Barry Van Veen, et. Al. “Beamforming Techniques for Spatial Filtering.”

Beamforming Techniques for Spatial Filtering

Barry Van Veen
University of Wisconsin

Kevin M. Buckley
Villanova University

61.1 Introduction

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, desired and interfering signals often originate from different spatial locations. This spatial separation can be exploited to separate signal from interference using a spatial filter at the receiver.

A beamformer is a processor used in conjunction with an array of sensors to provide a versatile form of spatial filtering. The term beamforming derives from the fact that early spatial filters were designed to form pencil beams (see polar plot in Fig. 61.5(c)) 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 section we discuss formation of beams for reception, providing an overview of beamforming from a signal processing perspective. Data independent, statistically optimum, adaptive, and partially adaptive beamforming are discussed.

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. A single sensor such as an antenna, sonar transducer, or microphone collects impinging energy over a continuous aperture, providing spatial filtering by summing coherently waves that are in phase across the aperture while destructively combining waves that are not. An array of sensors provides a discrete sampling across its aperture. When the spatial sampling is discrete, the processor that performs the spatial filtering is termed a beamformer. Typically a 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 because 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.

This section begins with the definition of basic terminology, notation, and concepts. Succeeding sections cover data-independent, statistically optimum, adaptive, and partially adaptive beamforming. We then conclude with a summary.

Throughout this section we use methods and techniques from FIR filtering to provide insight into various aspects of spatial filtering with 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 multidimensional. 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.

61.2 Basic Terminology and Concepts

In this section we introduce terminology and concepts employed throughout. We begin by defining the beamforming operation and discussing spatial filtering. Next we introduce second order statistics of the array data, developing representations for the covariance of the data received at the array and discussing distinctions between narrowband and broadband beamforming. Last, we define various types of beamformers.

61.2.1 Beamforming and Spatial Filtering

Figure 61.1 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 :

$$y(k) = \sum_{l=1}^{J} w_l^* x_l(k) \tag{61.1}$$

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

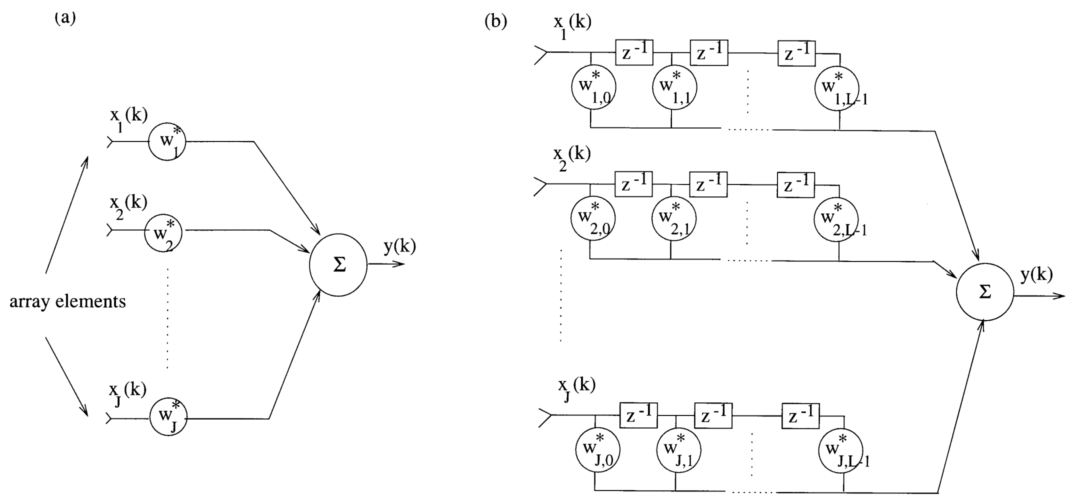

FIGURE 61.1: A beamformer forms a linear combination of the sensor outputs. In (a), sensor outputs are multiplied by complex weights and summed. This beamformer is typically used with narrowband signals. A common broadband beamformer is illustrated in (b).

The second beamformer in Fig. 61.1 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

$$y(k) = \sum_{l=1}^{J} \sum_{p=0}^{K-1} w_{l,p}^* x_l(k-p) \tag{61.2}$$

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 that permits us to treat both beamformers in Fig. 61.1 simultaneously. Note that Eqs. (61.1) and (61.2) can be written as

$$y(k) = \mathbf{w}^H \mathbf{x}(k) \tag{61.3}$$

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 Eq. (61.2) and $N = J$ when referring to Eq. (61.1). Except for Section 61.5 on adaptive algorithms, we will drop the time index and assume that its presence is understood throughout the remainder of the paper. Thus, Eq. (61.3) is written as $y = \mathbf{w}^H\mathbf{x}$. Many of the techniques described in this section are applicable to continuous time as well as discrete time beamforming.

The frequency response of an FIR filter with tap weights w_p^*, $1 \leq p \leq J$ and a tap delay of T seconds is given by

$$r(\omega) = \sum_{p=1}^{J} w_p^* e^{-j\omega T(p-1)} . \tag{61.4}$$

Alternatively

$$r(\omega) = \mathbf{w}^H \mathbf{d}(\omega) \tag{61.5}$$

where $\mathbf{w}^H = [w_1^* \; w_2^* \; ...w_J^* \;]$ and $\mathbf{d}(\omega) = [1 \; e^{j\omega T} \; e^{j\omega 2T} \; ...e^{j\omega(J-1)T} \;]^H$. $r(\omega)$ represents the response of the filter[1] 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 section we do not consider range. Figure 61.2 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 lth sensor. Substitution into Eq. (61.2) results in the beamformer output

$$y(k) = e^{j\omega k} \sum_{l=1}^{J} \sum_{p=0}^{K-1} w_{l,p}^* e^{-j\omega[\Delta_l(\theta)+p]} = e^{j\omega k} \, r(\theta\omega) \tag{61.6}$$

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

$$r(\theta, \omega) = \mathbf{w}^H \mathbf{d}(\theta, \omega) . \tag{61.7}$$

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

$$\mathbf{d}(\theta, \omega) = [1 \; e^{j\omega\tau_2(\theta)} \; e^{j\omega\tau_3(\theta)} \; ...e^{j\omega\tau_N(\theta)} \;]^H . \tag{61.8}$$

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 ith weight is applied. We refer to $\mathbf{d}(\theta, \omega)$ as the array response vector. It is also known as the steering vector, direction vector, or array manifold vector. Nonideal 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 somewhat

[1] An FIR filter is by definition linear, so an input sinusoid produces at the output a sinusoid of the same frequency. The magnitude and argument of $r(\omega)$ are, respectively, the magnitude and phase responses.

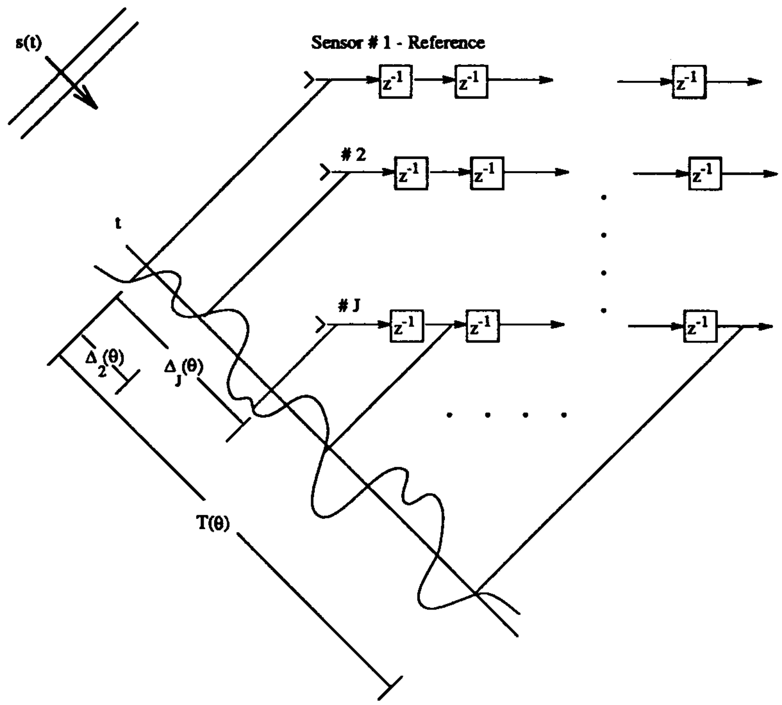

FIGURE 61.2: An array with attached delay lines provides a spatial/temporal sampling of propagating sources. This figure illustrates this sampling of a signal propagating in plane waves from a source located at DOA θ. With J sensors and K samples per sensor, at any instant in time the propagating source signal is sampled at JK nonuniformly 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 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 front 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 lth sensor.

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_o and the array geometry is linear and equi-spaced as illustrated in Fig. 61.3. 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 $\mathbf{d}(\theta, \omega_o)$ (beamformer) as $\omega = \omega_o(d/c)sin\theta$. Thus, temporal frequency in an FIR filter 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 (61.3) 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

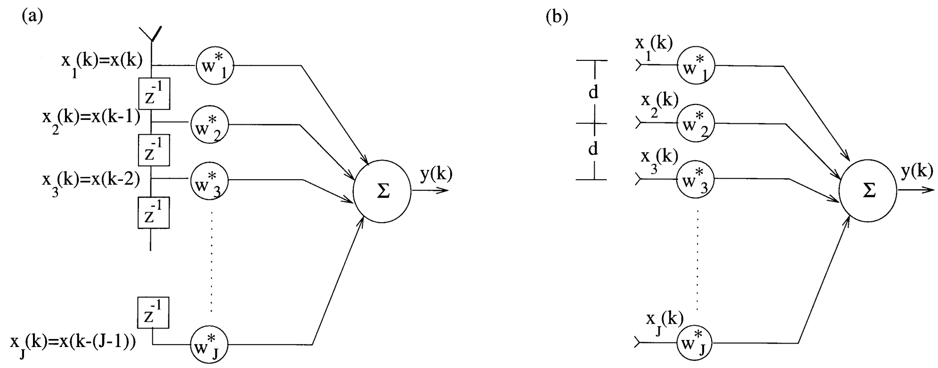

FIGURE 61.3: The analogy between an equi-spaced omni-directional narrowband line array and a single-channel FIR filter is illustrated in this figure.

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 $\mathbf{d}(\theta, \omega)$ 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 $\mathbf{d}(\theta_2, \omega_2)$.

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 $\mathbf{d}(\theta_1, \omega_o)$ and $\mathbf{d}(\theta_2, \omega_o)$. 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 sensor prevents this particular ambiguity.)

A primary focus of this section is on designing response via weight selection; however, (61.7) 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.

61.2.2 Second Order Statistics

Evaluation of beamformer performance usually involves power or variance, so the second order statistics of the data play an important role. We assume the data received at the sensors are zero mean throughout this section. 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 are 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 nonstationary data, the wide sense stationary assumption is used in developing statistically optimal beamformers and in evaluating steady state performance.

Suppose **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

$$\mathbf{R}_x = \frac{1}{2\pi} \int_{\omega_a}^{\omega_b} S(\omega)\ \mathbf{d}(\omega)\ \mathbf{d}^H(\omega)\ d\omega \tag{61.9}$$

with $\mathbf{d}(\omega)$ as defined for (61.5). Now assume the array data **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

$$\mathbf{R}_x = \frac{1}{2\pi} \int_{\omega_a}^{\omega_b} S(\omega)\ \mathbf{d}(\theta, \omega)\ \mathbf{d}^H(\theta, \omega)\ d\omega \tag{61.10}$$

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

$$\mathbf{R}_x = \sigma_s^2\ \mathbf{d}(\theta, \omega_o)\ \mathbf{d}^H(\theta, \omega_o) \tag{61.11}$$

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 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 (see Buckley [2]).

An array provides an effective temporal aperture over which a source is observed. Figure 61.2 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. 61.4. 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 usually termed frequency domain beamforming. The frequency domain beamformer outputs can be made equivalent to the DFT of the broadband beamformer output depicted in Fig. 61.1(b) with proper selection of beamformer weights and careful data partitioning.

61.2.3 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"

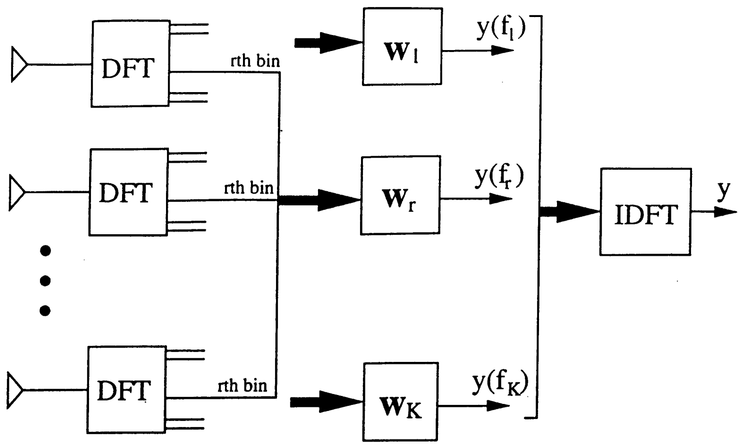

FIGURE 61.4: Beamforming is sometimes performed in the frequency domain when broadband signals are of interest. This figure illustrates transformation of the data at each sensor into the frequency domain. Weighted combinations of data at each frequency (bin) are performed. An inverse discrete Fourier transform produces the output time series.

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. 61.5.

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 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.

61.3 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 [8]). We shall exploit the analogies between beamforming and FIR filtering where possible in developing an understanding of the design problem. We also discuss aspects of the design problem 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.

61.3.1 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. 61.3. 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 [5].

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 because the output of each sensor is phase shifted prior to summation. Figure 61.5(b) depicts the magnitude of the actual response when $\mathbf{w} = \mathbf{T}\mathbf{d}(\theta_o, \omega_o)$, where $\mathbf{T}$ implements a common Dolph-Chebyshev tapering function. 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. The equivalence of the narrowband linear equi-spaced array and FIR filter (see Fig. 61.3) implies that the same techniques for choosing taper functions are applicable to either problem. Methods for choosing tapering weights also exist for more general array configurations.

61.3.2 General Data Independent Response Design

The methods discussed in this section 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 bandstop 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 desired response $r_d(\theta, \omega)$. Ad hoc techniques similar to those employed in FIR filter design can be used for selecting $\mathbf{w}$. Alternatively, formal optimization design methods can be employed (see, for example, Parks and Burrus [8]). Here, to illustrate the general optimization design approach, we only consider choosing $\mathbf{w}$ to minimize the weighted averaged square of the difference between desired and actual response.

Consider minimizing the squared error between the actual and desired response at P points $(\theta_i, \omega_i),\ 1 < i < P$. If $P > N$, then we obtain the overdetermined least squares problem

$$\min_{\mathbf{w}} \ |\mathbf{A}^H \mathbf{w} - \mathbf{r}_d|^2 \tag{61.12}$$

where

$$\mathbf{A} = [\mathbf{d}(\theta_1, \omega_1), \mathbf{d}(\theta_2, \omega_2) ... \mathbf{d}(\theta_P, \omega_P)]\ ; \tag{61.13}$$

$$\mathbf{r}_d = [r_d(\theta_1, \omega_1), r_d(\theta_2, \omega_2) ... r_d(\theta_P, \omega_P)]^H\ . \tag{61.14}$$

Provided $\mathbf{A}\mathbf{A}^H$ is invertible (i.e., $\mathbf{A}$ is full rank), then the solution to Eq. (61.12) is given as

$$\mathbf{w} = \mathbf{A}^+ \mathbf{r}_d \tag{61.15}$$

where $\mathbf{A}^+ = (\mathbf{A}\mathbf{A}^H)^{-1}\mathbf{A}$ is the pseudo-inverse of $\mathbf{A}$.

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

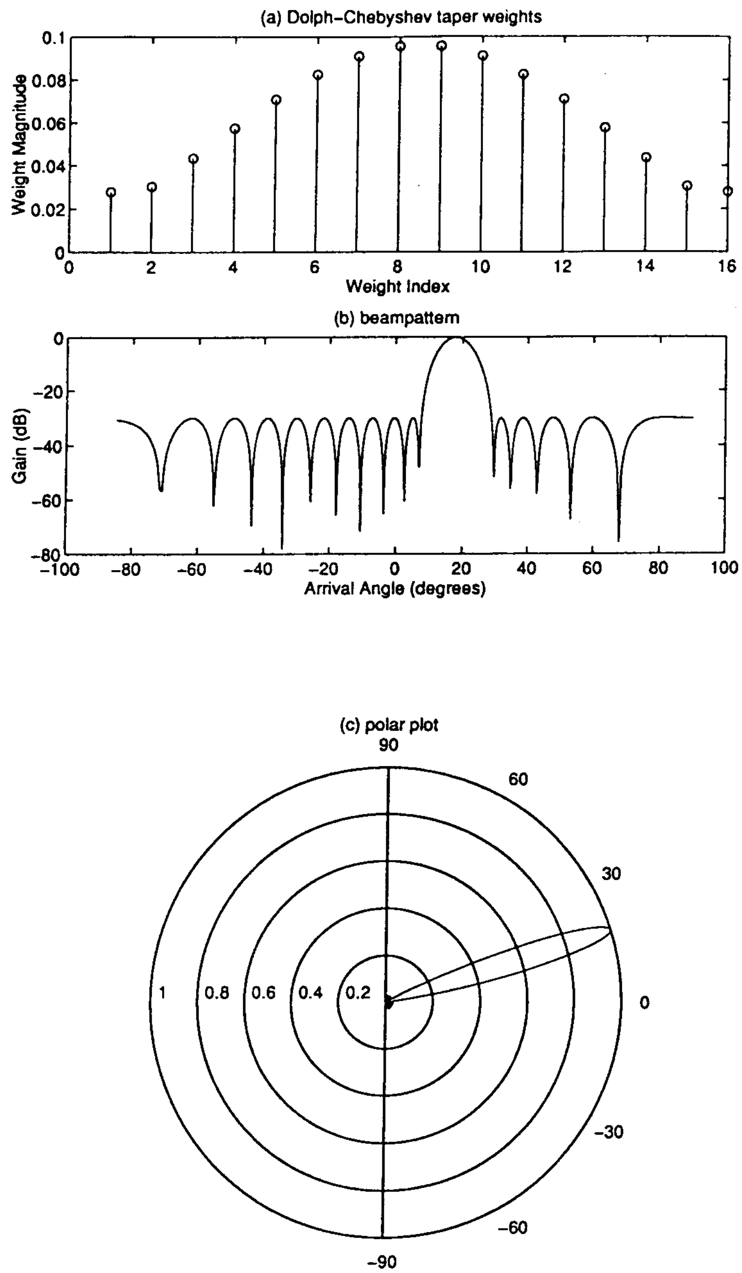

FIGURE 61.5: Beamformers come in both data independent and statistically optimum varieties. In (a) through (e) of this figure we consider an equi-spaced narrowband array of 16 sensors spaced at one-half wavelength. In (a), (b), and (c) the magnitude of the weights, the beampattern, and the beampattern, in polar coordinates are shown, respectively, for a Dolph-Chebyshev beamformer with -30 dB sidelobes. In (d) and (e) beampatterns are shown of statistically optimum beamformers which were 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 (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 in the interferer directions. The interferers in (e) are located between 20° and 23°, again with relative power of 30 dB. Again deep nulls are formed at the interferer directions; however, the sidelobe levels are significantly higher at other directions. (f) depicts the broadband 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 $-17.5°$ in the presence of white noise.

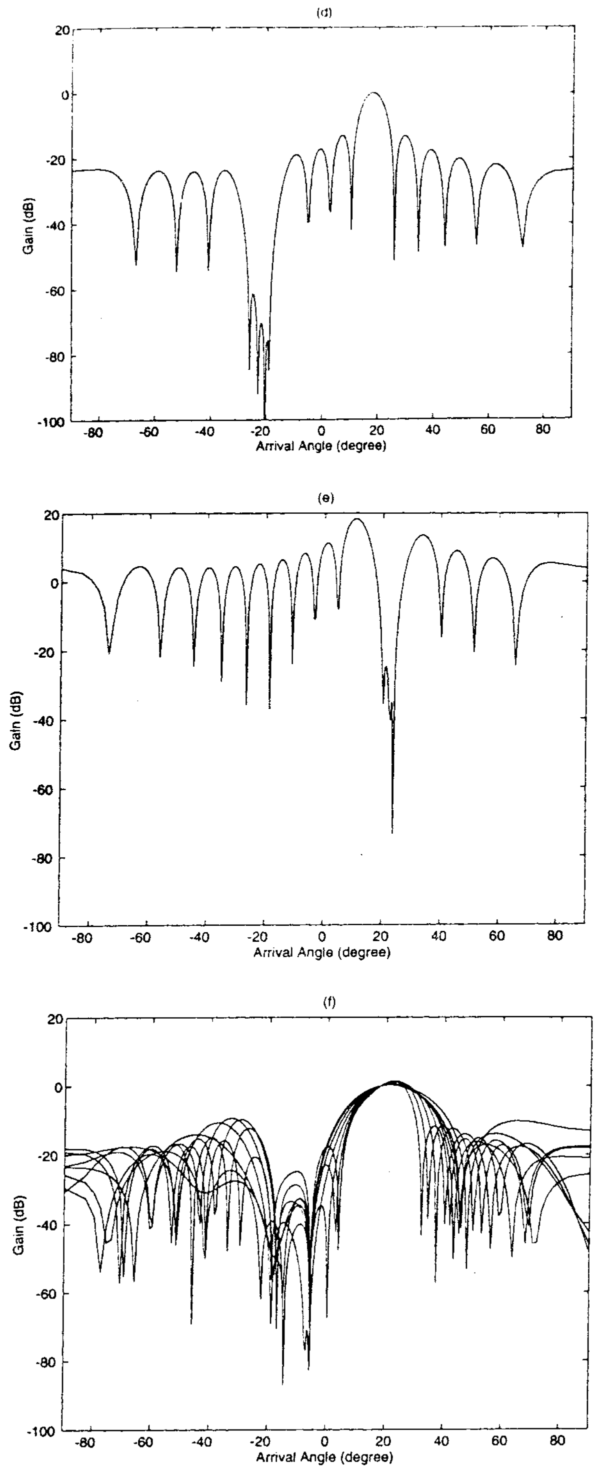

FIGURE 61.5: *(continued)* The interferers have a white spectrum on $[2\pi/5, 4\pi/5]$ and have powers of 40 dB and 30 dB relative to the white noise, respectively. The constraints are designed to present a unit gain and linear phase over $[2\pi/5, 4\pi/5]$ at a DOA of 18°. The array is linear equi-spaced with 16 sensors spaced at one-half wavelength for frequency $4\pi/5$ and five tap FIR filters are used in each sensor channel.

TABLE 61.1 Summary of Optimum Beamformers

Type	MSC	Reference signal	Max SNR	LCMV
Definitions	$\mathbf{x}_a$ — auxiliary data $\mathbf{y}_m$ — primary data $\mathbf{r}_{ma} = E\{\mathbf{x}_a \mathbf{y}_m^*\}$ $\mathbf{R}_a = E\{\mathbf{x}_a \mathbf{x}_a^H\}$ output: $\mathbf{y} = \mathbf{y}_m - \mathbf{w}_a^H \mathbf{x}_a$	$\mathbf{x}$ — array data $\mathbf{y}_d$ — desired signal $\mathbf{r}_{xd} = E\{\mathbf{x}\mathbf{y}_d^*\}$ $\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^H\}$ output: $\mathbf{y} = \mathbf{w}^H \mathbf{x}$	$\mathbf{x} = \mathbf{s} + \mathbf{x}$ — array data $\mathbf{s}$ — signal component $\mathbf{n}$ — noise component $\mathbf{R}_s = E\{\mathbf{s}\mathbf{s}^H\}$ $\mathbf{R}_n = E\{\mathbf{n}\mathbf{n}^H\}$ output: $\mathbf{y} = \mathbf{w}^H \mathbf{x}$	$\mathbf{x}$ — array data $\mathbf{C}$ — constraint matrix $\mathbf{f}$ — response vector $\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^H\}$ output: $\mathbf{y} = \mathbf{w}^H \mathbf{x}$
Criterion	$\min_{\mathbf{w}_a} E\{\lvert\mathbf{y}_m - \mathbf{w}_a^H \mathbf{x}_a\rvert^2\}$	$\min_{\mathbf{w}} E\{\lvert\mathbf{y} - \mathbf{y}_d\rvert^2\}$	$\max_{\mathbf{w}} \frac{\mathbf{w}^H \mathbf{R}_s \mathbf{w}}{\mathbf{w}^H \mathbf{R}_n \mathbf{w}}$	$\min_{\mathbf{w}} \{\mathbf{w}^H \mathbf{R}_x \mathbf{w}\} s.t. \mathbf{C}^H \mathbf{w} = \mathbf{f}$
Optimum weights	$\mathbf{w}_a = \mathbf{R}_a^{-1} \mathbf{r}_{ma}$	$\mathbf{w}_a = \mathbf{R}_x^{-1} \mathbf{r}_{rd}$	$\mathbf{R}_n^{-1} \mathbf{R}_s \mathbf{w} = \lambda_{\max} \mathbf{w}$	$\mathbf{w} = \mathbf{R}_x^{-1} \mathbf{C}[\mathbf{C}^H \mathbf{R}_x^{-1} \mathbf{C}]^{-1} f$
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 $\mathbf{R}_s$ and $\mathbf{R}_n$ Solve generalized eigenproblem for weights	Computation of constrained weight vector
References	Applebaum [1976]	Widrow [1967]	Monzingo and Miller [1980]	Frost [1972]

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 because 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 effective numerical dimension of the space spanned by the $\mathbf{d}(\theta_i, \omega_i)$, $1 \le i \le P$, 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 Eq. (61.12) by selection of the sample points (θ_i, ω_i) and/or unequally weighting of the error at each (θ_i, ω_i). Parks and Burrus [8] discuss this in the context of FIR filtering.

61.4 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 interfering signals. We discuss several different criteria for choosing statistically optimum beamformer weights. Table 61.1 summarizes these different approaches. Where possible, equations describing the criteria and weights are confined to Table 61.1. 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.

61.4.1 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. 61.6(a). The main channel can be either a single high gain antenna or a data independent beamformer (see Section 61.3). 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. 61.6(b). 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 61.1.

Choosing the weights to minimize output power can cause cancellation of the desired signal because 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 then be adapted in the absence of the desired signal and frozen when it is present.

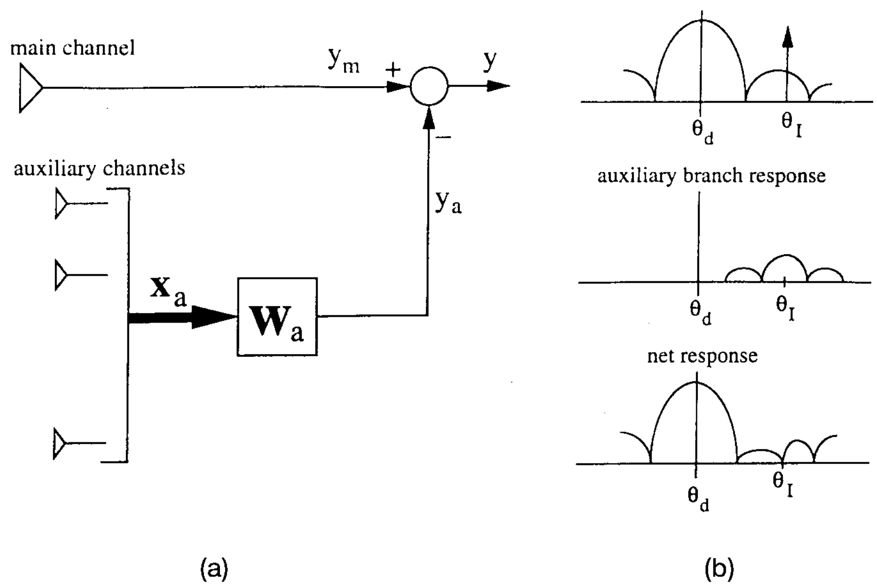

FIGURE 61.6: The multiple sidelobe canceller (MSC) consists of a main channel and several auxiliary channels as illustrated in (a). The auxiliary channel weights are chosen to "cancel" interference entering through sidelobes of the main channel. (b) Depicts the main channel, auxiliary branch, and overall system response when an interferer arrives from direction θ_I.

61.4.2 Use of a 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 61.1, 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 $\mathbf{x}$ 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 $\mathbf{x}$. The fact that the direction of the desired signal does not need to be known is a distinguishing feature of the reference signal approach. For this reason it is sometimes termed "blind" beamforming. Other closely related blind beamforming techniques choose weights by exploiting properties of the desired signal such as constant modulus, cyclostationarity, or third and higher order statistics.

61.4.3 Maximization of Signal-to-Noise Ratio

Here the weights are chosen to directly maximize the signal-to-noise ratio (SNR) as indicated in Table 61.1. A general solution for the weights requires knowledge of both the desired signal, $\mathbf{R}_s$, and noise, $\mathbf{R}_n$, covariance matrices. The attainability of this knowledge depends on the application. For example, in an active radar system $\mathbf{R}_n$ can be estimated during the time that no signal is being transmitted and $\mathbf{R}_s$ 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 $\mathbf{R}_s = \sigma^2 \mathbf{d}(\theta, \omega)\mathbf{d}^H(\theta, \omega)$ from the results in Section 61.2. In this case, the weights are obtained as

$$\mathbf{w} = \alpha \mathbf{R}_n^{-1} \mathbf{d}(\theta, \omega) \tag{61.16}$$

where the α is some non-zero complex constant. Substitution of Eq. (61.16) into the SNR expression shows that the SNR is independent of the value chosen for α.

61.4.4 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 section 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_o be unity.

In Section 61.2 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 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

$$\min_{\mathbf{w}} \quad \mathbf{w}^H \mathbf{R}_x \mathbf{w} \quad \text{subject to} \quad \mathbf{d}^H(\theta, \omega)\mathbf{w} = g^* \; . \tag{61.17}$$

The method of Lagrange multipliers can be used to solve Eq. (61.17) resulting in

$$\mathbf{w} = g^* \frac{\mathbf{R}_x^{-1}\mathbf{d}(\theta, \omega)}{\mathbf{d}^H(\theta, \omega)\mathbf{R}_x^{-1}\mathbf{d}(\theta, \omega)} \ . \tag{61.18}$$

Note that, in practice, the presence of uncorrelated noise will ensure that $\mathbf{R}_x$ is invertible. If $g = 1$, then Eq. (61.18) is often termed the minimum variance distortionless response (MVDR) beamformer. It can be shown that Eq. (61.18) is equivalent to the maximum SNR solution given in Eq. (61.16) by substituting $\sigma^2\mathbf{d}(\theta, \omega)\mathbf{d}^H(\theta, \omega) + \mathbf{R}_n$ for $\mathbf{R}_x$ in Eq. (61.18) and applying the matrix inversion lemma.

The single linear constraint in Eq. (61.17) is easily generalized to multiple linear constraints for added control over the beampattern. For example, if there is 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

$$\begin{bmatrix} \mathbf{d}^H(\theta, \omega) \\ \mathbf{d}^H(\phi, \omega) \end{bmatrix} \mathbf{w} \ = \ \begin{bmatrix} g^* \\ 0 \end{bmatrix} . \tag{61.19}$$

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 61.1.

Several different philosophies can be employed for choosing the constraint matrix and response vector. Specifically point, derivative, and eigenvector constraint approaches are popular. 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. See Van Veen and Buckley [11] or Van Veen [12] for a more in-depth discussion on this topic.

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 MSC and LCMV beamforming. Essentially, the GSC is a mechanism for changing a constrained minimization problem into unconstrained form.

Suppose we decompose the weight vector $\mathbf{w}$ into two orthogonal components $\mathbf{w}_o$ and $-\mathbf{v}$ (i.e., $\mathbf{w} = \mathbf{w}_o - \mathbf{v}$) that lie in the range and null spaces of $\mathbf{C}$, respectively. The range and null spaces of a matrix span the entire space so this decomposition can be used to represent any $\mathbf{w}$. Since $\mathbf{C}^H\mathbf{v} = \mathbf{0}$, we must have

$$\mathbf{w}_o = \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{f} \tag{61.20}$$

if $\mathbf{w}$ is to satisfy the constraints. Equation (61.20) is the minimum L_2 norm solution to the underdetermined equivalent of Eq. (61.12). The vector $\mathbf{v}$ is a linear combination of the columns of an N by M $(M = N-L)$ matrix $\mathbf{C}_n$ (i.e., $\mathbf{v} = \mathbf{C}_n\mathbf{w}_M$) 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 $\mathbf{w} = \mathbf{w}_o - \mathbf{C}_n\mathbf{w}_M$ is depicted in block diagram form in Fig. 61.7. The choice for $\mathbf{w}_o$ and $\mathbf{C}_n$ implies that $\mathbf{w}$ satisfies the constraints independent of $\mathbf{w}_M$ and reduces the LCMV problem to the unconstrained problem

$$\min_{\mathbf{w}_M} \ [\mathbf{w}_o - \mathbf{C}_n\mathbf{w}_M]^H\mathbf{R}_x[\mathbf{w}_o - \mathbf{C}_n\mathbf{w}_M] \, . \tag{61.21}$$

The solution is

$$\mathbf{w}_M = (\mathbf{C}_n^H\mathbf{R}_x\mathbf{C}_n)^{-1}\mathbf{C}_n^H\mathbf{R}_x\mathbf{w}_o \, . \tag{61.22}$$

The primary implementation advantages of this alternate but equivalent formulation stem from the facts that the weights $\mathbf{w}_M$ are unconstrained and a data independent beamformer $\mathbf{w}_o$ is implemented as an integral part of the optimum 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 Section 61.4.5).

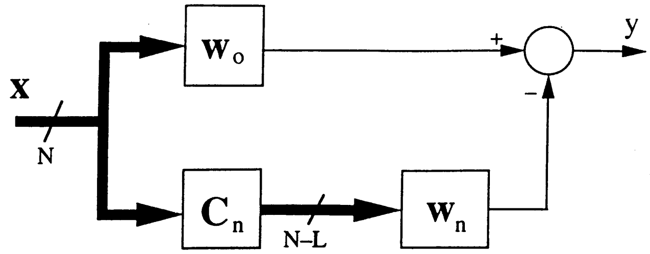

FIGURE 61.7: The generalized sidelobe canceller (GSC) represents an implementation of the LCMV beamformer in which the adaptive weights are unconstrained. It consists of a preprocessor composed of a fixed beamformer $\mathbf{w}_o$ and a blocking matrix $\mathbf{C}_n$, and a standard adaptive filter with unconstrained weight vector $\mathbf{w}_M$.

As an example, assume the constraints are as given in Eq. (61.17). Equation (61.20) implies $\mathbf{w}_o = g^*\mathbf{d}(\theta,\omega)/[\mathbf{d}^H(\theta,\omega)\mathbf{d}(\theta,\omega)]$. $\mathbf{C}_n$ satisfies $\mathbf{d}^H(\theta,\omega)\mathbf{C}_n = \mathbf{0}$ so each column $[\mathbf{C}_n]_i$; $1 < i < N - L$, can be viewed as a data independent beamformer with a null in direction θ at frequency ω: $\mathbf{d}^H(\theta,\omega)[\mathbf{C}_n]_j = 0$. 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 describing $\mathbf{C}_n$. 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 $\mathbf{w}_M$. Signals from directions and frequencies over which the response is not constrained will pass through the upper branch in Fig. 61.7 with some response determined by $\mathbf{w}_o$. The lower branch chooses $\mathbf{w}_M$ 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. 61.6).

61.4.5 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. Several approaches have been suggested to reduce this degradation (e.g., Cox et al. [3]).

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. [13], Shan and Kailath [10]).

61.5 Adaptive Algorithms for Beamforming

The optimum beamformer weight vector equations listed in Table 61.1 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. 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 common.

Among notable adaptive algorithms proposed for beamforming are the Howells-Applebaum adaptive loop developed in the late 1950s and reported by Howells [7] and Applebaum [1], and the Frost LCMV algorithm [4]. Rather than recapitulating adaptive algorithms for each optimum beamformer listed in Table 61.1, we take a unifying approach using the standard adaptive filter configuration illustrated on the right side of Fig. 61.7.

In Fig. 61.7 the weight vector $\mathbf{w}_M$ is 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

$$J(\mathbf{w}_M) = E\{|y_d - \mathbf{w}_M^H \mathbf{u}|^2\} = \sigma_d^2 - \mathbf{w}_M^H \mathbf{r}_{ud} - \mathbf{r}_{ud}^H \mathbf{w}_M + \mathbf{w}_M^H \mathbf{R}_u \mathbf{w}_M \ , \tag{61.23}$$

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

$$\mathbf{w}_{opt} = \mathbf{R}_u^{-1} \mathbf{r}_{ud} \ . \tag{61.24}$$

Comparison of (61.23) and the criteria listed in Table 61.1 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 Fig. 61.7 and Eqs. (61.22), and (61.24) 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 cannot in general be represented by Fig. 61.7 and Eq. (61.24). However, it was noted after (61.18) that if the desired signal is narrowband, then the maximum SNR and the LCMV beamformers are equivalent.

The block adaptation approach solves (61.24) using estimates of $\mathbf{R}_u$ and $\mathbf{r}_{ud}$ formed from K samples of $\mathbf{u}$ and y_d: $\mathbf{u}(k)$, $y_d(k)$; $0 < k < K-1$. The most common are the sample covariance matrix

$$\hat{\mathbf{R}}_u = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{u}(k) \mathbf{u}^H(k) \tag{61.25}$$

and sample cross-covariance vector

$$\hat{\mathbf{r}}_{ud} = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{u}(k) y_d^*(k) \ . \tag{61.26}$$

TABLE 61.2 Comparison of the LMS and RLS Weight Adaptation Algorithms

Algorithm	LMS	RLS
Initialization	$\mathbf{w}_M(0) = 0$ $\mathbf{y}(0) = \mathbf{y}_d(0)$ $0 < \mu < \frac{1}{\text{Trace}[\mathbf{R}_u]}$	$\mathbf{w}_M(0) = 0$ $\mathbf{P}(0) = \delta^{-1}\mathbf{I}$ δ small, 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 $\mathbf{R}_u$.

Performance analysis and guidelines for selecting the block size K are provided in Reed et al. [9].

Continuous adaptation algorithms are easily developed in terms of Fig. 61.7 and Eq. (61.23). 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 vector $\mathbf{w}_{opt}$ 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

$$\nabla_{\mathbf{w}_M(k)} = \frac{\partial}{\partial \mathbf{w}_M} J(\mathbf{w}_M)\bigg|_{\mathbf{w}_M = \mathbf{w}_m(k)} = -2\mathbf{r}_{ud} + 2\mathbf{R}_u\mathbf{w}_M(k) \tag{61.27}$$

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 $y(k) = y_d(k) - \mathbf{w}_M^H\mathbf{u}(k)$, we have

$$\mathbf{w}_M(k+1) = \mathbf{w}_M(k) + \mu\ \mathbf{u}(k)\ y^*(k)\ . \tag{61.28}$$

The gain constant μ controls convergence characteristics of the random vector sequence $\mathbf{w}_M(k)$. Table 61.2 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 [14], Haykin [6], and others).

One alternative to LMS is the exponentially weighted recursive least squares (RLS) algorithm. At the Kth time step, $\mathbf{w}_M(K)$ is chosen to minimize a weighted sum of past squared errors

$$\min_{\mathbf{w}_M(K)} \sum_{k=0}^{K} \lambda^{K-k} \left| y_d(k) - \mathbf{w}_M^H(K)\mathbf{u}(k) \right|^2 \ . \tag{61.29}$$

λ is a positive constant less than one which determines how quickly previous data are deemphasized. The RLS algorithm is obtained from (61.29) by expanding the magnitude squared and applying the matrix inversion lemma. Table 61.2 summarizes both the LMS and RLS algorithms.

61.6 Interference Cancellation and Partially Adaptive Beamforming

The computational requirements of each update in adaptive algorithms are proportional to either the weight vector dimension M (e.g., LMS) or dimension squared M^2 (e.g., RLS). If M is large, this requirement is quite severe and for practical real time implementation it is often necessary to reduce M. Furthermore, the rate at which an adaptive algorithm converges to the optimum solution may be very slow for large M. Adaptive algorithm convergence properties can be improved by reducing M.

The concept of "degrees of freedom" is much more relevant to this discussion than the number of weights. 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 $\mathbf{w}_M$. There are M degrees of freedom in the structure of Fig. 61.7. 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. However, there is a performance penalty associated with reducing degrees of freedom. A partially adaptive beamformer cannot generally 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. 61.7 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 the GSC, $\mathbf{w}_o$ and $\mathbf{C}_n$ are defined as in Section 61.4.4 and $\mathbf{T}_M$ is an $N-L$ by M matrix that reduces degrees of freedom ($M < N - L$).

The goal of partially adaptive beamformer design is to choose $\mathbf{T}_M$ (or $\mathbf{T}$) such that good interference cancellation properties are retained even though M is small. To see that this is possible in principle, consider the problem of simultaneously cancelling two narrowband sources from direction θ_1 and θ_2 at frequency ω_o. Perfect cancellation of these sources requires $\mathbf{w}^H\mathbf{d}(\theta_1, \omega_o) = 0$ and $\mathbf{w}^H\mathbf{d}(\theta_2, \omega_o) = 0$ so we must choose $\mathbf{w}_M$ to satisfy

$$\mathbf{w}_M^H[\mathbf{T}^H\mathbf{d}(\theta_1, \omega_o) \ \ \mathbf{T}^H\mathbf{d}(\theta_2, \omega_o)] = [g_1, g_2] \tag{61.30}$$

where $g_i = \mathbf{w}_o^H \mathbf{d}(\theta_i, \omega_o)$ is the response of the $\mathbf{w}_o$ branch to the ith interferer. Assuming $\mathbf{T}^H \mathbf{d}(\theta_1, \omega_o)$ and $\mathbf{T}^H \mathbf{d}(\theta_2, \omega_o)$ are linearly independent and nonzero, and provided $M \geq 2$, then at least one $\mathbf{w}_M$ exists that satisfies (61.30). Extending 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. In general, M narrowband interferers can be cancelled using M adaptive degrees of freedom with relatively mild restrictions on $\mathbf{T}$.

No such rule exists in the broadband case. Here complete cancellation of a single interferer requires choosing $\mathbf{T}\,\mathbf{w}_M$ so that the response of the adaptive branch, $\mathbf{w}_M^H \mathbf{T}^H \mathbf{d}(\theta_1, \omega)$, matches the response of the $\mathbf{w}_o$ branch, $\mathbf{w}_o^H \mathbf{d}(\theta_1, \omega)$, over the entire frequency band of the interferer. In this case, the degree of cancellation depends on how well these two responses match and is critically dependent on the interferer direction, frequency content, and $\mathbf{T}$. Good cancellation can be obtained in some situations when $M = 1$, while in others even large values of M result in poor cancellation.

A variety of intuitive and optimization-based techniques have been proposed for designing $\mathbf{T}_M$ that acheive good interference cancellation with relatively small degrees of freedom. See Van Veen and Buckley [11] and Van Veen [12] for further review and discussion.

61.7 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 to 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 and response time considerations dictate the use of partially adaptive beamformers with arrays composed of large numbers of sensors.

61.8 Defining Terms

Beamformer: A device used in conjunction with an array of sensors to separate signals and interference on the basis of their spatial characteristics. The beamformer output is usually given by a weighted combination of the sensor outputs.

Array response vector: Vector describing the amplitude and phase relationships between propagating wave components at each sensor as a function of spatial direction and temporal frequency. Forms the basis for determining the beamformer response.

Beampattern: The magnitude squared of the beamformer's spatial filtering response as a function of spatial direction and possibly temporal frequency.

Data independent, statistically optimum, adaptive, and partially adaptive beamformers: The weights in a data independent beamformer are chosen independent of the statistics of the data. A statistically optimum beamformer chooses its weights to optimize some statistical function of the beamformer output, such as signal-to-noise ratio. An adaptive beamformer adjusts its weights in response to the data to accomodate unknown or time varying statistics. A partially adaptive beamformer uses only a subset of the available adaptive degrees of freedom to reduce the computational burden or improve the adaptive convergence rate.

Multiple sidelobe canceller: Adaptive beamformer structure in which the data received at low gain auxiliary sensors is used to adaptively cancel the interference arriving in the mainlobe or sidelobes of a spatially high gain sensor.

Linearly constrained minimum variance (LCMV) beamformer: Beamformer in which the weights are chosen to minimize the output power subject to a linear response constraint. The constraint preserves the signal of interest while power minimization optimally attenuates noise and interference.

Minimum variance distortionless response (MVDR) beamformer: A form of LCMV beamformer employing a single constraint designed to pass a signal of given direction and frequency with unit gain.

Generalized sidelobe canceller: Structure for implementing LCMV beamformers that separates the constrained and unconstrained components of the adaptive weight vector. The unconstrained components adaptively cancel interference that leaks through the sidelobes of a data independent beamformer designed to satisfy the constraints.

References

[1] Applebaum, S.P., Adaptive arrays, Syracuse University Research Corp., Report SURC SPL TR 66-001, Aug. 1966 (reprinted in *IEEE Trans. on AP,* AP-24, 585–598, Sept. 1976).

[2] Buckley, K.M., Spatial/spectral filtering with linearly-constrained minimum variance beamformers, *IEEE Trans. on ASSP,* ASSP-35, 249–266, Mar. 1987.

[3] Cox, H., Zeskind, R.M., and Owen, M.M., Robust adaptive beamforming, *IEEE Trans. on ASSP,* ASSP-35, 1365–1375, Oct. 1987.

[4] Frost III, O.L., An algorithm for linearly constrained adaptive array processing, *Proc. IEEE,* 60, 926–935, Aug. 1972.

[5] Harris, F.J., On the use of windows for harmonic analysis with the discrete Fourier transform, *Proc. IEEE,* 66, 51–83, Jan. 1978.

[6] Haykin, S., *Adaptive Filter Theory,* 3rd ed., Prentice-Hall, Englewood Cliffs, NJ, 1996.

[7] Howells, P.W., Explorations in fixed and adaptive resolution at GE and SURC, *IEEE Trans. on AP,* AP-24, 575–584, Sept. 1976.

[8] Parks, T.W. and Burrus, C.S., *Digital Filter Design,* Wiley-lnterscience, New York, 1987.

[9] Reed, I.S., Mallett, J.D., and Brennen, L.E., Rapid convergence rate in adaptive arrays, *IEEE Trans. on AES,* AES-10, 853–863, Nov. 1974.

[10] Shan, T. and Kailath, T., Adaptive beamforming for coherent signals and interference, *IEEE Trans. on ASSP,* ASSP-33, 527–536, June 1985.

[11] Van Veen, B. and Buckley, K., Beamforming: a versatile approach to spatial filtering, *IEEE ASSP Magazine,* 5(2), 4–24, Apr. 1988.

[12] Van Veen, B., Minimum Variance Beamforming, in *Adaptive Radar Detection and Estimation,* Haykin, S. and Steinhardt, A., Eds., John Wiley & Sons, New York, Chap. 4, 161–236, 1992.

[13] Widrow, B., Duvall, K.M., Gooch, R.P., and Newman, W.C., Signal cancellation phenomena in adaptive arrays: causes and cures, *IEEE Trans. on AP,* AP30, 469–478, May 1982.

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

Further Reading

For further information, we refer the reader to the following books,

[1] Compton, Jr., R.T., *Adaptive Antennas: Concepts and Performance,* Prentice-Hall, Englewood Cliffs, NJ, 1988.

[2] Haykin, S., Ed., *Array Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
[3] Johnson, D. and Dudgeon, D., *Array Signal Processing: Concepts and Techniques*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
[4] Monzingo, R. and Miller, T., *Introduction to Adaptive Arrays*, John Wiley & Sons, New York, 1980.

a tutorial article,

[5] Gabriel, W.F., Adaptive arrays: an introduction, *Proc. IEEE*, 64, 239–272, Aug. 1976.

and bibliography

[6] Marr, J., A selected bibliography on adaptive antenna arrays, *IEEE Trans. on AES*, AES-22, 781–798, Nov. 1986.

Several special journal issues have been devoted to beamforming — *IEEE Transactions on Antennas and Propagation*, September 1976 and March 1986, and the *Journal of Ocean Engineering* 1987. Papers devoted to beamforming are often found in the *IEEE Transactions on: Antennas and Propagation*, *Signal Processing*, *Aerospace and Electronic Systems*, and in the *Journal of the Acoustical Society of America*.