponding to nonzero eigenvalues. Thus, $\boldsymbol{V}$ represents a basis for the space spanned by the interferer at the output of $\boldsymbol{T}$. $\boldsymbol{\Sigma}$ describes the distribution of the interference power in the space described by $\boldsymbol{V}$. $\boldsymbol{U}$ represents an orthonormal basis for the columns of $\boldsymbol{\Gamma} \boldsymbol{A}_\theta^H \boldsymbol{T}$. Let $[\boldsymbol{B}]_i$ denote the ith column of a matrix $\boldsymbol{B}$. Now

$$\left[\boldsymbol{\Gamma} \boldsymbol{A}_\theta^H \boldsymbol{T}\right]_i = \boldsymbol{\Gamma} \boldsymbol{A}_\theta^H [\boldsymbol{T}]_i. \tag{A.8}$$

$\boldsymbol{A}_\theta^H [\boldsymbol{T}]_i$ is a vector whose elements describe the response of the ith column of $\boldsymbol{T}$ in direction θ at the frequency sampling points $[\omega_1, \omega_2, \cdots, \omega_M]$. Premultiplication by $\boldsymbol{\Gamma}$ simply scales the response proportionally to the power spectral density of the source. Thus, each column of $\boldsymbol{\Gamma} \boldsymbol{A}_\theta^H \boldsymbol{T}$ represents the frequency distribution of the interferer at the output of the corresponding column of $\boldsymbol{T}$. We can conclude then that $\boldsymbol{U}$ is a basis for the space spanned by the interferer frequency distributions at the output of $\boldsymbol{T}$, just as $\boldsymbol{V}$ represented the interferer frequency distributions at the output of $\boldsymbol{C}_n$. Continuing with the derivation, substituting (A.7) into (A.6) yields

$$\boldsymbol{w}_a = \left[\boldsymbol{V} \boldsymbol{\Sigma}^2 \boldsymbol{V}^H + \sigma_w^2 \boldsymbol{I}\right]^{-1} \boldsymbol{V} \boldsymbol{\Sigma} \boldsymbol{U}^H \boldsymbol{g}, \tag{A.9}$$

with

$$\boldsymbol{g} = \boldsymbol{\Gamma} \boldsymbol{A}_\theta^H \boldsymbol{w}_q. \tag{A.10}$$

Following the discussion in the previous paragraph, g represents the frequency distribution of the interferer at the fixed beamformer output. Applying the matrix inversion lemma to (A.9) and simplifying gives

$$w_a = \boldsymbol{V}\boldsymbol{D}\,\boldsymbol{U}^H g, \tag{A.11}$$

where $\boldsymbol{D}$ is a diagonal matrix with entries

$$\boldsymbol{D}_{ii} = \frac{\sigma_i}{\sigma_i^2 + \sigma_w^2}. \tag{A.12}$$

Here σ_i represents the ith diagonal element of $\boldsymbol{\Sigma}$. Equation (A.10) indicates that the weight vector lies in the space spanned by $\boldsymbol{V}$, which is the space spanned by the interferer after transformation by $\boldsymbol{T}$. The coordinates of w_a in this space are given by $\boldsymbol{D}\,\boldsymbol{U}^H g$. Rewriting (A.10) for the fully adaptive case gives $w_a = \boldsymbol{V}_f \boldsymbol{D}_f \boldsymbol{U}_f^H g$. Eigenstructure designs utilise the fact that the fully adaptive weight vector lies in the space spanned by $\boldsymbol{V}_f$. The columns of $\boldsymbol{V}_f$ are the eigenvectors corresponding to nonzero eigenvalues of the spatially/temporally correlated portion of the interference covariance matrix at the output of $\boldsymbol{C}_n$. For the interference environment described by (4.34), $\boldsymbol{V}_f$ corresponds to the eigenvectors of $\boldsymbol{C}_n^H \boldsymbol{Q}(\theta)\,\boldsymbol{C}_n$ associated with nonzero eigenvalues. In an arbitrary interference environment, $\boldsymbol{V}_f$ corresponds to the eigenvectors of $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$ associated with nonzero eigenvalues where $\boldsymbol{R}_n$ represents only the spatially/temporally correlated terms (white noise excluded).

The final step in the derivation is to express $\boldsymbol{U}^H g$ in terms of angles between vectors. Define the generalised angle between two vectors x and y as [48]

$$\cos^2(x, y) = \frac{\left(x^H y\right)^2}{\left(x^H x\right)\left(y^H y\right)}. \tag{A.13}$$

Applying (A.13) to (A.11) allows us to identify the ith component of $\boldsymbol{D}\boldsymbol{U}^H g$ as

$$\left[\boldsymbol{D}\boldsymbol{U}^H g\right]_i = \frac{\sigma_i}{\sigma_i^2 + \sigma_w^2}\left(g^H g\right)^{1/2} \cos\phi_i\, e^{j\eta_i}, \tag{A.14}$$

where

$$\cos\phi_i\, e^{j\eta_i} = \frac{[\boldsymbol{U}]_i\, g}{\left([\boldsymbol{U}]_i^H\,[\boldsymbol{U}]_i\, g^H g\right)^{1/2}}. \tag{A.15}$$

The weighting placed on $[\boldsymbol{V}]_i$ therefore depends upon the normalised inner product between the interferer frequency distribution at the fixed beamformer output and the ith basis vector for the interferer frequency distribution at the output of $\boldsymbol{T}_n$ ($\cos\phi_i\, e^{j\eta_i}$), the interferer to white noise level in the ith mode, and the fixed beamformer output power due to the interferer ($g^H g$). As the white noise level decreases, the higher order modes are weighted more heavily. We may now rewrite the output power as

$$P_{out} = w_q^H \boldsymbol{R}_s w_q + \sigma_w^2 w_q^H w_q + g^H g - g^H \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{D}\,\boldsymbol{U}^H g, \tag{A.16}$$

which may be alternatively expressed as

$$P_{out} = w_q^H R_s w_q + w_q^H R_n w_q - g^H g \sum_{i=1}^{J} \frac{\sigma_i}{\sigma_i^2 + \sigma_w^2} \cos^2 \phi_i. \quad \text{(A.17)}$$

$w_q^H R_s w_q$ represents the signal power at the output, $w_q^H R_n w_q$ represents the interference plus white noise power at the fixed beamformer output, and the last term represents the reduction in output power resulting from the J adaptive weights. More general cancellation analysis can be performed following the arguments above; however, the analysis provides little insight, except in cases where the design region becomes a single point.

Appendix B
Multivariate statistics

This appendix will introduce some of the properties of multivariate normal distributions, and discuss their application to the sample covariance matrix introduced in chapter 5. The theorems listed below will be stated without proof. If required, proofs of these may be found in any multivariate analysis book. Muirhead [82], Mardia [86], Eaton [87], Siotani [88], Chatfield [89] and Bickel [90] are recommended. We begin by defining the multivariate normal distribution.

Definition B.1 *The random vector* $\boldsymbol{X}$ $(m \times 1)$ *is said to have an* m*–*variate normal distribution *if, for any vector* $\boldsymbol{\alpha} \in \mathbb{C}^m$*, the distribution of* $\boldsymbol{\alpha}^H \boldsymbol{X}$ *is univariate normal. The multivariate normal distribution is denoted by* $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and defines a* $m \times 1$ *random vector with mean vector* $\boldsymbol{\mu}$*, and covariance* $\boldsymbol{\Sigma}$*.* □

Theorem B.1 *If* $\boldsymbol{X}$ *is* $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *then the marginal distribution of any* k $(k < m)$ *components of* $\boldsymbol{X}$ *is* k*–variate normal.* □

A consequence of this theorem is that the marginal distribution of each component of $\boldsymbol{X}$ is univariate normal. The converse is not true in general, the fact that each component of a random vector is normal does not imply that the vector has a multivariate normal distribution.

Theorem B.2 *Suppose that* $\boldsymbol{X}$ *is* $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, and* $\boldsymbol{X}$ *is partitioned as*

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix},$$

with corresponding partitions

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

then if $\boldsymbol{\Sigma}_{22}$ *is full rank so that* $\boldsymbol{\Sigma}_{22}^{-1}$ *exists, the* conditional distribution *of* $\boldsymbol{X}_1$ *given* $\boldsymbol{X}_2$ *is multivariate normal with expected value*

$$E\{\boldsymbol{X}_1 | \boldsymbol{X}_2\} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2).$$

□

Theorem B.3 *Suppose that* $\boldsymbol{X}$ *is a random vector, then the mean of the conditional expectations of* $\boldsymbol{X}$ *given a random vector* $\boldsymbol{Y}$ *is*

$$E\{E\{\boldsymbol{X}|\boldsymbol{Y}\}\} = E\{\boldsymbol{X}\}.$$

□

Theorem B.4 *If $h(\boldsymbol{X})$ is bounded and $r(\boldsymbol{X})$ has finite mean, then*

$$E\{r(\boldsymbol{X})h(\boldsymbol{Y})\} = E\{h(\boldsymbol{Y})E\{r(\boldsymbol{X})|\boldsymbol{Y}\}\}.$$

□

Theorem B.5 *If $h(\boldsymbol{X})$ is bounded and $r(\boldsymbol{X})$ has finite mean, then*

$$E\{r(\boldsymbol{X})h(\boldsymbol{Y})|\boldsymbol{Y}=y\} = h(y)E\{r(\boldsymbol{X})|\boldsymbol{Y}=y\}.$$

□

Theorems B.4 and B.5 are commonly called the *product expectation* and *substitution* formulae for conditional expectations [90, pp. 5]. When used in conjunction with theorems B.2 and B.3 they provide a powerful method of manipulating functions of multivariate random variables.

Definition B.2 *If $\boldsymbol{S}$ $(m \times m)$ can be written $\boldsymbol{S} = \boldsymbol{Z}^H\boldsymbol{Z}$, where $\boldsymbol{Z}$ $(m \times n)$ is a data matrix from $N_m(\boldsymbol{0}, \boldsymbol{\Sigma})$, then $\boldsymbol{S}$ is said to have a* Wishart distribution *with scale matrix $\boldsymbol{\Sigma}$ and degrees of freedom parameter n. This is written $\boldsymbol{S} \sim W_m(n, \boldsymbol{\Sigma})$. When $\boldsymbol{\Sigma} = \boldsymbol{I}_m$, the distribution is said to be in standard form.* □

Note that when $m = 1$, the $W_1(n, \sigma^2)$ distribution is the same as the $\sigma^2\chi_n^2$ (chi–squared) distribution. The scale matrix $\boldsymbol{\Sigma}$ plays the same role as σ^2 does in the $\sigma^2\chi_n^2$ distribution. Note also that $\boldsymbol{S}$ is singular for $n < m$, so the density function of $W_m(n, \boldsymbol{\Sigma})$ will exist only for values of n greater than or equal to m. The density function of $W_m(n, \boldsymbol{\Sigma})$ was first derived by Wishart [91], hence the name given to the distribution. The properties of Wishart matrices are of considerable interest to the beamforming community. If $\boldsymbol{Z}$ is as a matrix of array data, then the matrix $(1/n)\,\boldsymbol{S}$ is termed the sample covariance matrix and represents the estimate of the data covariance matrix formed from n snapshots. The first moment of $W_m(n, \boldsymbol{\Sigma})$ is given by

$$E\{\boldsymbol{S}\} = n\boldsymbol{\Sigma}. \tag{B.1}$$

Theorem B.6 *If $\boldsymbol{S}$ is $W_m(n, \boldsymbol{\Sigma})$ and $\boldsymbol{M}$ is $n \times k$ of rank k, then $\boldsymbol{M}^H\boldsymbol{S}\boldsymbol{M}$ is $W_k(n, \boldsymbol{\Sigma})$.* □

An interesting example of this theorem is the case when $k = 1$, i.e. $\boldsymbol{M}$ is a vector, then the ratio $\boldsymbol{M}^H\boldsymbol{S}\boldsymbol{M}/\boldsymbol{M}^H\boldsymbol{\Sigma}\boldsymbol{M}$ has the χ_n^2 distribution, provided $\boldsymbol{M}^H\boldsymbol{\Sigma}\boldsymbol{M} > 0$. The form of $\boldsymbol{M}^H\boldsymbol{S}\boldsymbol{M}$ is familiar in beamforming, since it gives the power output for a given weighting vector. In [83] Capon and Goodman used the Goodman [79] theory on complex Wishart distributions to find probability distributions of several estimators, similar in form to this ratio. Later, Reed *et al.* [84] applied similar ideas to predict the convergence of an adaptive clutter cancellation filter. Many other papers have examined expressions similar to this [73,80,92–104].

Theorem B.7 *If $\boldsymbol{S}$ is $W_m(n, \boldsymbol{\Sigma})$ and $\boldsymbol{S}$ and $\boldsymbol{\Sigma}$ are partitioned as*

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

where $\boldsymbol{S}_{11}$ *and* $\boldsymbol{\Sigma}_{11}$ *are* $k \times k$, *then* $\boldsymbol{S}_{11}$ *is* $W_k(n, \boldsymbol{\Sigma}_{11})$. □

This theorem stems directly from theorem B.1. In general any diagonal submatrix of $\boldsymbol{S}$ is Wishart distributed, although their distributions will not in general be independent.

Definition B.3 *If* $\boldsymbol{S}$ *is* $W_m(n, \boldsymbol{\Sigma})$ *then the distribution of* $\boldsymbol{S}^{-1}$ *is called the* inverted Wishart distribution, *denoted by* $W_m^{-1}(n, \boldsymbol{\Sigma})$. □

The expectation of $\boldsymbol{S}^{-1}$ is

$$E\left\{\boldsymbol{S}^{-1}\right\} = \frac{1}{n-m-1}\boldsymbol{\Sigma}^{-1}. \tag{B.2}$$

This is one of the most useful expressions relating to the use of sample covariance matrices, since the adaptive weights in many adaptive algorithms (e.g. maximised signal–to–noise ratio) are formed by inverting an estimated covariance matrix. This distribution, through the mean, allows predictions to be made about the expected convergence performance.

Theorem B.8 *If* $\boldsymbol{S}$ *is* $W_m(n, \boldsymbol{\Sigma})$, *and* $\boldsymbol{M}$ *is* $n \times k$ *of rank* k, *then* $\left(\boldsymbol{M}^H \boldsymbol{S}^{-1} \boldsymbol{M}\right)^{-1}$ *has distribution given by* $W_k\left(n-m+k, (\boldsymbol{M}^H \boldsymbol{\Sigma}^{-1} \boldsymbol{M})^{-1}\right)$. □

Until now, no definition has been given for the density function of a Wishart matrix. The actual expression has a rather cumbersome form, however, for completeness the complex Wishart probability density function is defined as [4, pp. 302]

$$p(\boldsymbol{S}) = \frac{|\boldsymbol{S}|^{n-m-1} \exp\left\{-\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)\right\}}{\pi^{1/2(m+1)m}\Gamma(n)\Gamma(n-1)\cdots\Gamma(n-m)\,|\boldsymbol{\Sigma}|^n}, \qquad n \geq m, \tag{B.3}$$

where

$$\Gamma(k) = (k-1)!$$

This completes our brief review of multivariate statistics.

Appendix C

Original publications

The work described in this thesis has been reported in the following publications:

- † I. Scott and B. Mulgrew, "A Sparse Approach in Partially Adaptive Linearly Constrained Arrays", in Proceedings *ICASSP-94*, Adelaide, Australia, vol. IV, pages 541-544, April, 1994.
- † I. Scott and B. Mulgrew, "A Sparse Approach to Partially Adaptive Airborne Radar", in Proceedings *EUSIPCO-94*, Edinburgh, Scotland, pages 351-354, September, 1994.

Paper submitted:

- I. Scott and B. Mulgrew, "Sparse LCMV Beamformer Design for Suppression of Ground Clutter in Airborne Radar", *IEEE Trans. Signal Processing.*

In preparation:

- I. Scott and B. Mulgrew, "Convergence analysis of partially adaptive generalised sidelobe canceller", *IEEE Trans. Aerospace and Electronic Systems.*

† Reprinted in this appendix.

A SPARSE APPROACH IN PARTIALLY ADAPTIVE LINEARLY CONSTRAINED ARRAYS

Iain Scott and Bernard Mulgrew

Department of Electrical Engineering,
The University of Edinburgh,
Mayfield Road, Edinburgh EH9 3JL

ABSTRACT

In conventional partially adaptive linearly constrained minimum variance (LCMV) beamformer design the approach has been to represent the noise subspace with some reduced set of vectors, typically the eigenvectors associated with the largest eigenvalues of the noise covariance matrix. This, whilst yielding good performance, will not give the optimum performance for a given partially adaptive dimension.

This paper presents an alternative method for selecting the "best" degrees of freedom to be retained in a partially adaptive design. The iterative algorithm described selects those degrees of freedom which minimize the beamformer output mean square error. This approach leads to a sparse structure for the transformation matrix, which when implemented in a generalized sidelobe canceller (GSC) structure will reduce the computational load. This approach also allows a reduction in adaptive dimension as compared to the eigenvector based approach. An illustrative example demonstrates the effectiveness of this method.

I. INTRODUCTION

This paper is concerned with developing a simple technique for the selection of the adaptive degrees of freedom (DOF) to be retained in a partially adaptive beamforming algorithm. The computational complexity associated with linearly constrained adaptive arrays quickly becomes prohibitive when element numbers are increased, forcing consideration of techniques which employ only a subset of the available DOF. Fortunately, most interference suppression problems are rank deficient in nature, that is they require less adaptive DOF than are offered by the array. By appropriate design, the additional DOF which are not required can be discarded so that only those that are important are retained. This is termed partially adaptive beamforming.

The array studied is assumed to be a linearly constrained array having N elements each with L temporal samples processed. Array response is subject to a total of K linear constraints. As a result the total adaptive DOF will be $NL-K$, which may be much greater than the number that can be implemented in the processor. This limitation may arise because of restricted computational hardware or the desire for real–time performance. The necessary reduction in adaptive dimension is performed by inserting a transformation matrix before the adaptive weights, which maps the fully adaptive space to some reduced dimension space. Previously reported techniques have fallen into three groups, a sub–optimum sequential approach based on output power minimization [1], eigenstructure based schemes [2], and a reduction in available DOF by the addition of linear constraints [3].

In this paper we describe a sparse approach for specifying the DOF to be retained in a partially adaptive design. The primary results reported here are the reduction in required adaptive dimension facilitated by the approach, and the reduced computational load engendered by a sparse solution. As such it offers several advantages over eigenvector approaches which are based upon the structure of the noise subspace. Simulation of a partially adaptive GSC illustrates the effectiveness of this approach.

II. PARTIALLY ADAPTIVE BEAMFORMER MSE

For an adaptive array, the output $y(n)$ can be expressed as the inner product of a data vector $\mathbf{x}(n)$, and a weight vector $\mathbf{w}(n)$, that is

$$y(n) = \mathbf{w}^{\dagger}(n)\,\mathbf{x}(n), \tag{1}$$

where the dagger $\dagger$ denotes Hermitian transpose. For generality both the data vector and the weight vector are assumed complex valued. For a broadband linear array of N elements each with L taps, $\mathbf{x}(n)$ and $\mathbf{w}(n)$ are NL dimensional vectors. In a LCMV beamformer the array output power is minimized subject to a set of K linear constraints. This is expressed mathematically as

$$\min_{\mathbf{w}} \mathbf{w}^{\dagger}\mathbf{R}_{\mathbf{xx}}\mathbf{w} \quad \text{subject to} \quad \mathbf{C}^{\dagger}\mathbf{w} = \mathbf{f}, \tag{2}$$

where $\mathbf{R}_{\mathbf{xx}}$ denotes the covariance matrix of $\mathbf{x}(n)$, the constraint matrix $\mathbf{C}$ contains K column vectors whilst the response vector $\mathbf{f}$ contains the scalar constraint value for each vector. The solution to (2) is obtained via Lagrange multipliers as

$$\mathbf{w}_{opt} = \mathbf{R}_{\mathbf{xx}}{}^{-1}\,\mathbf{C}\left(\mathbf{C}^{\dagger}\,\mathbf{R}_{\mathbf{xx}}{}^{-1}\,\mathbf{C}\right)^{-1}\mathbf{f} \tag{3}$$

The constraint equations in (2) describe a $NL-K$ dimensional hyperplane in the NL dimensional space which is termed the *adaptive weight vector solution space*. This terminology reflects the fact that all weight vectors must lie

in this subspace. The adaptive weight vector will converge to an optimum weight vector which lies on the hyperplane and simultaneously minimizes the output power. A large array will have a large value NL, but may not have a correspondingly large number of constraints K, so that the solution space may be overly large meaning the adaptive processor may have a long reaction time and require prohibitively large computation. The desire for fast convergence rates and real time operation leads to the study of arrays which employ reduced dimensional weight vectors - so called partially adaptive arrays. The objective in this paper is to reduce the dimension of the solution space in some sparse manner whilst maintaining an adequate level of array performance.

The LCMV beamformer can be implemented in either of two equivalent structures. In the first the adaptive weights are computed then applied directly to the array data. Frost [4] proposed an efficient gradient-based algorithm for updating the adaptive weight vector. In a generalised sidelobe canceller (GSC) implementation the weight vector $\mathbf{w}$ is decomposed into two orthogonal components, one which lies in the range space of $\mathbf{C}$ called $\mathbf{w}_q$, and one which lies in the null space of $\mathbf{C}$ given as $-\mathbf{C}_n\mathbf{w}_a$. The orthogonality ensures the lower branch's response to the desired signal is zero. The signal blocking matrix $\mathbf{C}_n$ is full rank and satisfies $\mathbf{C}_n^\dagger\mathbf{C} = \mathbf{0}$. The adaptive weight vector $\mathbf{w}_a$ is chosen as a solution of the unconstrained minimization problem

$$\min_{\mathbf{w}_a}(\mathbf{w}_q - \mathbf{C}_n\mathbf{w}_a)^\dagger\mathbf{R}_{\mathbf{xx}}(\mathbf{w}_q - \mathbf{C}_n\mathbf{w}_a), \tag{4}$$

the optimal weight vector being given by

$$\mathbf{w}_a = (\mathbf{C}_n^\dagger\mathbf{R}_{\mathbf{xx}}\mathbf{C}_n)^{-1}\mathbf{C}_n^\dagger\mathbf{R}_{\mathbf{xx}}\mathbf{w}_q \tag{5}$$

The important feature of the beamformer outlined in [4] was the use of broad-band steering delays at each element. The desired signal was identified by time-delay steering the sensor outputs so that any signal incident from the direction of interest (look direction) would appear as an identical replica at the output of the steering delays. All other signals which did not have this property would be processed as noise or interference. As far as the desired signal is concerned the array then appears as a single tapped delay line, with tap weights given by the sum of the corresponding elemental tap weights for that particular tap. The pre-steering delays therefore allow control of the frequency response in the look direction.

More importantly broad-band steering delays allow a very simple structure for the constraint and signal blocking matrices $\mathbf{C}$ and $\mathbf{C}_n$. The constraint matrix will be of the form

$$\mathbf{C} = \begin{bmatrix} \mathbf{1}_N & \mathbf{0}_N & \cdots & \\ \mathbf{0}_N & \mathbf{1}_N & & \\ \vdots & \vdots & \vdots & \mathbf{0}_N \\ & & & \mathbf{1}_N \end{bmatrix} \tag{6}$$

and the response vector $\mathbf{f}$ will specify the equivalent tapped delay line weights. In (6) the column vector $\mathbf{1}_N$ consists of N unity entries and the vector $\mathbf{0}_N$ contains N zeros. With the above structure for the constraint matrix the problem of finding an orthogonal signal blocking matrix is greatly simplified. One example is the diagonal matrix with L submatrices $\mathbf{w}$ shown below.

$$\mathbf{C}_n = \begin{bmatrix} \mathbf{w} & & & \\ & \mathbf{w} & & \\ & & \ddots & \\ & & & \mathbf{w} \\ & & & & \mathbf{w} \end{bmatrix} \quad ; \quad \mathbf{w}^\dagger = \begin{bmatrix} w_1^\dagger \\ w_2^\dagger \\ \vdots \\ w_{N-1}^\dagger \end{bmatrix} \tag{7}$$

The columns w_i of $\mathbf{w}$ should sum to zero and be mutually orthogonal to ensure that $\mathbf{C}_n$ is full rank. A physically simple and elegant example of $\mathbf{w}$ is shown below. In this example each of the columns involves only a simple difference operation between adjacent elemental outputs, ensuring that the beamformer hardware will be simple to implement.

$$\mathbf{w} = \begin{bmatrix} 1 & & & & & & \\ -1 & 1 & & & \mathbf{0} & & \\ & -1 & 1 & & & & \\ & & & \ddots & & & \\ & & & & -1 & 1 & \\ & & \mathbf{0} & & & -1 & 1 \\ & & & & & & -1 \end{bmatrix} \tag{8}$$

A signal blocking matrix whose elements consist entirely of ones and zeros has several computational advantages. Firstly a matrix multiplication which involves a sparse matrix can be implemented more efficiently than one which does not. Secondly, if the only non-zero entries are either 1 or -1 then the matrix multiplication consists wholly of addition/subtraction operations. These are faster and simpler to implement than multiplications. If we can then subsequently reduce the adaptive dimension, also in some sparse manner, then we can still exploit these desirable features. The inputs to the adaptive processor would therefore be derived through simple differencing of adjacent elements.

In a partially adaptive generalized sidelobe canceller [2] a $NL - K \times J$ transformation matrix $\mathbf{T}_n$, which maps the fully adaptive problem to a lower (J) dimensional space, is inserted after the signal blocking matrix. The problem now becomes that a designing $\mathbf{T}_n$ so as to minimize any degradation in array performance. The performance measure employed in this paper is output mean square error.

The beamformer output mean square error is defined as [5]

$$\mathrm{MSE} = E\left[\left|s - \mathbf{w}^\dagger\mathbf{x}\right|^2\right] = \mathbf{w}^\dagger\mathbf{R}_{\mathbf{nn}}\mathbf{w}, \tag{9}$$

in which the desired signal $s = \mathbf{w}_q^\dagger\mathbf{s}$ is the beamformer output in the absence of noise for the desired signal vector $\mathbf{s}$, and $\mathbf{R}_{\mathbf{nn}}$ is the covariance matrix of noise and interference. For the partially adaptive GSC the mean square error can be shown to be

$$\begin{aligned}\mathrm{MSE} = {} & \mathbf{w}_q^\dagger\mathbf{R}_{\mathbf{nn}}\mathbf{w}_q \\ & -\mathbf{w}_q^\dagger\mathbf{R}_{\mathbf{nn}}\mathbf{C}_n\mathbf{T}_n(\mathbf{T}_n^\dagger\mathbf{C}_n^\dagger\mathbf{R}_{\mathbf{nn}}\mathbf{C}_n\mathbf{T}_n)^{-1}\mathbf{T}_n^\dagger\mathbf{C}_n^\dagger\mathbf{R}_{\mathbf{nn}}\mathbf{w}_q\end{aligned} \tag{10}$$

Our task now is to minimize this by appropriate design of $\mathbf{T}_n$.

III. A SPARSE APPROACH

Analytic minimization of the mean square error over all interference scenarios of interest presents a formidable problem, and has been discussed [2]. Our problem in a sparse design of $\mathbf{T}_n$, is that of choosing which degrees of freedom should be retained, and which should be discarded. Here a sub–optimum iterative approach based upon minimization of the output MSE of the partially adaptive array is presented. Once the desired adaptive dimension is specified the algorithm iteratively searches for degrees of freedom which will best minimize the output MSE. Recall that the solution space has dimension $NL-K$. A sparse solution means that a reduced number of dimensions, those that have most influence upon the adaption, will be selected for the optimization. This results in a transformation matrix that will be composed of a selection of unit vectors, the non–zero entries indicating the degrees of freedom selected. Table 1 summarizes the proposed algorithm.

If we denote the transformation matrix of dimension $NL-K \times i$ as $\mathbf{T}_n^i$, and the set of allowed degrees of freedom as $\{\mathbf{e}^j\}$, $1 \le j \le NL-K$, and the selected degrees of freedom as $\{\mathbf{i}^j\}$, $1 \le j \le J$, then our selection procedure can be described as follows.

Initially the algorithm selects the first of the set of allowed vectors and forms $\mathbf{T}_n^1$ as a matrix with the single column $\mathbf{e}^1$. The output mean square error is computed for this transformation matrix and stored. The column $\mathbf{e}^1$ is now replaced with the second vector in the allowed set $\mathbf{e}^2$, and the output mean square is evaluated once more. This procedure is repeated until all the vectors in the allowed set have been tried. The column to be selected is chosen as the allowed vector which resulted in the smallest output MSE. This selected column $\mathbf{i}^1$, is then deleted from the allowed set and the algorithm commences upon a search for additional vectors. At any stage the transformation matrix can be partitioned into two portions - the previously selected columns and the allowed vector for which the output MSE is currently being evaluated. This iterative search for vectors which best minimize the output MSE is continued until all the allowed columns have been added, or until the output MSE reaches an acceptable level. The simulations presented later show that often only a small number of the allowed columns are required, and that addition of further columns does little to further improve the output MSE performance.

In the fully adaptive case $\mathbf{T}_n$ will be an identity matrix (or some column–wise permutation), but in the partially adaptive case the columns of $\mathbf{T}_n$ will be those degrees of freedom that have most influence upon the output MSE. At each step the algorithm searches the remaining DOF for one that results in the greatest reduction in output MSE.

This technique is similar to the orthogonal least squares algorithm described by Chen *et al.* [6]. In this radial basis function centres are chosen one by one so that each additional centre minimizes the least square error. After a centre is chosen the remaining basis vectors (columns) are made orthogonal to the chosen vector. This procedure is iterated until the output error is sufficiently small. The algorithm described above is simpler because the allowed set of vectors are already mutually orthogonal, removing the need to orthogonalise the remaining columns after one is selected. The iterative nature of these algorithms does lead to a "good" set of vectors, but not may not yield the optimum selection. In the future some form of back–tracking may need to be added to the algorithm to allow an improved set to be chosen.

- At the first step, for $1 \le j \le NL-K$, compute

  $\mathbf{T}_n^1 = \mathbf{e}^j$

  $\mathrm{MSE}^{(\mathbf{e}^j)} = \mathbf{w}_q^\dagger \mathbf{R}_{nn}\mathbf{w}_q - \mathbf{w}_q^\dagger \mathbf{R}_{nn}\mathbf{C}_n\mathbf{T}_n^1(\mathbf{T}_n^{1\dagger}\mathbf{C}_n^\dagger\mathbf{R}_{nn}\mathbf{C}_n\mathbf{T}_n^1)^{-1}\mathbf{T}_n^{1\dagger}\mathbf{C}_n^\dagger\mathbf{R}_{nn}\mathbf{w}_q$

  Find $\mathrm{MSE}_1^{(\mathbf{i}^1)} = \min\{\mathrm{MSE}^{(\mathbf{e}^j)},\ 1 \le j \le NL-K\}$

  then select $\mathbf{T}_n^1 = \mathbf{i}^1$.

- At the k^{th} step, $k \ge 2$, for $1 \le j \le NL-K$, $\mathbf{e}^j \ne \mathbf{i}^1, \ldots, \mathbf{i}^{k-1}$, compute

  $\mathbf{T}_n^k = \left[\mathbf{T}_n^{k-1} : \mathbf{e}^j\right]$

  $\mathrm{MSE}^{(\mathbf{e}^j)} = \mathbf{w}_q^\dagger \mathbf{R}_{nn}\mathbf{w}_q - \mathbf{w}_q^\dagger \mathbf{R}_{nn}\mathbf{C}_n\mathbf{T}_n^k(\mathbf{T}_n^{k\dagger}\mathbf{C}_n^\dagger\mathbf{R}_{nn}\mathbf{C}_n\mathbf{T}_n^k)^{-1}\mathbf{T}_n^{k\dagger}\mathbf{C}_n^\dagger\mathbf{R}_{nn}\mathbf{w}_q$

  Find $\mathrm{MSE}_k^{(\mathbf{i}^k)} = \min\{\mathrm{MSE}^{(\mathbf{e}^j)},\ 1 \le j \le NL-K, \mathbf{e}^j \ne \mathbf{i}^1, \ldots, \mathbf{i}^{k-1}\}$

  then select $\mathbf{T}_n^k = \left[\mathbf{T}_n^{k-1} : \mathbf{i}^k\right]$

- The procedure is terminated at the J^{th} step when

  $$\frac{\mathrm{MSE}_{NL-K}}{\mathrm{MSE}_J} > \rho$$

  where $0 < \rho < 1$ is some performance measure. This gives a beamformer of adaptive dimension J.

Table 1: *The selection algorithm*

IV. EXAMPLE

The performance of the new algorithm is now examined for the suppression of ground clutter received at an airborne array radar. A computer simulation shows the performance of this sparse algorithm as compared to a more conventional eigenstructure based technique. The clutter returns at a 16 element array with each element having 8 taps were computed. A GSC implementation with a 8 ($K = L$) linear constraints was employed, giving a solution space of fully adaptive dimension 120. The transformation matrix was designed with a covariance matrix that would result from an omnidirectional transmit pattern, i.e. one in which all Doppler frequencies along the diagonal of the Doppler–cosine azimuth are illuminated equally. Figure 1 shows the eigenspectra for the clutter covariance matrix. The step–like nature of the eigenspectra is useful in this application because the rank

of the interference covariance matrix is well defined. In this case the rank of the correlated portion of the interference covariance matrix is approximately 33. For partially adaptive dimensions less than this number we might expect the new algorithm to have superior performance.

Figure 2 compares the output MSE of the new sparse algorithm with that of the eigenstructure based technique during the training procedure. As can be seen, for low partially adaptive dimensions, the new algorithm has a lower output MSE than that of the eigenvector technique. What is also interesting is that the new algorithm tends to the fully adaptive MSE more quickly than the eigenvector approach. In fact, for an output MSE of -35.5dB only 43 DOF are required in the sparse design, as opposed to 79 in the eigenvector case. It should be remebered though, that the eigenstructure based design does this at a considerably greater implementational expense.

SUMMARY

The problem of selecting the degrees of freedom to be retained in a partially adaptive beamformer has been investigated. It has been shown that an iterative method which gives a sparse solution for the transformation matrix can reduce the required partially adaptive dimension, whilst also reducing the computational complexity of the partially adaptive beamformer. The usefulness of this technique has been demonstrated through computer simulation.

ACKNOWLEDGMENT

This work was supported by the Science and Engineering Research Council and by GEC Marconi Avionics Ltd, under grant number 91309742.

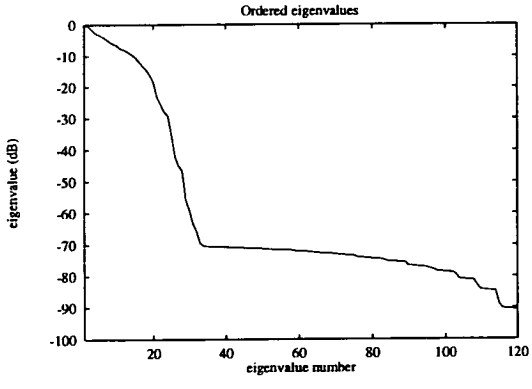

Figure 1: *Interference eigenstructure*

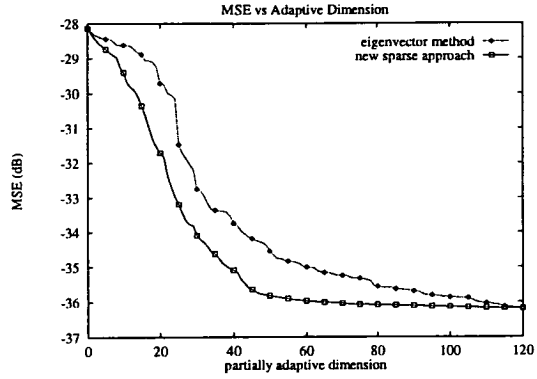

Figure 2: *Output mean square error for new sparse design and eigenstructure approach*

REFERENCES

[1] B.D. Van Veen, "Improved power minimization based partially adaptive beamformer design", presented at *ICASSP-88*, pages 2793–2796, April 1988.

[2] B.D. Van Veen, "Eigenstructure based partially adaptive array design", *IEEE Trans. Antennas and Propagation*, vol. AP-36, no. 3, pp. 357–362, March 1988.

[3] T.T. Ma and L.J. Griffiths, "A solution space approach to achieving partially adaptive arrays", presented at *ICASSP-88*, pages 2869–2872, April 1988.

[4] O.L. Frost III, "An algorithm for linearly constrained adaptive array processing", *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, August 1972.

[5] F. Qian and B.D. Van Veen, "Quadratically constrained adaptive beamforming in coherent interference environments", presented at *ICASSP-93*, pages 528–531, April 1993.

[6] S. Chen, C.F.N. Cowan, and P.M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks", *IEEE Trans. Neural Networks*, vol. NN-2, no. 2, pp. 302–309, March 1991.

A Sparse Approach to Partially Adaptive Airborne Radar

Iain SCOTT and Bernard MULGREW

Dept. of Electrical Eng., The University of Edinburgh, Edinburgh EH9 3JL,
Scotland, U.K., Tel/Fax: +44 [31] 650 5655 / 650 6554. E-Mail: is@uk.ac.ed.ee

Abstract. This paper is concerned with linearly constrained minimum variance (LCMV) beamforming. Partially adaptive LCMV beamformers are designed by determining a transformation which maps the fully adaptive weight space into a lower dimension partially adaptive weight space, usually so that some set of performance measures is optimised. One common method is to utilise the eigenvectors associated with the interference data covariance matrix. An iterative design technique which satisfies the dual goals of mimimum output mean squared error (MSE), and reduced adaptive dimension was first presented in [1]. This paper extends these results by considering the convergence performance of the resultant beamformer. Simulation results demonstrate that this iterative approach leads to a lower converged MSE whilst retaining simplicity in the beamforming structure.

1. Introduction

The computational requirement of each update in adaptive beamforming algorithms increases rapidly with the number of elements in the array. In many situations the beamformer will have an overly large number of degrees of freedom, "degrees of freedom" denoting the number of unconstrained or "free" weights that must be computed. For example, an LCMV beamformer with L constraints upon N elements has $N - L$ degrees of freedom, the generalised sidelobe canceller (GSC) [2] implementation would separate these into an unconstrained adaptive weight vector $\boldsymbol{w}_m$. A fully adaptive beamformer uses all of these degrees of freedom whilst a *partially adaptive* beamformer will utilise only a subset of these degrees of freedom. When the system has too many degrees of freedom several undesirable results arise:

(i) the system will require many iterations before convergence; and
(ii) the computational burden per iteration will increase quickly as the number of weights.

It is therefore of great importance that we reduce the number of degrees of freedom available to the processor. Fortunately, most interference suppression problems are rank deficient in nature, i.e. they require less adaptive degrees of freedom than are offered by the array. By appropriate design the additional degrees of freedom can be discarded so that only those that are important are retained. This is the goal in partially adaptive beamforming.

This paper will consider a method for designing partially adaptive beamformers first reported in [1]. Although only applied to the GSC in this paper, this technique can be used in a variety of prediction/estimation problems in which low-rank representations of signals are required. The case considered here is the suppression of ground clutter received at an airborne pulse–Doppler radar. This interference is two dimensional in nature, the clutter returns being a function of both azimuth (bearing) and Doppler (frequency).

In this paper we describe an iterative algorithm for specifying the degrees of freedom to be retained in a partially adaptive design. Additionally we examine the convergence performance using a simple least squares algorithm. The primary results reported here are the reduction in required adaptive dimension facilitated by an iterative solution and the improved convergence performance achieved by this structure. As such it offers several advantages over the eigenstructure approaches which are based upon the structure of the interference subspace. Simulation of a partially adaptive GSC illustrates the effectiveness of this structure.

2. Background

Let the N-dimensional vector $\boldsymbol{x}(k)$ denote the received data in the beamformer structure. The beamformer output $y(k)$ is formed as a linear combination of the components of $\boldsymbol{x}(k)$, i.e. $y(k) = \boldsymbol{w}^H\boldsymbol{x}(k)$. Here $\boldsymbol{w}$ is the weight vector and is typically chosen to minimise the output power whilst maintaining a specified response to the desired signal. Formally this is,

$$\min_{\boldsymbol{w}} \boldsymbol{w}^H\boldsymbol{R}_x\boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{C}^H\boldsymbol{w} = \boldsymbol{f}, \tag{1}$$

in which $\boldsymbol{R}_x$ is the data covariance matrix, $\boldsymbol{C}$ is an $N \times L$ constraint matrix, and $\boldsymbol{f}$ is an $N \times 1$ response vector. The superscript H indicates Hermitian transpose. The GSC implementation decomposes the weight vector $\boldsymbol{w}$

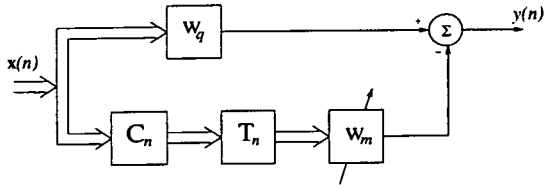

Figure 1: *Partially adaptive generalised sidelobe canceller.*

into two orthogonal components,

$$w = w_q - C_n w_m, \quad (2)$$

where $w_q \in range(C)$ satisfies $C^H w_q = f$, and the full rank $N \times (N-L)$ signal blocking matrix C_n satisfies $C_n^H C = 0$. The orthogonality ensures the desired signal is excluded from the adaptive portion of the beamformer. The $(N-L)$ dimensional weight vector w_m represents the available adaptive degrees of freedom and satisfies the minimisation problem

$$\min_{w_m} (w_q - C_n w_m)^H R_x (w_q - C_n w_m). \quad (3)$$

The optimal weight vector is given by

$$w_m = (C_n^H R_x C_n)^{-1} C_n^H R_x w_q. \quad (4)$$

In a partially adaptive GSC the number of available adaptive weights is reduced from $(N-L)$ to K. This is done by inserting the $(N-L) \times K$ matrix T_n after the signal blocking matrix (as in Fig. 1). If $T = C_n T_n$ is full rank then the partially adaptive GSC weight vector is $w = w_q - T w_m$, for which the optimal solution is

$$w_m = (T^H R_x T)^{-1} T^H R_x w_q. \quad (5)$$

The data present in the array consists of a portion $s(k)$ due to the desired signal and a portion $n(k)$ due to the interference and noise, i.e. $x(k) = s(k) + n(k)$. Let $s(k)$ denote the desired signal at the beamformer output. We assume the constraints $C^H w = f$ are chosen to ensure that w passes the desired signal with unit gain. Thus, $s(k) = w_q^H s(k) = w^H s(k)$. In other words, $s(k)$ is the beamformer output in the absence of noise and interference, $n(k) = 0$. Now, the output mean squared error is defined as

$$e_n = E\left\{\left|s(k) - w^H x(k)\right|^2\right\}. \quad (6)$$

Substitution of $x(k) = s(k) + n(k)$ and application of the constraint $s(k) = w^H s(k)$ yield

$$\begin{aligned} e_n &= E\left\{\left|s(k) - w^H s(k) - w^H n(k)\right|^2\right\} \\ &= E\left\{\left|w^H n(k)\right|^2\right\} \\ &= w^H R_n w, \end{aligned} \quad (7)$$

in which R_n is the covariance matrix of noise and interference. For the partially adaptive GSC the MSE is given by

$$\begin{aligned} e_n &= w_q^H R_n w_q \\ &\quad - w_q^H R_n T \left(T^H R_n T\right)^{-1} T^H R_n w_q. \end{aligned} \quad (8)$$

The output MSE therefore consists of two components which are related specifically to the adaptive and non-adaptive processing paths in the GSC.

3. Beamformer Design

Partially adaptive techniques are used to reduce the implemental and computational complexity of adaptive beamformers. It is therefore of primary importance that we maintain a simple structure for the matrix operations shown in Fig. 1. Otherwise the reduction in adaptive dimension will be negated by the over-complexity of the network effecting this reduction. The two important elements of Fig. 1 are the signal blocking matrix C_n and the transformation matrix T_n – together they form the transformation T. If each of the components C_n and T_n are simple then the overall beamforming structure will be easy to implement.

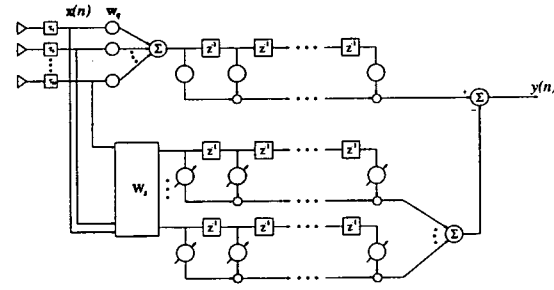

Figure 2: *The GSC broadband beamformer.*

Fig. 2 depicts a physical realisation of the basic GSC structure. Each circle indicates a weight and each square box a time delay. The wideband steering delays τ_i steer the elemental outputs so that the desired signal appears identically at the input to w_q and W_s. The quiescent weight vector w_q is a simple summation ($w_q = 1$). The subsequent tap weights are used to identify the Doppler of target signals. The matrix W_s performs the signal blocking operation by simply differencing the outputs of the steering delays. Typically the matrix W_s has the form

$$W_s = \begin{bmatrix} 1 & & & & & & \\ -1 & 1 & & & 0 & & \\ & -1 & 1 & & & & \\ & & & \ddots & & & \\ & & & & -1 & 1 & \\ & & 0 & & & -1 & 1 \\ & & & & & & -1 \end{bmatrix} \quad (9)$$

This is a sparse bidiagonal matrix in which the only non–zero entries are either 1 or -1. The iteratively designed transformation matrix described in [1] identifies those adaptive weights in the lower path that have greatest influence on the output, and sets the remainder to zero. This implies that $\boldsymbol{T}_n$ will be a matrix in which each column has only one non–zero entry of value 1.

A large array will have a transformation matrix which contains a large number of free elements. Global optimisation over each element is therefore an unrealistic proposition. Van Veen and others considered this in several papers [3,4] and suggested employing iterative techniques. The algorithm described in [1] is one example of these iterative techniques which is particularly pertinent to our current problem. A mathematical definition of the algorithm can be found in this reference, but it will be useful to summarise it here.

At the outset the matrix $\boldsymbol{T}_n$ has dimension zero, i.e. no columns (weights) have been selected. At each iteration the algorithm appends each of the remaining weights in turn to $\boldsymbol{T}_n$ and evaluates the output MSE. The weight that achieves the best reduction in output MSE is then selected. This process is iterated until either the output MSE has reached an acceptable level, or until the required partially adaptive dimension has been reached. For comparison, we will also consider the eigenstructure based beamformer described in [4]. In this design the transformation is non–sparse, being formed from the eigenvectors of the matrix $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$.

4. Training

The scenario considered within this paper is that of an airborne look down pulse–Doppler radar. The aim is to identify low flying aircraft in both bearing and frequency. Ground clutter echoes are generated as the summation of independent identically distributed Gaussian scatterers on a single radar range ring. A GSC beamformer with sixteen elements in a linear equispaced geometry, with eight tap FIR filters in each channel is employed. Fig. 3 shows the ordered eigenspectrum for the matrix $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$. In this case the rank would appear to be between 20 and 30.

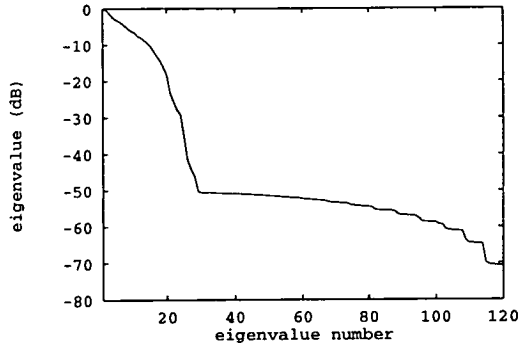

Figure 3: *Eigenspectrum for the correlated portion of* $\boldsymbol{C}_n^H \boldsymbol{R}_n \boldsymbol{C}_n$.

Fig. 4 compares the output MSE of the iterative approach with that of the eigenstructure technique during the training procedure. The iterative technique performs significantly better - for an output MSE of -35.5dB only 43 degrees of freedom are required in the iterative design, as opposed to 78 in the eigenstructure beamformer.

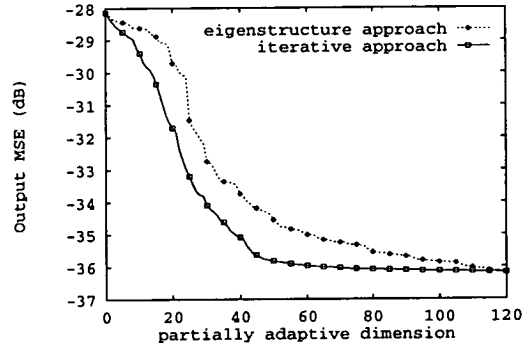

Figure 4: *Output mean square error for iterative design and eigenstructure approach during training phase.*

5. Convergence Performance

In practice the covariance matrix $\boldsymbol{R}_n$ is unknown and must be estimated from the data. Assume that there are M data vectors $\boldsymbol{n}(n)$, $n = 1, 2, \cdots, M$ available. The sample covariance matrix estimate of $\boldsymbol{R}_n$ is

$$\hat{\boldsymbol{R}}_n = \frac{1}{M} \sum_{n=1}^{M} \boldsymbol{n}(n) \boldsymbol{n}^H(n). \quad (10)$$

Additionally we define the data matrix as $\boldsymbol{X} = [\boldsymbol{n}(1)\,\boldsymbol{n}(2)\cdots\boldsymbol{n}(M)]$, then (10) can be rewritten in matrix form as

$$\hat{\boldsymbol{R}}_n = \frac{1}{M}\boldsymbol{X}\boldsymbol{X}^H. \tag{11}$$

Substitution of (11) in (8) yields the sample MSE $\hat{e}_n$ as

$$\hat{e}_n = \frac{1}{M}\left\{\boldsymbol{w}_q^H\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{w}_q - \boldsymbol{w}_q^H\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{T}\left(\boldsymbol{T}^H\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{T}\right)^{-1}\boldsymbol{T}^H\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{w}_q\right\}. \tag{12}$$

These expressions can be used as the basis of an unweighted least squares problem. In [5] an expression for the mean value of the sample MSE was given as

$$E\{\hat{e}_n\} = \frac{M-K}{M}e_n, \tag{13}$$

where K is the adaptive dimension as before. The ratio $(M-K)/M$ determines the adaptive convergence of the mean when viewed as a function of the number of snapshots M. Eqn. (13) shows that the expected value of the excess MSE will be within 3dB of the optimum after $M = 2K$ data vectors. However, the eigenspectrum of Fig. 3 demonstrates the rank deficient nature of this problem. The adaptive weight vector will lie in a space whose dimension is no larger than the input data. We would therefore expect $E\{\hat{e}_n\}$ to be within 3dB of the optimum after approximately 50 snapshots.

Sample MSE is plotted against M for a single realisation in Fig. 5. Both beamformer types have an adaptive dimension of 43. For comparison $E\{\hat{e}_n\}$ is drawn for the case $K = 23$. The estimated covariance matrix is rendered invertible by augmenting the leading diagonal by a small term as follows

$$\hat{\boldsymbol{R}}_n = \lambda\boldsymbol{I} + \frac{1}{M}\boldsymbol{X}\boldsymbol{X}^H. \tag{14}$$

In practice λ should be made equal to the mean value of uncorrelated noise. All three curves have attained near optimum performance after 46 snapshots. Both beamformer types differ only slightly from the mean, verifying the initial estimate of the clutter dimensionality made in section 4. The iteratively designed beamformer has a lower final value of MSE validating it as a design technique.

6. Summary

The design of a partially adaptive beamformer which is effective in cancelling two dimensional interferences has been considered. It has been shown that an iterative method which gives a sparse solution for the transformation matrix can reduce the required partially adaptive dimension, whilst also ensuring a simple beamforming structure. The convergence performance of the resultant beamformer has also been examined and demonstrated through computer simulation.

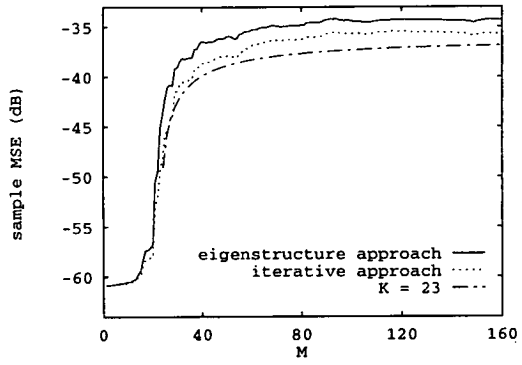

Figure 5: *Output sample mean square error of iterative and eigenstructure beamformers versus data matrix size.*

Acknowledgement

The authors wish to thank the Science and Engineering Research Council and GEC Marconi Avionics Ltd, Edinburgh, for their financial support.

References

[1] I. Scott and B. Mulgrew, "A sparse approach in partially adaptive linearly constrained arrays", presented at *ICASSP-94*, April 1994.

[2] L.J. Griffiths and C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming", *IEEE Trans. Antennas and Propagation*, vol. AP-30, no. 1, pp. 27–34, January 1982.

[3] B.D. Van Veen and R.A. Roberts, "Partially adaptive beamformer design via output power minimization", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-35, no. 11, pp. 1524–1532, November 1987.

[4] B.D. Van Veen, "Eigenstructure based partially adaptive array design", *IEEE Trans. Antennas and Propagation*, vol. AP-36, no. 3, pp. 357–362, March 1988.

[5] B.D. Van Veen, "Adaptive convergence of linearly constrained beamformers based on the sample covariance matrix", *IEEE Trans. Signal Processing*, vol. 39, no. 6, pp. 1470–1473, June 1991.

Appendix D

Additional results

This appendix contains additional simulation results for the selection algorithm presented in chapter 4. Four additional scenarios are considered. A variety of beamformer dimensions are examined, along with the influence of range, velocity and depression angle. Table D.1 summarises the simulation parameters. All other radar parameters are as in Table 3.1. For brevity, the estimated rank of the correlated portion of the matrix $\boldsymbol{C}_n^H \bar{\boldsymbol{R}}_n \boldsymbol{C}_n$ has been included in Table D.1.

	Scenario 4	**Scenario 5**	**Scenario 6**	**Scenario 7**
NL, K	4, 4	8, 4	8, 8	16, 8
range	$1100m$	$2000m$	$3000m$	$1200m$
depression angle	65.4°	30°	19.5°	56.4°
velocity	$150ms^{-1}$	$100ms^{-1}$	$100ms^{-1}$	$150ms^{-1}$
rank $\boldsymbol{C}_n^H \hat{\boldsymbol{R}}_n \boldsymbol{C}_n$	8	14	19	35

Table D.1: Parameters for additional training scenarios.

	Eigenstructure			**Sparse**			**Saving %**	
Scenario	J_e	Multiplies	Additions/ Subtract.	J_s	Multiplies	Additions/ Subtract.	Multiplies	Additions/ Subtract.
4	11	11011	11011	11	4323	7612	61	31
5	25	107625	107625	24	39168	65856	64	39
6	50	857800	857800	30	151800	279420	83	68
7	90	5763240	5763240	45	834165	1584990	86	73

Table D.2: Operational expense of eigenstructure and iterative beamformers - Scenarios 4–7.

The performance of the design procedures are depicted in Figures D.1–D.4. It is apparent that all the additional scenarios exhibit similar performance to that found in chapter 4. The new sparse selection approach attains a lower output MSE at virtually all partially adaptive dimensions, which suggests that this technique is robust to changing radar parameters and

beamformer dimensions. The greatest improvements in performance occur for the larger beamformers. Table D.2 shows the operational expense of the sparse and eigenstructure beamformers found for scenarios 4–7. The operational expenses were computed from (4.40) and (4.41).

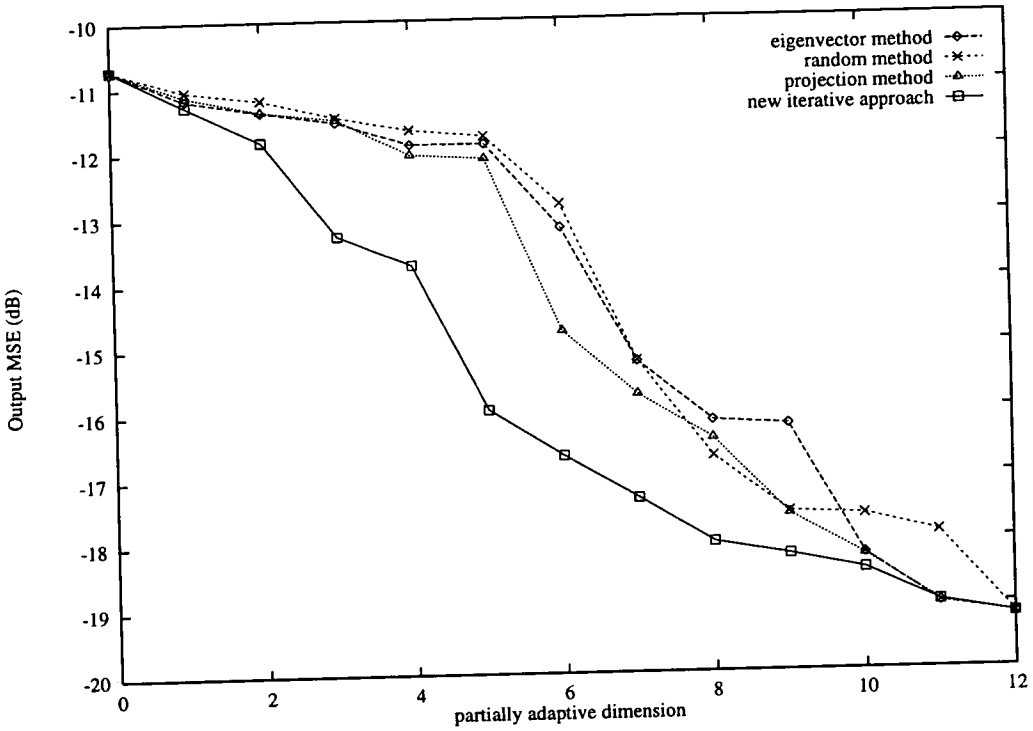

Figure D.1: *Output mean squared error for new iterative design and existing techniques during training phase - Scenario 4.*

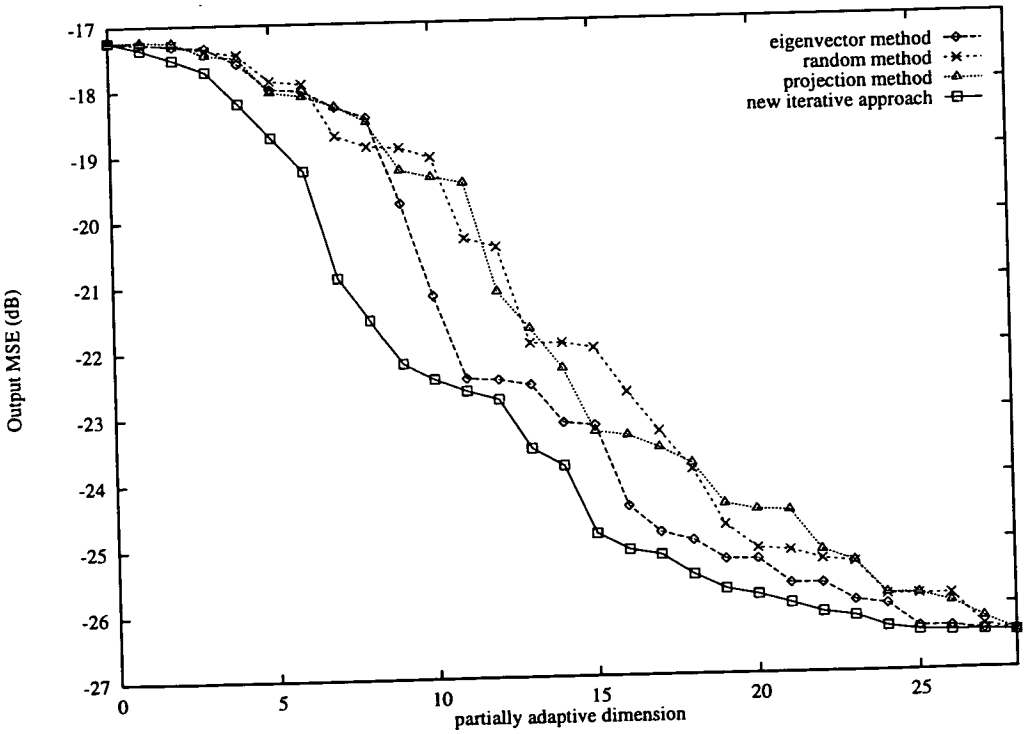

Figure D.2: *Output mean squared error for new iterative design and existing techniques during training phase - Scenario 5.*

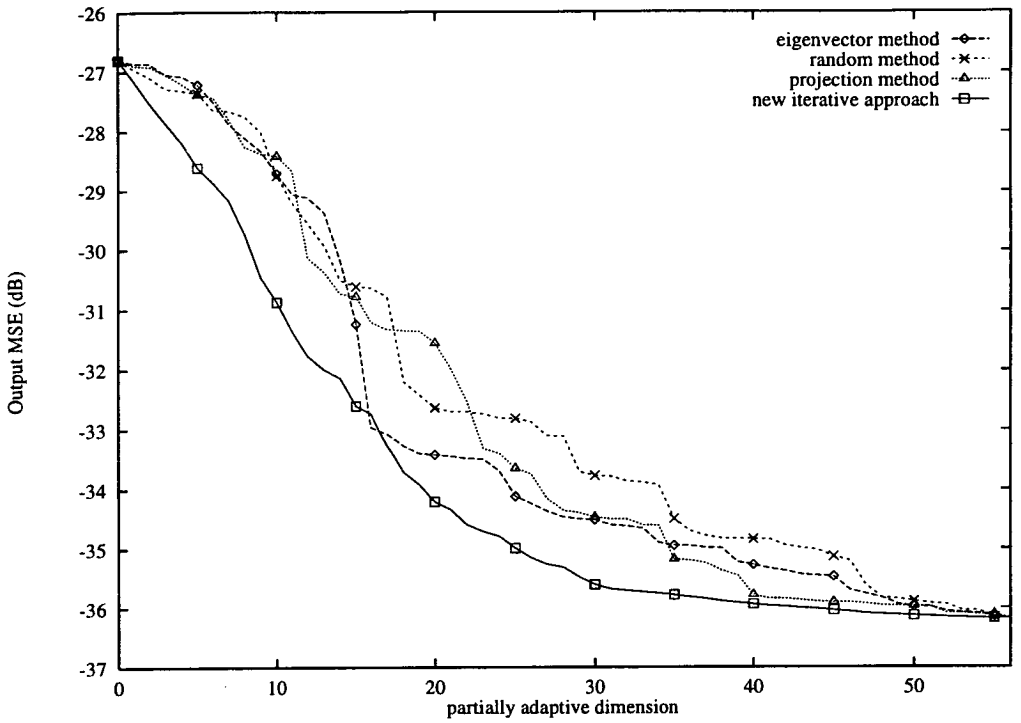

Figure D.3: *Output mean squared error for new iterative design and existing techniques during training phase - Scenario 6.*

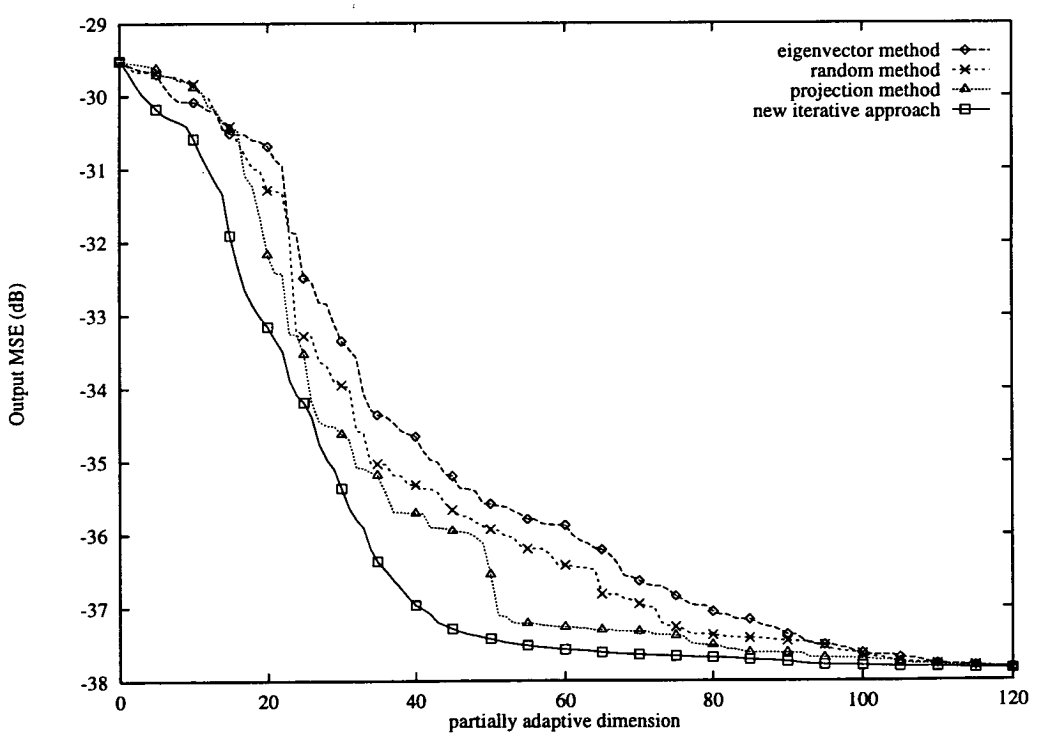

Figure D.4: *Output mean squared error for new iterative design and existing techniques during training phase - Scenario 7.*