

Beamforming: A Versatile Approach to Spatial Filtering

Barry D. Van Veen and Kevin M. Buckley

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 wave fields, which are processed by the beamformer. The objective is to estimate the signal arriving from a desired direction in the presence of noise and interfering signals. A beamformer performs spatial filtering to separate signals that have overlapping frequency content but originate from different spatial locations. This paper provides an overview of beamforming from a signal processing perspective, with an emphasis on recent research. Data independent, statistically optimum, adaptive, and partially adaptive beamforming are discussed.

I. INTRODUCTION

The term *beamforming* derives from the fact that early spatial filters were designed to form pencil beams (see polar plot in Fig. 1.1) in order to receive a signal radiating from a specific location and attenuate signals from other locations. "Forming beams" seems to indicate radiation of energy; however, beamforming is applicable to either radiation or reception of energy. In this paper we discuss formation of beams for reception.

Systems designed to receive spatially propagating signals often encounter the presence of interference signals. If the desired signal and interferers occupy the same temporal frequency band, then temporal filtering cannot be used to separate signal from interference. However, the desired and interfering signals usually originate from different spatial locations. This spatial separation can be exploited to separate signal from interference using a spatial filter at the receiver. Implementing a temporal filter requires processing of data collected over a temporal aperture. Similarly, implementing a spatial filter requires processing of data collected over a spatial aperture.

Several applications that employ spatial filtering of data are listed in Table 1.1. Fig. 1.1 illustrates a microwave communications antenna that employs a continuous spatial aperture to accomplish spatial filtering with a single antenna. Fig. 1.2 depicts a low frequency-towed sonar array in which the spatial aperture is obtained through a discrete spatial sampling by an array of sensors. When the spatial sampling is discrete, the processor that performs the spatial filtering is termed a *beamformer*. Typically a

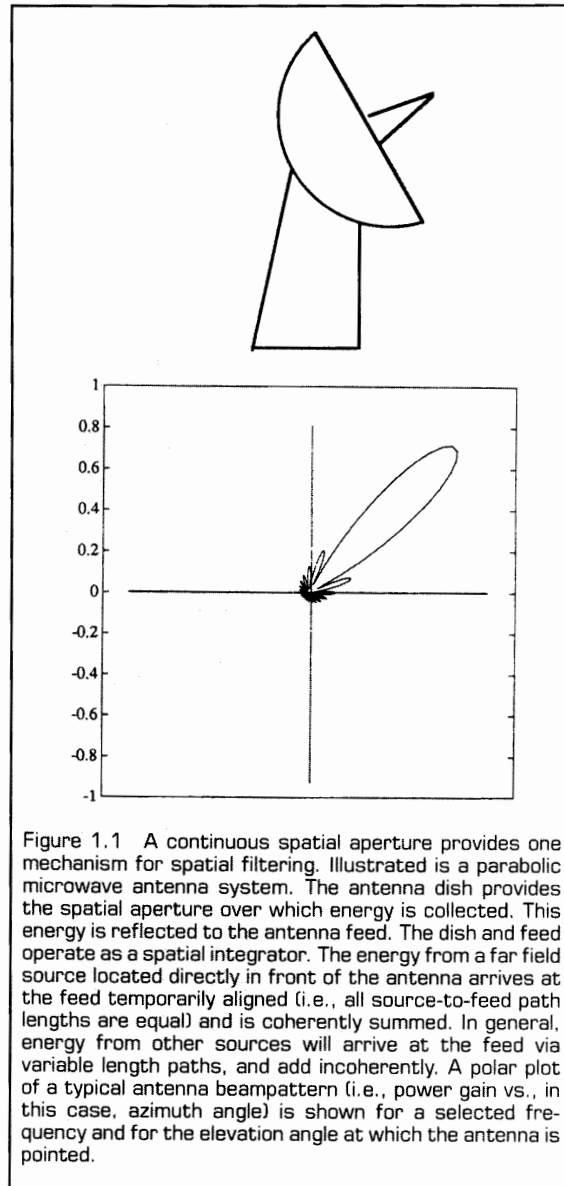


Figure 1.1 A continuous spatial aperture provides one mechanism for spatial filtering. Illustrated is a parabolic microwave antenna system. The antenna dish provides the spatial aperture over which energy is collected. This energy is reflected to the antenna feed. The dish and feed operate as a spatial integrator. The energy from a far field source located directly in front of the antenna arrives at the feed temporarily aligned (i.e., all source-to-feed path lengths are equal) and is coherently summed. In general, energy from other sources will arrive at the feed via variable length paths, and add incoherently. A polar plot of a typical antenna beam pattern (i.e., power gain vs., in this case, azimuth angle) is shown for a selected frequency and for the elevation angle at which the antenna is pointed.

0740-7467/88/0400-0004\$01.00©1988IEEE

TABLE 1.1

ARRAYS AND BEAMFORMERS PROVIDE AN EFFECTIVE AND VERSATILE MEANS OF SPATIAL FILTERING. THIS TABLE LISTS A NUMBER OF APPLICATIONS OF SPATIAL FILTERING, GIVES EXAMPLES OF ARRAYS AND BEAMFORMERS, AND PROVIDES A FEW KEY REFERENCES.

Application	Description	References
RADAR	phased-array RADAR; air traffic control; synthetic aperture RADAR	Brookner [1985]; Haykin [1985]; Munson et al. [1983]
SONAR	source localization and classification	Knight et al. [1981]; Owsley [1985]
Communications	directional transmission and reception; sector broadcast in satellite communications	Mayhan [1976]; Compton [1978]; Adams et al. [1980]
Imaging	ultrasonic; optical; tomographic	Macovski [1983]; Pratt [1978]; Kak [1985]
Geophysical Exploration	earth crust mapping; oil exploration	Justice [1985]
Astrophysical Exploration	high resolution imaging of the universe	Readhead [1982]; Yen [1985]
Biomedical	fetal heart monitoring; tissue hyperthermia; hearing aids	Widrow et al. [1975]; Gee et al. [1984]; Peterson et al. [1987]

beamformer linearly combines the spatially sampled time series from each sensor to obtain a scalar output time series in the same manner that an FIR filter linearly combines temporally sampled data. Two principal advantages of spatial sampling with an array of sensors are discussed below.

Spatial discrimination capability depends on the size of the spatial aperture; as the aperture increases, discrimination improves. The absolute aperture size is not important, rather its size in wavelengths is the critical parameter. A single physical antenna (continuous spatial aperture) capable of providing the requisite discrimination is often practical for high frequency signals since the wavelength is short. However, when low frequency signals are of interest, an array of sensors can often synthesize a much larger spatial aperture than that practical with a single physical antenna.

A second very significant advantage of using an array of sensors, relevant at any wavelength, is the spatial filtering versatility offered by discrete sampling. In many application areas it is necessary to change the spatial filtering function in real time to maintain effective suppression of interfering signals. This change is easily implemented in a discretely sampled system by changing the way in which the beamformer linearly combines the sensor data. Changing the spatial filtering function of a continuous aperture antenna is impractical.

The purpose of this paper is to describe beamforming from a signal processing perspective, provide an overview of beamformer design, and briefly discuss performance and implementation issues with an emphasis on recent research. The paper begins with a section devoted to

defining basic terminology, notation, and concepts. Subsequent sections cover data-independent, statistically optimum, adaptive, and partially adaptive beamforming. We then provide a brief discussion of implementation issues and conclude with a summary.

Throughout the paper we use familiar methods and techniques from FIR filtering to provide insight into various aspects of spatial filtering with a beamformer. However, in some ways beamforming differs significantly from FIR filtering. For example, in beamforming a source of energy has several parameters that can be of interest: range, azimuth and elevation angles, polarization, and temporal frequency content. Different signals are often mutually correlated as a result of multipath propagation. The spatial sampling is often nonuniform and multi-dimensional. Uncertainty must often be included in characterization of individual sensor response and location, motivating development of robust beamforming techniques. These differences indicate that beamforming represents a more general problem than FIR filtering and as a result, more general design procedures and processing structures are common.

Rather than making a futile attempt at attributing developments due to many different researchers in beamforming, we refer the reader to the following references: books—J. W. R. Griffiths et al., ed. [1973], Hudson [1981], Monzingo and Miller [1980], Haykin, ed. [1985], Compton [1988]; special issues—*IEEE Transactions on Antennas and Propagation* [1976], [1986], *Journal of Ocean Engineering* [1987]; tutorial—Gabriel [1976]; and bibliography—Marr [1986]. Papers devoted to beamforming are often found in the *IEEE Transactions on: Antennas and Propagation*,

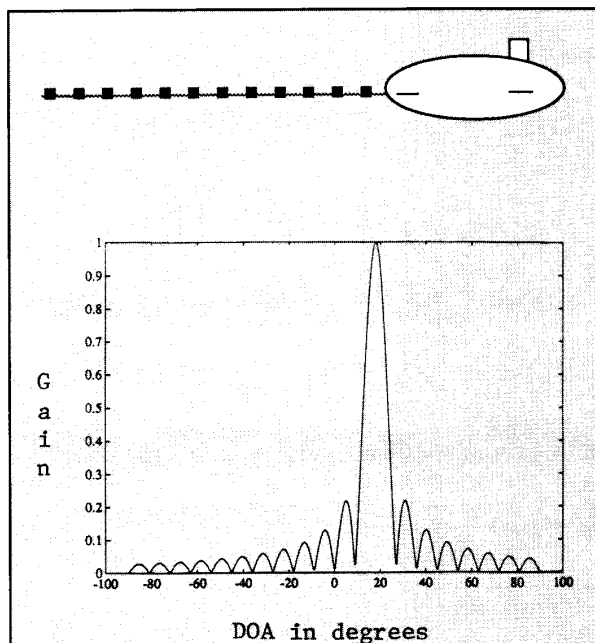


Figure 1.2 This figure exemplifies spatial filtering with a discrete aperture from a towed SONAR array of equi-spaced sensors. The signals received at the sensors due to far field sources located at array broadside (i.e., from along the plane perpendicular to the array axis) will be temporarily aligned across the array, assuming the sensors are identical and propagation is isotropic. Summing the sensor outputs results in a coherent gain similar to that demonstrated in Figure 1.1. If a source from another Direction-Of-Arrival (*DOA*) is of interest, delays can be electronically implemented at the sensor outputs to compensate for that *DOA*'s intersensor propagation delays, thereby temporarily aligning the source. The delays and summing device constitute a beamformer. A more versatile approach is to use filters at the sensor outputs and sum the filter outputs. A rectangular plot of the magnitude vs. *DOA* is shown for a phase shift and sum beamformer. A linear equi-spaced array of 16 omnidirectional sensors is employed. The beamformer is "steered" to 18°.

Acoustics, Speech, and Signal Processing, Aerospace and Electronic Systems, and in the *Journal of the Acoustical Society of America*. There is a vast body of literature on various aspects of beamforming and we can only reference a subset in support of our discussions. We often refer to the FIR filtering literature in our discussions of beamforming, since their histories are both parallel and overlapping.

II. BASIC TERMINOLOGY AND CONCEPTS

In this section we introduce terminology and concepts employed throughout the paper. We begin with a subsection that defines the beamforming operation and discusses spatial filtering. The next subsection, entitled "Second Order Statistics," develops representations for the covariance of the data received at the array and discusses dis-

tinctions between narrowband and broadband beamforming. The final subsection defines various types of beamformers.

A. Beamforming and Spatial Filtering

Fig. 2.1 depicts two beamformers. The first, which samples the propagating wave field in space, is typically used for processing narrowband signals. The output at time k , $y(k)$, is given by a linear combination of the data at the J sensors at time k :

$$y(k) = \sum_{i=1}^J w_i^* x_i(k) \quad (2.1)$$

where $*$ represents complex conjugate. It is conventional to multiply the data by conjugates of the weights to simplify notation. We assume throughout that the data and weights are complex since in many applications a quadrature receiver is used at each sensor to generate in phase and quadrature (I and Q) data. Each sensor is assumed to

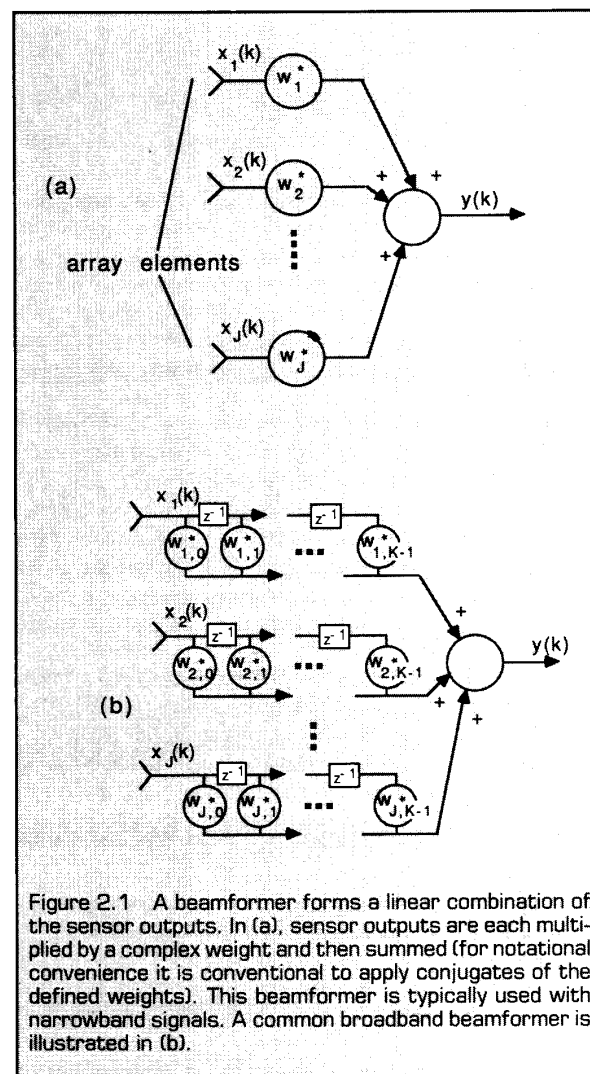


Figure 2.1 A beamformer forms a linear combination of the sensor outputs. In (a), sensor outputs are each multiplied by a complex weight and then summed (for notational convenience it is conventional to apply conjugates of the defined weights). This beamformer is typically used with narrowband signals. A common broadband beamformer is illustrated in (b).

have any necessary receiver electronics and an A/D converter if beamforming is performed digitally.

The second beamformer in Fig. 2.1 samples the propagating wave field in both space and time and is often used when signals of significant frequency extent (broadband) are of interest. The output in this case can be expressed as

$$y(k) = \sum_{l=1}^J \sum_{p=0}^{K-1} w_{l,p}^* x_l(k-p) \quad (2.2)$$

where $K-1$ is the number of delays in each of the J sensor channels. If the signal at each sensor is viewed as an input, then a beamformer represents a multi-input single output system.

It is convenient to develop notation which permits us to treat both beamformers in Fig. 2.1 simultaneously. Note that (2.1) and (2.2) can be written as

$$y(k) = \mathbf{w}^H \mathbf{x}(k). \quad (2.3)$$

by appropriately defining a weight vector $\mathbf{w}$ and data vector $\mathbf{x}(k)$. We use lower and upper case boldface to denote vector and matrix quantities respectively, and let superscript H represent Hermitian (complex conjugate) transpose. Vectors are assumed to be column vectors. Assume that $\mathbf{w}$ and $\mathbf{x}(k)$ are N dimensional; this implies that $N = KJ$ when referring to (2.2) and $N = J$ when referring to (2.1). Except for Section V on adaptive algorithms, we will drop the time index and assume that its presence is understood throughout the remainder of the paper. Thus (2.3) is written as $y = \mathbf{w}^H \mathbf{x}$. Many of the techniques described in this paper are applicable to continuous time as well as discrete time beamforming.

The frequency response of an FIR filter with tap weights w_p^* , $0 \leq p \leq J$ and a tap delay of T seconds is given by

$$r(\omega) = \sum_{p=1}^J w_p^* e^{-j\omega T(p-1)}. \quad (2.4a)$$

Alternatively

$$r(\omega) = \mathbf{w}^H \mathbf{d}(\omega) \quad (2.4b)$$

where $\mathbf{w}^H = [w_1^* w_2^* \dots w_J^*]$ and $\mathbf{d}(\omega) = [1 e^{j\omega T} e^{j2\omega T} \dots e^{j(J-1)\omega T}]^H$. $r(\omega)$ represents the response of the filter to a complex sinusoid of frequency ω and $\mathbf{d}(\omega)$ is a vector describing the phase of the complex sinusoid at each tap in the FIR filter relative to the tap associated with w_1 .

Similarly, beamformer response is defined as the amplitude and phase presented to a complex plane wave as a function of location and frequency. Location is in general a three dimensional quantity, but often we are only concerned with one or two dimensional direction of arrival (DOA). Throughout the remainder of the paper we do not consider range. Fig. 2.2 illustrates the manner in which an array of sensors samples a spatially propagating signal. Assume that the signal is a complex plane wave with DOA θ and frequency ω . For convenience let the phase be zero at the first sensor. This implies $x_1(k) = e^{j\omega k}$ and $x_l(k) = e^{j\omega[k - \Delta_l(\theta)]}$, $2 \leq l \leq J$. $\Delta_l(\theta)$ represents the time delay due to propagation from the first to the l^{th} sensor. Substitution into (2.2) results in the beamformer output

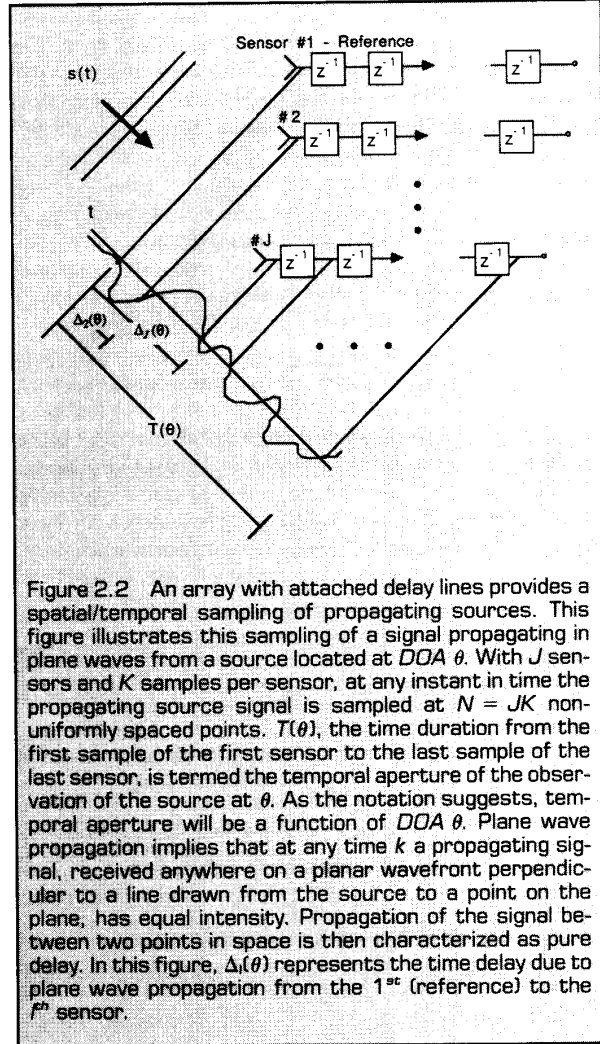


Figure 2.2 An array with attached delay lines provides a spatial/temporal sampling of propagating sources. This figure illustrates this sampling of a signal propagating in plane waves from a source located at DOA θ . With J sensors and K samples per sensor, at any instant in time the propagating source signal is sampled at $N = JK$ non-uniformly spaced points. $T(\theta)$, the time duration from the first sample of the first sensor to the last sample of the last sensor, is termed the temporal aperture of the observation of the source at θ . As the notation suggests, temporal aperture will be a function of DOA θ . Plane wave propagation implies that at any time k a propagating signal, received anywhere on a planar wavefront perpendicular to a line drawn from the source to a point on the plane, has equal intensity. Propagation of the signal between two points in space is then characterized as pure delay. In this figure, $\Delta_l(\theta)$ represents the time delay due to plane wave propagation from the 1st (reference) to the l^{th} sensor.

$$y(k) = e^{j\omega k} \sum_{l=1}^J \sum_{p=0}^{K-1} w_{l,p}^* e^{-j\omega[\Delta_l(\theta) + p]} = e^{j\omega k} r(\theta, \omega) \quad (2.5)$$

where $\Delta_1(\theta) = 0$. $r(\theta, \omega)$ is the beamformer response and can be expressed in vector form as

$$r(\theta, \omega) = \mathbf{w}^H \mathbf{d}(\theta, \omega). \quad (2.6)$$

The elements of $\mathbf{d}(\theta, \omega)$ correspond to the complex exponentials $e^{-j\omega[\Delta_l(\theta) + p]}$. In general it can be expressed as

$$\mathbf{d}(\theta, \omega) = [1 e^{j\omega\tau_2(\theta)} e^{j\omega\tau_3(\theta)} \dots e^{j\omega\tau_N(\theta)}]^H \quad (2.7)$$

where the $\tau_i(\theta)$, $2 \leq i \leq N$, are the time delays due to propagation and any tap delays from the zero phase reference to the point at which the i^{th} weight is applied. We refer to $\mathbf{d}(\theta, \omega)$ as the array response vector. It is also known as the steering vector or direction vector. Non-ideal sensor characteristics can be incorporated into $\mathbf{d}(\theta, \omega)$ by multiplying each phase shift by a function $a_i(\theta, \omega)$, which describes the associated sensor response as a function of frequency and direction.

The *beampattern* is defined as the magnitude squared of $r(\theta, \omega)$. Note that each weight in $\mathbf{w}$ affects both the temporal and spatial response of the beamformer. Historically, use of FIR filters has been viewed as providing frequency dependent weights in each channel. This interpretation is accurate but somewhat incomplete since the coefficients in each filter also influence the spatial filtering characteristics of the beamformer. As a multi-input single output system, the spatial and temporal filtering that occurs is a result of mutual interaction between spatial and temporal sampling.

The correspondence between FIR filtering and beamforming is closest when the beamformer operates at a single temporal frequency ω_0 and the array geometry is linear and equi-spaced as illustrated in Fig. 2.3. Letting the sensor spacing be d , propagation velocity be c , and θ represent DOA relative to broadside (perpendicular to the array) we have $\tau_i(\theta) = (i - 1)(d/c) \sin \theta$. In this case we identify the relationship between temporal frequency ω in $\mathbf{d}(\omega)$ (FIR filter) and direction θ in $\mathbf{d}(\theta, \omega_0)$ (beamformer) as $\omega = \omega_0(d/c) \sin \theta$. Thus, temporal frequency in an FIR fil-

ter corresponds to the sine of direction in a narrowband, linear equi-spaced beamformer. Complete interchange of beamforming and FIR filtering methods is possible for this special case provided the mapping between frequency and direction is accounted for.

The vector notation introduced in (2.6) suggests a vector space interpretation of beamforming. This point of view is useful both in beamformer design and analysis. We use it here in consideration of spatial sampling and array geometry. The weight vector $\mathbf{w}$ and the array response vectors $\mathbf{d}(\theta, \omega)$ are vectors in an N dimensional vector space. The angles between $\mathbf{w}$ and $\mathbf{d}(\theta, \omega)$ determine the response $r(\theta, \omega)$. For example, if for some (θ, ω) the angle between $\mathbf{w}$ and $\mathbf{d}(\theta, \omega)$ is 90° (i.e., if $\mathbf{w}$ is orthogonal to $\mathbf{d}(\theta, \omega)$), then the response is zero. If the angle is close to 0° , then the response magnitude will be relatively large. The ability to discriminate between sources at different locations and/or frequencies, say (θ_1, ω_1) and (θ_2, ω_2) , is determined by the angle between their array response vectors, $\mathbf{d}(\theta_1, \omega_1)$ and $\mathbf{d}(\theta_2, \omega_2)$ (Cox [1973]).

The general effects of spatial sampling are similar to temporal sampling. Spatial aliasing corresponds to an ambiguity in source locations. The implication is that sources at different locations have the same array response vector, e.g., for narrowband sources $\mathbf{d}(\theta_1, \omega_0) = \mathbf{d}(\theta_2, \omega_0)$. This can occur if the sensors are spaced too far apart. If the sensors are too close together, spatial discrimination suffers as a result of the smaller than necessary aperture; array response vectors are not well dispersed in the N dimensional vector space. Another type of ambiguity occurs with broadband signals when a source at one location and frequency cannot be distinguished from a source at a different location and frequency, i.e., $\mathbf{d}(\theta_1, \omega_1) = \mathbf{d}(\theta_2, \omega_2)$. For example, this occurs in a linear equi-spaced array whenever $\omega_1 \sin \theta_1 = \omega_2 \sin \theta_2$. (The addition of temporal samples at one sample prevents this particular ambiguity.)

A primary focus of this paper is on designing response via weight selection; however, (2.6) indicates that response is also a function of array geometry (and sensor characteristics if the ideal omnidirectional sensor model is invalid). In contrast with single channel filtering where A/D converters provide a uniform sampling in time, there is no compelling reason to space sensors regularly. Sensor locations provide additional degrees of freedom in designing a desired response and can be selected so that over the range of (θ, ω) of interest the array response vectors are unambiguous and well dispersed in the N dimensional vector space. Utilization of these degrees of freedom can become very complicated due to the multidimensional nature of spatial sampling and the nonlinear relationship between $r(\theta, \omega)$ and sensor locations. References discussing array geometry design for response synthesis include Unz [1956], Harrington [1961], Ishimaru [1962], Lo [1963], and Skolnik et al. [1964].

B. Second Order Statistics

Evaluation of beamformer performance usually involves power or variance, so the second order statistics of the

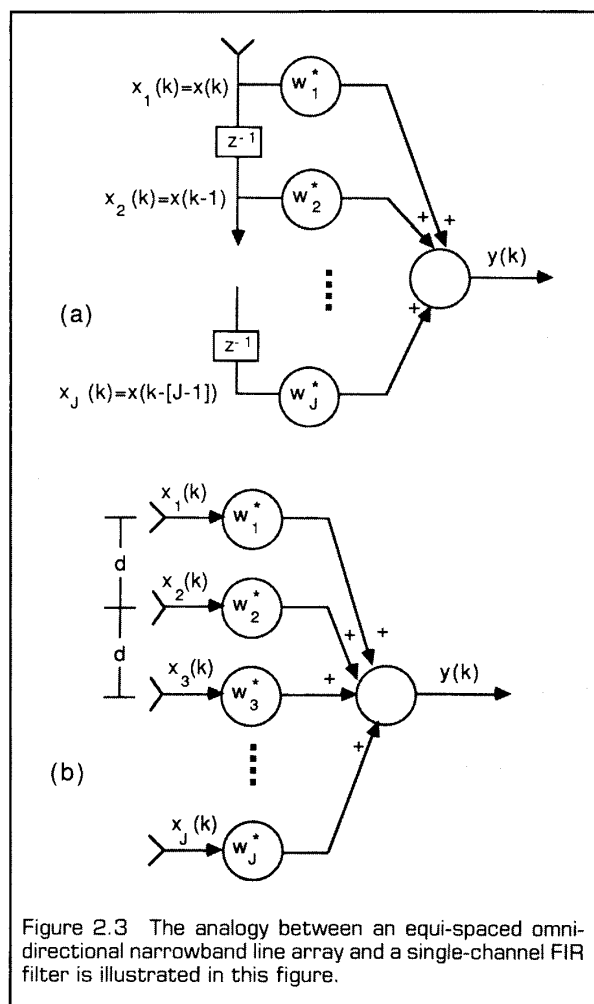


Figure 2.3 The analogy between an equi-spaced omnidirectional narrowband line array and a single-channel FIR filter is illustrated in this figure.

data play an important role. We assume the data received at the sensors is zero mean throughout the paper. The variance or expected power of the beamformer output is given by $E\{|y|^2\} = \mathbf{w}^H E\{\mathbf{x}\mathbf{x}^H\} \mathbf{w}$. If the data is wide sense stationary, then $\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^H\}$, the data covariance matrix, is independent of time. Although we often encounter non-stationary data, the wide sense stationary assumption is used in developing statistically optimal beamformers and in evaluating steady state performance.

Suppose $\mathbf{x}$ represents samples from a uniformly sampled time series having a power spectral density $S(\omega)$ and no energy outside of the spectral band $[\omega_a, \omega_b]$. $\mathbf{R}_x$ can be expressed in terms of the power spectral density of the data using the Fourier transform relationship as

$$\mathbf{R}_x = \frac{1}{2\pi} \int_{\omega_a}^{\omega_b} S(\omega) \mathbf{d}(\omega) \mathbf{d}^H(\omega) d\omega \quad (2.8)$$

with $\mathbf{d}(\omega)$ as defined for (2.4b). Now assume the array data $\mathbf{x}$ is due to a source located at direction θ . In like manner to the time series case we can obtain the covariance matrix of the array data as

$$\mathbf{R}_x = \frac{1}{2\pi} \int_{\omega_a}^{\omega_b} S(\omega) \mathbf{d}(\theta, \omega) \mathbf{d}^H(\theta, \omega) d\omega. \quad (2.9)$$

A source is said to be narrowband of frequency ω_0 if $\mathbf{R}_x$ can be represented as the rank one outer product

$$\mathbf{R}_x = \sigma_s^2 \mathbf{d}(\theta, \omega_0) \mathbf{d}^H(\theta, \omega_0) \quad (2.10)$$

where σ_s^2 is the source variance or power.

The conditions under which a source can be considered narrowband depend on both the source bandwidth and the time over which the source is observed. To illustrate this, consider observing an amplitude modulated sinusoid or the output of a narrowband filter driven by white noise on an oscilloscope. If the signal bandwidth is small relative to the center frequency (i.e., if it has a small fractional bandwidth), and the time intervals over which the signal is observed are short relative to the inverse of the signal bandwidth, then each observed waveform has the shape of a sinusoid. Note that as the observation time interval is increased, the bandwidth must decrease for the signal to remain sinusoidal in appearance. It turns out, based on statistical arguments, that the observation time bandwidth product (TBWP) is the fundamental parameter that determines whether a source can be viewed as narrowband (Buckley [1987], Compton [1988]).

An array provides an effective temporal aperture over which a source is observed. Fig. 2.2 illustrates this temporal aperture $T(\theta)$ for a source arriving from direction θ . Clearly the TBWP is dependent on the source DOA. An array is considered narrowband if the observation TBWP is much less than one for all possible source directions.

Narrowband beamforming is conceptually simpler than broadband since one can ignore the temporal frequency variable. This fact, coupled with interest in temporal frequency analysis for some applications, has motivated implementation of broadband beamformers with a narrowband decomposition structure, as illustrated in

Fig. 2.4. The narrowband decomposition is often performed by taking a discrete Fourier transform (DFT) of the data in each sensor channel using an FFT algorithm. The data across the array at each frequency of interest are processed by their own beamformer. This is often termed frequency domain beamforming. The frequency domain beamformer outputs can be made equivalent to the DFT of the broadband beamformer output depicted in Fig. 2.1b with proper selection of beamformer weights and careful data partitioning. This equivalence corresponds to implementing FIR filters via circular convolution with the DFT.

C. Beamformer Classification

Beamformers can be classified as either data independent or statistically optimum, depending on how the weights are chosen. The weights in a data independent beamformer do not depend on the array data and are chosen to present a specified response for all signal/interference scenarios. The weights in a statistically optimum beamformer are chosen based on the statistics of the array data to "optimize" the array response. In general, the statistically optimum beamformer places nulls in the directions of interfering sources in an attempt to maximize the signal to noise ratio at the beamformer output. A comparison between data independent and statistically optimum beamformers is illustrated in Fig. 2.5.

The next four sections cover data independent, statistically optimum, adaptive, and partially adaptive beamforming. Data independent beamformer design techniques are often used in statistically optimum beamforming (e.g., constraint design in linearly constrained minimum variance beamforming). The statistics of the array data are not usually known and may change over time so adaptive algorithms are typically employed to determine the weights. The adaptive algorithm is designed so the beamformer

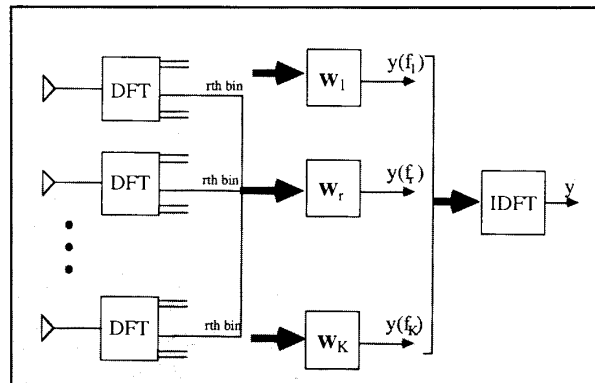


Figure 2.4 Beamforming is sometimes performed in the frequency domain when broadband signals are of interest. This figure illustrates transformation of the data at each sensor into the frequency domain. Weighted combinations of the data at each frequency (bin) are formed. An inverse discrete Fourier transform produces the output time series.

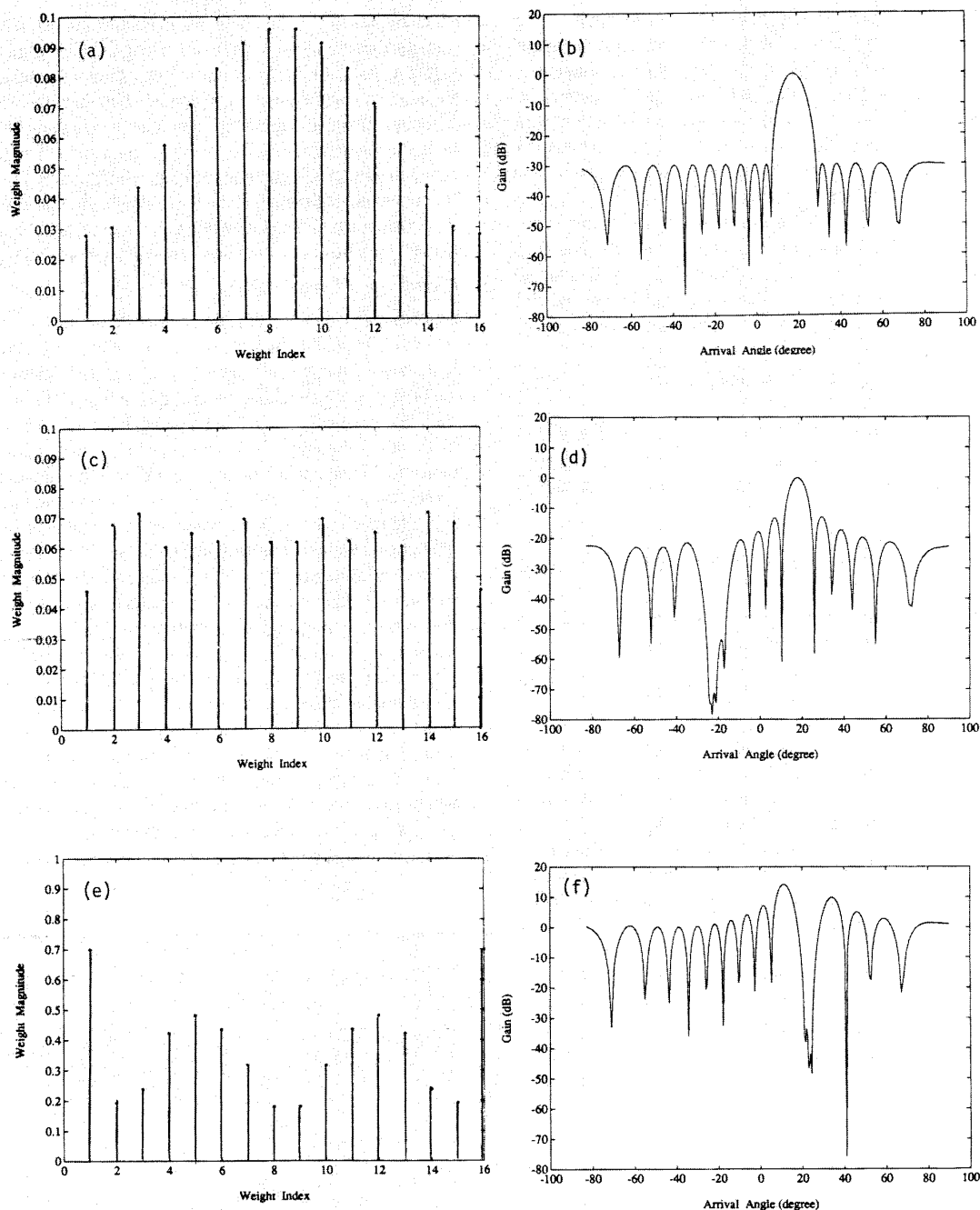


Figure 2.5 Beamformers come in both data independent and statistically optimum varieties. In this figure we consider an equi-spaced narrowband line array of sixteen sensors spaced at one-half wavelength. The magnitude of the weights and beampatterns are shown for a data independent and two statistically optimum beamformers. In (a) and (b) a Dolph-Chebyshev beamformer (Dolph [1946]) with -30 dB sidelobes is represented. In (c) through (f) statistically optimum beamformers are designed to minimize output power subject to a constraint that the

response be unity for an arrival angle of 18°. Energy is assumed to arrive at the array from several interference sources. In (c) and (d) several interferers are located between -20° and -23°, each with power of 30 dB relative to the uncorrelated noise power at a single sensor. Deep nulls are formed at the interferer directions. The interferers in (e) and (f) are located between 20° and 23°, again with relative power levels of 30 dB. Again deep nulls are formed at the interferer directions; however the sidelobe levels are significantly higher at other directions.

response converges to a statistically optimum solution. Partially adaptive beamformers reduce the adaptive algorithm computational load at the expense of a loss (designed to be small) in statistical optimality.

III. DATA INDEPENDENT BEAMFORMING

The weights in a data independent beamformer are designed so the beamformer response approximates a desired response independent of the array data or data statistics. This design objective—approximating a desired response—is the same as that for classical FIR filter design (see, for example, Parks and Burrus [1987]). We shall exploit the analogies between beamforming and FIR filtering where possible in developing an understanding of the design problem and in presenting design procedures. We also discuss design problems specific to beamforming.

The first part of this section discusses forming beams in a classical sense, i.e., approximating a desired response of unity at a point of direction and zero elsewhere. Methods for designing beamformers having more general forms of desired response are presented in the second part.

A. Classical Beamforming

Consider the problem of separating a single complex frequency component from other frequency components using the J tap FIR filter illustrated in Fig. 2.3. If frequency ω_o is of interest, then the desired frequency response is unity at ω_o and zero elsewhere. A common solution to this problem is to choose $\mathbf{w}$ as the vector $\mathbf{d}(\omega_o)$. This choice can be shown to be optimal in terms of minimizing the squared error between the actual response and desired response. The actual response is characterized by a main lobe (or beam) and many sidelobes. Since $\mathbf{w} = \mathbf{d}(\omega_o)$, each element of $\mathbf{w}$ has unit magnitude. Tapering or windowing the amplitudes of the elements of $\mathbf{w}$ permits trading of main lobe or *beam* width against sidelobe levels to *form* the response into a desired shape. Let $\mathbf{T}$ be a J by J diagonal matrix with the real-valued taper weights as diagonal elements. The tapered FIR filter weight vector is given by $\mathbf{T}\mathbf{d}(\omega_o)$. A detailed comparison of a large number of tapering functions is given in Harris [1978].

In spatial filtering one is often interested in receiving a signal arriving from a known location point θ_o . Assuming the signal is narrowband (frequency ω_o), a common choice for the beamformer weight vector is the array response vector $\mathbf{d}(\theta_o, \omega_o)$. The resulting array and beamformer is termed a phased array since the output of each sensor is phase shifted prior to summation. Fig. 1.2 depicts the magnitude of the actual response when $\mathbf{w} = \mathbf{d}(\theta_o, \omega_o)$. As in the FIR filter discussed above, beam width and sidelobe levels are the important characteristics of the response. Amplitude tapering can be used to control the shape of the response, i.e., to form the beam. Figs. 2.5a and b illustrate the effect of amplitude tapering on the response.

The equivalence of the narrowband linear equi-spaced array and FIR filter (see Fig. 2.3) implies that the same techniques for choosing taper functions are applicable to

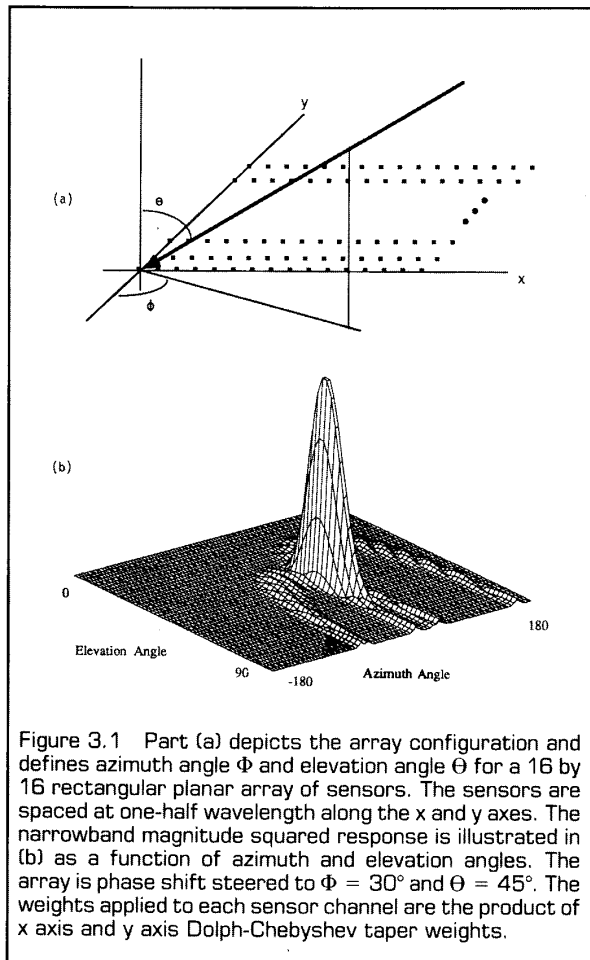
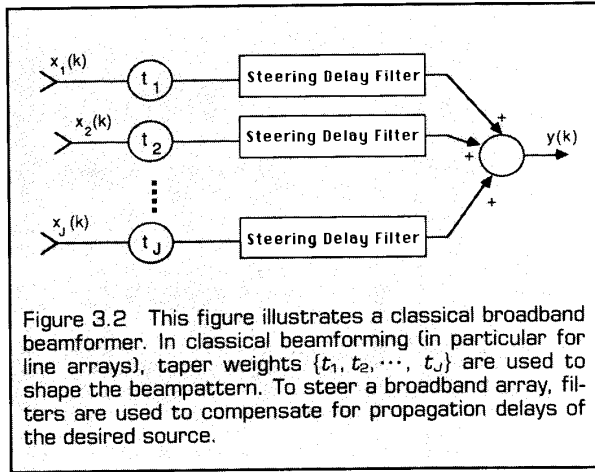


Figure 3.1 Part (a) depicts the array configuration and defines azimuth angle Φ and elevation angle Θ for a 16 by 16 rectangular planar array of sensors. The sensors are spaced at one-half wavelength along the x and y axes. The narrowband magnitude squared response is illustrated in (b) as a function of azimuth and elevation angles. The array is phase shift steered to $\Phi = 30^\circ$ and $\Theta = 45^\circ$. The weights applied to each sensor channel are the product of x axis and y axis Dolph-Chebyshev taper weights.

either problem (Schelkunoff [1943]). Methods for choosing tapering weights also exist for more general array configurations. If the array is narrowband and the sensors lie on a line, then methods evolving from continuous spatial aperture design can be employed. A desirable smooth amplitude distribution function from a continuous aperture is approximated by a step distribution in a discrete aperture (see Ishimaru [1962] for relationship between continuous and discrete apertures).

If the array is planar and factorable, then line array techniques can be used to synthesize the overall response as the product of two linear array responses. A planar array in the xy plane is factorable if its response can be "factored" into the product of responses due to line arrays in the x and y directions. Fig. 3.1 depicts the beam pattern for a 16×16 narrowband planar array. The sensors are spaced by one-half wavelength in both the x and y directions and the direction of interest is 30° bearing and 45° elevation. The response is synthesized as the product of Dolph-Chebyshev tapered line arrays in the x and y directions.

If the beamformer is broadband and employs FIR filters, then tapering can be applied independently to both the



sensor outputs and FIR filters as illustrated in Fig. 3.2. The taper weights are chosen to shape the spatial response and the FIR filter coefficients to present a desired temporal response. As noted in Section II, spatial and temporal response interact so the spatial and temporal responses cannot be synthesized completely independently. One example of where the structure of Fig. 3.2 is used is delay sum beamforming. Here the FIR filters approximate the propagation delays (linear phase over the frequency band of interest) and the taper weights are chosen to shape the main beam and sidelobe structure of the spatial response.

B. General Data Independent Response Design

The methods discussed in this subsection apply to design of beamformers that approximate an arbitrary desired response. This is of interest in several different applications. For example, we may wish to receive any signal arriving from a range of directions, in which case the desired response is unity over the entire range. As another example, we may know that there is a strong source of interference arriving from a certain range of directions, in which case the desired response is zero in this range. These two examples are analogous to bandpass and band-stop FIR filtering. Although we are no longer "forming beams," it is conventional to refer to this type of spatial filter as a beamformer.

Consider choosing $\mathbf{w}$ so the actual response $r(\theta, \omega) = \mathbf{w}^H \mathbf{d}(\theta, \omega)$ approximates a desired response $r_d(\theta, \omega)$. *Ad hoc* techniques similar to those employed in FIR filter design can be used for selecting $\mathbf{w}$; however, here we only consider choosing $\mathbf{w}$ to minimize the weighted L_p norm of the difference between desired and actual response. Weighted L_p approximation is utilized in several established FIR filter design techniques. The most commonly used norms are L_∞ (minimax) and L_2 (least squares). Specific techniques include (see Parks and Burrus [1987]):

- 1) Windowing of an ideal filter's unit pulse response (minimizes L_2 norm over continuous ω);
- 2) Frequency response sampling and linear weighted least squares (minimized L_2 norm over discrete ω);

- 3) Minimax design with the Remez exchange algorithm (minimizes L_∞ norm over discrete ω);
- 4) Minimax complex and magnitude response design (minimizes L_∞ norm over discrete ω).

FIR filter design corresponds to a polynomial approximation problem since the frequency response (2.4b) is the discrete Fourier transform of the FIR filter weight sequence. Several of the above methods exploit this polynomial structure.

Excluding the cases for which beamformer design can be reduced to equi-spaced line array geometries, beamformer design is not a polynomial approximation problem. In general, the response in (2.6) is a weighted sum of exponentials raised to non-integer powers. Thus, the L_∞ methods (3 and 4) are not applicable since they are based on the alteration theorem of polynomial approximation. The windowing method (1) is based on the discrete time Fourier transform and is also not applicable. However, the L_2 procedure using linear weighted least squares (2) is applicable.

To illustrate data independent beamformer design via L_2 optimization, consider minimizing the squared error between the actual and desired response at P points (θ_i, ω_i) , $1 \leq i \leq P$. If $P > N$, then we obtain the overdetermined least squares problem

$$\min_{\mathbf{w}} \|\mathbf{A}^H \mathbf{w} - \mathbf{r}_d\|^2 \quad (3.1)$$

where

$$\mathbf{A} = [\mathbf{d}(\theta_1, \omega_1) \mathbf{d}(\theta_2, \omega_2) \cdots \mathbf{d}(\theta_P, \omega_P)];$$

$$\mathbf{r}_d = [r_d(\theta_1, \omega_1) r_d(\theta_2, \omega_2) \cdots r_d(\theta_P, \omega_P)]^H.$$

Provided $\mathbf{A}\mathbf{A}^H$ is invertible (i.e., $\mathbf{A}$ is full rank), then the solution to (3.1) is given as

$$\mathbf{w} = \mathbf{A}^+ \mathbf{r}_d \quad (3.2)$$

where $\mathbf{A}^+ = (\mathbf{A}\mathbf{A}^H)^{-1}\mathbf{A}$ is the pseudo inverse of $\mathbf{A}$. Fig. 3.3 depicts the response of a beamformer design using (3.2).

A note of caution is in order at this point. The *white noise gain* of a beamformer is defined as the output power due to unit variance white noise at the sensors. Thus, the norm squared of the weight vector, $\mathbf{w}^H \mathbf{w}$, represents the white noise gain. If the white noise gain is large, then the accuracy by which $\mathbf{w}$ approximates the desired response is a moot point, since the beamformer output will have a poor SNR due to white noise contributions. If $\mathbf{A}$ is ill-conditioned, then $\mathbf{w}$ can have a very large norm and still approximate the desired response. The matrix $\mathbf{A}$ is ill-conditioned when the numerical dimension of the space spanned by the $\mathbf{d}(\theta^i, \omega^i)$, $1 \leq i \leq P$, is less than N . For example, if only one source direction is sampled, then the numerical rank of $\mathbf{A}$ is approximately given by the TBWP for that direction. Low rank approximates of $\mathbf{A}$ and $\mathbf{A}^+$ should be used whenever the numerical rank is less than N . This ensures that the norm of $\mathbf{w}$ will not be unnecessarily large.

Specific directions and frequencies can be emphasized in (3.1) by selection of the sample points (θ_i, ω_i) and/or

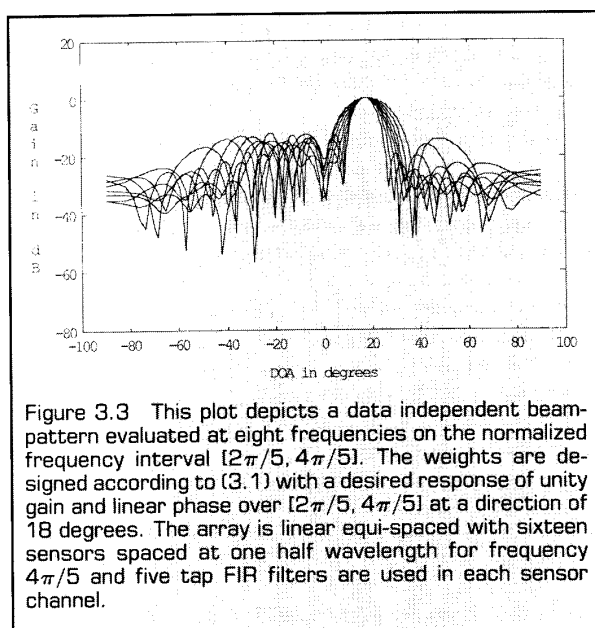


Figure 3.3 This plot depicts a data independent beam-pattern evaluated at eight frequencies on the normalized frequency interval $[2\pi/5, 4\pi/5]$. The weights are designed according to (3.1) with a desired response of unity gain and linear phase over $[2\pi/5, 4\pi/5]$ at a direction of 18 degrees. The array is linear equi-spaced with sixteen sensors spaced at one half wavelength for frequency $4\pi/5$ and five tap FIR filters are used in each sensor channel.

unequally weighting of the error at each (θ_i, ω_i) . Parks and Burrus [1987] discuss this in the context of FIR filtering. Kumar and Murthy [1977] consider unequal weighting of the error to obtain minimax response design for beam-former weights in a linear narrowband array. In general, guidelines for selection of the error weighting and (θ_i, ω_i) are not available.

There are several alternatives to L_p optimization for general data independent response design, including methods discussed in Butler and Unz [1967] and Sanzgiri and Butler [1971].

IV. STATISTICALLY OPTIMUM BEAMFORMING

In statistically optimum beamforming the weights are chosen based on the statistics of the data received at the array. Loosely speaking, the goal is to "optimize" the beamformer response so the output contains minimal contributions due to noise and signals arriving from directions other than the desired signal direction. We discuss below several different criteria for choosing statistically optimum beamformer weights. Table 4.1 summarizes these different approaches. Where possible, equations describing the criteria and weights are confined to Table 4.1. Throughout the section we assume that the data is wide-sense stationary and that its second order statistics are known. Determination of weights when the data statistics are unknown or time varying is discussed in the following section on adaptive algorithms.

A. Multiple Sidelobe Canceller

The multiple sidelobe canceller (MSC) is perhaps the earliest statistically optimum beamformer. An MSC consists of a "main channel" and one or more "auxiliary channels" as depicted in Fig. 4.1a. The main channel can be either a single high gain antenna or a data independent

TABLE 4.1
SUMMARY OF OPTIMUM BEAMFORMERS

Type	MSC	Reference Signal	Max SNR	LCMV
Definitions	x_a —auxiliary data y_m —primary data $r_{ma} = E\{x_a y_m^*\}$ $R_a = E\{x_a x_a^H\}$ output: $y = y_m - w_a^H x_a$	x —array data y_d —desired signal $r_{xd} = E\{x y_d^*\}$ $R_x = E\{x x^H\}$ output: $y = w^H x$	$x = s + n$ —array data s —signal component n —noise component $R_s = E\{s s^H\}$ $R_n = E\{n n^H\}$ output: $y = w^H x$	x —array data C —constraint matrix f —response vector $R_x = E\{x x^H\}$ output: $y = w^H x$
Criterion	$\min_{w_a} E\{ y_m - w_a^H x_a ^2\}$	$\min_w E\{ y - y_d ^2\}$	$\max_w \frac{w^H R_s w}{w^H R_n w}$	$\min_w \{w^H R_x w\} \text{ s.t. } C^H w = f$
Optimum Weights	$w_a = R_a^{-1} r_{ma}$	$w = R_x^{-1} r_{xd}$	$R_n^{-1} R_s w = \lambda_{\max} w$	$w = R_x^{-1} [C^H R_x^{-1} C]^{-1} f$
Advantages	Simple	Direction of desired signal can be unknown	True maximization of SNR	Flexible and general constraints
Disadvantages	Requires absence of desired signal from auxiliary channels for weight determination	Must generate reference signal	Must know R_s and R_n , Solve generalized eigenproblem for weights	Computation of constrained weight vector
References	Applebaum [1976]	Widrow [1967]	Monzingo and Miller [1980]	Frost [1972]

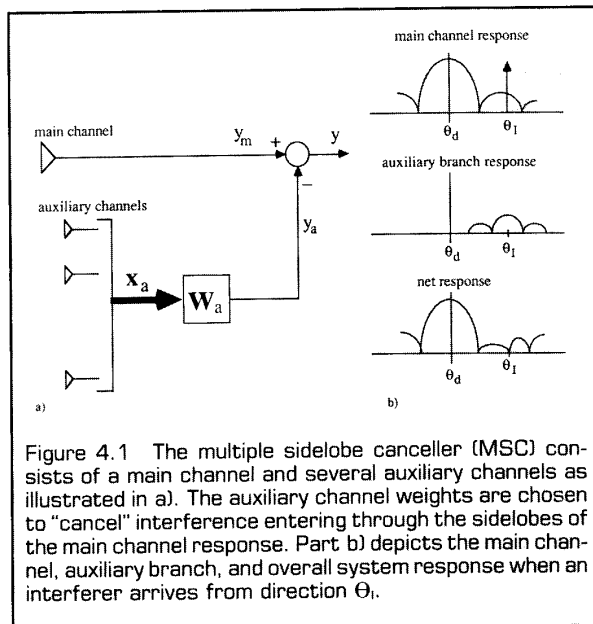


Figure 4.1 The multiple sidelobe canceller (MSC) consists of a main channel and several auxiliary channels as illustrated in a). The auxiliary channel weights are chosen to "cancel" interference entering through the sidelobes of the main channel response. Part b) depicts the main channel, auxiliary branch, and overall system response when an interferer arrives from direction θ_l .

beamformer (see Section III). It has a highly directional response, which is pointed in the desired signal direction. Interfering signals are assumed to enter through the main channel sidelobes. The auxiliary channels also receive the interfering signals. The goal is to choose the auxiliary channel weights to cancel the main channel interference component. This implies that the responses to interferers of the main channel and linear combination of auxiliary channels must be identical. The overall system then has a response of zero as illustrated in Fig. 4.1b. In general, requiring zero response to all interfering signals is either not possible or can result in significant white noise gain. Thus, the weights are usually chosen to trade off interference suppression for white noise gain by minimizing the expected value of the total output power as indicated in Table 4.1.

Choosing the weights to minimize output power can cause cancellation of the desired signal, since it also contributes to total output power. In fact, as the desired signal gets stronger it contributes to a larger fraction of the total output power and the percentage cancellation increases. Clearly this is an undesirable effect. The MSC is very effective in applications where the desired signal is very weak (relative to the interference), since the optimum weights will not pay any attention to it, or when the desired signal is known to be absent during certain time periods. The weights can be adapted in the absence of the desired signal and frozen when it is present.

B. Use of Reference Signal

If the desired signal were known, then the weights could be chosen to minimize the error between the beamformer output and the desired signal. Of course, knowledge of the desired signal eliminates the need for beamforming. However, for some applications enough may be known about the desired signal to generate a signal that closely

represents it. This signal is called a reference signal. As indicated in Table 4.1, the weights are chosen to minimize the mean square error between the beamformer output and the reference signal.

The weight vector depends on the cross covariance between the unknown desired signal present in x and the reference signal. Acceptable performance is obtained provided this approximates the covariance of the unknown desired signal with itself. For example, if the desired signal is amplitude modulated, then acceptable performance is often obtained by setting the reference signal equal to the carrier. It is also assumed that the reference signal is uncorrelated with interfering signals in x . The fact that the direction of the desired signal does not need to be known is a distinguishing feature of the reference signal approach.

C. Maximization of Signal to Noise Ratio

Here the weights are chosen to directly maximize the signal to noise ratio (SNR) as indicated in Table 4.1. A general solution for the weights requires knowledge of both the desired signal, R_s , and noise, R_n , covariance matrices. The attainability of this knowledge depends on the application. For example, in an active radar system R_n can be estimated during the time that no signal is being transmitted and R_s can be obtained from knowledge of the transmitted pulse and direction of interest. If the signal component is narrowband, of frequency ω , and direction θ , then $R_s = \sigma^2 d(\theta, \omega) d^H(\theta, \omega)$ from the results in Section II. In this case the weights are obtained as

$$w = \alpha R_n^{-1} d(\theta, \omega) \quad (4.1)$$

where the α is some non-zero complex constant. Substitution of (4.1) into the SNR expression shows that the SNR is independent of the value chosen for α .

D. Linearly Constrained Minimum Variance Beamforming

In many applications none of the above approaches is satisfactory. The desired signal may be of unknown strength and may always be present, resulting in signal cancellation with the MSC and preventing estimation of signal and noise covariance matrices in the maximum SNR processor. Lack of knowledge about the desired signal may prevent utilization of the reference signal approach. These limitations can be overcome through the application of linear constraints to the weight vector. Use of linear constraints is a very general approach that permits extensive control over the adapted response of the beamformer. In this subsection we illustrate how linear constraints can be employed to control beamformer response, discuss the optimum linearly constrained beamforming problem, and present the generalized sidelobe canceller structure.

The basic idea behind linearly constrained minimum variance (LCMV) 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 minimize output variance or power subject to the response constraint. This has the effect of

preserving the desired signal while minimizing contributions to the output due to interfering signals and noise arriving from directions other than the direction of interest. The analogous FIR filter has the weights chosen to minimize the filter output power subject to the constraint that the filter response to signals of frequency ω_0 be unity.

In Section II we saw that the beamformer response to a source at angle θ and temporal frequency ω is given by $\mathbf{w}^H \mathbf{d}(\theta, \omega)$. Thus, by linearly constraining the weights to satisfy $\mathbf{w}^H \mathbf{d}(\theta, \omega) = g$, where g is a complex constant, we ensure that any signal from angle θ and frequency ω is passed to the output with response g . Minimization of contributions to the output from interference (signals not arriving from θ with frequency ω) is accomplished by choosing the weights to minimize the expected value of output power or variance $E\{|y|^2\} = \mathbf{w}^H \mathbf{R}_x \mathbf{w}$. The LCMV problem for choosing the weights is thus written

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_x \mathbf{w} \quad \text{subject to} \quad \mathbf{d}^H(\theta, \omega) \mathbf{w} = g^*. \quad (4.2)$$

The method of Lagrange multipliers can be used to solve (4.2) resulting in

$$\mathbf{w} = g^* \frac{\mathbf{R}_x^{-1} \mathbf{d}(\theta, \omega)}{\mathbf{d}^H(\theta, \omega) \mathbf{R}_x^{-1} \mathbf{d}(\theta, \omega)}. \quad (4.3)$$

Note that in practice the presence of uncorrelated noise will ensure that $\mathbf{R}_x$ is invertible. If $g = 1$, then (4.3) is often termed the minimum variance distortionless response (MVDR) beamformer. It can be shown that (4.3) is equivalent to the maximum SNR solution given in (4.1) by substituting $\sigma^2 \mathbf{d}(\theta, \omega) \mathbf{d}^H(\theta, \omega) + \mathbf{R}_n$ for $\mathbf{R}_x$ in (4.3) and applying the matrix inversion lemma.

The single linear constraint in (4.2) is easily generalized to multiple linear constraints for added control over the beampattern. For example, if there is a fixed interference source at a known direction ϕ , then it may be desirable to force zero gain in that direction in addition to maintaining the response g to the desired signal. This is expressed as

$$\begin{bmatrix} \mathbf{d}^H(\theta, \omega) \\ \mathbf{d}^H(\phi, \omega) \end{bmatrix} \mathbf{w} = \begin{bmatrix} g^* \\ 0 \end{bmatrix}. \quad (4.4)$$

If there are $L < N$ linear constraints on $\mathbf{w}$, we write them in the form $\mathbf{C}^H \mathbf{w} = \mathbf{f}$ where the N by L matrix $\mathbf{C}$ and L dimensional vector $\mathbf{f}$ are termed the constraint matrix and response vector. The constraints are assumed to be linearly independent so $\mathbf{C}$ has rank L . The LCMV problem and solution with this more general constraint equation are given in Table 4.1.

Constraint Design. Several different philosophies can be employed for choosing the constraint matrix and response vector. We discuss point (Kelly and Levin [1964]), derivative (Owsley [1973], Er and Cantoni [1983]), and eigenvector (Buckley [1987]) constraints below. In many 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 with L constraints there are only $N - L$ degrees of freedom available for minimizing variance.

Point constraints fix the beamformer response at points of spatial direction and temporal frequency. Equation (4.4) represents an example of two point constraints on $\mathbf{w}$. The number of points at which response can be constrained is limited to N . If N constraints are used then there are no degrees of freedom left for power minimization and a data independent beamformer is obtained.

Derivative constraints are employed to influence response over a region of direction and/or frequency by forcing the derivatives of the beamformer response at some point of direction and frequency to be zero. They are usually employed in conjunction with point constraints. An example where derivative constraints are useful is when the desired signal direction is only known approximately. If the signal arrives near the direction at which a point constraint is employed, then application of a derivative constraint at that point prevents the beamformer from synthesizing a response of zero to the desired signal.

Eigenvector constraints are based on a least squares approximation to the desired response and are typically used to control beamformer response over regions of direction and/or frequency. Constraining the beamformer response in a least squares sense ensures that the mean square error between desired and actual beamformer response over a region is minimized for a given number of constraints. In this sense eigenvector constraints are efficient. Consider designing a set of constraints which will control the beamformer response to a source from direction θ_0 over the frequency band $[\omega_a, \omega_b]$. The dimension of the span of $\mathbf{d}(\theta_0, \omega)$ over this band is approximately given by the source TBWP discussed in Section II. Eigenvector constraints are derived from (3.1) by choosing P significantly greater than the TBWP. The ω_i then oversample $[\omega_a, \omega_b]$ and $\mathbf{A}$ is ill-conditioned. A rank L approximation of $\mathbf{A}$ is obtained from its singular value decomposition

$$\mathbf{A}_L = \mathbf{V} \mathbf{\Sigma}_L \mathbf{U}^H \quad (4.5)$$

where $\mathbf{\Sigma}_L$ is an L by L diagonal matrix containing the largest singular values of $\mathbf{A}$, and the L columns of $\mathbf{V}$ and $\mathbf{U}$ are respectively the left and right singular vectors of $\mathbf{A}$ corresponding to these singular values. Replacing $\mathbf{A}$ in (3.1) by its rank L approximate (4.5) and bringing $\mathbf{U} \mathbf{\Sigma}_L$ to the right side (the pseudo inverse of $\mathbf{U}$ is $\mathbf{U}^H$), yields

$$\mathbf{V}^H \mathbf{w} = \mathbf{\Sigma}_L^{-1} \mathbf{U}^H \mathbf{r}_d. \quad (4.6)$$

Equation (4.6) has the same form as the constraint equation $\mathbf{C}^H \mathbf{w} = \mathbf{f}$. The columns of $\mathbf{V}$ correspond to the eigenvectors of $\mathbf{A} \mathbf{A}^H$; hence the name eigenvector constraints. (Note that $\mathbf{A} \mathbf{A}^H$ represents an approximation of $\mathbf{R}_x$ in (2.9) if $S(\omega) = 1$.)

Example. Fig. 4.2 depicts the response of an LCMV beamformer at eight equally spaced frequencies on the interval $[2\pi/5, 4\pi/5]$ when two interferers arrive from -17.5 and -5.75 degrees in the presence of white noise. The interferer power levels relative to the white noise are

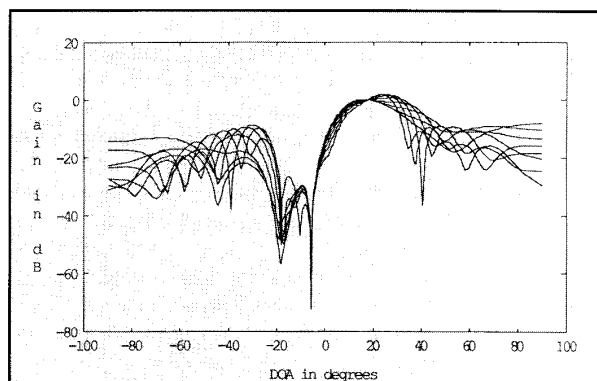


Figure 4.2 This plot depicts the LCMV beamformer magnitude response at eight frequencies on the normalized frequency interval $[2\pi/5, 4\pi/5]$ when two interferers arrive from directions -5.75 and -17.5 degrees in the presence of white noise. The interferers have a white spectrum on $[2\pi/5, 4\pi/5]$ and have powers of 40 and 30 dB relative to the white noise, respectively. The constraints are designed according to (4.6) to present unit gain and linear phase over $[2\pi/5, 4\pi/5]$ at a DOA of 18 degrees. The array is the same as that of Fig. 3.3.

30 and 40 dB, respectively, and the interferers have a flat spectrum on $[2\pi/5, 4\pi/5]$. The array has sixteen linear equi-spaced sensors with five taps per sensor. The tap spacing is normalized to one second and the sensors are spaced at one-half wavelength corresponding to frequency $4\pi/5$. The response is constrained to pass signals arriving from 18 degrees in the band $[2\pi/5, 4\pi/5]$, with unit gain and linear phase using ten eigenvector constraints designed from (4.6). The effectiveness of the constraints is evident, since all the frequency curves pass through zero dB at 18 degrees. The response has nulls in the directions of the interferers with the deeper null corresponding to the stronger interferer. The response as a function of frequency for the interferer directions is plotted in Fig. 6.2. The array gain is 50 dB for this example.

Generalized Sidelobe Canceller. The generalized sidelobe canceller (GSC) represents an alternative formulation of the LCMV problem, which provides insight, is useful for analysis, and can simplify LCMV beamformer implementation. It also illustrates the relationship between the MSC and LCMV beamforming. Essentially, the GSC is a mechanism for changing a constrained minimization problem into unconstrained form. Perhaps the first reference to this concept is in Hanson and Lawson [1969], where a procedure for transforming constrained least squares problems to unconstrained least squares problems is given. Griffiths and Jim [1982] applied the same concept to LCMV beamforming and coined the term GSC. Similar techniques were discussed in Applebaum and Chapman [1976].

Suppose we decompose the weight vector $\mathbf{w}$ into two orthogonal components $\mathbf{w}_o$ and $-\mathbf{v}$ ($\mathbf{w} = \mathbf{w}_o - \mathbf{v}$) that lie in the range and null space of $\mathbf{C}$, respectively. The range and null space of a matrix span the entire space so this

decomposition can be used to represent any $\mathbf{w}$. Since $\mathbf{C}^H\mathbf{v} = 0$, we must have

$$\mathbf{w}_o = \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{f} \quad (4.7)$$

if $\mathbf{w}$ is to satisfy the constraints. (4.7) is the minimum L_2 norm solution to the underdetermined equivalent of (3.1). The vector $\mathbf{v}$ is a linear combination of the columns of an N by $N-L$ matrix $\mathbf{C}_n$ ($\mathbf{v} = \mathbf{C}_n\mathbf{w}_n$) provided the columns of $\mathbf{C}_n$ form a basis for the null space of $\mathbf{C}$. $\mathbf{C}_n$ can be obtained from $\mathbf{C}$ using any of several orthogonalization procedures such as Gram-Schmidt, QR decomposition, or singular value decomposition. The weight vector $\mathbf{w} = \mathbf{w}_o - \mathbf{C}_n\mathbf{w}_n$ is depicted in block diagram form in Fig. 4.3. The choice for $\mathbf{w}_o$ and $\mathbf{C}_n$ implies that $\mathbf{w}$ satisfies the constraints independent of $\mathbf{w}_n$ and reduces the LCMV problem to the unconstrained problem

$$\min_{\mathbf{w}_n} [\mathbf{w}_o - \mathbf{C}_n\mathbf{w}_n]^H \mathbf{R}_x [\mathbf{w}_o - \mathbf{C}_n\mathbf{w}_n]. \quad (4.8)$$

The solution is

$$\mathbf{w}_n = (\mathbf{C}_n^H \mathbf{R}_x \mathbf{C}_n)^{-1} \mathbf{C}_n^H \mathbf{R}_x \mathbf{w}_o. \quad (4.9)$$

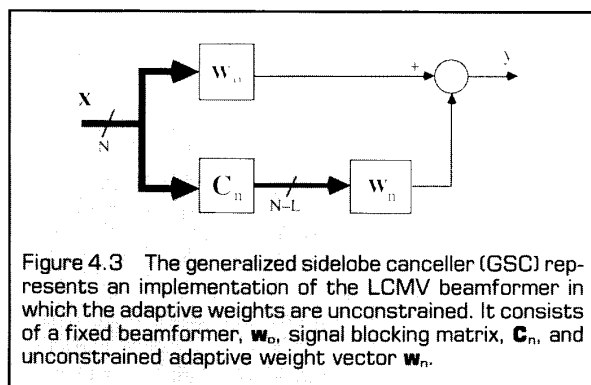


Figure 4.3 The generalized sidelobe canceller (GSC) represents an implementation of the LCMV beamformer in which the adaptive weights are unconstrained. It consists of a fixed beamformer, $\mathbf{w}_o$, signal blocking matrix, $\mathbf{C}_n$, and unconstrained adaptive weight vector $\mathbf{w}_n$.

The primary implementation advantages of this alternate but equivalent formulation stem from the facts that the weights $\mathbf{w}_n$ are unconstrained and a data independent beamformer $\mathbf{w}_o$ is implemented as an integral part of the adaptive beamformer. The unconstrained nature of the adaptive weights permits much simpler adaptive algorithms to be employed and the data independent beamformer is useful in situations where adaptive signal cancellation occurs (see subsection E).

As an example, assume the constraints are as given in (4.2). (4.7) implies $\mathbf{w}_o = \mathbf{g}^* \mathbf{d}(\theta, \omega) / [\mathbf{d}^H(\theta, \omega) \mathbf{d}(\theta, \omega)]$. $\mathbf{C}_n$ satisfies $\mathbf{d}^H(\theta, \omega) \mathbf{C}_n = 0$ so each column $[\mathbf{C}_n]_i$; $1 \leq i \leq N-L$, can be viewed as a data independent beamformer with a null in direction θ at frequency ω : $\mathbf{d}^H(\theta, \omega) [\mathbf{C}_n]_i = 0$. Thus, a signal of frequency ω and direction θ arriving at the array will be blocked or nulled by the matrix $\mathbf{C}_n$. In general, if the constraints are designed to present a specified response to signals from a set of directions and frequencies, then the columns of $\mathbf{C}_n$ will block those directions and frequencies. This characteristic has led to the term "blocking matrix" for $\mathbf{C}_n$. These signals are only processed by $\mathbf{w}_o$ and since $\mathbf{w}_o$ satisfies the constraints, they are presented

with the desired response independent of $\mathbf{w}_n$. Signals from directions and frequencies over which the response is not constrained will pass through the upper branch in Fig. 4.3 with some response determined by $\mathbf{w}_o$. The lower branch chooses $\mathbf{w}_n$ to estimate the signals at the output of $\mathbf{w}_o$ as a linear combination of the data at the output of the blocking matrix. This is similar to the operation of the MSC, in which weights are applied to the output of auxiliary sensors in order to estimate the primary channel output (see Fig. 4.1).

E. Signal Cancellation in Statistically Optimum Beamforming.

Optimum beamforming requires some knowledge of the desired signal characteristics, either its statistics (for maximum SNR or reference signal methods), its direction (for the MSC), or its response vector $\mathbf{d}(\theta, \omega)$ (for the LCMV beamformer). If the required knowledge is inaccurate, the optimum beamformer will attenuate the desired signal as if it were interference. Cancellation of the desired signal is often significant, especially if the SNR of the desired signal is large (Cox [1973]). Several approaches have been suggested to reduce this degradation (e.g., Jablon [1986], Cox et al. [1987]).

A second cause of signal cancellation is correlation between the desired signal and one or more interference signals. This can result either from multipath propagation of a desired signal or from smart (correlated) jamming. When interference and desired signals are uncorrelated the beamformer attenuates interferers to minimize output power. However, with a correlated interferer the beamformer minimizes output power by processing the interfering signal in such a way as to cancel the desired signal. If the interferer is partially correlated with the desired signal, then the beamformer will cancel the portion of the desired signal that is correlated with the interferer. Methods for reducing signal cancellation due to correlated interference have been suggested (e.g., Widrow et al. [1982], Shan and Kailath [1985], Yang and Kaveh [1987]).

V. ADAPTIVE ALGORITHMS FOR BEAMFORMING

The optimum beamformer weight vector equations listed in Table 4.1 require knowledge of second order statistics. These statistics are usually not known, but with the assumption of ergodicity, they (and therefore the optimum weights) can be estimated from available data. Statistics may also change over time (e.g., due to moving interferers). To solve these problems, weights are typically determined by adaptive algorithms.

There are two basic adaptive approaches: 1) block adaptation, where statistics are estimated from a temporal block of array data and used in an optimum weight equation; and 2) continuous adaptation, where the weights are adjusted as the data is sampled such that the resulting weight vector sequence converges to the optimum solution. If a nonstationary environment is anticipated, block adaptation can be used, provided that the weights are recomputed periodically. Adams, et al. [1980] and others

have described applications of block data processing. Continuous adaptation is usually preferred when statistics are time-varying or (for computational reasons) when the number of adaptive weights M is moderate to large (values of $M > 50$ are not uncommon).

Among notable adaptive algorithms proposed for beamforming are the Howells-Applebaum adaptive loop developed in the late 1950's and reported by Howells [1966, 1976] and Applebaum [1966], and the Frost LCMV algorithm [1972]. Rather than recapitulating adaptive algorithms for each optimum beamformer listed in Table 4.1 (for this see texts by Monzingo and Miller [1980], Hudson [1981] and Compton [1988]), we take a unifying approach using the standard adaptive filter problem illustrated in Fig. 5.1.

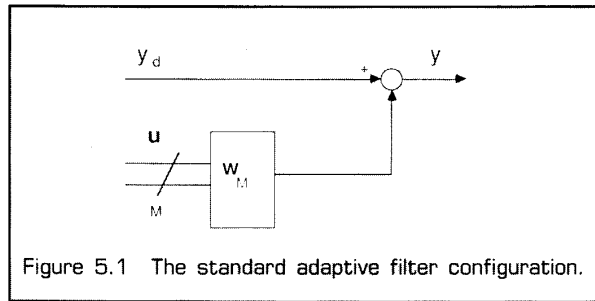


Figure 5.1 The standard adaptive filter configuration.

In Fig. 5.1 the weights are chosen to estimate the desired signal y_d as linear combination of the elements of the data vector $\mathbf{u}$. We select $\mathbf{w}_M$ to minimize the MSE

$$J(\mathbf{w}_M) = E\{|\mathbf{y}_d - \mathbf{w}_M^H \mathbf{u}|^2\} \\ = \sigma_d^2 - \mathbf{w}_M^H \mathbf{r}_{ud} - \mathbf{r}_{ud}^H \mathbf{w}_M + \mathbf{w}_M^H \mathbf{R}_u \mathbf{w}_M, \quad (5.1)$$

where $\sigma_d^2 = E\{|\mathbf{y}_d|^2\}$, $\mathbf{r}_{ud} = E\{\mathbf{u}\mathbf{y}_d^*\}$ and $\mathbf{R}_u = E\{\mathbf{u}\mathbf{u}^H\}$. (5.1) is minimized by

$$\mathbf{w}_{opt} = \mathbf{R}_u^{-1} \mathbf{r}_{ud}. \quad (5.2)$$

Comparison of (5.1) and the criteria listed in Table 4.1 indicates that this standard adaptive filter problem is equivalent to both the MSC beamformer problem (with $y_d = y_m$ and $\mathbf{u} = \mathbf{x}_a$) and the reference signal beamformer problem (with $\mathbf{u} = \mathbf{x}$). The LCMV problem is apparently different. However closer examination of Figs. 4.3, 5.1, and Equations (4.9), (5.2) reveals that the standard adaptive filter problem is equivalent to the LCMV problem implemented with the GSC structure. Setting $\mathbf{u} = \mathbf{C}_n^H \mathbf{x}$ and $y_d = \mathbf{w}_o^H \mathbf{x}$ implies $\mathbf{R}_u = \mathbf{C}_n^H \mathbf{R}_x \mathbf{C}_n$, and $\mathbf{r}_{ud} = \mathbf{C}_n^H \mathbf{R}_x \mathbf{w}_o$. The maximum SNR beamformer can not in general be represented by Fig. 5.1 and Equation (5.2). However, it was noted after (4.3) that if the desired signal is narrowband, then the maximum SNR and the LCMV beamformers are equivalent.

The block adaptation approach solves (5.2) using estimates of $\mathbf{R}_u$ and $\mathbf{r}_{ud}$ formed from K samples of $\mathbf{u}$ and y_d : $\mathbf{u}(k)$, $y_d(k)$; $0 \leq k \leq K-1$. The most common are the sample covariance matrix

$$\hat{\mathbf{R}}_u = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{u}(k) \mathbf{u}^H(k)$$

and sample cross-covariance vector

$$\hat{\mathbf{r}}_{ud} = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{u}(k) y_d^*(k).$$

Performance analysis and guidelines for selecting the block size K are provided in Reed et al. [1974].

Continuous adaptation algorithms are easily developed in terms of Fig. 5.1 and Equation (5.1). Note that $J(\mathbf{w}_M)$ is a quadratic error surface. Since the quadratic surface's "Hessian" $\mathbf{R}_u$ is the covariance matrix of noisy data, it is positive definite. This implies that the error surface is a "bowl." The shape of the bowl is determined by the eigenstructure of $\mathbf{R}_u$. The optimum weight $\mathbf{w}_{opt}$ corresponds to the bottom of the bowl.

One approach to adaptive filtering is to envision a point on the error surface that corresponds to the present weight vector $\mathbf{w}_M(k)$. We select a new weight vector $\mathbf{w}_M(k+1)$ so as to descend on the error surface. The gradient vector

$$\nabla_{\mathbf{w}_M(k)} = \left. \frac{\partial}{\partial \mathbf{w}_M} J(\mathbf{w}_M) \right|_{\mathbf{w}_M = \mathbf{w}_M(k)} = -2\mathbf{r}_{ud} + 2\mathbf{R}_u \mathbf{w}_M(k) \quad (5.3)$$

tells us the direction in which to adjust the weight vector. Steepest descent, i.e., adjustment in the negative gradient direction, leads to the popular least-mean-square (LMS)

adaptive algorithm. The LMS algorithm replaces $\nabla_{\mathbf{w}_M(k)}$ with the instantaneous gradient estimate $\hat{\nabla}_{\mathbf{w}_M(k)} = -2[\mathbf{u}(k)y_d^*(k) - \mathbf{u}(k)\mathbf{u}^H(k)\mathbf{w}_M(k)]$. Denoting $y(k) = y_d(k) - \mathbf{w}_M^H(k)\mathbf{u}(k)$, we have

$$\mathbf{w}_M(k+1) = \mathbf{w}_M(k) + \mu \mathbf{u}(k)y^*(k). \quad (5.4)$$

The gain constant μ controls convergence characteristics of the random vector sequence $\mathbf{w}_M(k)$. Table 5.1 provides guidelines for its selection.

The primary virtue of the LMS algorithm is its simplicity. Its performance is acceptable in many applications; however, its convergence characteristics depend on the shape of the error surface and therefore the eigenstructure of $\mathbf{R}_u$. When the eigenvalues are widely spread, convergence can be slow and other adaptive algorithms with better convergence characteristics should be considered. Alternative procedures for searching the error surface have been proposed in addition to algorithms based on least-squares and Kalman filtering. Roughly speaking, these algorithms trade-off computational requirements with speed of convergence to $\mathbf{w}_{opt}$. We refer you to texts on adaptive filtering for detailed descriptions and analysis (Widrow and Stearns [1985], Haykin [1986], Alexander [1986], Treichler et al. [1987], and others).

One alternative to LMS is the exponentially weighted recursive least squares (RLS) algorithm. At the K^{th} time

TABLE 5.1
COMPARISON OF THE LMS AND RLS WEIGHT ADAPTATION ALGORITHMS

Algorithm	LMS	RLS
Initialization	$\mathbf{w}_M(0) = 0$ $y(0) = y_d(0)$ $0 < \mu < \frac{1}{\text{Trace}[\mathbf{R}_u]}$	$\mathbf{w}_M(0) = 0$ $\mathbf{P}(0) = \delta^{-1}\mathbf{I}$ δ small, $\mathbf{I}$ identity matrix
Update Equations	$\mathbf{w}_M(k) = \mathbf{w}_M(k-1) + \mu \mathbf{u}(k-1)y^*(k-1)$ $y(k) = y_d(k) - \mathbf{w}_M^H(k)\mathbf{u}(k)$	$\mathbf{v}(k) = \mathbf{P}(k-1)\mathbf{u}(k)$ $\mathbf{k}(k) = \frac{\lambda^{-1}\mathbf{v}(k)}{1 + \lambda^{-1}\mathbf{u}^H(k)\mathbf{v}(k)}$ $\alpha(k) = y_d(k) - \mathbf{w}_M^H(k-1)\mathbf{u}(k)$ $\mathbf{w}_M(k) = \mathbf{w}_M(k-1) + \mathbf{k}(k)\alpha^*(k)$ $\mathbf{P}(k) = \lambda^{-1}\mathbf{P}(k-1) - \lambda^{-1}\mathbf{k}(k)\mathbf{v}^H(k)$
Multiplies per update	$2M$	$4M^2 + 4M + 2$
Performance Characteristics	Under certain conditions, convergence of $\mathbf{w}_M(k)$ to the statistically optimum weight vector $\mathbf{w}_{opt}$ in the mean-square sense is guaranteed if μ is chosen as indicated above. The convergence rate is governed by the eigenvalue spread of $\mathbf{R}_u$. For large eigenvalue spread, convergence can be very slow.	The $\mathbf{w}_M(k)$ represents the least-squares solution at each instant k and are optimum in a deterministic sense. Convergence to the statistically optimum weight vector $\mathbf{w}_{opt}$ is often faster than that obtained using the LMS algorithm because it is independent of the eigenvalue spread of $\mathbf{R}_u$.

step, $\mathbf{w}_M(K)$ is chosen to minimize a weighted sum of past squared errors

$$\min_{\mathbf{w}_M(K)} \sum_{k=0}^K \lambda^{K-k} |y_d(k) - \mathbf{w}_M^H(K) \mathbf{u}(k)|^2. \quad (5.5)$$

λ is a positive constant less than one which determines how quickly previous data are deemphasized. The RLS algorithm is obtained from (5.5) by expanding the magnitude squared and applying the matrix inversion lemma. Table 5.1 summarizes both the LMS and RLS algorithms.

VI. INTERFERENCE CANCELLATION AND PARTIALLY ADAPTIVE BEAMFORMING

The computational requirements of each update in adaptive algorithms are proportional to either the weight vector dimension M or dimension squared (M^2). If M is large, this requirement is quite severe and for practical real time implementation it is often necessary to reduce M .

The expression *degrees of freedom* refers to the number of unconstrained or "free" weights in an implementation. For example, an LCMV beamformer with L constraints on N weights has $N-L$ degrees of freedom; the GSC implementation separates these as the unconstrained weight vector $\mathbf{w}_o$. There are M degrees of freedom in the structure of Fig. 5.1. A *fully adaptive* beamformer uses all available degrees of freedom and a *partially adaptive* beamformer uses a reduced set of degrees of freedom. Reducing degrees of freedom lowers computational requirements and often improves adaptive response time.¹ However, there is a performance penalty associated with reducing degrees of freedom. A partially adaptive beamformer cannot converge to the same optimum solution as the fully adaptive beamformer. The goal of partially adaptive beamformer design is to reduce degrees of freedom without significant degradation in performance.

The discussion in this section is general, applying to different types of beamformers although we borrow much of the notation from the GSC. We assume the beamformer is described by the adaptive structure of Fig. 5.1 where the desired signal y_d is obtained as $y_d = \mathbf{w}_o^H \mathbf{x}$ and the data vector $\mathbf{u}$ as $\mathbf{u} = \mathbf{T}^H \mathbf{x}$. Thus, the beamformer output is $y = \mathbf{w}^H \mathbf{x}$ where $\mathbf{w} = \mathbf{w}_o - \mathbf{T} \mathbf{w}_M$. In order to distinguish between fully and partially adaptive implementations we decompose $\mathbf{T}$ into a product of two matrices $\mathbf{C}_n \mathbf{T}_M$. The definition of $\mathbf{C}_n$ depends on the particular beamformer and $\mathbf{T}_M$ represents the mapping which reduces degrees of freedom. The MSC and GSC are obtained as special cases of this representation. In the MSC $\mathbf{w}_o$ is an N vector that selects the primary sensor, $\mathbf{C}_n$ is an N by $N-1$ matrix that selects the $N-1$ possible auxiliary sensors from the complete set of N sensors, and $\mathbf{T}_M$ is an $N-1$ by M matrix that selects the M auxiliary sensors actually utilized. In terms of

¹Adaptive algorithm convergence characteristics have not been discussed in this paper. Generally, more data are required to derive an accurate estimate of a larger optimum weight vector with block adaptive processing or RLS [Reed et al., 1974]. With LMS, convergence is governed by the covariance matrix eigenvalue spread, which tends to be larger for larger dimensional problems.

the GSC, $\mathbf{w}_o$ and $\mathbf{C}_n$ are defined as in Section IV and $\mathbf{T}_M$ is an $N-L$ by M matrix that reduces degrees of freedom ($M < N-L$).

This section begins by considering the interference cancellation process in these general beamformer implementations. This develops the intuition required for understanding why and how the number of adaptive weights can be reduced. We conclude this section by surveying different partially adaptive beamformer design philosophies.

A. Interference Cancellation Vs Degrees of Freedom.

The results in this subsection depend on $\mathbf{T}$ and are independent of the individual terms $\mathbf{C}_n$ and $\mathbf{T}_M$. We assume that the beamformer does not cancel the desired signal (see Section IV.E.) and that the optimum weights affect only interferers and uncorrelated noise. This simplifies the analysis by permitting us to exclude consideration of the desired signal.

Suppose a narrowband interfering source of frequency ω_o arrives at the array from direction θ_1 . The response of the $\mathbf{w}_o$ branch is $g_1 = \mathbf{w}_o^H \mathbf{d}(\theta_1, \omega_o)$. Perfect cancellation of this source requires $\mathbf{w}_M^H \mathbf{d}(\theta_1, \omega_o) = 0$ so we must choose $\mathbf{w}_M$ to satisfy

$$\mathbf{w}_M^H \mathbf{T}^H \mathbf{d}(\theta_1, \omega_o) = g_1. \quad (6.1)$$

If we assume that $\mathbf{T}^H \mathbf{d}(\theta_1, \omega_o)$ is nonzero, (6.1) represents a system of one equation in M unknowns (elements of $\mathbf{w}_M$) for which a solution always exists. To simultaneously cancel a second interferer located at θ_2 , $\mathbf{w}_M$ must satisfy

$$\mathbf{w}_M^H [\mathbf{T}^H \mathbf{d}(\theta_1, \omega_o) \quad \mathbf{T}^H \mathbf{d}(\theta_2, \omega_o)] = [g_1 \quad g_2] \quad (6.2)$$

where $g_2 = \mathbf{w}_o^H \mathbf{d}(\theta_2, \omega_o)$. Assuming $\mathbf{T}^H \mathbf{d}(\theta_2, \omega_o)$ and $\mathbf{T}^H \mathbf{d}(\theta_1, \omega_o)$ are linearly independent and nonzero, and provided $M \geq 2$, then at least one $\mathbf{w}_M$ exists that satisfies (6.2). Continuing this reasoning, we see that $\mathbf{w}_M$ can be chosen to cancel M narrowband interferers (assuming the $\mathbf{T}^H \mathbf{d}(\theta_i, \omega_o)$ are linearly independent and nonzero), independent of $\mathbf{T}$. Total cancellation occurs if $\mathbf{w}_M$ is chosen so the response of $\mathbf{T} \mathbf{w}_M$ perfectly matches the $\mathbf{w}_o$ branch response to the interferers.

So far we have only considered narrowband point interferers. Uncorrelated noise will be present in any real system and contributes to the output power. In an optimum beamformer $\mathbf{w}_M$ is chosen to minimize the overall output power. Recall that the output power due to uncorrelated noise is proportional to the L_2 norm squared of the overall weight vector $\mathbf{w}$ (white noise gain). The norm of $\mathbf{w}$ can become large when $\mathbf{w}_M$ is chosen to provide total interference cancellation, depending on the choice for $\mathbf{T}$ and the interferer locations. Thus, although in principle point sources of energy in direction and frequency can be totally canceled with one weight per interferer independent of $\mathbf{T}$, the presence of uncorrelated noise results in the degree of cancellation being dependent on the mapping described by $\mathbf{T}$.

Now consider interferers that are spatial point sources but emit broadband energy on the band $\omega_a \leq \omega \leq \omega_b$. The response of the $\mathbf{w}_o$ branch to an interferer at θ_1 is $\mathbf{w}_o^H \mathbf{d}(\theta_1, \omega) = g_1(\omega)$. To achieve total cancellation $\mathbf{w}_M$ must

be chosen to satisfy

$$\mathbf{w}_M^H \mathbf{T}^H \mathbf{d}(\theta_1, \omega) = g_1(\omega) \quad \omega_a \leq \omega \leq \omega_b. \quad (6.3a)$$

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

$$f_i(\omega) = [\mathbf{T}]_i^H \mathbf{d}(\theta_1, \omega) \quad 1 \leq i \leq M \quad (6.3b)$$

where $[\mathbf{T}]_i$ denotes the i th column of $\mathbf{T}$. (6.3a) requires $g_1(\omega)$ to be expressed as a linear combination of the $f_i(\omega)$, $1 \leq i \leq M$, on $\omega_a \leq \omega \leq \omega_b$. In general, this cannot be accomplished and we conclude that total cancellation of broadband interference cannot be obtained. The output power due to the broadband interferer can be expressed as the integral over frequency of the magnitude squared of the difference between the $\mathbf{w}_0$ branch and adaptive branch responses weighted by the interferer power spectrum. The degree of cancellation can vary dramatically and is critically dependent on the interferer direction, frequency content, and choice for $\mathbf{T}$. Good cancellation can be obtained in some situations when $M = 1$, while in others even large values of M result in poor cancellation. These conclusions are also valid for narrowband sources that are broad in direction (spatially distributed radiation).

B. Partially Adaptive Beamformer Design.

The preceding discussion indicates that the degree of interference cancellation is critically dependent on the ability of the adaptive channel ($\mathbf{T}\mathbf{w}_M$) to match the main beam response over the interferer frequency extent. This provides a means by which to evaluate partially adaptive beamformers. The majority of work reported on partially adaptive beamforming has been concerned with narrowband environments. We begin with a discussion of several narrowband approaches and briefly discuss their extension to broadband situations. We then consider techniques that are directly applicable to broadband situations. Some techniques select $\mathbf{T}_M$ for a given $\mathbf{C}_n$ while others select $\mathbf{T}$ directly.²

Several approaches to reducing degrees of freedom are based on processing a subset of the outputs of the matrix $\mathbf{C}_n$. This implies that $\mathbf{T}_M$ is a sparse matrix of zeros and ones. The outputs of $\mathbf{C}_n$ in the MSC are simply auxiliary sensor outputs. Morgan [1978] evaluated partially adaptive beamformer performance when $\mathbf{T}_M$ selected various subsets of the auxiliary outputs. This is termed an element space approach, since a subset of the sensor element outputs is utilized. Several investigators, including Vural [1977], Adams, et al. [1980], and Gabriel [1986a] have considered choosing the columns of $\mathbf{T}$ to form beams. This is traditionally termed a beam space approach. The columns of $\mathbf{C}_n$ are designed as data independent beamformers, each steered to a different direction, and $\mathbf{T}_M$ can be used to select a subset of the beam outputs. The objective is to direct a beam at each interfering source so that it can be subtracted from the output of the $\mathbf{w}_0$ branch. One way to

accomplish this is by selecting enough beams to cover all possible directions from which interferers might arrive. Another is to utilize source direction finding techniques to select which beams correspond to estimated interferer directions. The biggest advantage of the element space approach is the simplicity of implementation. Improved performance obtained using beam space processing is especially evident for interference due to either spatially distributed sources or sources with appreciable temporal bandwidth. However, this improvement is obtained at the expense of implementing the required beams.

Chapman [1976] and Owsley [1978] have considered choosing the columns of $\mathbf{T}$ to select subarrays, i.e., each column involves only a subset of the sensors in the array. The weightings applied to each subarray (elements of $\mathbf{T}$) can be chosen in various ways, one of which is to use the subarray to form a beam. Performance depends on the number of sensors in each subarray, which sensors are used in each subarray, and the weightings used to combine the sensor outputs in each subarray. Note that each column of $\mathbf{T}$ will have zeros in locations corresponding to sensors excluded from that subarray, so the overall $\mathbf{T}$ is of sparse structure.

Owsley [1985] suggests a narrowband method for the GSC in which the columns of $\mathbf{T}_M$ are chosen as a basis for the space spanned by the fully adaptive weight vectors. The dimension of this space is given by the rank of the spatially correlated component of the interference covariance matrix. Van Veen [1988a] extended this approach to the broadband case. The dimension of the fully adaptive space can be large in this case since it is given by the rank of the correlated components of the broadband interference covariance matrix.

The above approaches are capable of satisfactory performance with narrowband interference since cancellation requires about one degree of freedom per interferer to match the main beam response at each interferer direction. However, with broadband interference the main beam response must be matched over a range of frequency at each interferer direction. Although the narrowband approaches can be extended, it is difficult to do this and keep the number of adaptive weights M small. For example, several banks of beams could be designed to span the range of directions, with each bank operating at different frequency. The problem is that the number of beams required can become large as the frequency bandwidth increases. We seek a $\mathbf{T}$ that is efficient, i.e., provides good cancellation with a minimum of columns.

Van Veen and Roberts [1987a] have considered an optimization based approach for choosing the columns of $\mathbf{T}$ in the context of LCMV beamforming. The matrix $\mathbf{C}_n$ is designed to meet the constraints, reducing the problem to an unconstrained optimization over the elements of $\mathbf{T}_M$. $\mathbf{T}_M$ is chosen to minimize the average interference output power over a range of likely interference environments. Let a vector $\boldsymbol{\alpha}$ parameterize the interference environment of interest. In general $\boldsymbol{\alpha}$ can represent interferer locations, power levels, spectral distributions, numbers of in-

²In principle one can generate auxiliary constraints in an LCMV beamformer to reduce the number of adaptive weights in a GSC implementation (Griffiths [1987]). Here we assume all constraints are already specified in partially adaptive LCMV beamforming.

terferers, etc. Defining $P_I(\alpha)$ as the interference output power, T_M is chosen according to

$$\min_{T_M} \int_{\alpha_a}^{\alpha_b} P_I(\alpha) d\alpha \quad (6.4)$$

where $[\alpha_a, \alpha_b]$ denotes the set of interference scenarios of interest. Since output power corresponds to the error between the w_o branch and adaptive channel responses, in effect (6.4) selects T_M to provide the best response match possible for interference environments in $[\alpha_a, \alpha_b]$.

Equation (6.4) represents a design problem that is nonlinear in the design parameters (elements of T_M). A suboptimal approach to solving (6.4), in which T_M is sequentially designed one column at a time, has been shown to be effective in achieving near fully adaptive interference cancellation using a small adaptive dimension. The problem is still nonlinear; however an effective approximate solution is obtained by solving a linear least squares problem. An interpretation of this approximate solution is given in Van Veen [1988b].

Fig. 6.1 illustrates the magnitude response at several frequencies for a partially adaptive beamformer designed using this approach for the same interference environment and array geometry as the fully adaptive example in Section IV (see Fig. 4.2). The partially adaptive beamformer has 28 adaptive weights compared to 70 for the fully adaptive beamformer. Fig. 6.2 compares the magnitude response of the fully and partially adaptive beamformers as a function of frequency at the interferer DOA's. T_M is designed assuming α is a two-dimensional vector with each element corresponding to the DOA of an interferer. Both interferers have white spectra on $2\pi/5 \leq \omega \leq 4\pi/5$ and power levels of 40 dB relative to white noise. The

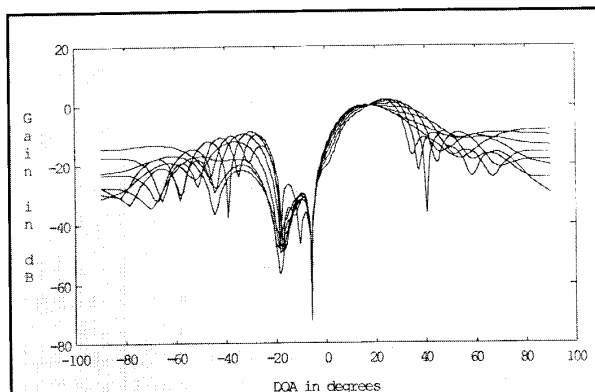


Figure 6.1 This plot depicts the magnitude response of a partially adaptive beamformer designed to minimize average output power at eight frequencies on the normalized frequency interval $[2\pi/5, 4\pi/5]$. The beamformer uses 28 adaptive weights and was designed assuming two interferers of power 40 dB relative to white noise arriving between -45 and 45 degrees. The actual interference environment, constraints, and array geometry are the same as in Fig. 4.2. Note that the beamformer of Fig. 4.2 uses 70 adaptive weights.

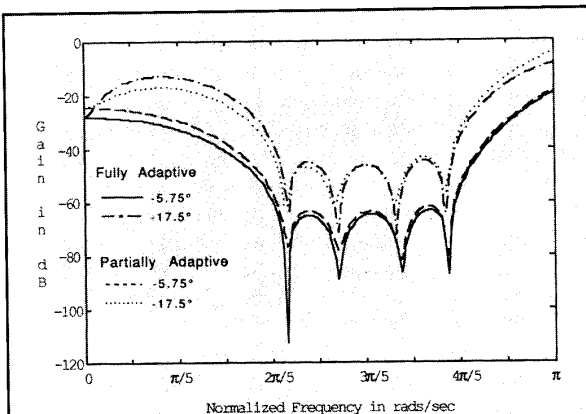


Figure 6.2 This plot depicts the magnitude response of the partially and fully adaptive beamformers corresponding to Figs. 4.2 and 6.1 as a function of frequency at the interferer DOA's.

region $[\alpha_a, \alpha_b]$ includes all DOA's between -45 and 45 degrees. Partially and fully adaptive array gain was evaluated at 190 points in this region; the maximum difference between partially adaptive and fully adaptive array gain was 2.4 dB and the average .7 dB for this particular design.

VII. BEAMFORMER IMPLEMENTATIONS

In its simplest form, a beamformer represents a linear combination of the sensor data. However, there are many different ways of implementing the weighted sum, each with its own performance and complexity characteristics. The section begins by introducing a general description for implementations and then briefly overviews several beamformer implementations. For additional detail the reader is encouraged to consult the references given in this section.

In essence, most implementations can be represented by decomposing w into a product of matrices and a vector

$$w = \left[\prod_{i=1}^{v_o} V_i \right] w_v \quad (7.1)$$

V_i , $1 \leq i \leq v_o$, are a series of matrix transformations of conformable dimensions and w_v is a vector. As a general rule, the matrix transformations are chosen to enhance performance and/or lower computational complexity. The FFT implementation of the DFT is analogous to (7.1), since the DFT matrix can be expressed as a product of permutation and butterfly matrices, representing a series of very simple computations. The GSC is an example of (7.1) where $v_o = 1$, $V_1 = [w_o | C_n]$, and $w_v = [1, w_n^H]^H$. V_1 simplifies adaptive algorithm implementation by permitting unconstrained adaptation of w_n . In order to simplify multiplication of the data by w_o and C_n , several researchers have studied decomposition of $[w_o | C_n]$ into products of simple matrices (e.g., Kalson and Yao [1985] and Ward, et al. [1986]).

As discussed in Section II, broadband beamforming can be performed in the frequency domain or time domain. The weights used to obtain the outputs at each frequency for the frequency domain beamformer depicted in Fig. 2.4 are easily represented in terms of (7.1) by employing the matrix representation of the DFT. Frequency domain optimum beamformers usually choose the weights at each frequency based solely on the data at the frequency. This partitioning of the data influences both performance and computational complexity. Discussion and evaluation of frequency domain beamforming is given in Owsley [1985], Vural [1977], and Gabriel [1986b].

Systolic implementations of optimum beamformers have been studied by a number of investigators. They are usually designed to both compute and implement the adaptive weights. In general, systolic implementations can be described in terms of (7.1) where each V_i has a structure amenable to parallel computation and local communication. McWhirter [1983] (see also Haykin [1986] ch. 10) has developed a systolic array that computes the beamformer output without explicit computation of the adaptive weight vector. Additional references to systolic implementations include Schreiber and Kuekes [1985], Ward, et al. [1986], Owsley [1987], and Van Veen and Roberts [1987b].

The study of beamformer implementations is an evolving research area. Future developments will result from advances in VLSI and parallel computing technologies.

VIII. SUMMARY

A beamformer forms a scalar output signal as a weighted combination of the data received at an array of sensors. The weights determine the spatial filtering characteristics of the beamformer and enable separation of signals having overlapping frequency content if they originate from different locations. The weights in a data independent beamformer are chosen to provide a fixed response independent of the received data. Statistically optimum beamformers select the weights to optimize the beamformer response based on the statistics of the data. The data statistics are often unknown and may change with time so adaptive algorithms are used to obtain weights that converge to the statistically optimum solution. Computational considerations dictate the use of partially adaptive beamformers with arrays composed of large numbers of sensors. Many different approaches have been proposed for implementing optimum beamformers. Future work will likely address signal cancellation problems, further reductions in computational load for large arrays and improved structures for implementation. Beamforming truly represents a versatile approach to spatial filtering.

REFERENCES

- Adams, R. N., Horowitz, L. L. and Senne, K. D. [1980], "Adaptive main-beam nulling for narrow-beam antenna arrays," *IEEE Trans. on AES*, Vol. 16, pp. 509–516, Jul, 1980.
- Alexander, S. T. [1986], *Adaptive Signal Processing: Theory and Applications*, Springer-Verlag, New York, 1986.
- Applebaum, S. P. [1966], "Adaptive arrays," Syracuse Un. Research Corp., Report SURC SPL TR 66-001, Aug. 1966 (reprinted in *IEEE Trans. on AP*, Vol. AP-24, pp. 585–598, Sept. 1976).
- Applebaum, S. P. and Chapman, D. J. [1976], "Adaptive arrays with main beam constraints," *IEEE Trans. on AP*, Vol. AP-24, pp. 650–662, Sept. 1976.
- Brookner, E. [1985], "Phased-array radar," *Scientific American*, Vol. 252, pp. 94–102, Feb. 1985.
- Buckley, K. M. [1987], "Spatial/spectral filtering with linearly-constrained minimum variance beamformers," *IEEE Trans. on ASSP*, Vol. ASSP-35, pp. 249–266, Mar. 1987.
- Butler, J. K. and Unz, E. [1967], "Beam efficiency and gain optimization of antenna arrays with nonuniform spacings," *Radio Sci.*, Vol. 2, pp. 711–720, 1967.
- Chapman, D. J. [1976], "Partial adaptivity for the large arrays," *IEEE Trans. on AP*, Vol. 24, pp. 685–696, Sept. 1976.
- Compton, Jr., R. T. [1978], "An adaptive array in a spread-spectrum communication system," *Proc. IEEE*, Vol. 66, pp. 289–298, Mar. 1978.
- Compton, Jr., R. T. [1988], *Adaptive Antennas: Concepts and Performance*, Prentice-Hall, Englewood Cliffs, New Jersey, 1988.
- Cox, H. [1973], "Resolving power and sensitivity to mismatch of optimum array processors," *Jour. Acoust. Soc. Amer.*, Vol. 54, No. 3, pp. 771–785, 1973.
- Cox, H., Zeskind, R. M. and Owen, M. M. [1987], "Robust adaptive beamforming," *IEEE Trans. on ASSP*, Vol. ASSP-35, pp. 1365–1375, Oct. 1987.
- Dolph, C. L. [1946], "A current distribution for broadside arrays which optimizes the relationship between beam width and side-lobe level," *Proc. of IRE*, Vol. 34, pp. 335–348, June 1946.
- Er, M. H. and Cantoni, A. [1983], "Derivative constraints for broad-band element space antenna array processors," *IEEE Trans. on ASSP*, Vol. ASSP-31, pp. 1378–1393, Dec. 1983.
- Frost III, O. L. [1972], "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, Vol. 60, pp. 926–935, Aug. 1972.
- Gabriel, W. F. [1986a], "Using spectral estimation techniques in adaptive processing antenna systems," *IEEE Trans. on AP*, Vol. 34, pp. 291–300, Mar. 1986.
- Gabriel, W. F. [1986b], "Adaptive digital processing investigation of DFT subbanding vs. transversal filter canceller," *NRL Report 8981*, July, 1986.
- Gabriel, W. F. [1976], "Adaptive arrays—an introduction," *Proc. IEEE*, Vol. 64, pp. 239–272, Aug. 1976.
- Gee, W., Lee, S., Bong, N. K., Cain, C. A., Mittra, R. and Magin, R. L. [1984], "Focused array hyperthermia applicator: theory and experiment," *IEEE Trans. on BME*, Vol. BME-31, pp. 38–45, Jan. 1984.
- Griffiths, J. W. R., et al., ed. [1973], *Signal Processing*, Academic Press, 1973.
- Griffiths, L. J. and Jim, C. W. [1982], "An alternative ap-

- proach to linearly constrained adaptive beamforming," *IEEE Trans. on AP*, Vol. AP-30, pp. 27–34, Jan. 1982.
- Griffiths, L. J. [1987], "A new approach to partially adaptive arrays," *Proc. ICASSP-87*, pp. 1999–2002, Apr. 1987.
- Hanson, R. L. and Lawson, C. L. [1969], "Extensions and applications of the Householder algorithm for solving linear least squares problems," *Math. Comp.*, Vol. 23, pp. 917–926, 1969.
- Harrington, R. F. [1961], "Sidelobe reduction by non-uniform element spacing," *IRE Trans. on AP*, Vol. AP-9, pp. 187–192, Mar. 1961.
- Harris, F. J. [1978], "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. of IEEE*, Vol. 66, pp. 51–83, Jan. 1978.
- Haykin, S., ed. [1985], *Array Signal Processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 1985.
- Haykin, S. [1986], *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, New Jersey, 1986.
- Howells, P. W. [1966], "Intermediate frequency sidelobe canceller," *U.S. Patent 3202990*, Aug. 1966.
- Howells, P. W. [1976], "Explorations in fixed and adaptive resolution at GE and SURC," *IEEE Trans. on AP*, Vol. AP-24, pp. 575–584, Sept. 1976.
- Hudson, J. E. [1981], *Adaptive Array Principles*, Peter Peregrinus Ltd., London, 1981.
- Ishimaru, A. [1962], "Theory of unequally-spaced arrays," *IRE Trans. on AP*, Vol. AP-10, pp. 691–702, Nov. 1962.
- Jablon, N. K. [1968], "Steady state analysis of the generalized sidelobe canceller by adaptive noise cancelling techniques," *IEEE Trans. on AP*, Vol. AP-34, pp. 330–337, March 1986.
- Justice, J. H. [1985], in *Array Signal Processing*, S. Haykin, ed., Prentice-Hall, Englewood Cliffs, New Jersey, 1985.
- Kak, A. C. [1985], in *Array Signal Processing*, S. Haykin, ed., Prentice-Hall, Englewood Cliffs, New Jersey, 1985.
- Kalson, S. and Yao, K. [1985], "A systolic array for linearly constrained least squares filtering," *Proc. ICASSP-85*, pp. 977–980, Mar. 1985.
- Kelly, Jr., E. L. and Levin, M. L. [1964], "Signal parameter estimation for seismometer arrays," *Mass. Inst. Tech., Lincoln Lab., Tech. Report 339*, Jan. 1964.
- Knight, W. C., Pridham, R. G. and Kay, S. M. [1981], "Digital signal processing for sonar," *Proc. IEEE*, Vol. 69, pp. 1451–1506, Nov. 1981.
- Kumar, A. and Murthy, P. K. [1977], "Synthesis of equally excited linear arrays," *IRE Trans. on AP*, Vol. AP-25, pp. 425–428, May 1977.
- Lo, Y. T. [1963], "Sidelobe level in nonuniformly spaced antenna arrays," *IEEE Trans. on AP*, Vol. AP-11, pp. 511–512, July 1963.
- Macovski, A. [1983], *Medical Imaging*, Prentice-Hall, Englewood Cliffs, New Jersey, 1983.
- Marr, J. [1986], "A selected bibliography on adaptive antenna arrays," *IEEE Trans. on AES*, Vol. AES-22, pp. 781–798, Nov. 1986.
- Mayhan, J. T. [1976], "Nulling limitations for a multiple-beam antenna," *IEEE Trans. on AP*, Vol. 24, pp. 769–779, Nov. 1976.
- McWhirter, J. G. [1983], "Recursive least-squares minimization using a systolic array," *Proc. SPIE 431*, pp. 105–112, 1983.
- Monzingo, R. and Miller, T. [1980], *Introduction to Adaptive Arrays*, Wiley and Sons, New York, 1980.
- Munson, Jr., D. C., O'Brian, J. D. and Jenkins, W. K. [1983], "A tomographic formulation of spot-light mode synthetic aperture radar," *Proc. IEEE*, Vol. 71, pp. 917–925, Aug. 1983.
- Morgan, D. R. [1978], "Partially adaptive array techniques," *IEEE Trans. on AP*, Vol. 26, pp. 823–833, Nov. 1978.
- Owsley, N. L. [1973], "A recent trend in adaptive spatial processing for sensor arrays: constrained adaptation," in *Signal Processing*, ed., J. W. R. Griffiths, et al., pp. 591–604, Academic Press, 1973.
- Owsley, N. L. [1985], in *Array Signal Processing*, S. Haykin, ed., Prentice-Hall, Englewood Cliffs, New Jersey, 1985.
- Owsley, N. L. [1987], "Systolic array adaptive beamforming," *Naval Underwater Systems Center Report 7971*, April, 1987.
- Parks, T. W. and Burrus, C. S. [1987], *Digital Filter Design*, Wiley-Interscience, New York, 1987.
- Peterson, P. M., Durlach, N. I., Rabinowitz, W. M. and Zurek, P. M. [1987], "Multimicrophone adaptive beamforming for interference reduction in hearing aids," *Jour. of Rehab. R&D*, Vol. 24, Fall 1987.
- Pratt, W. K. [1978], *Digital Image Processing*, Wiley and Sons, New York, 1978.
- Readhead, A. [1982], "Radio astronomy by very-long-baseline interferometry," *Scientific American*, Vol. 246, pp. 52–61, June 1982.
- Reed, I. S., Mallett, J. D. and Brennan, L. E. [1974], "Rapid convergence rate in adaptive arrays," *IEEE Trans. on AES*, Vol. AES-10, pp. 853–863, Nov. 1974.
- Sanzgiri, S. M. and Butler, J. K. [1971], "Constrained optimization of the performance indices of arbitrary array antennas," *IEEE Trans. on AP*, Vol. AP-19, pp. 725–730, July 1971.
- Schelkunoff, S. A. [1943], "A mathematical theory of linear arrays," *Bell Sys. Tech. Jour.*, Vol. 22, pp. 80–107, Jan. 1943.
- Schreiber, R. and Kuekes, P. J. [1985], in *VLSI and Modern Signal Processing*, S. Y. Kung, H. J. Whitehouse and T. Kailath, eds., Prentice-Hall, Englewood Cliffs, New Jersey, 1985.
- Shan, T. and Kailath, T. [1985], "Adaptive beamforming for coherent signals and interference," *IEEE Trans. on ASSP*, Vol. ASSP-33, pp. 527–536, June 1985.
- Skolnik, M. I., Nemhauser, G. and Sherman, J. W. [1964], "Dynamic programming applied to unequally spaced arrays," *IEEE Trans. on AP*, Vol. AP-12, pp. 35–43, Jan. 1964.
- Treichler, J. R., Johnson, Jr., C. R. and Larimore, M. G. [1987], *Theory and Design of Adaptive Filters*, Wiley-Interscience, New York, 1987.
- Unz, H. [1956], "Linear arrays with arbitrarily distributed elements," *U. of Cal., Berkeley, Elect. Reac. Lab. Rept. 168*, Nov. 1956.
- Van Veen, B. D. and Roberts, R. A. [1987a], "Partially adap-

tive beamformer design via output power minimization," *IEEE Trans. on ASSP*, Vol. ASSP-35, pp. 1534–1532, Nov. 1987.

Van Veen, B. D. and Roberts, R. A. [1987b], "Systolic arrays for partially adaptive beamforming," *21-st Asilomar Conference on Signals, Systems, and Computers*, pp. 584–588, Nov. 1987.

Van Veen, B. D. [1988a], "Eigenstructure based partially adaptive array design," *IEEE Trans. on AP*, Vol. AP-36, pp. 357–362, Mar. 1988.

Van Veen, B. D. [1988b], "Improved power minimization based partially adaptive beamforming," *ICASSP-88*, Apr. 1988.

Vural, A. M. [1977], "A comparative performance study of adaptive array processors," *Proc. ICASSP-77*, pp. 695–700, Apr. 1977.

Ward, C. R., Hargrave, P. J. and McWhirter, J. G. [1986], "A novel algorithm and architecture for adaptive digital beamforming," *IEEE Trans. on AP*, Vol. AP-34, pp. 338–346, Mar. 1986.

Widrow, B., Mantey, P. E., Griffiths, L. J. and Goode, B. B. [1967], "Adaptive antenna systems," *Proc. IEEE*, Vol. 55, pp. 2143–2159, Dec. 1967.

Widrow, B., Glover, Jr., J. R., McCool, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., Zeidler, J. R., Dong, Jr., E. and Goodlin, R. C. [1975], "Adaptive noise cancelling: principles and applications," *Proc. IEEE*, Vol. 63, pp. 1692–1716, Dec. 1975.

Widrow, B., Duvall, K. M., Gooch, R. P. and Newman, W. C. [1982], "Signal cancellation phenomena in adaptive arrays: causes and cures," *IEEE Trans. on AP*, Vol. AP-30, pp. 469–478, May 1982.

Widrow, B. and Stearns, S. [1985], *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 1985.

Yang, J. and Kaveh, M. [1987], "Wide-band adaptive arrays based on the coherent signal-subspace transformation," *Proc. ICASSP-87*, pp. 2011–2013, Apr. 1987.

Yen, J. L. [1985], in *Array Signal Processing*, S. Haykin ed., Prentice-Hall, Englewood Cliffs, New Jersey, 1985.



Kevin M. Buckley (M'81) was born in Washington, DC, on December 4, 1954. He received the B.S. and M.S. degrees in electrical engineering from Villanova University, Philadelphia, PA, in 1976 and 1980, respectively, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1986.

From 1980 to 1982, he worked at Sonic Sciences Inc., Warminster, PA, on communication and sonar applications of digital signal processing. In 1982–1983 he worked as a full time Instructor at Villanova University. He is currently an Assistant Professor in the Department of Electrical Engineering, University of Minnesota, Minneapolis. His primary research and teaching interests are in the theory and application of sensor array processing, spectrum estimation, and adaptive algorithms.



Barry D. Van Veen was born in Green Bay, WI, on June 20, 1961. He received the B.S. degree from Michigan Technological University in 1983 and the Ph.D. degree from the University of Colorado in 1986, both in electrical engineering. He was an ONR Fellow while working on the Ph.D. degree. In the spring of 1987 he was with the Department of Electrical and Computer Engineering at the University of Colorado–Boulder. Since August of 1987 he has been with the Department of Electrical and Computer Engineering at the University of Wisconsin–Madison as an Assistant Professor. His research interests include array processing, spectral estimation, and adaptive algorithms.