N. K. Bose and C. R. Rao, eds., *Handbook of Statistics, Vol. 10*

High-Resolution Direction Finding*

Stephan V. Schell and William A. Gardner

1. Introduction

Although the term *direction finding* can refer to determining which way to travel, this act often involves estimating the direction from which some information arrives. For example, humans instinctively use a two-sensor acoustic array – the ears – to determine the direction in which to look when a signal – a shout – is received, and one might then decide to run toward or away from the source depending on the attractiveness or offensiveness of the message and its sender. This direction-finding mechanism works fine when only one transmitter (speaker) is present or, if many transmitters are present, when they are widely separated. One need only stand in the thick of a crowded room of people to experience its failure. Although the underlying direction-finding algorithm, which is implemented in the brain, cannot be consciously modified, much research on the design of improved algorithms for man-made direction-finding systems is being conducted.

In this chapter, *direction finding* refers to the process of estimating the directions of arrival of propagating signals as they impinge upon a man-made receiver (see Figure 1). The signals of primary interest are acoustic or electromagnetic, and the sophisticated electronic control and processing that can be applied to such received signals have resulted in a wide variety of both techniques and applications. Some of these applications are listed in Table 1. Due to the abundance of applications and correspondingly enormous number of algorithms, only a small subset of direction-finding techniques are discussed here. Starting points for more detailed investigations include Pillai (1989), Gabriel (1980), Johnson (1982), Wax (1985), Roy (1987), Barabell et al. (1984) and the references therein.

The problem of particular interest here is to estimate the directions of arrival of propagating signals from two or more closely-spaced sources, that is, to perform high-resolution direction finding. The methods discussed here perform

* This work was supported in part by the National Science Foundation under Grant No. MIP-88-12902, and by the United States Army Research Office under contract DAAL03-89-C-0035 sponsored by the Army Communications Electronics Center for Signals Warfare.

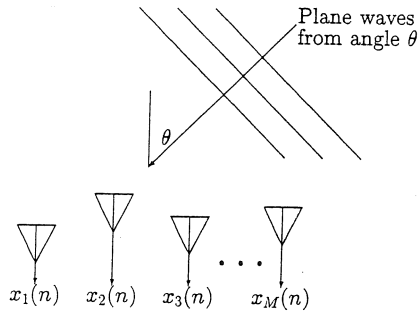

Fig. 1. Example of a sensor array receiving a plane wave arriving from angle θ.

passive direction finding, in contrast to active methods which *transmit* a known signal to illuminate the region of interest and then process the reflected signals. Emphasis is placed on techniques that use an array of closely-spaced sensors having known locations and assume that the array response does not vary significantly over the bandwidth of the signals – a property that loosely characterizes the narrowband assumption, which is described in more detail in Section 3.1 – although a brief discussion of methods for wideband signals also is included. Within this already restricted scope, only a few representative approaches are considered in this chapter.

Wherever possible, the direction-finding (DF) methods are presented in the heuristically pleasing yet rigorous framework of linear combining, that is, discrete spatial filtering, which includes beamforming and null-steering. It is shown how this framework leads naturally to subspace fitting, which is prevalent in the recent research literature. It should be noted that the explanations based on the framework of spatial filtering are not presented as replacements for those based on the framework of subspace fitting but are intended to offer

Table 1
Example application areas for direction finding and corresponding references

Application	Description	References
Intelligence	Covert location of transmitters and signal interception	Wiley (1985), Ball (1989)
RADAR	Air traffic control and target acquisition	Evans et al. (1982), Haykin (1980, 1985)
Commercial communications	Multipath in mobile radio	Matthews and Mohebbi (1989)
Geophysics	Seismology	Justice (1985),
	Atmospheric electromagnetics	Hayakawa et al. (1990)
Astronomy	Radio astronomy	Readhead (1982)
SONAR	Source localization	Wagstaff and Baggeroer (1985)
Signal extraction	Adaptive antenna arrays	Van Veen and Buckley (1988), Compton Jr (1988), Monzingo and Miller (1980)

an alternate motivation for the newest super-resolution methods and to give an alternate insight into *why* the newest super-resolution methods can perform better than other methods. It is believed that this alternate approach can be valuable not only to the nonspecialist as an accessible description of high-resolution direction finding but also to the specialist as a means for unifying many diverse methods. It should also be noted that although the interpretations in terms of spatial filtering involve *linear* operations on the data (i.e., data from different sensors are linearly combined to enhance signal components arriving from some directions and/or to attenuate signal components arriving from other directions), the weighting factors applied to the data before summing are *nonlinear* functions of the data. Although the concept of subspace fitting has suggested many new algorithms and analytical techniques, its abstract nature tends to mask the more physically motivated concept of spatial filtering. Some review of spatial filtering is presented, and the reader interested in more detail on this is referred to the wide body of array signal processing literature including Widrow and Stearns (1985), Monzingo and Miller (1980), Haykin (1980), Compton Jr (1988), Marr (1986), Van Veen and Buckley (1988), Gabriel (1976, 1986), and references therein. Unnecessary abstraction is further avoided by using the time-average framework rather than the probabilistic (ensemble-average) framework for deriving methods that use data correlations (cf. Gardner, 1987b).

In Section 2, systems that do not use sensor arrays and systems that use time-difference of arrival (TDOA) for a pair of sensors are discussed briefly to put the array-based systems treated in this chapter in perspective. In Section 3, the mathematical model of the narrowband signal and a brief review of spatial filtering, antenna patterns, and directional ambiguities are presented. In Sections 4 and 5, several high-resolution direction-finding algorithms are discussed with emphasis on those exhibiting super-resolution. In Section 6, some results on performance limits for narrowband direction-finding algorithms are discussed. In Section 7, several departures from ideality, in which prior knowledge required by the algorithms is in error or unavailable altogether, are considered. Finally, a brief summary of the chapter and topics of continuing research interest are presented in Section 8.

2. Possible approaches

Before narrowing the scope to array-based high-resolution DF algorithms, several alternate approaches which could be more appropriate in some applications are discussed. First, spinning DF, multilobe amplitude comparison, and phase interferometry systems, which typically require less signal processing than do array-based high-resolution methods, are considered. Then, systems that use the time-difference of arrival (TDOA) of a received signal between two widely-spaced sensors are discussed briefly.

Spinning DF systems use a rotating directional antenna to estimate directions

of arrival by estimating the received power as a function of direction: directions in which locally maximum power is received are then taken to be the direction estimates. Since the antenna cannot simultaneously scan in all directions of interest, some transient signals can escape detection. However, if the width of the beam were widened to accommodate a wider angular range and to reduce the chances of missing a signal, the ability to resolve two closely spaced sources would decrease. Due to the physical size and weight of some spinning DF antennas, these systems are more likely to be found in land- or ship-based installations rather than, say, airplanes or satellites. For example, a typical naval ship unit, the AS-899F, is seven feet high, three feet in diameter, and weighs 270 pounds (Wiley, 1985). Although the signal processing requirements consist mainly of estimating the power of a signal, mechanical constraints on durability and vibration must be met.

In contrast, multilobe amplitude comparison systems use several (typically four or eight) fixed directional antennas each pointing in a different direction. Given a signal impinging on the antennas, the two antennas having the largest and next largest responses are identified. Then, the ratio of the two responses (along with knowledge of the antenna patterns) is used to interpolate between the two pointing directions of the antennas to obtain the direction estimate. Although this approach entails more computation than the spinning DF approach, it does not have as many mechanical requirements and is less prone to missing transient signals.

Yet another approach is two-sensor phase interferometry for DF which obtains the direction estimate from the relative phase between two sensor outputs. Consider two sensors separated by a distance d that receive a single-tone (sinusoidal) signal having frequency f and arriving from angle θ (see Figure 1). The relative phase between the two sensors is given by

$$\phi = 2\pi \frac{f}{c} d \sin(\theta) . \tag{1}$$

The DOA estimate is then obtained simply by inverting the relationship (1). It can be seen from (1) that ambiguity can result if the distance d is greater than half of the wavelength c/f.

As with the multilobe amplitude comparison approach, phase interferometry needs no moving parts and is responsive to transient signals. However, the amplitude ratio or phase relationship on which these two methods are based no longer holds if the signal paths to the two sensors are radically different, or if multiple signals or interference are present, or if the received signals are not nearly sinusoidal, thereby necessitating other approaches.

For some applications (e.g., astronomy) that require very long baselines to achieve the desired resolution and in others (e.g., underwater acoustics) in which the propagation medium can alter the attenuation or phase shift of one signal path relative to another, methods based on measuring the TDOA of the received signals can be useful (Carter, 1981). For example, the generalized cross correlation (GCC) method (Knapp and Carter, 1976) is based on the

property that the cross correlation between a signal and a delayed version of the signal is maximized when the lag used in the cross correlation equals the TDOA. However, this property does not hold in the presence of interference and multiple signals of interest. Since the peak in the cross-correlation function occurs at the TDOA when a single signal is present, and since the width of the peak is approximately equal to the coherence time of the signal, it follows that the peaks due to multiple signals merge into one peak if the spacing between adjacent TDOAs of the signals is less than the coherence times (reciprocal bandwidths) of the signals. Consequently, since the TDOAs are proportional to the separation between sensors, one solution to this problem is to increase the separation. However, the separation between sensors can be fixed in some applications (e.g., the sensors are mounted on one aircraft), or the separation required to resolve the signals can be otherwise prohibitively large. Thus, multiple signals of interest can be difficult (or impossible) to accommodate. Nonetheless, the GCC method can still operate properly in some environments because, before correlating, it can, effectively, filter the received data to reject out-of-band interference. Of course, this technique does not perform the desired interference rejection if the interference and desired signals are completely spectrally overlapping. However, TDOA methods that exploit the cyclostationarity that some desired signals (such as communication and telemetry signals) exhibit can alleviate these problems by being signal-selective even for signals that are spectrally overlapping. Therefore, these methods apply to much more difficult environments (Gardner and Chen, 1991; Chen and Gardner, 1992). Also, even in environments in which conventional methods can operate properly, the methods that exploit cyclostationarity can yield more accurate TDOA estimates.

3. Narrowband sensor arrays

Array-based approaches to the DF problem exploit properties of the received data that are different than those exploited by simple energy detectors such as spinning DF systems or by long-delay estimators such as TDOA systems. Here, the properties of interest are the relative amplitudes and phases of the sensor outputs as a function of direction of arrival (DOA). The mathematical model for these and for the data are briefly summarized here. Also, as with most successful models, an intuitive approach to DF (the conventional beamforming method) is suggested by the model and is used here to convey underlying concepts shared by all of the array-based methods considered, namely, the concept of weighting and summing the sensor outputs to amplify some received signals and attenuate others.

3.1. Narrowband model

The direction-finding behavior of each technique depends explicitly on the model presented here, which describes the received data at the sensor outputs

as a function of the impinging signal waveforms, directions of arrival, and sensor characteristics. In addition, array-based DF depends on the array geometry because it, too, affects the vector $\boldsymbol{x}(n)$ of output signals $x_1(n), \ldots, x_M(n)$ from the individual sensors. Before proceeding to models for realistic signals, the special case of an array receiving a sinusoidal signal is considered, since the subsequent models are based on this simple case.

Consider the analytic signal (which contains only the positive-frequency component) $\exp(\mathrm{j}(\omega t + \phi))$ corresponding to a real sine wave having frequency ω and phase ϕ and arriving at the array from angle θ. For simplicity, assume that the sensors in the array and the signal source are coplanar so that ordered pairs and a single angle suffice to describe the positions of the sensors and the direction of arrival of the signal, respectively. If the propagation medium does not significantly affect the signal as it propagates from one end of the array to the other, then the signal received at one sensor differs from the signal received at another sensor only by a delay. As suggested by Figure 1, the dependence of the delay on the geometry of the sensors and on the angle of arrival can be determined by using elementary geometry. Specifically, if we assume that the coordinates of the M sensors are $(q_1, r_1), \ldots, (q_M, r_M)$, then it can be shown that the delay t_m of the signal at the m-th sensor relative to the signal at the origin of the coordinate system can be expressed as $t_m = -[q_m \sin(\theta) + r_m \cos(\theta)]/c$, where c is the propagation speed and θ is measured clockwise from the r axis. However, since the signal is sinusoidal, the propagation delay t_m is equivalent to a phase shift by the amount $\psi_m = -\omega t_m$, which is in turn equivalent (since the signal is a complex sinusoid) to multiplication by $\exp(\mathrm{j}\psi_m)$. Thus, the signal received by the array can be expressed in the vector form

$$\begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \begin{bmatrix} \mathrm{e}^{\mathrm{j}\psi_1} \\ \vdots \\ \mathrm{e}^{\mathrm{j}\psi_M} \end{bmatrix} \mathrm{e}^{\mathrm{j}(\omega t + \phi)}, \tag{2}$$

where

$$\psi_m(\omega, \theta) = [q_m \sin(\theta) + r_m \cos(\theta)]\omega/c\,. \tag{3}$$

More generally, the sensors can have differing directional and frequency-dependent characteristics, which can be modeled by applying differing gains and phases to the elements of the vector in (2). Denoting the gain and phase of the m-th sensor by $g_m(\omega, \theta)$ and $\phi_m(\omega, \theta)$, respectively, the analytic signal at the outputs of the sensors can be expressed as

$$\begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \begin{bmatrix} g_1(\omega, \theta)\, \mathrm{e}^{\mathrm{j}\phi_1(\omega,\theta)}\, \mathrm{e}^{\mathrm{j}\psi_1(\omega,\theta)} \\ \vdots \\ g_M(\omega, \theta)\, \mathrm{e}^{\mathrm{j}\phi_M(\omega,\theta)}\, \mathrm{e}^{\mathrm{j}\psi_M(\omega,\theta)} \end{bmatrix} \mathrm{e}^{\mathrm{j}(\omega t + \phi)} = \boldsymbol{a}(\omega, \theta)\, \mathrm{e}^{\mathrm{j}(\omega t + \phi)}\,. \tag{4}$$

In the more general (and interesting) case in which multiple nonsinusoidal signals arrive at the array, the data can be modeled by decomposing it in the frequency domain and using linear superposition:

$$\begin{aligned} \boldsymbol{x}(\omega) &= \sum_{l=1}^{L} \boldsymbol{a}(\omega, \theta_l) s_l(\omega) + \boldsymbol{i}(\omega) \\ &= [\boldsymbol{a}(\omega, \theta_1) \cdots \boldsymbol{a}(\omega, \theta_L)][s_1(\omega) \cdots s_L(\omega)]^{\mathrm{T}} + \boldsymbol{i}(\omega) \\ &= \boldsymbol{A}(\omega, \Theta)\boldsymbol{s}(\omega) + \boldsymbol{i}(\omega) , \end{aligned} \tag{5}$$

where L signals $s_1(\omega), \ldots, s_L(\omega)$ arrive from angles $\theta_1, \ldots, \theta_L$ and $\boldsymbol{i}(\omega)$ represents interference and noise components (e.g., thermal noise from the sensors and associated electronics, background noise from the environment, and spatially diffuse sources of man-made interference such as cities). That is, the array data is linear with respect to the signals and is linear (in the frequency domain) with respect to the vector $\boldsymbol{a}(\omega, \theta)$. The vector $\boldsymbol{a}(\omega, \theta)$ is referred to in this chapter as the *array response vector*, but the terms *aperture vector*, *array vector*, *array manifold vector*, *DOA vector*, *direction vector*, and *steering vector* also appear in the literature. The collection of array response vectors for all angles θ and all frequencies ω of interest is referred to as the *array manifold*.

However, if only a relatively narrow frequency band is of interest (e.g., if prior knowledge regarding the center frequencies and bandwidths of the signals of interest is available to select the narrow band of interest), then it is advantageous to reject signal components and noise that lie outside the band. If this band is sufficiently narrow that the array response vector $\boldsymbol{a}(\omega, \theta)$ is approximately constant with respect to ω over the band of interest for all angles θ (e.g., if the reciprocal of the time required for the signal to propagate across the array is much less than the bandwidth of the signal, and if the sensor characteristics do not vary significantly across this bandwidth), then the dependence on ω can be dropped and the array data can be modeled in the time domain as the analytic signal

$$\begin{aligned} \boldsymbol{x}(t) &= \sum_{l=1}^{L} \boldsymbol{a}(\theta_l) s_l(t) + \boldsymbol{i}(t) \\ &= [\boldsymbol{a}(\theta_1) \cdots \boldsymbol{a}(\theta_L)][s_1(t) \cdots s_L(t)]^{\mathrm{T}} + \boldsymbol{i}(t) \\ &= \boldsymbol{A}(\Theta)\boldsymbol{s}(t) + \boldsymbol{i}(t) , \end{aligned} \tag{6}$$

where $\boldsymbol{s}(t)$ and $\boldsymbol{i}(t)$ are analytic signals. Although the signals $s_l(t)$ are not sinusoids, the spatial characteristics of the array response can be approximately modeled as if they were. A more detailed discussion of the conditions under which this assumption is valid, as well as a detailed investigation of the representation of wideband array data, can be found in Buckley (1987).

In this chapter and in much of the literature on direction finding, the sampled complex envelope of the array data is used in the description and analysis of the various algorithms because the algorithms are typically implemented on a digital computer and therefore operate on sampled data. Since the complex envelope of a bandlimited analytic signal can be obtained by performing a complex down-conversion (i.e., by multiplying the data by $\exp(-j\omega t)$ for some appropriate ω), the corresponding model for the sampled complex envelope is essentially the same as in (6), except that $x(t)$, $s(t)$, and $i(t)$ denote the complex envelopes of the array data, the signals, and the noise, respectively, and t is replaced with n:

$$x(n) = \sum_{l=1}^{L} a(\theta_l)s_l(n) + i(n) = A(\Theta)s(n) + i(n) . \tag{7}$$

In practice, the array manifold is approximated using calibration data taken at discrete angles, although it can be known analytically in some situations (e.g., a uniform linear array (ULA) of identical sensors). Note that, in general, a search over the array manifold will be required in the application of a DF method because typically no analytical expression for DOA as a function of array response is available.

Since one objective of high-resolution DF is to distinguish among signals from several different directions, the array manifold must be free of ambiguity. For example, if the array response due to a signal from θ_1 is identical to the response due to a signal from θ_2, then a DF method is unable to distinguish between two signals arriving from angles θ_1 and θ_2, respectively, and might even conclude erroneously that only one signal is received. Also, even if only one signal arrives, this ambiguous array response prevents a DF method from determining whether a single signal arrives from θ_1 or θ_2. This type of ambiguity is a rank 1 ambiguity. In general, a rank $K-1$ ambiguity (cf. Schmidt, 1981) exists if and only if there exist angles $\theta_1, \ldots, \theta_K$ such that the corresponding array vectors are linearly dependent, and no lower rank ambiguities exist. For example, a uniform linear array (ULA) of M sensors can be shown to be free of rank M ambiguities in the range $\theta \in [-90°, 90°]$ if the interelement spacing is less than half of the wavelength of the highest frequency in the receiver band. This requirement for a ULA can be interpreted as a spatial analogue of the Nyquist sampling criterion that allows the reconstruction of a continuous time waveform from its discrete-time samples. The lack of ambiguities is a desirable property because ambiguities reduce the number of signals that can be properly accommodated by a DF algorithm.

In addition to there being possible ambiguities in the array manifold, partially or fully correlated signals can arrive from different directions due to multipath propagation (e.g., the signal propagates from the emitter to the sensor array by multiple spatially distinct paths) or smart jamming (e.g., a hostile emitter receives the desired signal and retransmits a possibly corrupted version so as to reduce the likelihood that the intended receiver can properly

demodulate the signal), thus creating another source of difficulty for a DF algorithm. The temporal properties of the received signals are typically ignored by the DF algorithms discussed in this chapter (i.e., the algorithms operate as if the data samples are independent), so two signals are said in this chapter to be fully correlated if and only if they are scaled versions of each other. If one signal is a delayed version of the other, then the correlation between the two signals is only partial.

3.2. Conventional beamforming

Before discussing the conventional beamforming method of direction finding, the basic concepts of discrete spatial filtering, of which beamforming and null-steering are examples, are reviewed.

Given knowledge of the array manifold (and, at the least, assuming it to be free of low-rank ambiguities), an array can be steered electronically just as a fixed antenna can be steered mechanically. However, the array pattern can change shape in addition to changing orientation. A weight vector w can be used to linearly combine the output signals from the sensors to form a single output signal $y(n)$,

$$y(n) = w^{\mathrm{H}} x(n) , \tag{8}$$

where superscript H denotes the Hermitian (conjugate transpose) operation, and the response of this spatial filter can be described by its effective antenna pattern $P(\theta)$,

$$P(\theta) = |w^{\mathrm{H}} a(\theta)|^2 . \tag{9}$$

That is, $P(\theta)$ equals the average power of the output of the spatial filter when a single, unity-power signal arrives from angle θ. To perform beamforming or null-steering, the gains and phases in w can be adjusted so that beams can be formed in some directions and nulls in others. For example, the effective antenna pattern shown in Figure 2 (and discussed in more detail later in this section) contains beams at 0 degrees and near ±40 degrees and contains nulls near ±25 degrees and ±55 degrees. Thus, a signal arriving from an angle near 25 degrees (one of the nulls) would be attenuated or completely rejected, whereas a signal arriving from an angle near 0 degrees would be amplified. Note that the effective array pattern differs radically from the patterns of the individual sensors, which in this example are isotropic (i.e., the gains of the individual sensors are flat horizontal lines when plotted as a function of direction of arrival).

The total average power in the output of the spatial filter can be expressed as

$$\begin{aligned} P_{\text{total}} &= \langle |y(n)|^2 \rangle_N \\ &= w^{\mathrm{H}} \langle x(n) x^{\mathrm{H}}(n) \rangle_N w \\ &= w^{\mathrm{H}} \boldsymbol{R}_{xx} w , \end{aligned} \tag{10}$$

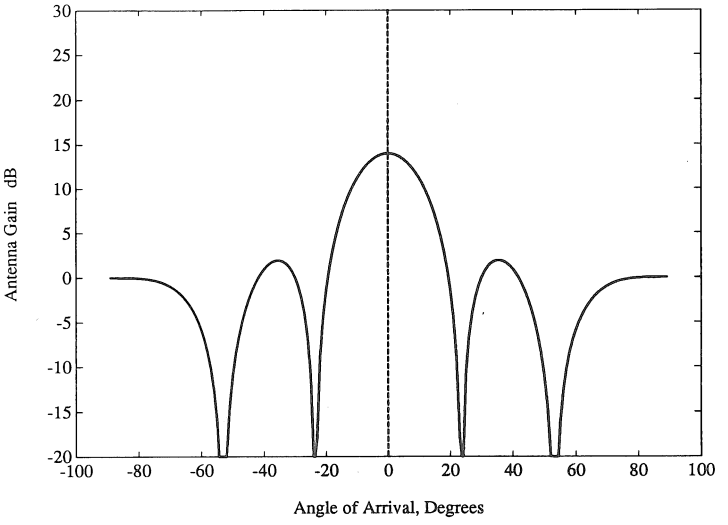

Fig. 2. Antenna pattern of 5-element ULA beamformed to 0 degrees.

where $\langle \cdot \rangle_N$ denotes time averaging over N time samples, and $\boldsymbol{R}_{xx}$ is the measured spatial autocorrelation matrix of the received data defined in (10). This expression for the total average received power plays a central role in the DF algorithms described below because the spatial autocorrelation matrix $\boldsymbol{R}_{xx}$ contains information about the array response vectors and the signals that can be interpreted and used in a variety of ways:

$$\begin{aligned} \boldsymbol{R}_{xx} &= \langle \boldsymbol{x}(n)\boldsymbol{x}^{\mathrm{H}}(n) \rangle_N \\ &= \langle (\boldsymbol{A}(\Theta)\boldsymbol{s}(n) + \boldsymbol{i}(n))(\boldsymbol{A}(\Theta)\boldsymbol{s}(n) + \boldsymbol{i}(n))^{\mathrm{H}} \rangle_N \\ &\rightarrow \boldsymbol{A}(\Theta)\boldsymbol{R}_{ss}\boldsymbol{A}^{\mathrm{H}}(\Theta) + \boldsymbol{R}_{ii} \quad \text{for } N \rightarrow \infty . \end{aligned} \tag{11}$$

Unlike the spinning DF systems, which scan the angular region of interest by physically rotating an antenna, the conventional beamforming technique scans the angular region *electronically*. In doing so the technique explicitly exploits some properties of the underlying narrowband-signal model.

For any particular direction θ_0, the antenna pattern formed using the weight vector $w_b = a(\theta_0)$ has the highest gain in direction θ_0 of any possible weight vector of the same magnitude. This is so because w_b aligns the phases of the signal components arriving from θ_0 at the sensors, causing them to add constructively, and optimally adjusts the amplitudes. Mathematically, this can be shown using the Cauchy–Schwarz inequality:

$$|w^{\mathrm{H}}a(\theta_0)|^2 \leq \|w\|^2 \|a(\theta_0)\|^2 \tag{12}$$

for all vectors w, with equality holding if and only if w is proportional to $a(\theta_0)$.

In the absence of a rank-1 ambiguity, the effective pattern (9) will have a global maximum at θ_0.

Thus, one approach to DF is to scan such a beam over the angular region of interest and to identify those angles where the received power exhibits local maxima. Specifically, in the conventional beamforming approach, the beam is scanned over an angular region of interest (usually in discrete steps), and for each look direction θ the average power output $P_b(\theta)$ of the steered (beamformed) array is measured, where

$$\begin{aligned} P_b(\theta) &= \langle |y_b(n)|^2 \rangle_N \\ &= \langle |\boldsymbol{w}_b^H \boldsymbol{x}(n)|^2 \rangle_N \\ &= \boldsymbol{w}_b^H \boldsymbol{R}_{xx} \boldsymbol{w}_b \qquad (13) \\ &= \boldsymbol{a}^H(\theta) \boldsymbol{R}_{xx} \boldsymbol{a}(\theta) \,. \qquad (14) \end{aligned}$$

Locations of locally maximum average power output are then taken to be direction estimates, just as they are in the spinning DF systems.

This method has some important advantages over the spinning DF systems. Note that $\boldsymbol{R}_{xx}$ need be computed only once and can then be processed as desired. Computing $P_b(\theta)$ for one range of θ does not prevent the algorithm from subsequently computing $P_b(\theta)$ for another range using the same data: the spatial characteristics of the data for all directions are compactly represented by $\boldsymbol{R}_{xx}$. Thus, the conventional beamforming method does not have blind spots in time during which transient signals away from the look direction can appear intermittently and fail to be detected. Another advantage is that by steering the antenna electronically rather than mechanically, mechanical design constraints are more easily met, and the speed of the scan through a region of interest is limited by computational speed instead of mechanical speed.

However, this approach exhibits many of the same drawbacks as the spinning DF systems in addition to some new ones. The width of the beam and the height of the sidelobes limit the effectiveness when multiple signals are present because the signals over a wide angular region contribute to the measured average power at each look direction. For example, a local maximum of average output power can be shifted away from the true DOA of a weak signal by a strong interferer in the vicinity. Alternatively, two closely spaced signals can result in only one peak, or two barely discernible peaks in the wrong locations, in the average output power. This phenomenon is illustrated by a computer simulation of a ULA, having 5 elements with interelement spacing equal to half of the carrier wavelength, receiving two signals each having 10 dB SNR and arriving from 0 and 15 degrees. As shown in Figure 2, when the antenna is beamformed in the direction of the signal from 0 degrees, it still exhibits significant gain in the direction of the signal from 15 degrees. Consequently, the power from the signal at 15 degrees contributes to the measured power at 0 degrees, thereby yielding a low

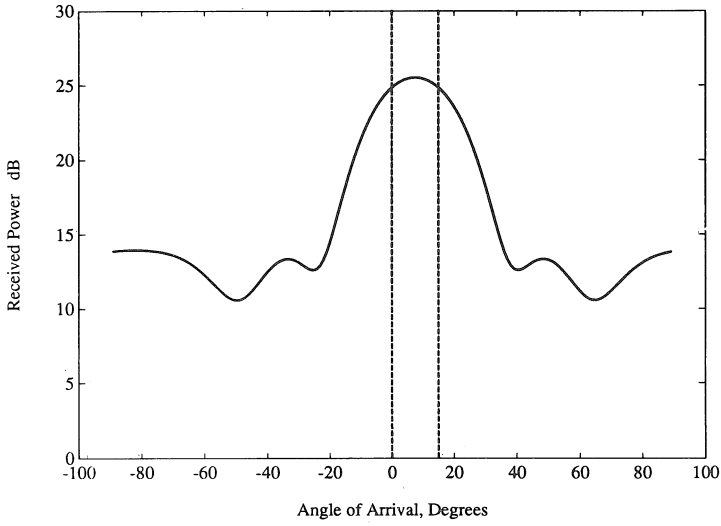

Fig. 3. Average output power obtained by conventional beamforming method for two signals having 10 dB SNR arriving from 0 and 15 degrees (denoted by dashed lines).

angular-resolution measurement of average power at 0 degrees, as shown in Figure 3. So, despite the fact that the antenna array is forming the strongest possible beam in any given direction as it scans, the conventional beamforming method is unable to resolve the two closely spaced signals. Incidentally, the measured average power shown in Figure 3 (and in all subsequent figures unless stated otherwise) is idealized since the averaging time N is allowed to approach infinity. That is, no effects due to measurement error or finite averaging time are present; these effects are discussed briefly in Section 7.1.

Although it can be easier to add sensor elements to decrease the beam width in the array-based approach than it would be to increase the physical size or alter the geometry of the fixed antenna in a spinning DF system, increasing the number of sensor elements increases the number of receivers and the amount of storage required for the calibration data. Furthermore, the need to store array calibration data is not shared by the spinning DF system.

4. A high-resolution array-based approach

One source of performance degradation in the conventional beamforming approach results from the (not-always-correct) notion that the strongest beam in a particular direction yields the best estimate of power arriving from that direction. That is, this approach uses all available degrees of freedom (equal in number to one less than the number of sensors) to strengthen the beam in the look direction, the motivation being that the output power is maximized when the look direction coincides with the true direction of arrival of the signal. This

is appropriate when there is only one signal present. However, when multiple signals are present, a more accurate estimate can be obtained by using some degrees of freedom to form a beam in the look direction and simultaneously using the remaining degrees of freedom to form nulls in other directions in order to reject other signals. In terms of the array processor output power, forming nulls in the directions from which other signals arrive can be accomplished by minimizing the output power and simultaneously constraining a beam (or at least maintaining unity gain) in the look direction to prevent the processor from using the trivial solution $\boldsymbol{w} = \boldsymbol{0}$. Thus, for a particular look direction, Capon's method (Capon, 1969, 1979) uses all but one of the degrees of freedom to minimize the array processor output power while using the remaining degree of freedom to constrain the gain in the look direction to be unity:

$$\min_{\boldsymbol{w}} \langle |y(n)|^2 \rangle_N \quad \text{subject to } \boldsymbol{w}^{\mathrm{H}}\boldsymbol{a}(\theta) = 1 \,. \tag{15}$$

The weight vector chosen in this way is often referred to as the minimum-variance distortionless-response (MVDR) beamformer since, for a particular look direction, it minimizes the variance (average power) of the array processor output signal $y(n)$ while passing a signal arriving from the look direction with no distortion (unity gain and zero phase shift). The resulting weight vector $\boldsymbol{w}_{\mathrm{c}}(\theta)$ can be shown to be given by

$$\boldsymbol{w}_{\mathrm{c}}(\theta) = (\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{R}_{xx}^{-1}\boldsymbol{a}(\theta))^{-1}\boldsymbol{R}_{xx}^{-1}\boldsymbol{a}(\theta) \,. \tag{16}$$

In order to estimate the DOAs, Capon's method searches over θ to find the directions for which the measured received power,

$$\begin{aligned} P_{\mathrm{c}}(\theta) &= \boldsymbol{w}_{\mathrm{c}}^{\mathrm{H}}(\theta)\boldsymbol{R}_{xx}\boldsymbol{w}_{\mathrm{c}}(\theta) \\ &= (\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{R}_{xx}^{-1}\boldsymbol{a}(\theta))^{-1} \,, \end{aligned} \tag{17}$$

is maximized. Although it is not the maximum likelihood estimator of θ, Capon's method is sometimes referred to as Capon's maximum likelihood method for the following reason. For any choice of θ, $P_{\mathrm{c}}(\theta)$ (evaluated using the estimated autocorrelation matrix) is the maximum likelihood estimate of the power of a signal arriving from angle θ in the presence of temporally white Gaussian noise having arbitrary spatial characteristics (Capon, 1979) (see also exercise 15 in chapter 6 of Gardner, 1987b). In other words, $P_{\mathrm{c}}(\theta)$ is the point-wise maximum likelihood estimate of the spatial spectrum or angular density of received power.

The performance improvement of Capon's method relative to the conventional beamforming method is illustrated here with a computer simulation. For the same environment used for Figure 3, Capon's method successfully resolves the two signals, as shown in the plot of measured ($N \rightarrow \infty$) power in Figure 4.

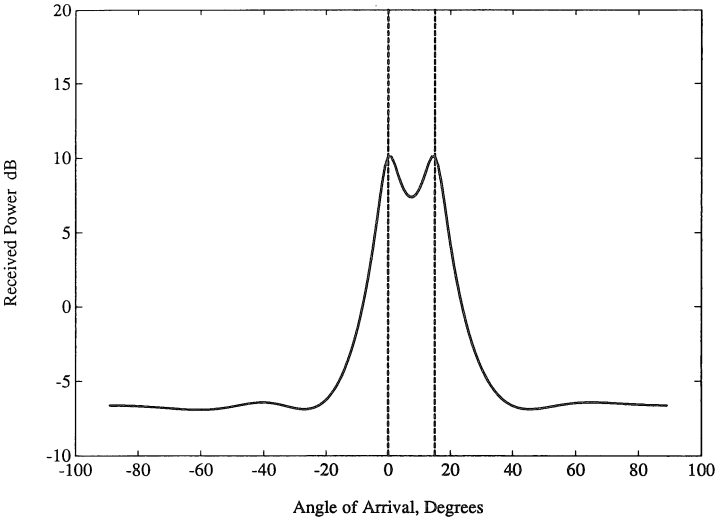

Fig. 4. Average power output of Capon's method for two signals arriving from 0 and 15 degrees with 10 dB SNR.

As predicted by the preceding analysis, Capon's method succeeds in this example because it severely attenuates the signal arriving from 15 degrees while it is looking in the direction of the signal from 0 degrees, as shown by the antenna pattern in Figure 5. However, if the signals are spaced only 10 degrees apart, then Capon's method cannot resolve them, as revealed in Figure 6. As

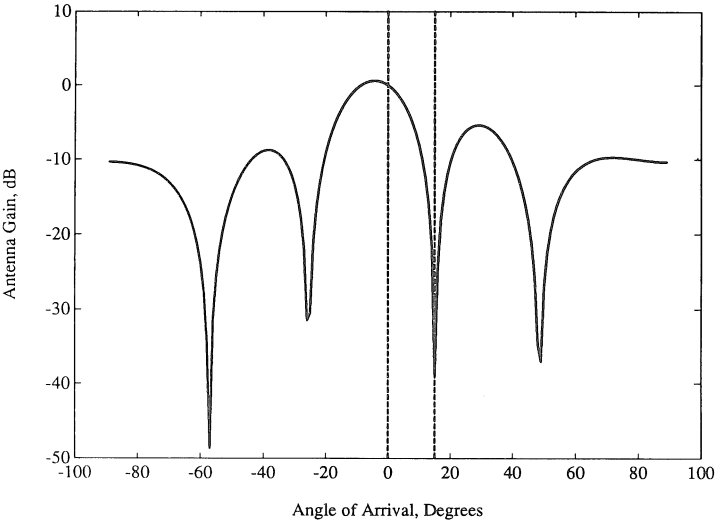

Fig. 5. Antenna pattern of Capon's method for look direction of 0 degrees in the presence of signals arriving from 0 and 15 degrees.

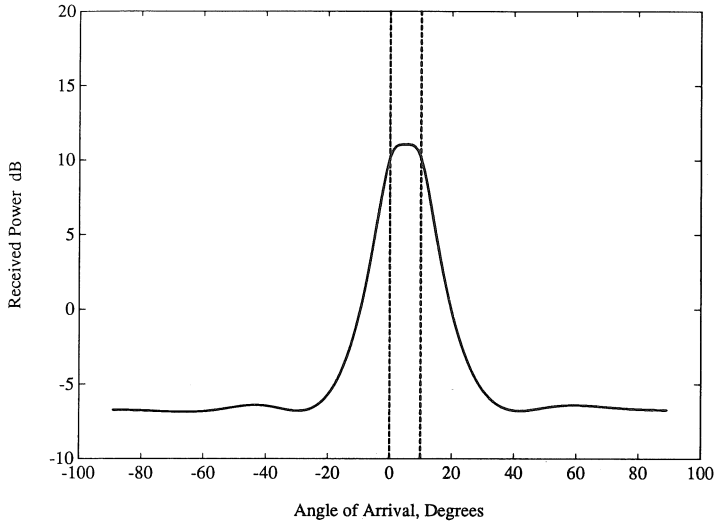

Fig. 6. Average power output of Capon's method for two signals arriving from 0 and 10 degrees with 10 dB SNR. Only one peak is present, located at 4 degrees.

illustrated in Figure 7, although the method attenuates the signal from 10 degrees while it looks in the direction of the signal from 0 degrees, the location of the beam is noticeably displaced from the look direction. Furthermore, when the look direction is between 0 and 10 degrees neither signal is nulled, creating the broad peak shown in Figure 6.

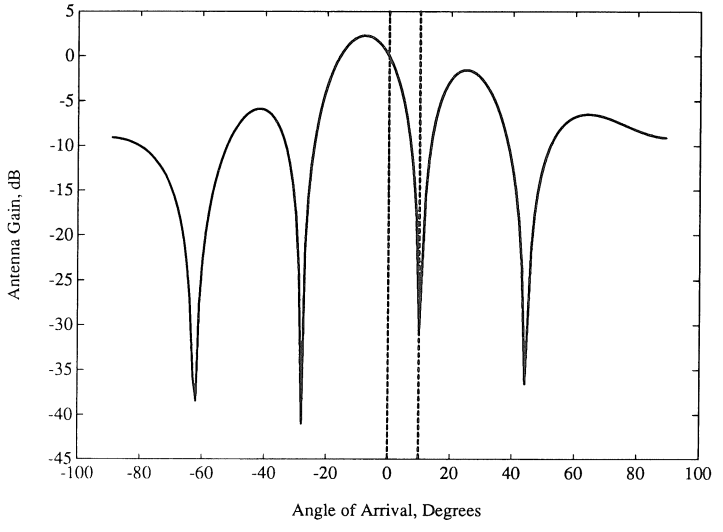

Fig. 7. Antenna pattern of Capon's method for look direction of 0 degrees in the presence of signals arriving from 0 to 10 degrees.

Capon's method also fails if other signals that are correlated with the signal of interest are present because it inadvertently uses that correlation to reduce the processor output power: the processor attenuates or even cancels the signal of interest without having to spatially null the signal. Thus, despite having unity gain in the direction of the desired signal, the processor uses remaining degrees of freedom to obtain a low output power in this case; consequently, the method yields an inaccurate estimate or fails altogether.

This behavior is illustrated as follows. For the same environment as that used for Figure 4, the two signals arriving from 0 and 15 degrees are now made perfectly correlated (i.e., they are scaled versions of each other). Therefore, Capon's method does not spatially null either signal because it can simply add them destructively to reduce the final output power. As shown in Figure 8, the peaks in the average power output in Figure 4 merge to form a single peak halfway between the true DOAs, due to the failure of the method to spatially null either signal. The antenna pattern corresponding to the 0-degree look direction is shown in Figure 9 to illustrate this.

Capon's method can be interpreted as a method for estimating the spatial spectrum (an angular decomposition of average power received by the array) and is related to a high-resolution spectrum estimation method (for estimating the frequency decomposition of average power in a signal) of the same name. Other high-resolution spectrum estimation techniques that have application to the DF problem include the maximum entropy method (Burg, 1972; Kesler, 1982; Burg, 1967) and various other autoregressive (AR) modeling and linear prediction techniques (e.g., Tufts and Kumaresan, 1982). A survey of spectrum estimation methods can be found in Haykin and Cadzow (1982) and, within the

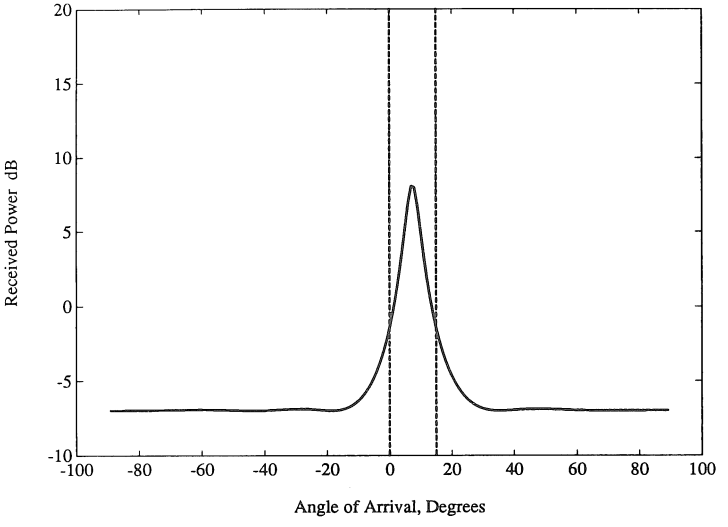

Fig. 8. Average power output of Capon's method for two perfectly correlated signals arriving from 0 and 15 degrees with 10 dB SNR.

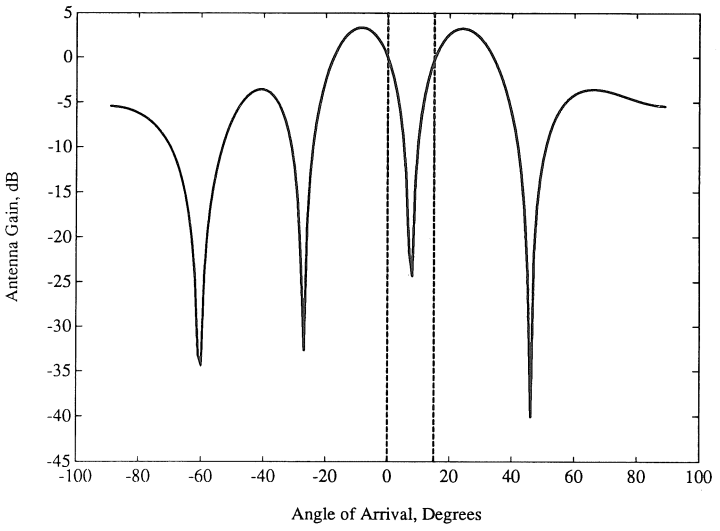

Fig. 9. Antenna pattern of Capon's method for look direction of 0 degrees in the presence of perfectly correlated signals arriving from 0 and 15 degrees.

time-average framework used in this chapter, in Gardner (1987b). The application of high-resolution spectrum estimation methods to the DF problem is explicitly addressed in Gabriel (1980), Johnson (1982).

5. Super-resolution array-based approaches

The failure of the previously discussed methods in the presence of closely-spaced sources or correlated sources can be traced to two causes. First, the received-data autocorrelation matrix is composed of a full-rank autocorrelation matrix of the interference and noise plus a lower-rank autocorrelation matrix due to the desired signal components. By ignoring this structure, these methods must make a three-way tradeoff among beamforming in a look direction, attenuating other signals, and maintaining low sidelobes to attenuate noise. This is analogous to the tradeoff among passband gain, rolloff, and stopband attenuation in finite-impulse-response temporal filters (e.g., see Oppenheim and Schafer, 1975). Second, by not distinguishing between signal rejection due to spatial nulls and signal rejection due to correlated signals from different directions being combined destructively, the previous methods perform poorly or fail completely when highly correlated signals are present (e.g., when multipath or smart jamming occurs).

Methods for transcending the preceding resolution limits can attain even higher resolution than that of Capon's method and are sometimes referred to as *super-resolution* methods. A commonality among them is that they all

exploit the structure of the autocorrelation matrix and jointly estimate DOAs of all signals of interest rather than estimating them one at a time. Furthermore, in doing so, some of them perform beamforming and nulling in subspaces, restricted spaces, or transformed spaces of the received data space instead of operating directly on the data. Certain of these methods can also operate properly in the presence of multipath and smart jamming, both of which can result in highly or perfectly correlated signals.

Although these methods are derived here from the perspective of spatial filtering, the resulting algorithms lead naturally to derivations based on the concept of subspace fitting. Recently, it has been shown (Viberg and Ottersten, 1991) that most of the existing super-resolution methods (including those described here) are special cases within a general subspace-fitting framework. The concept of vector space and the properties of eigenvalues and eigenvectors (and singular values and singular vectors) are used heavily in the literature in deriving these methods and understanding their performance, but their interpretation in terms of the more physically motivated actions of beamforming and null-steering, which is emphasized here, is usually not mentioned.

5.1. *MUSIC*

One of the earliest methods proposed for superresolution DF is the currently popular multiple signal classification (MUSIC) method (Schmidt, 1979, 1986, 1981; Bienvenu and Kopp, 1980). Although MUSIC is often motivated in the literature solely by some observations regarding the properties of the eigenvalues and eigenvectors of the spatial autocorrelation matrix of the received data, it is shown here that the MUSIC algorithm can be obtained using the aforementioned idea of locating signals using beamforming and nulling in a restricted data space.

Consider the case of a known autocorrelation matrix $\boldsymbol{R}_{ii}$ of the interference and noise. Substracting $\boldsymbol{R}_{ii}$ from the autocorrelation matrix $\boldsymbol{R}_{xx}$ of the received data leaves only the components due to the desired signals. Thus, for any given weight vector w, the processor output power P due only to the desired signals can be expressed as (cf. (11))

$$P = w^{\mathrm{H}}(\boldsymbol{R}_{xx} - \boldsymbol{R}_{ii})w = w^{\mathrm{H}}A(\Theta)\boldsymbol{R}_{ss}A^{\mathrm{H}}(\Theta)w\,. \tag{18}$$

Given a set of K weight vectors $w_1, \ldots, w_K$, the sum of the average powers of the K corresponding processor outputs due only to the desired signals can be expressed as

$$P_{\text{total}} = \sum_{k=1}^{K} P_k = \operatorname{tr}\{W^{\mathrm{H}}(\boldsymbol{R}_{xx} - \boldsymbol{R}_{ii})W\}\,, \tag{19}$$

where $\operatorname{tr}\{\cdot\}$ denotes the matrix trace operation and $W = [w_1 \ \cdots \ w_K]$. As mentioned before, simultaneous beamforming and null-steering can limit res-

olution (analogous to limiting the narrowness of the transition band of an FIR temporal filter by imposing simultaneous passband and stopband requirements.) The alternate approach used here is to attempt only one of these tasks. For example, consider the approach of nulling all of the signals simultaneously. That this is possible can be seen as follows. Since the power P is computed in this restricted signals-only data space, it is as if there were no interfering signals and noise present. Consequently, if fewer than M signals are present and none of them are fully correlated (so that they can be rejected only by forming spatial nulls), weight vectors can be chosen that perfectly null all of the signals simultaneously. Since only $K = M - L$ such vectors can be found while still being linearly independent of each other, at most K such vectors need to be found. Therefore, to implement this approach it is necessary to find an $M \times K$ matrix W for which each column w_k, for $k = 1, \ldots, K$, nulls *all* of the desired signals. This can be accomplished by minimizing the total average output power P_{total} subject to the constraint that the columns of W be linearly independent (to prevent trivial or redundant solutions). The constraint is typically expressed by the more specific constraint $W^{\text{H}}W = I$, which yields a convenient solution (and is often the normalization chosen in the analysis of principal components, e.g. (Johnson and Wichern, 1988), which is closely related to the action of the MUSIC method). That is, the resulting method is summarized by choosing W_{null} equal to the solution of

$$\min_{W} P_{\text{total}}$$

$$\Rightarrow \min_{W} \text{tr}\{W^{\text{H}}(R_{xx} - R_{ii})W\} \quad \text{subject to } W^{\text{H}}W = I\,. \tag{20}$$

The angles of arrival can now be estimated by searching over θ for those array vectors $a(\theta)$ for which

$$\|W_{\text{null}}^{\text{H}} a(\theta)\|^2 = 0\,, \tag{21}$$

that is, by searching for those directions where a null is present simultaneously in all $K = M - L$, weight vectors.

Alternatively, to find the W for which the total average output power P_{total} is maximized, corresponding to beamforming simultaneously in all desired-signal directions, choose W_{beam} equal to the solution of

$$\max_{W} P_{\text{total}}$$

$$\Rightarrow \max_{W} \text{tr}\{W^{\text{H}}(R_{xx} - R_{ii})W\} \quad \text{subject to } W^{\text{H}}W = I\,, \tag{22}$$

where $K = L$. Then, the angles of arrival can be estimated by searching over θ for those array vectors $a(\theta)$ for which $\|W_{\text{beam}}^{\text{H}} a(\theta)\|^2$ is maximized, that is, by searching for those directions in which the cumulative array pattern $\|W_{\text{beam}}^{\text{H}} a(\theta)\|^2$ exhibits a beam.

In order to implement this method, means of determining W_{null} and/or W_{beam} are required. The solutions to these two optimization problems can be expressed in terms of the eigenvalues and eigenvectors of the signals-only spatial autocorrelation matrix:

$$W_{\text{null}} = [\mathbf{e}_{L+1} \quad \cdots \quad e_M] \tag{23}$$

and

$$W_{\text{beam}} = [e_1 \quad \cdots \quad e_L], \tag{24}$$

where $\{e_m\}$ are the eigenvectors defined by the equation

$$(R_{xx} - R_{ii})e_m = \lambda_m e_m, \tag{25}$$

and are ordered according to the associated eigenvalues λ_m, which are real and nonnegative,

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_L > \lambda_{L+1} = \cdots = \lambda_M = 0. \tag{26}$$

In practice, the noise-and-interference autocorrelation matrix might be known only to within an unknown multiplicative constant σ^2, say $R_{ii} = \sigma^2 Q_{ii}$. Then the eigenvalue problem is more appropriately expressed as

$$R_{xx}e_m = \lambda_m Q_{ii}e_m, \tag{27}$$

where the ordered eigenvalues satisfy

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_L > \lambda_{L+1} = \cdots = \lambda_M = \sigma^2. \tag{28}$$

Thus, the MUSIC algorithm is often described in the literature as consisting of the following steps. First, the ideal ($N \to \infty$) spatial autocorrelation matrix R_{xx} is estimated (using finite N). Second, the equation (27) is solved. Finally, the DOA estimates are given by either the maxima of

$$\|W_{\text{beam}}^{\text{H}} a(\theta)\|^2 \tag{29}$$

or the minima of

$$\|W_{\text{null}}^{\text{H}} a(\theta)\|^2. \tag{30}$$

An important point to note here is that the maximization of (29) and the minimization of (30) can be shown to be mathematically equivalent, which refutes the often-advanced argument that nulls are sharper than beams and should therefore yield higher resolution.

The conditions needed for the MUSIC method to work properly can be seen to include the requirements that (i) the number L of signals be less than the number M of sensors, (ii) $\boldsymbol{R}_{ii}$ be known to within a multiplicative constant, and (iii) the autocorrelation matrix of the transmitted signals have full rank, $\text{rank}\{\boldsymbol{R}_{ss}\} = L$. Otherwise, (i) no weight vectors exist that null all of the signals, (ii) the modified data-space in which the nulling or beamforming occurs is not the signals-only data space, and (iii) signals can be effectively nulled without being spatially nulled, respectively.

A significant conceptual difference between MUSIC and the previously discussed methods is that MUSIC finds the weight vectors of interest first and then searches for beams or nulls in the resulting antenna patterns, whereas the previous methods must compute a new weight vector for each search direction. Thus, in some sense, MUSIC jointly processes all of the desired signals and then applies some simple post-processing to locate them, whereas the previous methods estimate the direction of one signal and then ignore that information while searching for other signals. However, it should be noted that although the nulling vectors are all found simultaneously (which is characteristic of a multidimensional optimization problem), the search over θ proceeds in one dimension. This notion of joint processing helps to explain the superior performance of MUSIC (and the other methods presented in this section) as compared with the performance of Capon's method and conventional beamforming.

To illustrate this superiority, consider the environment processed by Capon's method corresponding to Figures 6 and 7. In this environment, the cumulative antenna patterns $\|\boldsymbol{W}_{\text{null}}^{\text{H}}\boldsymbol{a}(\theta)\|^2$ and $\|\boldsymbol{W}_{\text{beam}}^{\text{H}}\boldsymbol{a}(\theta)\|^2$ found by the MUSIC method are shown in Figures 10 and 11, respectively. Clearly, the null-steering pattern shows that MUSIC resolves the two signals; although it is not visible from the plot of the beam-steering pattern, there are indeed two peaks and they are located at the true DOAs. In fact, given the ideal ($N \rightarrow \infty$) spatial autocorrelation matrix and perfect calibration data as in the examples presented so far in this chapter, MUSIC can resolve two signals regardless of how close together they are, in stark contrast to the resolution limits of Capon's method and the conventional beamforming method.

If the user wishes to avoid estimating the directions of undesired signals, then the MUSIC algorithm requires a description of the undesired signals, namely, their spatial autocorrelation matrix $\boldsymbol{Q}_{ii}$. In the example above, if $\boldsymbol{Q}_{ii}$ were taken to be $\boldsymbol{I} + 10\boldsymbol{a}(0°)\boldsymbol{a}^{\text{H}}(0°)$, then the DOA of only the signal arriving from 10 degrees would remain to be estimated because the spatial characteristics of the signal arriving from 0 degrees having 10 dB SNR are included in $\boldsymbol{Q}_{ii}$ and are thus excluded from $\boldsymbol{W}_{\text{beam}}$ and $\boldsymbol{W}_{\text{null}}$. On the other hand, if $\boldsymbol{Q}_{ii}$ is in error, the performance of the MUSIC algorithm can be severely degraded because MUSIC is no longer beamforming and null-steering in the signals-only data space – interference and noise are also present.

An alternate interpretation of MUSIC is that of subspace fitting. Under this interpretation, the signals-only component of the autocorrelation matrix de-

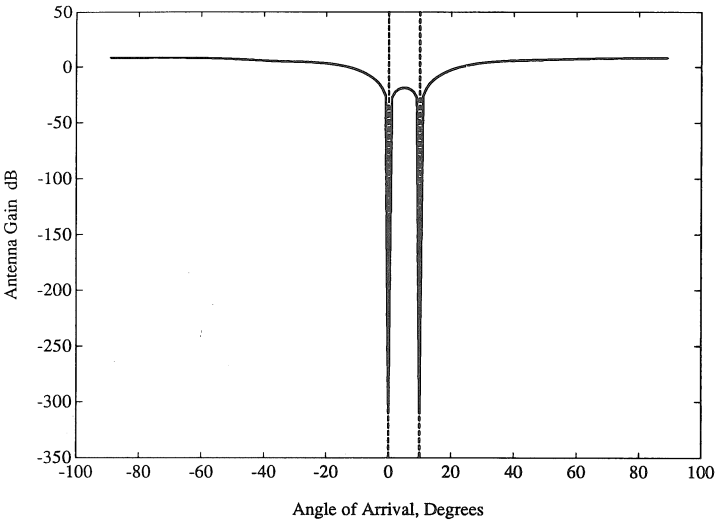

Fig. 10. Cumulative null-steering antenna pattern of MUSIC in the presence of two signals arriving from 0 and 10 degrees.

fines a *signal subspace* which is spanned by the eigenvectors W_{beam}, and the orthogonal complement of that subspace is sometimes referred to as the *noise subspace* which is spanned by the eigenvectors W_{null}. If R_{ss} is of full rank, then the signal subspace is also spanned by the array vectors of the desired signals, and consequently the noise subspace is orthogonal to these array vectors. Thus,

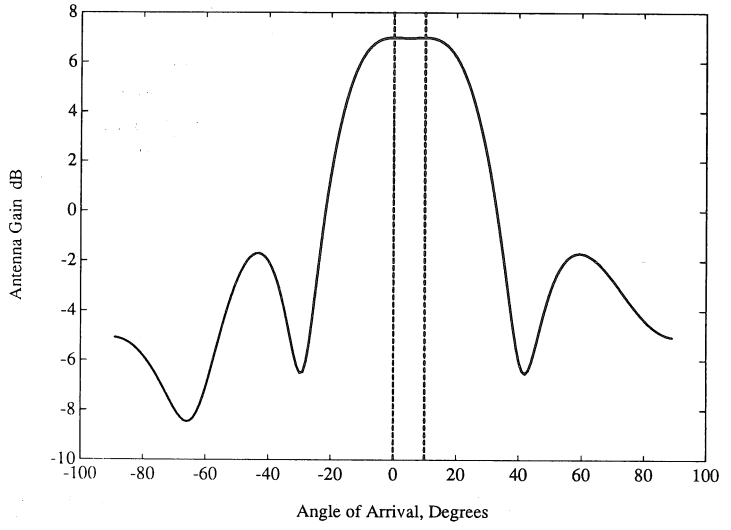

Fig. 11. Cumulative beam-steering antenna pattern of MUSIC in the presence of two signals arriving from 0 and 10 degrees. (Although it is not clearly visible here, there are indeed two peaks at the true DOAs).

the MUSIC algorithm searches for those DOAs for which the array vectors span the signal subspace, which corresponds to the simultaneous beamforming interpretation. In fact, each column of W_{beam} is a linear combination of conventional beamformers. Equivalently, it searches for those DOAs for which the array vectors are orthogonal to the noise subspace, which corresponds to the simultaneous null-steering interpretation. In other words, MUSIC attempts to fit an estimated signal subspace (which is spanned by the array response vectors corresponding to the estimated DOAs) to the observed signal subspace (which is derived from $\boldsymbol{R}_{xx}$).

These two interpretations can also be helpful for understanding the behavior of MUSIC in the presence of fully correlated multipath or smart jamming. Under the beamforming/null-steering interpretation, simultaneous nulling can occur without forcing a spatial null by destructively adding the correlated signal components, just as in Capon's method. This type of signal rejection is not detectable merely by examining the magnitude of the antenna pattern, and thus creates the impression that there are actually fewer signals impinging on the array. Alternatively, under the subspace-fitting interpretation, the reduction in dimension of the signal subspace causes an identical increase in dimension of the noise subspace, so fewer linearly independent array vectors can be found that are orthogonal to the noise subspace. Again, this creates the impression that fewer signals are impinging on the array.

Several modifications to MUSIC that enable it to operate properly in the presence of fully correlated sources have been proposed. The spatial smoothing approach (Shan et al., 1985; Reddi, 1987) requires that the array have a uniform linear geometry and reinterprets the array as a set of identical overlapping subarrays. By averaging the spatial autocorrelation matrices of these subarrays, a spatially smoothed autocorrelation matrix is formed in which the signals are no longer fully correlated. The standard MUSIC algorithm can then be applied to this smoothed autocorrelation matrix. Another approach, suggested in Schmidt (1981) and developed more fully in Zoltowski and Haber (1986), increases the dimensionality of the MUSIC search to allow for linear combinations of array vectors being orthogonal to the noise subspace. Yet another modification Haber and Zoltowski (1986) depends on the array being in motion to impart a unique Doppler shift to each signal. Averaging over several Doppler cycles essentially decorrelates the signals and allows MUSIC to be applied to the averaged spatial autocorrelation matrix.

5.2. ESPRIT

One of the acknowledged problems with the MUSIC algorithm is the need to obtain, store, and periodically check the calibration data, or to know the analytical expression for the array manifold. To sample the array manifold every θ_{res} degrees in azimuth and elevation over the whole sphere of possible arrivals with B-bit accuracy for the real and imaginary parts for each of M elements requires $(16200)(M)(B/\theta_{\text{res}}^2)$ bytes. For example, with spatial resolu-

tion of 0.1 degree, 16-bit accuracy, and an array of 10 sensors, the resulting requirement of measuring, storing, and accessing 250 megabytes of data can be quite costly if not impractical. Although sampling only in azimuth substantially reduces the storage to $(90)(M)(B/\theta_{\text{res}})$ bytes, the parameters in the preceding example still yield a requirement of 144 000 bytes. Also, the actual array manifold can fluctuate over time due to perturbations of the sensor locations, weather, nearby reflective and absorptive bodies, and so forth. Furthermore, DOA estimation requires a search over that calibration data, and this can be computationally expensive.

The technique for the estimation of signal parameters via rotational invariance (ESPRIT) (Paulraj et al., 1985; Roy, 1987; Roy and Kailath, 1989) avoids these requirements by imposing a particular structure on the array geometry and then exploiting that structure to great advantage. Specifically, the array is assumed to consist of two identical subarrays, one of which is spatially translated by a known distance with respect to the other. If the direction of translation is unknown, then the resulting DOA estimates will all be shifted by the same unknown amount from the true values. Examples of this doublet geometry are shown in Figure 12.

A simplified interpretation of ESPRIT is that it is a generalized interferometer that accommodates multiple signals by using more than two sensors. From Section 2 recall that two-sensor interferometry for the narrowband signal model in the absence of noise consists simply of measuring the phase between the two sensor output signals, say $x_1(n)$ and $x_2(n)$, and obtaining the DOA directly:

$$R_{x_1x_1} = \lambda R_{x_1x_2}\,, \tag{31}$$

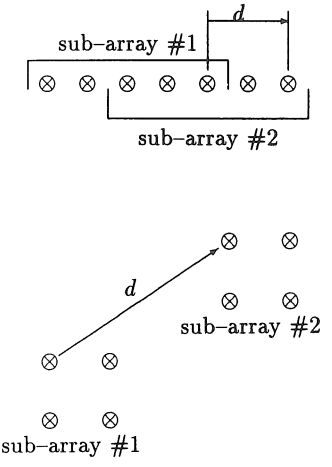

Fig. 12. Two examples of the doublet geometry. One array consists of two overlapping subarrays, whereas the other consists of two identical and disjoint subarrays.

where the DOA θ can be found by inverting the relation

$$\lambda = \exp\left(\mathrm{j}2\pi \frac{f}{c} d \sin\theta\right). \tag{32}$$

This idea can be extended to accommodate two subarrays, having sensor output signals $x_1(n)$ (for subarray 1) and $x_2(n)$ (for subarray 2), as follows. First, as in the MUSIC algorithm, L spatial filters W_{beam} (22), (24) that maximize the average processor output power (where L is the number of signals present) are found. As discussed previously, each column of W_{beam} is a weighted sum of conventional beamformers, where the conventional beamformer is defined for the doublet geometry as $[\boldsymbol{a}^{\mathrm{T}}(\theta),\ \boldsymbol{a}^{\mathrm{T}}(\theta)\exp(\mathrm{j}2\pi(f/c)\, d\sin\theta)]^{\mathrm{T}}$, in which $\boldsymbol{a}(\theta)$ is the array response vector for subarray 1, and thus aligns the phases and adjusts the gains of a desired signal at the different sensors. That is,

$$W_{\text{beam}} = \begin{bmatrix} A(\Theta) \\ A(\Theta)\Phi(\Theta) \end{bmatrix} T \tag{33}$$

for some $L \times L$ full-rank matrix T, where $\Phi(\Theta)$ is the diagonal matrix induced by the doublet geometry and has elements of the form (32),

$$\Phi(\Theta) = \mathrm{diag}\left\{\exp\left(\mathrm{j}2\pi \frac{f}{c} d \sin\theta_l\right)\right\}_{l=1}^{L}. \tag{34}$$

When only one signal is present and two sensors are used (so that $A(\Theta)$ and T are both scalars), the phase difference $\Phi(\Theta)$ is similar to the phase difference in the usual two-sensor interferometer, and can thus be inverted to obtain the estimated DOA of the desired signal. When multiple signals are present and more than two sensors are used, a more complicated procedure for obtaining the estimated DOAs is required.

The beamforming matrix W_{beam} can be expressed as

$$W_{\text{beam}} = \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}, \tag{35}$$

where E_1 and E_2 are $(M/2) \times L$ matrices, and M is the total number of sensors. Using the fact from (33) that $A(\Theta) = E_1 T^{-1}$, it can be shown that

$$E_2 = E_1 T^{-1} \Phi(\Theta) T = E_1 \Psi, \tag{36}$$

where $\Psi = T^{-1}\Phi(\Theta)T$. Thus, since $\Phi(\Theta)$ and Ψ are related by a similarity transformation, they have the same eigenvalues, which then yield the DOAs. However, since E_1 and E_2 are replaced in practice by their estimates $\hat{E}_1$ and $\hat{E}_2$, respectively, typically no matrix $\hat{\Psi}$ will satisfy

$$\hat{E}_2 = \hat{E}_1 \hat{\Psi}. \tag{37}$$

Therefore, a least-squares solution is required. Unlike the conventional least squares problem, both matrices $\hat{E}_1$ and $\hat{E}_2$ contain errors. Following Golub and van Loan (1989), the total least squares (TLS) solution for the $L \times L$ matrix $\hat{\Psi}$ can be found by computing the singular value decomposition

$$[\hat{E}_1 \quad \hat{E}_2] = [\hat{U}_1 \quad \hat{U}_2] \begin{bmatrix} \hat{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{21} & \hat{V}_{22} \end{bmatrix} \tag{38}$$

and forming

$$\hat{\Psi} = -\hat{V}_{12}\hat{V}_{22}^{-1} . \tag{39}$$

The eigenvalues of $\hat{\Psi}$ are the TLS-ESPRIT estimates of DOAs of the desired signals. That is, the eigenvalues can be interpreted loosely as being the phase differences from a generalized interferometer.

Notice that no search over θ is required, which means that no calibration data (apart from the doublet spacing d) is needed and that less computation can be required than for MUSIC. In addition to these implementation benefits, ESPRIT can outperform MUSIC in some environments (Roy, 1987; Roy and Kailath, 1989).

In contrast to the preceding interpretation which is based on spatial filtering, the subspace-fitting interpretations in Roy, (1987), Viberg and Ottersten (1991) show that TLS-ESPRIT fits a set of array response vectors to the signal subspace of the received data. However, rather than being constrained to lie in a completely known array manifold, the set of array response vectors is constrained only to have the form

$$\begin{bmatrix} A \\ A\Phi \end{bmatrix}, \tag{40}$$

where Φ is an $L \times L$ diagonal matrix having entries of the form (32).

The freedom from performing a search over θ does not come without disadvantages. One drawback of ESPRIT is the restriction that the array geometry be of the doublet variety. Difficulties in fabricating pairs of identical sensors and in placing numerous such pairs to form a doublet geometry can prevent ESPRIT from operating properly, akin to the performance degradation of MUSIC in the presence of calibration errors. Another limitation is that ESPRIT, like MUSIC, fails in the presence of fully correlated sources because of correlated signals from different directions being combined destructively by the spatial filters or, equivalently, because of loss of rank of the signal subspace.

5.3. Maximum likelihood

MUSIC and ESPRIT can in principle perform quite well in many signal environments and entail either a search having dimension equal to the number

of parameters per signal to be estimated (e.g., one for a search over azimuth only) or no search, respectively. However, the probability of failure increases as the degree of correlation among signals increases, and, in the presence of perfectly correlated signals, the algorithms fail entirely. As mentioned before, the signal nulling due to correlated signals from different directions being destructively combined (spatial-filtering perspective) and loss of dimension of the signal subspace (subspace-fitting perspective) preclude proper operation of MUSIC and ESPRIT under these conditions.

Motivated by this drawback, a different approach to superresolution DF is described here. This alternative approach is also presented from the spatial-filtering perspective but again leads naturally to a subspace-fitting point of view.

Given the received data sequence $\{\boldsymbol{x}(n)\}$, it is desired to reconstruct the components of the data due only to the desired signals. The parameter values for which the reconstruction approximates the received data with maximal accuracy are then taken to be the DOA and desired-signal waveform estimates. The approach taken here is to subtract from $\boldsymbol{x}(n)$ an estimate $A(\hat{\Theta})\hat{s}(n)$ of the signal components $A(\Theta)s(n)$. If the estimates $\hat{\Theta}$ and $\hat{s}(n)$ are sufficiently good, then the residual $\boldsymbol{x}(n) - A(\hat{\Theta})\hat{s}(n)$ will consist primarily of noise and interference (cf. (7)). Thus, one would expect that minimizing the energy in this residual by proper choice of $\hat{\Theta}$ and $\hat{s}(n)$ would result in accurate estimates of Θ and $s(n)$. In fact, this technique does perform well even in the cases where transmitted signals are perfectly correlated. To see why this should be so, consider a single signal $s(n)$ impinging on the array from two different directions θ_1 and θ_2 (e.g., zero-delay specular multipath),

$$\boldsymbol{x}(n) = [\boldsymbol{a}(\theta_1) + \boldsymbol{a}(\theta_2)]s(n) + \boldsymbol{i}(n) \,. \tag{41}$$

Unless there exist ambiguities in the array manifold, there is no scalar $\hat{\theta}$ for which $\boldsymbol{a}(\hat{\theta}) = \boldsymbol{a}(\theta_1) + \boldsymbol{a}(\theta_2)$, thus preventing the interpretation that fewer signals are impinging on the array. Therefore, despite the reduced dimension of the signal subspace (i.e., one instead of two), only the appropriate choice of two angle-estimates can account for all signal components in $\boldsymbol{x}(n)$.

The method can be stated mathematically in a least-squares form as

$$\min_{\hat{\Theta},\hat{s}(n)} \langle \| \boldsymbol{x}(n) - A(\hat{\Theta})\hat{s}(n) \|^2 \rangle_N \,, \tag{42}$$

for which the best least-squares fit between the received data and a reconstruction of the signal components of this data is sought. It can be shown that the solution for $\hat{s}(n)$ in terms of any $\hat{\Theta}$ is given by

$$\begin{aligned} \hat{s}(n) &= (A^{\mathrm{H}}(\hat{\Theta})A(\hat{\Theta}))^{-1}A^{\mathrm{H}}(\hat{\Theta})\boldsymbol{x}(n) \,, \\ &= W^{\mathrm{H}}\boldsymbol{x}(n) \,, \end{aligned} \tag{43}$$

which is the output of a spatial filter. Substituting this signal estimate back into the residual and minimizing over the vector of DOA estimates $\hat{\Theta}$ can be shown to be equivalent to maximizing a matrix trace as follows:

$$\max_{\hat{\Theta}} \operatorname{tr}\{P_A(\hat{\Theta})R_{xx}\}\,, \tag{44}$$

where $P_A(\hat{\Theta})$ is the projection matrix for the space spanned by the columns of $A(\hat{\Theta})$,

$$P_A(\hat{\Theta}) = A(\hat{\Theta})(A^{\mathrm{H}}(\hat{\Theta})A(\hat{\Theta}))^{-1}A^{\mathrm{H}}(\hat{\Theta})\,. \tag{45}$$

As predicted, for $\hat{\Theta} = \Theta$ the reconstructed data $A(\hat{\Theta})\hat{s}(n)$ is equal to the true signal components (since $P_A(\hat{\Theta})A(\Theta) = A(\Theta)$ when $\hat{\Theta} = \Theta$) plus residual noise,

$$\begin{aligned} x(n) - A(\hat{\Theta})\hat{s}(n) &= x(n) - P_A(\Theta)x(n) \\ &= (A(\Theta)s(n) + i(n)) - (A(\Theta)s(n) + P_A(\Theta)i(n)) \\ &= (I - P_A(\Theta))i(n)\,. \end{aligned} \tag{46}$$

Assuming that the interference and noise $i(n)$ is spatially uncorrelated (which rules out the usual type of interference of interest), $R_{ii} = I$, then the average power in the residual is the minimum attainable value compared with that obtained from using any other choice of $\hat{\Theta}$. Notice that the pitfall of destructively combining correlated signals arriving from different directions to reduce the average power in the residual is avoided here. That is, the elimination of energy in the residual due to a signal component $a(\theta_l)s_l(n)$ cannot be accomplished by subtracting a signal component $a(\theta_j)s_j(n)$ from a different direction, even if $s_l(n)$ and $s_j(n)$ are perfectly correlated. However, unlike the preceding methods, the DOAs must be found jointly (by means of a multidimensional search as indicated by (44)) instead of individually (by means of a one-dimensional search).

The natural progression to the framework of subspace fitting suggested by the appearance of the projection matrix yields the interpretation that the method fits to the received data a subspace spanned by vectors from the array manifold. However, the goodness of fit is measured differently here than in MUSIC and ESPRIT.

Another interpretation is that $\hat{\Theta}$ and $\hat{s}(n)$ obtained by this method are the maximum likelihood (ML) estimators of Θ and $s(n)$ under the assumption that $i(n)$ is a zero-mean white Gaussian time series. An advantage of this statistically oriented interpretation is that many existing analytical techniques and results can be applied to evaluate the bias and root mean-squared error (RMSE) of the estimates. In addition, this facilitates the computation of a lower bound on the RMSE of the estimates obtained by *any* algorithm that yields unbiased estimates when applied to this model of the received data, as discussed in Section 6.

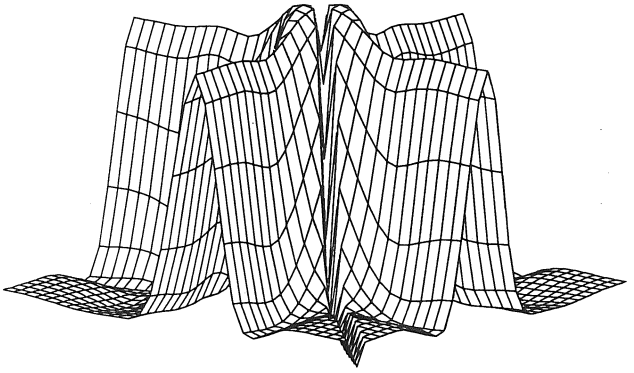

Fig. 13. Surface plot of the ML function in (44) for two uncorrelated signals arriving from 0 and 10 degrees.

The ability to operate properly in the presence of closely spaced signals is illustrated here by an example. The same environment used to create Figures 6 and 7 is used here: two uncorrelated signals arrive from 0 and 10 degrees with 10 dB SNR. The function to be maximized in (44) is evaluated over its two-dimensional domain in which $\hat{\theta}_1$ and $\hat{\theta}_2$ range from -90 to 90 degrees. The result is shown as a three-dimensional surface and as a contour plot in Figures 13 and 14, respectively, and the antenna patterns formed by $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are shown in Figure 15. The two peaks in the surface correspond to angle estimates of (0, 10) and (10, 0), order not being important. One of these peaks is marked by the intersection of the dashed lines in the contour plot in Figure 14.

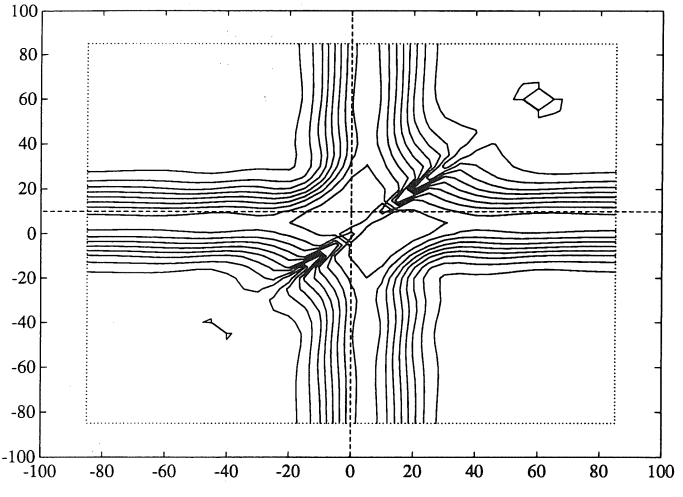

Fig. 14. Contour plot of the ML function in (44) for two uncorrelated signals arriving from 0 and 10 degrees. The intersection of the dashed lines marks the peak at (0, 10).

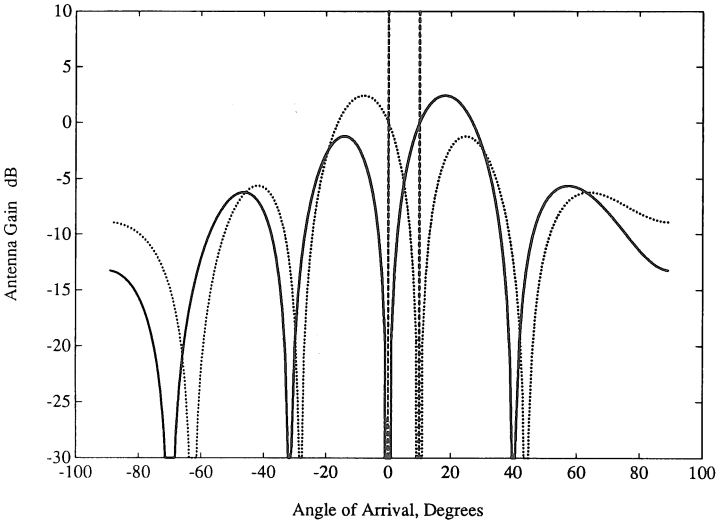

Fig. 15. Antenna patterns of the ML processor for two uncorrelated signals arriving from 0 and 10 degrees (denoted by the dashed lines).

Notice that each antenna pattern in Figure 15 exhibits high gain in the direction of one signal and a null in the direction of the other signal, a strategy which is referred to in Section 5 as one that potentially limits resolving capability. However, the fact that Capon's method is unsuccessful at using this strategy whereas the ML method succeeds is due largely to the signal-by-signal action of Capon's method, in contrast to the joint estimation technique of the ML method. That is, Capon's method estimates the DOA of only one signal at a time and forgets about prior estimates as it continues to scan for more signals. In contrast, the ML method simultaneously estimates the DOAs of all signals so as to best account for the spatial characteristics of the data as represented by $\boldsymbol{R}_{xx}$.

If, instead, the two signals from 0 and 10 degrees are perfectly correlated, the ML processor is still able to resolve them, as shown in Figures 16–18. Notice that the two peaks in the surface are not as prominent as in the case of uncorrelated signals. Also, as shown by the antenna patterns, the resulting waveform estimates $\hat{s}_1(n)$ and $\hat{s}_2(n)$ are not maximum-SNR solutions. The maximum-SNR processor would exploit the correlation between the two signals to improve the SNR, whereas the ML processor does not, as illustrated by the spatial null in each pattern in the direction of one of the signals. Since the ML method cannot reduce its average residual output power by destructively combining correlated signals, the fact that the signals are perfectly correlated does not prevent the ML method from choosing the correct DOA estimates, unlike Capon's method, MUSIC, and ESPRIT.

As is obvious from the figures, a major drawback of the ML algorithm is the

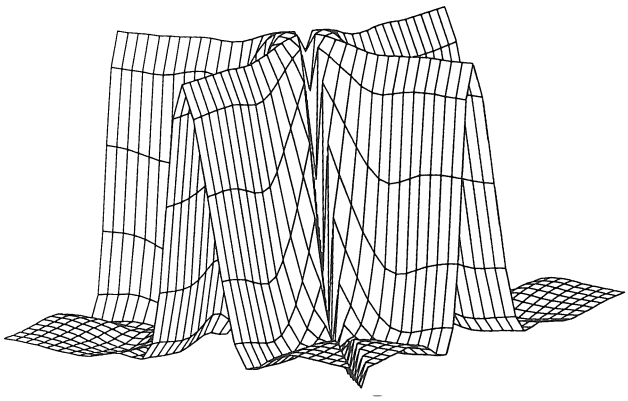

Fig. 16. Surface plot of the ML function in (44) for two correlated signals arriving from 0 and 10 degrees.

need to perform a multidimensional optimization which can be prohibitive in environments containing even a few signals. Furthermore, the matrix inversion (or QR decomposition) needed to compute $\boldsymbol{P}_A(\hat{\Theta})$ for each new estimate $\hat{\Theta}$ complicates matters. However, in Ziskind and Wax (1988) an efficient algorithm based on the alternating directions approach to multidimensional optimization is presented. This algorithm takes advantage of certain properties of projection matrices to eliminate the need to recompute $\boldsymbol{P}_A(\hat{\Theta})$ from scratch at every new iteration on the estimate $\hat{\Theta}$. However, the multidimensional search must still be performed, and the alternating directions approach is not guaranteed to be convergent in general to the global maximum in (44). Alternatively, the optimization problem can be reexpressed to allow a modified Gauss–

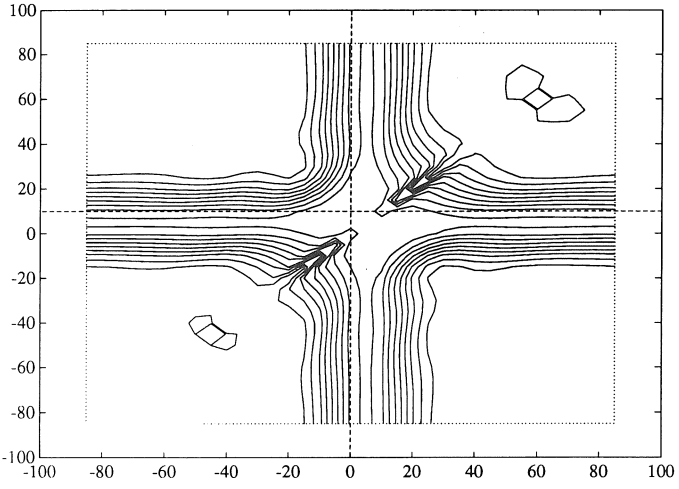

Fig. 17. Contour plot of the ML function in (44) for two correlated signals arriving from 0 and 10 degrees. The intersection of the dashed lines marks the peak at $(0, 10)$.

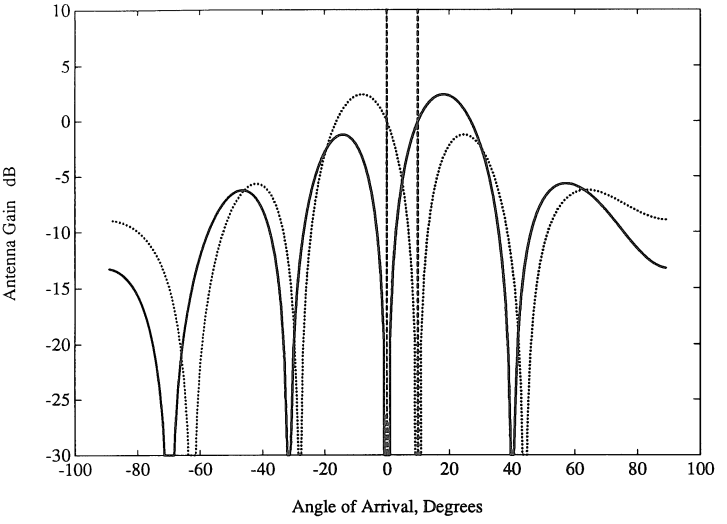

Fig. 18. Antenna patterns of the ML processor for two correlated signals arriving from 0 and 10 degrees (denoted by the dashed lines).

Newton method to be applied as suggested in Viberg and Ottersten (1991), and this can greatly accelerate convergence but requires a good initial estimate.

An alternative approach that works well for ULAs is the iterative quadratic maximum likelihood (IQML) algorithm in Bresler and Macovski (1986), which exploits the fact that the ML DOA estimates for a ULA are given by the roots of a polynomial.

5.4. *Weighted subspace fitting*

Although the MUSIC, ESPRIT, and ML methods seem to approach the DF problem very differently regardless of whether one interprets the methods from the spatial-filtering perspective or from the subspace-fitting perspective, each method can in fact be obtained as a solution to a special case of the basic subspace-fitting problem (Viberg and Ottersten, 1991)

$$\min_{\hat{\Theta},T} \|D - A(\hat{\Theta})T\|_F^2 , \tag{47}$$

where D is an $M \times q$ matrix that is obtained from the data (e.g., it can be the data itself, a choice that can avoid explicit computation of the autocorrelation matrix and can yield better performance when SNR is low), $\hat{\Theta}$ is the $p \times 1$ vector of DOA-estimation variables, and T is a $p \times q$ matrix of variables. Using standard methods, this problem can be shown to be equivalent to

$$\max_{\hat{\Theta}} \operatorname{tr}\{P_A(\hat{\Theta})DD^{\mathrm{H}}\} . \tag{48}$$

In other words, the basic subspace-fitting problem seeks the DOA estimates $\hat{\Theta}$ that adequately describe the spatial characteristics of the data. In Viberg and Ottersten (1991) it is shown that different choices of $\boldsymbol{D}$ or $\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}}$, the dimension p of $\hat{\Theta}$, and the form of $\boldsymbol{A}(\hat{\Theta})$ can yield the ML, MUSIC, and ML-ESPRIT methods and others, including the conventional beamforming method. For example, choosing $\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}} = \boldsymbol{R}_{xx}$ and setting p equal to the number L of signals yields the ML method, whereas choosing $\boldsymbol{D} = \boldsymbol{W}_{\mathrm{beam}}$ (the signal subspace eigenvectors of $\boldsymbol{R}_{xx}$) and setting $p = 1$ yields the MUSIC method.

These seemingly diverse methods are unified still further, for it is shown that the ML estimator (44) has the same asymptotic distribution as the estimator

$$\max_{\hat{\Theta}} \operatorname{tr}\{\boldsymbol{P}_{\boldsymbol{A}}(\hat{\Theta})\boldsymbol{E}_{\mathrm{s}}(\Lambda_{\mathrm{s}} - \sigma^2\boldsymbol{I})\mathbf{E}_{\mathrm{s}}^{\mathrm{H}}\}\,, \tag{49}$$

where $\boldsymbol{E}_{\mathrm{s}} = [\boldsymbol{e}_1 \cdots \boldsymbol{e}_L]$ (the signal subspace eigenvectors of $\boldsymbol{R}_{xx}$) and $\Lambda_{\mathrm{s}} = \operatorname{diag}\{\lambda_1 \cdots \lambda_L\}$ (the corresponding eigenvalues). Since MUSIC and ESPRIT already lend themselves naturally to the description

$$\max_{\hat{\Theta}} \operatorname{tr}\{\boldsymbol{P}_{\boldsymbol{A}}(\hat{\Theta})\boldsymbol{E}_{\mathrm{s}}\boldsymbol{W}\boldsymbol{E}_{\mathrm{s}}^{\mathrm{H}}\}\,, \tag{50}$$

the result (49) implies that , asymptotically, each subspace-fitting method can be expressed in the form of (50) for a particular choice of weighting matrix $\boldsymbol{W}$, dimension p of $\hat{\Theta}$, and constraints on $\boldsymbol{A}(\hat{\Theta})$.

A distinct advantage of this unified subspace-fitting perspective is that analyses of consistency and asymptotic covariance of estimator error for general $\boldsymbol{W}$ can be performed and the results can be specialized to the method of interest by the appropriate choice of $\boldsymbol{W}$. Consequently, it is possible to find the optimal $\boldsymbol{W}$ which minimizes the asymptotic covariance of estimator error. In Viberg and Ottersten (1991) it is shown that the optimal $\boldsymbol{W}$ is given by

$$\boldsymbol{W}_{\mathrm{WSF}} = (\Lambda_{\mathrm{s}} - \sigma^2\boldsymbol{I})^2\Lambda_{\mathrm{s}}^{-1} \tag{51}$$

or any consistent estimate thereof, and the algorithm (50)–(51) is referred to as the weighted subspace fitting (WSF) method. It seems unlikely that this optimal $\boldsymbol{W}$ could have been derived from the spatial-filtering perspective, further substantiating the usefulness of the subspace-fitting concept.

5.5. Methods for cyclostationary signals

In light of the optimality of the WSF method, one might question the need to investigate alternatives. However, by noting that the answers to questions such as

- Is performance degraded when $\boldsymbol{R}_{ii}$ is unknown or unequal to $\sigma^2\boldsymbol{I}$?
- When the amount of data available is limited, can other methods outperform the WSF method?
- Does the method fail when there are fewer sensors than signals?

- Must the method estimate the DOAs of all signals even if the DOAs of only a subset of signals are desired?

are affirmative for all previously discussed methods, then it becomes clear that alternative approaches are still required for some applications.

One class of methods addresses the above problems by being signal-selective to discriminate against undesired signals, interference, and noise. In particular, the methods discussed here are applicable to signals of interest that exhibit a time-domain statistical property known as *cyclostationarity* or, equivalently, a frequency-domain statistical property known as *spectral correlation*. Detailed treatments of the theory of cyclostationary signals are given in Gardner (1987b, 1989), Gardner and Brown (1991) and a brief tutorial survey is given in Gardner (1991). The great majority of communication and telemetry signals, including digital quadrature amplitude modulated (QAM) signals, phase-shift keyed (PSK) signals, frequency-shift keyed (FSK) signals, and others, exhibit this property as a result of periodic keying, gating, sampling, and mixing operations in the transmitter. Analyses of this property for specific modulation types for both analog and digital messages are given in Gardner (1987a), Gardner et al. (1987), and Gardner (1987b).

Briefly, a signal that exhibits second-order cyclostationarity has a periodically time-varying autocorrelation (which is equal to the additive periodic components of the lag product of the data) instead of the time-invariant autocorrelation that is characteristic of stationary signals, and the frequencies of this variation are referred to as the *cycle frequencies* of the signal. Equivalently, such a signal exhibits spectral correlation, meaning that the signal is correlated with frequency-shifted versions of itself; that is, the amplitude and phase fluctuations of narrowband signal components separated in frequency by a cycle frequency are correlated. For example, the cycle frequencies of digital communication signals can include the keying rate, the doubled sine-wave carrier frequency, and sums and differences of these.

Several DF algorithms that exploit cyclostationarity by discriminating among signals with different cycle frequencies are described next. More detailed discussion, including derivations and finite-time performance results, is given in Schell (1990, 1993 and references therein). Also, these techniques for DF and others for blind adaptive spatial filtering are the subject of a patent application (Gardner et al., 1988).

5.5.1. Cyclic MUSIC

The first signal-selective method that exploits the cyclostationarity property is the Cyclic MUSIC method (Gardner, 1988; Schell et al., 1989; see also Schell, 1990, 1993) and can be motivated from either the subspace-fitting perspective or the spatial-filtering perspective. By exploiting the fact that the signals with cycle frequency α contribute to a measurement of spectral correlation for frequencies separated by α while other signals, as well as noise, eventually become decorrelated in such a measurement, Cyclic MUSIC is able to select a subset of signals on which to perform the DF task as follows. Given a signal

$s(n)$ with cycle frequency α, the *cyclic autocorrelation function* $R_{ss}^{\alpha}(\tau)$ can be estimated using the finite-time average of the sinusoidally-weighted lag product

$$R_{ss}^{\alpha}(\tau) = \langle s(n + \tau/2)s^*(n - \tau/2)\, e^{-j2\pi\alpha n} \rangle_N , \tag{52}$$

and this converges to a non-zero complex value (for each value of lag τ within some range) in the limit as the sample size N goes to infinity. Moreover, if additive noise-and-interference $i(n)$, independent of $s(n)$, is present

$$x(n) = s(n) + i(n) , \tag{53}$$

and is not cyclostationary with cycle frequency α, then the cyclic autocorrelation function still converges to the same value,

$$\begin{aligned} R_{xx}^{\alpha}(\tau) &\rightarrow R_{ss}^{\alpha}(\tau) + R_{ii}^{\alpha}(\tau) \\ &= R_{ss}^{\alpha}(\tau) , \end{aligned} \tag{54}$$

since $R_{ii}^{\alpha}(\tau) \equiv 0$ for $N \rightarrow \infty$. This signal-selectivity property can be generalized to accommodate the vector signal at the antenna array receiver output as follows, under the assumption that the narrowband approximation holds. If a number L_α of signals $s_1(n), \ldots, s_{L_\alpha}(n)$ with cycle frequency α impinge on the array along with other signals, interference, and noise $i(n)$ not exhibiting cyclostationarity with cycle frequency α (e.g., amplifier noise, signals with other keying rates, etc.), then the cyclic autocorrelation matrix $\boldsymbol{R}_{xx}^{\alpha}(\tau)$ is given by

$$\begin{aligned} \boldsymbol{R}_{xx}^{\alpha}(\tau) &= \langle \boldsymbol{x}(n + \tau/2)\boldsymbol{x}^{H}(n - \tau/2)\, e^{-j2\pi\alpha n} \rangle_N \\ &\rightarrow \boldsymbol{A}(\Theta)\boldsymbol{R}_{ss}^{\alpha}(\tau)\boldsymbol{A}^{H}(\Theta) \quad \text{as } N \rightarrow \infty , \end{aligned} \tag{55}$$

where $\boldsymbol{A}(\Theta)$ is an $M \times L_\alpha$ matrix and $\boldsymbol{R}_{ss}^{\alpha}(\tau)$ is an $L_\alpha \times L_\alpha$ matrix. It should be noted that the often-used but overly strict single-tone idealization of the narrowband assumption (cf. Section 3.1) is inconsistent with the concept of spectral correlation, which inherently involves more than one frequency. However, all that is typically required for the conditions of the narrowband assumption to be satisfied is that the bandpass bandwidth of the received data be much less than the reciprocal of the propagation time across the array. For example, simulations of Cyclic MUSIC for a 7-element uniform linear array receiving two signals having different cycle frequencies indicate that the relative bandwidth (the bandpass bandwidth of the receiver divided by the carrier frequency) can be as large as 50% without substantial degradation in performance.

The most important property of $\boldsymbol{R}_{xx}^{\alpha}$ is that it converges to (55), which does not contain any contributions from undesired and interfering signals and noise. Therefore, $\boldsymbol{R}_{xx}^{\alpha}(\tau)$ can be used as $\boldsymbol{R}_{xx} - \boldsymbol{R}_{ii}$ is used in conventional MUSIC to find weight vectors that null certain signals. That is, weight vectors $\boldsymbol{w}$ can be

found to simultaneously null all of the signals that have cycle frequency α,

$$\boldsymbol{w}^{\mathrm{H}}\boldsymbol{R}_{xx}^{\alpha}(\tau)\boldsymbol{w}=0\,. \tag{56}$$

Since L_α signals are to be nulled simultaneously, $M-L_\alpha$ linearly independent weight vectors can be found.

From the subspace-fitting perspective, if $\boldsymbol{R}_{ss}^{\alpha}(\tau)$ has full rank (e.g., no perfectly correlated multipath), then the $M-L_\alpha$-dimensional null space of $\boldsymbol{R}_{xx}^{\alpha}(\tau)$ is orthogonal to the L_α columns of $\boldsymbol{A}(\Theta)$ corresponding to the desired signals, analogous to the similar result that holds for the $M-L$-dimensional noise subspace in conventional MUSIC. Cyclic MUSIC finds the set of array response vectors that are orthogonal to the space spanned by the null-space eigenvectors of $\boldsymbol{R}_{xx}^{\alpha}(\tau)$. Thus, the Cyclic MUSIC method solves

$$\boldsymbol{R}_{xx}^{\alpha}(\tau)\boldsymbol{w}_m=\lambda_m\boldsymbol{w}_m\,,\quad m=1,\ldots,M\,, \tag{57}$$

where the eigenvalues are ordered as

$$|\lambda_1|\geqslant\cdots\geqslant|\lambda_{L_\alpha}|\geqslant|\lambda_{L_\alpha+1}|=\cdots=|\lambda_M|=0\,, \tag{58}$$

and searches for the DOAs for which the array vectors are orthogonal to the null space,

$$\min_{\hat{\theta}}\|\boldsymbol{W}_{\mathrm{null}}^{\mathrm{H}}\boldsymbol{a}(\hat{\theta})\|^2\,, \tag{59}$$

where the null space $\boldsymbol{W}_{\mathrm{null}}$ is given by $\boldsymbol{W}_{\mathrm{null}}=[\boldsymbol{w}_{L_\alpha+1}\cdots\boldsymbol{w}_M]$. Equivalently, Cyclic MUSIC searches for the DOAs for which the corresponding array vectors lie entirely within the cyclic signal subspace spanned by $\boldsymbol{W}_{\mathrm{signal}}=[\boldsymbol{w}_1\cdots\boldsymbol{w}_{L_\alpha}]$. Note that the columns of $\boldsymbol{W}_{\mathrm{signal}}$ are not directly analogous to those of $\boldsymbol{W}_{\mathrm{beam}}$ of conventional MUSIC because the former are not orthogonal in general ($\boldsymbol{R}_{xx}^{\alpha}$ is not Hermitian) whereas the latter are guaranteed to be so ($\boldsymbol{R}_{xx}$ is Hermitian by definition). However, the singular vectors obtained from the singular value decomposition of $\boldsymbol{R}_{xx}^{\alpha}$ can (and typically should) be used instead of the eigenvectors, in which case the columns spanning the signal subspace *are* analogous to those of $\boldsymbol{W}_{\mathrm{beam}}$ of conventional MUSIC.

However, unlike conventional MUSIC, Cyclic MUSIC does not require knowledge of $\boldsymbol{R}_{ii}$ to obtain the desired signals-only correlation matrix and can be even more selective because, for a particular choice of α, only those signals having a cycle frequency equal to α contribute to (55).

The implications of this signal selectivity are numerous. For example, if the received signals can be grouped into sets of signals having the same cycle frequency, and each set contains fewer than M signals, then Cyclic MUSIC can be applied for each α of interest to cumulatively resolve many more signals than there are sensors. This works because each application of Cyclic MUSIC must null only the desired signals for that particular value of α, which it can do

as long as there are fewer of them than there are sensors. Also, DF can be performed simultaneously on multiple signals with different values of α by adding their cyclic autocorrelations, provided that there are fewer such signals than sensors. Another implication is that the amount of post-processing of the DOA estimates to eliminate those due to undesired signals is reduced or eliminated.

Yet another benefit of signal selectivity is an expanded notion of resolution, above and beyond the usual meaning. A conventional algorithm can exploit only spatial processing to resolve two closely spaced signals (present in the same spectral band and in the same time interval), even if only one of those is a desired signal, and thus can fail when the angular separation between the signals is small since angular resolution is limited by the amount of available data (cf. Sections 6 and 7). However, if the signals have different cycle frequencies, then Cyclic MUSIC must process only one signal at a time in the spatial domain – the signals are actually resolved in the cycle frequency domain (regardless of temporal and spectral overlap). Although cycle resolution is also limited by the amount of useful data available, this is essentially independent of the angular separation. Furthermore, Monte Carlo simulations have shown that the amount of data needed for cycle resolution can be much smaller than that needed for angular resolution (see Schell, 1990, and Schell et al., 1989).

In many applications involving communication and telemetry systems, $L_\alpha = 1$ for each α of interest. That is, multiple signals often do not share the same sine-wave carrier frequency, although there are important exceptions like code division multiplexing, in which both the sine-wave carrier frequency and the keying rate are the same for all users, and some other systems in which multiple users share the same keying rate but might have different sine-wave carriers (e.g., in frequency division multiplexing). When $L_\alpha = 1$, tremendous simplifications result in Cyclic MUSIC (and in other algorithms to be described subsequently).

As an example of the ability of Cyclic MUSIC to operate properly when more signals are received than there are sensors, consider a five-element ULA receiving six signals all having 10 dB SNR. Three of the signals have a cycle frequency α_1 and arrive from $-27°$, $0°$, and $10°$. The other three signals have a cycle frequency α_2 and arrive from $-47°$, $-3°$, and $35°$. The first group of signals does not have cycle frequency α_2, and the second group does not have cycle frequency α_1. For example, the signals in the first group are digital radio signals that all have keying rate equal to α_1, and the signals in the second group have keying rate equal to α_2, where α_1 and α_2 are unequal and are not harmonics of each other. The Cyclic MUSIC algorithm is applied with the cycle frequency parameter α equal to α_1. The antenna pattern resulting from the null-space eigenvectors of $\boldsymbol{R}_{xx}^{\alpha}$ is shown in Figure 19. Clearly, only the three signals having cycle frequency α_1 (denoted by the dashed lines) are nulled, whereas the other signals (denoted by dotted lines) are ignored. Similarly, if α is set equal to α_2, then only the three signals having cycle frequency α_2 are nulled. Thus, the Cyclic MUSIC algorithm successfully estimates the DOAs of

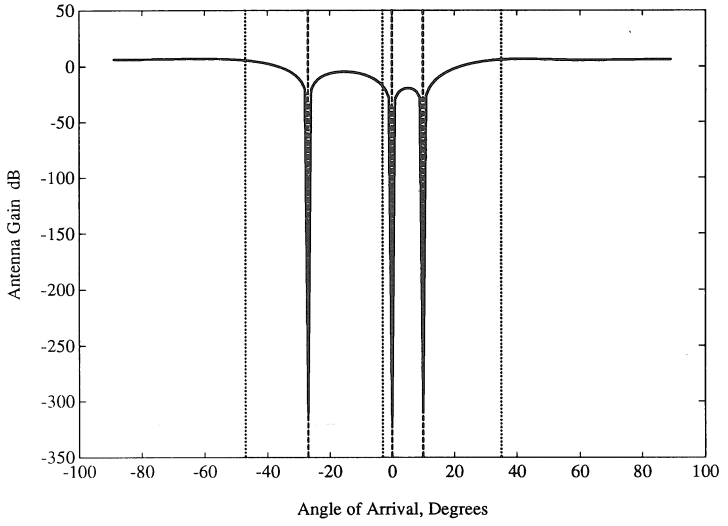

Fig. 19. Null-steering antenna pattern of the Cyclic MUSIC processor for six uncorrelated signals arriving at a five-element array. The signals arriving from -27, 0, and 10 degrees (denoted by dashed lines) all have cycle frequency α_1, and the signals arriving from -47, -3, and 35 degrees (denoted by dotted lines) do not have cycle frequency α_1. The cycle frequency parameter α in Cyclic MUSIC is set equal to α_1 to obtain the pattern shown here.

all six signals by processing only smaller subsets of signals. In contrast, the conventional MUSIC algorithm (without knowledge of the correlation matrix $\boldsymbol{R}_{ii}$ that incorporates some of the signals of interest) is guaranteed to fail because it can accommodate at most four signals. However, in order to obtain some results for comparison, MUSIC is operated under the assumption that there are four signals present, and the eigenvector corresponding to the smallest eigenvalue of $\boldsymbol{R}_{xx}$ is treated as if it were the noise subspace eigenvector. MUSIC does obtain four estimates that are in the neighborhood of the true DOAs but clearly does not obtain satisfactory results, as shown in Figure 20.

These benefits of signal selectivity are also shared by the additional four cyclostationarity-exploiting DF algorithms described next.

5.5.2. Phase-SCORE Cyclic MUSIC

The Phase-SCORE[1] Cyclic MUSIC method (see Schell, 1990, and Schell et al., 1989) is based on the following observation. For statistically independent zero-mean signals with cycle frequency α, both the cyclic and conventional autocorrelation matrices are diagonal, because both cyclic and conventional

[1] The acronym *SCORE* stands for *spectral coherence restoral* and represents a class of algorithms originally designed for blind adaptive signal extraction rather than DF (Agee et al., 1990). The Phase-SCORE method for signal extraction derives its name from the fact that it exploits the phase as well as the magnitude of the spectral coherence.

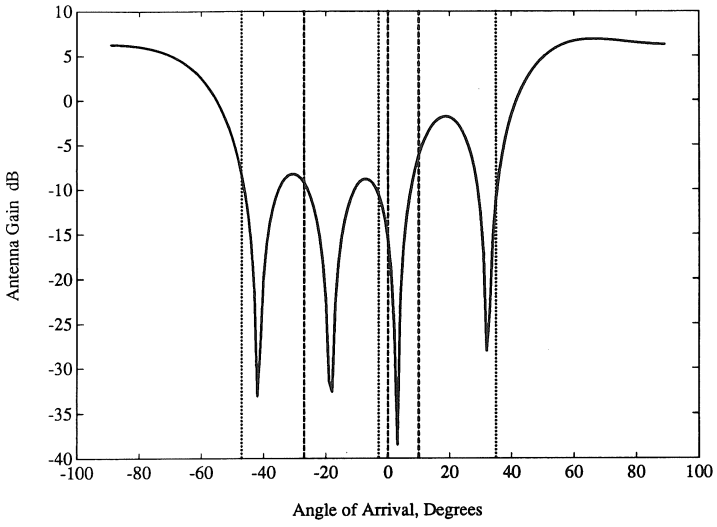

Fig. 20. Null-steering antenna pattern of the MUSIC processor operating in the same environment as used for Figure 19. Four signals are assumed to be present to render operable the MUSIC algorithm, although six signals are being received.

correlations between different signals are zero. Therefore, the cyclic autocorrelation matrix is proportional (via a diagonal matrix) to the conventional autocorrelation matrix,

$$\boldsymbol{R}_{ss}^{\alpha}(\tau) = \boldsymbol{R}_{ss}\Lambda\,. \tag{60}$$

However, each sensor receives some linear combination of the signals, so this property does not hold if $s(n)$ is replaced by $x(n)$. Replacing $s(n)$ by $y(n)$, where $y(n)$ is output vector of the matrix processor $W^{\mathrm{H}}x(n)$, might allow the desired property to hold,

$$\boldsymbol{R}_{yy}^{\alpha}(\tau) = \boldsymbol{R}_{yy}\Lambda\,, \tag{61}$$

if the set of spatial filters described by the columns of W is chosen appropriately. Furthermore, since additive noise and interference destroy this property by contributing only to the right-hand side of (61), then restoring this property to the output of the receiver processor ought to extract good estimates of the waveforms of the cyclostationary signals. This property restoral (61) requires the following equality to be enforced,

$$W^{\mathrm{H}}\boldsymbol{R}_{xx}^{\alpha}(\tau)W = W^{\mathrm{H}}\boldsymbol{R}_{xx}W\Lambda\,, \tag{62}$$

and this can be accomplished by solving the eigenvalue equation

$$\boldsymbol{R}_{xx}^{\alpha}(\tau)w_m = \lambda_m\boldsymbol{R}_{xx}w_m\,, \quad m = 1, \ldots, M\,. \tag{63}$$

It can be shown (Schell et al., 1989) that these eigenvectors can be used in the same manner as those in Cyclic MUSIC with the added benefits that the signal waveform estimates $\hat{s}_m(n) = w_m^H x(n)$, $m = 1, \ldots, L_\alpha$, using the signal subspace eigenvectors achieve nearly the maximum attainable signal-to-interference-and-noise ratio (Schell and Agee, 1988) without using any knowledge of the array manifold. Also, it can be shown that (63) is equivalent to applying Cyclic MUSIC to the spatially whitened data $R_{xx}^{-1/2} x(n)$, which implies that strong interfering signals are suppressed prior to the application of the Cyclic MUSIC algorithm. However, this also implies that strong desired signals are also suppressed.

5.5.3. *Cyclic ESPRIT*

A simplification of the ESPRIT method by exploitation of cyclostationarity to obtain signal selectivity was proposed in Gardner (1988) for the special case in which the number L_α of signals having cycle frequency α is equal to one. However, its use of a generalized eigenvalue equation in which both matrices are singular creates difficulties in implementation. To address this problem and to obtain a method that can accommodate $L_\alpha > 1$, the total least squares (TLS) ESPRIT method (which accommodates multiple signals and is numerically well-behaved) is generalized here to exploit cyclostationarity. Instead of computing the conventional autocorrelation matrix R_{xx} of the sensor output signals $x(n)$ and computing its eigenvalue decomposition to obtain W_{beam}, the cyclic autocorrelation matrix $R_{xx}^\alpha(\tau)$ is estimated, and its singular value decomposition is computed to obtain W_{beam}. By using L_α in place of L in the remainder of the TLS-ESPRIT procedure, the eigenvalues of the resulting total least squares solution for $\hat{\Psi}$ yield the estimated DOAs of only those signals having cycle frequency α. Thus Cyclic ESPRIT can be interpreted as a generalized interferometer that accommodates multiple signals by using multiple sensors but, unlike conventional TLS-ESPRIT, discriminates against interference and noise that does not have the desired cycle frequency.

As with the aforementioned DF methods that exploit cyclostationarity, Cyclic ESPRIT can obtain better performance than conventional TLS-ESPRIT when closely spaced signals have different cycle frequencies, and when noise and interference have unknown spatial characteristics. Also, since the number L_α of signals having cycle frequency α is often less than the number L of all signals present and in many cases is equal to one, Cyclic ESPRIT can often require much less computation and post-processing to classify the DOA estimates than conventional TLS-ESPRIT.

As with conventional TLS-ESPRIT, Cyclic ESPRIT requires the sensor array to have a doublet geometry and does not require any calibration data except for the translation vector separating the two subarrays. Also, it does not operate properly when perfectly correlated signals having the specified cycle frequency arrive from different directions, which can be caused by multipath propagation or smart jamming.

5.5.4. Cyclic Least Squares

The Cyclic MUSIC methods are attractive because they retain the computational simplicity of the one-dimensional search (like MUSIC) while addressing the many problems associated with the lack of signal selectivity inherent in MUSIC and almost all other algorithms that do not exploit cyclostationarity. Similarly, the Cyclic ESPRIT method avoids the search over the array manifold, but unlike conventional TLS-ESPRIT, exploits cyclostationarity to obtain better performance in some environments. However, these methods fail in the presence of fully correlated sources for essentially the same reasons that MUSIC and ESPRIT do, although failure occurs in the cyclostationarity-exploiting algorithms only when the signals having the specified cycle frequency are fully correlated. In response to this shortcoming, the Cyclic Least Squares (CLS) method exploits the cyclostationarity of the desired signals in a subspace-fitting approach more akin to that of the conventional ML method. That is, CLS attempts to estimate the DOAs and signal waveforms so as to minimize the average power in the residual between the received data and a frequency-shifted reconstruction of the cyclostationary signal components of the received data,

$$\min_{\hat{\Theta},\hat{s}(n)} \left\langle \left\| \boldsymbol{x}(n) - \boldsymbol{A}(\hat{\Theta})\boldsymbol{s}(n)\, \mathrm{e}^{-\mathrm{j}2\pi\alpha n} \right\|^2 \right\rangle_N , \tag{64}$$

where $\hat{\boldsymbol{s}}(n) = \boldsymbol{W}^{\mathrm{H}}\boldsymbol{x}(n)$. In Schell (1990) and Schell and Gardner (1990c) it is shown that the resulting $\boldsymbol{W}$ is given by

$$\boldsymbol{W} = \boldsymbol{R}_{xx}^{-1}\boldsymbol{R}_{xx}^{\alpha\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}(\hat{\Theta})(\boldsymbol{A}(\hat{\Theta})\boldsymbol{A}^{\mathrm{H}}(\hat{\Theta}))^{-1} \tag{65}$$

and extracts estimates $\hat{\boldsymbol{s}}(n)$ with maximum signal-to-interference-and-noise ratio when the signals are uncorrelated. The resulting simplified optimization problem can be expressed as

$$\max_{\hat{\Theta}} \mathrm{tr}\{\boldsymbol{P}_{\boldsymbol{A}}(\hat{\Theta})\boldsymbol{R}_{xx}^{\alpha}\boldsymbol{R}_{xx}^{-1}\boldsymbol{R}_{xx}^{\alpha\mathrm{H}}\} \tag{66}$$

and bears a resemblance in form to the conventional ML algorithm (cf. (44)). In return for its ability to estimate DOAs even when signals are perfectly correlated, the algorithm suffers from the same complexity due to multidimensional optimization as WSF, ML, and others. However, unlike these methods, the CLS algorithm is applicable when $\boldsymbol{R}_{ii}$ is unknown or unequal to $\sigma^2\boldsymbol{I}$ and needs only the weaker condition $L_\alpha < M$. Furthermore, since CLS estimates the DOAs of only the signals with cycle frequency α, the dimension and complexity of the multidimensional search is reduced correspondingly. For example, as discussed in Section 5.5.1, $L_\alpha = 1$ is common in practice.

5.5.5. Cyclic DF method of Xu and Kailath

Although the Cyclic MUSIC and CLS methods can operate properly on data

that is not strictly narrowband, performance can be degraded nonetheless. Also, it is not obvious how to exploit the cyclic correlation associated with multiple lag values τ, which can be necessary when desired signals having the same keying rate but different modulation types or keying envelopes are present. A method that is asymptotically exact for wideband data and efficiently exploits the cyclic correlation associated with multiple lag values is presented by Xu and Kailath (1992). This method is the generalization of the earlier Cyclic Phase Difference (CPD) method (Gardner and Chen, 1992) from two sensors to an array of sensors. That is, if the array contains only two sensors, the DF method of Xu and Kailath reduces to the explicit solution

$$\hat{\theta} = \sin^{-1}\left(-\frac{c}{2\pi\alpha d}\,\text{angle}\left\{\sum_{\tau} R^{\alpha}_{x_1x_1}(\tau)R^{\alpha *}_{x_2x_2}(\tau)\right\}\right), \tag{67}$$

where c is the propagation speed and d is the distance between the two sensors. Equation (67) is the CPD algorithm, which implements the solution to a least-squares phase-fitting problem.

This method differs from the previously discussed methods because it does not perform beam- and null-steering to enhance some signals and reject others, nor does it require cross-correlations between the data at different sensors. Instead, it operates as a type of multi-sensor TDOA method, except that the TDOA is not measured directly but is measured indirectly through the differences in phases of additive sinewave components (at frequency α) that appear in the lag products of the data, which are reflected in the phase differences of the cyclic autocorrelation, and the TDOA is explicitly parameterized by the direction of arrival. The method exploits the property that the cyclic autocorrelation of a delayed signal $y(t) = x(t - t_d)$ is a phase-shifted version of the cyclic autocorrelation of the original signal $x(t)$:

$$R^{\alpha}_{yy}(\tau) = R^{\alpha}_{xx}(\tau)\exp(-\text{j}2\pi\alpha t_d)\,. \tag{68}$$

Under the assumption that each signal received at one sensor is merely a delayed version of a corresponding signal received at another sensor, the phase differences between the cyclic autocorrelations at two or more sensors can be used to estimate the delay(s) and thus to estimate the direction of arrival.

When L_α signals having cycle frequency α are present, the vector of cyclic autocorrelations at the sensors can be expressed as a linear combination of L_α induced array response vectors,

$$\begin{bmatrix} R^{\alpha}_{x_1x_1}(\tau) \\ \vdots \\ R^{\alpha}_{x_Mx_M}(\tau) \end{bmatrix} = \sum_{l=1}^{L_\alpha} \begin{bmatrix} \exp(-\text{j}2\pi\alpha\,\dfrac{d_1}{c}\sin\theta_l) \\ \vdots \\ \exp(-\text{j}2\pi\alpha\,\dfrac{d_M}{c}\sin\theta_l) \end{bmatrix} R^{\alpha}_{s_ls_l}(\tau) = \sum_{l=1}^{L_\alpha} \boldsymbol{b}(\theta_l)R^{\alpha}_{s_ls_l}(\tau)\,. \tag{69}$$

Notice that the cycle frequency α plays the role of the center frequency in the induced array response vectors $\boldsymbol{b}(\theta_l)$, and that these vectors describe the phase characteristic induced by the property (68), where the values $(d_m/c)\sin\theta_l$ play the role of the TDOA t_d. More importantly, this representation resembles the narrowband signal model (with the left-hand side playing the role of the received data, the induced array response vectors playing the role of the usual array response vectors, the cyclic autocorrelations of the desired signals playing the role of the desired signal waveforms, and the lag value τ playing the role of the time index n) in the absence of interference and noise. Consequently, measuring the cyclic autocorrelations at the sensors for K different lag values $\tau_1, \ldots, \tau_K$ and collecting them into a matrix

$$\boldsymbol{R} = \begin{bmatrix} R^{\alpha}_{x_1x_1}(\tau_1) & \cdots & R^{\alpha}_{x_1x_1}(\tau_K) \\ \vdots & \ddots & \vdots \\ R^{\alpha}_{x_Mx_M}(\tau_1) & \cdots & R^{\alpha}_{x_Mx_M}(\tau_K) \end{bmatrix} \tag{70}$$

yields a matrix whose columns span the same space as the induced array response vectors.

In practice the matrix $\boldsymbol{R}$ must be estimated, and the relation (69) no longer holds exactly. Consequently, a least-squares fit of induced array response vectors to the column space of the estimated matrix $\boldsymbol{R}$ can be used to find the estimated angles of arrival:

$$\min_{\hat{\theta},W} \| \boldsymbol{R}\boldsymbol{W} - [\boldsymbol{b}(\hat{\theta}_1) \quad \cdots \quad \boldsymbol{b}(\hat{\theta}_{L_\alpha}] \|^2 . \tag{71}$$

This minimization problem can be shown to reduce to the simpler maximization problem:

$$\max_{\hat{\theta}} \| [\boldsymbol{u}_1 \quad \cdots \quad \boldsymbol{u}_{L_\alpha}]^{\mathrm{H}} \boldsymbol{b}(\hat{\theta}) \|^2 , \tag{72}$$

where $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{L_\alpha}$ are the left singular vectors corresponding to the L_α significant singular values of $\boldsymbol{R}$. The L_α highest peaks in the objective function are taken to be the direction estimates.

Thus, the method estimates the matrix $\boldsymbol{R}$, computes the L_α left singular vectors corresponding to the L_α significant singular values, and finds the angle estimates $\hat{\theta}_l$ for which the corresponding induced array response vectors most nearly lie in the space spanned by those singular vectors.

By exploiting the cyclic correlation associated with multiple lag values, this method can achieve much better performance (e.g., less RMSE) than Cyclic MUSIC or CLS, and does so without requiring potentially troublesome cross-correlations between sensors. It also benefits from signal-selectivity just as Cyclic MUSIC and CLS do, but extends this benefit more accurately to environments that cannot be adequately modeled as narrowband. However, the method is not applicable to environments in which the signals are partially

or fully correlated. Also, since the presence of ambiguities in the induced array manifold is determined by the product of the sensor spacing and the cycle frequency α, environments in which the desired signals *are* appropriately modeled as being narrowband but do not exhibit cyclostationarity associated with the carrier frequency (e.g., QAM or QPSK signals having small relative bandwidths) can require that the sensors be widely spaced, since α will then be relatively small. The need for this wide spacing has also been noticed in work on the CPD method for TDOA estimation (Gardner and Chen, 1992).

6. Performance limits

Given a number of uncorrelated signals less than the number of sensors and an unlimited supply of data, most of the preceding DF methods can uniquely and exactly locate the sources. However, the presence of too many signals, or the availability of only a finite amount of data, can cause any given DF algorithm to yield erroneous DOA estimates or to fail completely. Two performance limits are discussed here. The first limit is the largest number of signals that can be present such that the DOAs can be estimated uniquely, and the second limit is the Cramér–Rao lower bound on the RMSE of the DOA estimates.

6.1. Uniqueness

In Wax and Ziskind (1989) it is shown that certain conditions on the array manifold, the number of sensors, the number of signals, and the rank of the autocorrelation matrix of the signals determine whether or not the DOAs of the signals can be estimated uniquely. Depending on the strength of the conditions, uniqueness can be either guaranteed for every possible batch of received data or assured with probability one. Both cases require that the array manifold be known and that the array manifold vectors corresponding to M distinct DOAs be linearly independent for all possible choices of those DOAs.

The stronger condition states that uniqueness is guaranteed if the number L of signals is less than the average of the number M of sensors and the rank of the signal autocorrelation matrix:

$$L < \frac{M + \operatorname{rank}\{\boldsymbol{R}_{ss}\}}{2} . \tag{73}$$

This result is proven by showing that (73) implies that

$$A(\Theta)[s(0) \quad \cdots \quad s(N-1)] \neq A(\hat{\Theta})[\hat{s}(0) \quad \cdots \quad \hat{s}(N-1)] \tag{74}$$

for all $\hat{\Theta} \neq \Theta$ regardless of how $\{\hat{s}(n)\}$ is chosen. Consequently, the result applies to every possible choice of distinct angles $\theta_1, \ldots, \theta_L$ and every possible $\{s(n)\}$ for which rank $\{\boldsymbol{R}_{ss}\}$ has the desired value. For example, if the signals

are uncorrelated, then rank $\{\boldsymbol{R}_{ss}\} = L$ and (73) merely states the familiar condition $L < M$. However, the effect of correlated signals is to reduce the rank of $\boldsymbol{R}_{ss}$ and consequently to reduce the maximum number of sources that can be localized uniquely. For example, if all signals are fully correlated then rank $\{\boldsymbol{R}_{ss}\} = 1$, implying that the number of uniquely localizable sources is $L < (M+1)/2$.

Since the strong condition (73) might seem overly restrictive in the presence of multipath, it is fortunate that, if one is willing to accept uniqueness with probability one rather than guaranteed uniqueness, the following weaker condition is sufficient,

$$L < \frac{2\,\text{rank}\{\boldsymbol{R}_{ss}\}}{2\,\text{rank}\{\boldsymbol{R}_{ss}\}+1}\,M\,. \tag{75}$$

This condition also reduces to $L < M$ for uncorrelated signals, but it reduces to $L < \frac{2}{3}M$ when all of the signals are fully correlated.

Although these results say nothing of the accuracy of the DOA estimates in the presence of noise, they do imply that ambiguous solutions obtained while using any given batch of received data ought not to occur. However, the approach used in Wax and Ziskind (1989) to derive the results is based on the assumption that no extra knowledge of the signals is available. For example, if it were known that only a subset of the signals, say $\tilde{s}(n)$, were correlated with a known waveform (e.g., a short training signal inserted into the signals at periodic intervals before transmission), then that knowledge could be exploited to find the DOAs of only that subset of signals. In such a case it might be shown that rank $\{\boldsymbol{R}_{ss}\}$ in (73) and (75) could be replaced by rank $\{\boldsymbol{R}_{\tilde{s}\tilde{s}}\}$ and L by $\tilde{L}$.

A similar result is obtained in Schell (1990) and Schell and Gardner (1990c) for the case in which a subset of the signals exhibits spectral correlation with cycle frequency α. Specifically, it is shown for the Cyclic Least Squares algorithm that the condition

$$L_\alpha < \frac{M + \text{rank}\{\boldsymbol{R}_{ss}^{\alpha}\}}{2}\,, \tag{76}$$

guarantees that the signals can be resolved given sufficient data, where L_α is the number of signals exhibiting cyclostationarity with cycle frequency α. Two benefits of signal selectivity can make this condition on L_α much more easily satisfied than the condition on L. First, L_α is always less than or equal to the total number of signals. Second, although perfect cyclic correlation among the L_α *desired* signals reduces the rank of $\boldsymbol{R}_{ss}^{\alpha}$ (and typically results from perfect noncyclic correlation among them and hence is accompanied by a reduction in the rank of $\boldsymbol{R}_{ss}$), perfect correlation among the $L - L_\alpha$ *undesired* signals reduces the rank of $\boldsymbol{R}_{ss}$ but not of $\boldsymbol{R}_{ss}^{\alpha}$. That is, the number of sensors required and, conversely, the number of desired signals that can be accommodated are determined only by the properties of the desired signals.

6.2. Cramér–Rao lower bound

Given an array receiving signals from sources in fixed locations, how accurate is a particular method and what is the best possible accuracy? Or, more specifically, if a DF experiment is performed many times, what is the average squared error and what is a meaningful lower bound on it? The answers to these questions can be useful in choosing a DF method (e.g., by choosing the one with the smallest error) and in determining whether to continue the search for more accurate methods. Unfortunately, these questions are too vague to be answered precisely in general because the statistical properties of the signals (e.g., stationarity and whiteness) directly affect the statistical properties of the DOA estimates.

However, one can get a flavor for the performance of various methods by assuming that the noise is stationary, zero-mean, and Gaussian, and either that the signals themselves are, too, or that they are simply unknown. The former case is sometimes referred to as the stochastic or unconditional Gaussian signal model, whereas the latter is sometimes referred to as the deterministic or conditional Gaussian signal model. In either case, the probability density function of the received data is a Gaussian density function. Aside from the relative ease of working with the Gaussian density (as compared with most other densities), the fact that most DF methods use only second-order statistics (i.e., auto- and cross-correlation measurements) matches well with the property of zero-mean Gaussian signals that they are completely characterized by their second-order probabilistic parameters (ideal correlations). Furthermore, a recent result in Ottersten and Ljung (1989) shows that the stochastic ML DOA estimator for zero-mean independent stationary Gaussian signals when applied to any zero-mean uncorrelated stationary signals yields the same asymptotic (as the number N of data samples approaches infinity) variance regardless of the distribution of the signals.

Although the MSE of the DOA estimates obtained by specific DF methods is certainly of interest, the details of the analytical methods and results are quite diverse (even within the WSF framework) and, if presented here, would detract from the heuristic nature of this chapter. Examples of specific analyses are given in Viberg and Ottersten (1991), Ottersten and Ljung (1989), Barabell et al. (1984), Kaveh and Barabell (1986), Schmidt (1981), Stoica and Nehorai (1989), Wang and Kaveh (1986), Kesler and Shahmirian (1988), and Porat and Friedlander (1988). Instead, the topic of interest here is that of finding a lower bound on the MSE of any unbiased DOA estimator, and the particular bound discussed here is the Cramér–Rao lower bound (CRLB). Two cases are considered, yielding the stochastic or unconditional CRLB and the deterministic or conditional CRLB.

Derivations of the stochastic CRLB for the narrowband DF problem appear in Porat and Friedlander (1988), Barabell et al. (1984), Schmidt (1981), although the one in Bangs (1971) is the first. The derivation is not repeated here. Given the ideal spatial autocorrelation matrix $\boldsymbol{R}_{xx}$ of the received data,

and assuming that the noise is spatially white ($\boldsymbol{R}_{ii} = \sigma^2 \boldsymbol{I}$), the end result is that the covariance matrix of any unbiased estimator $\hat{\Theta}$ is greater than the inverse of the Fisher information matrix $\boldsymbol{J}$,

$$\operatorname{cov}\{\hat{\Theta}\} \geq \boldsymbol{J}^{-1}(\Theta)\,, \tag{77}$$

in the sense that $\operatorname{var}\{\boldsymbol{v}^{\mathrm{T}}\hat{\Theta}\} = \boldsymbol{v}^{\mathrm{T}} \operatorname{cov}\{\hat{\Theta}\}\boldsymbol{v} \geq \boldsymbol{v}^{\mathrm{T}}\boldsymbol{J}^{-1}(\Theta)\boldsymbol{v}$ for all vectors $\boldsymbol{v}$, where the elements of $\boldsymbol{J}$ are given by

$$\begin{aligned} J_{ij} &= N \operatorname{tr}\left\{ \boldsymbol{R}_{xx}^{-1} \frac{\partial \boldsymbol{R}_{xx}}{\partial \theta_i} \boldsymbol{R}_{xx}^{-1} \frac{\partial \boldsymbol{R}_{xx}}{\partial \theta_j} \right\} \\ &= -N \operatorname{tr}\left\{ \frac{\partial \boldsymbol{R}_{xx}^{-1}}{\partial \theta_i} \frac{\partial \boldsymbol{R}_{xx}}{\partial \theta_j} \right\}. \end{aligned} \tag{78}$$

In this context the unknown parameters θ_i are not restricted to be angles of arrival. In fact, to obtain a meaningful lower bound for most applications, the signal powers, cross-correlations among the signals, and noise powers (and perhaps the cross-correlations among the noises at different sensors), must also be considered as unknown parameters.

One intuitive aspect of the result is that the CRLB decreases as $1/N$, so that doubling the amount of collected data halves the minimum MSE. Also, in the special case of one signal arriving at the array and θ being the only unknown parameter, the covariance matrix in (77) is simply the MSE of any DOA estimate. Other than these two observations, the general intractability of proceeding analytically makes it difficult to obtain much more understanding of a heuristic nature.

Thus, due to the difficulties in obtaining more useful results analytically, a numerical example is presented. Consider the five-element ULA used in the previous examples in this chapter. Two signals arrive with equal power 10 dB above that of additive white Gaussian noise, and the DOA of one signal is held fixed at 0 degrees while the DOA of the other signal is allowed to vary. The number of data samples is assumed to be 100, although the results for any desired number N of samples can be obtained by vertically shifting the curve by $\log_{10} \sqrt{N/100}$. Intuitively, we expect the RMSE to increase as the angular separation decreases because the two distinct arrivals begin to appear as one. This expectation is confirmed by the results of the numerical evaluation displayed in Figure 21. Judging from the figure by the location of the intersection of the dotted line (RMSE = angular separation) with the solid line (the CRLB), the CRLB is greater than the angular separation itself when the angular separation is less than approximately 2 degrees. This point can be taken to be the resolution threshold for the given SNR and number of data samples. Thus, regardless of the method to be applied to this environment, more data (or higher SNR) is required to operate when the separation is less than 2 degrees. Notice that as the DOA of the second signal approaches the

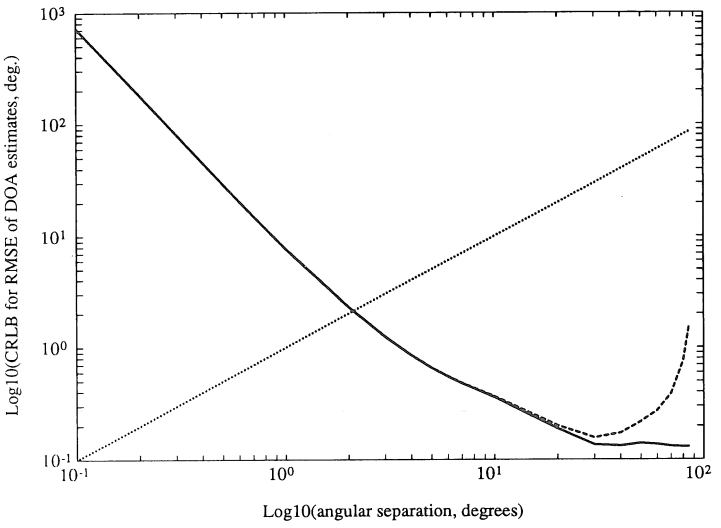

Fig. 21. CRLB for the DOA estimates of signals having 10 dB SNR and arriving from 0 degrees (solid line) and θ_2 degrees (dashed line), plotted as a function of θ_2. (In the region to the left of the intersection of these curves with the dotted line, the CRLB is greater than the angular separation, whereas the CRLB is less than the angular separation in the region to the right of the intersection.)

end-fire position ($\theta_2 \rightarrow 90°$), the CRLB rises dramatically because the array response vector becomes insensitive to small changes in DOA ($\mathrm{d}\boldsymbol{a}(\theta_2)/\mathrm{d}\theta_2 \rightarrow 0$).

It should be noted that the resolution threshold adopted here is a severe one that is intended to distinguish the region of excellent performance from the regions of possibly acceptable performance and definitely unacceptable performance. Also, this aspect of resolving signals is distinct from the typically necessary prerequisite of correctly detecting the number of signals present. Only when the two are taken together can a complete characterization of performance that is applicable to most problems be made.

More recently derived than the stochastic CRLB is the deterministic CRLB, presented in Stoica and Nehorai (1989). Unlike the stochastic CRLB, the deterministic CRLB depends on the signal waveforms rather than their probabilistic model. However, when the number N of data samples is sufficiently large, the deterministic CRLB can be expressed in terms of the limit ($N \rightarrow \infty$) autocorrelation matrix of the signals,

$$\mathrm{cov}\{\hat{\Theta}\} \geqslant \frac{\sigma^2}{2N} \left[\mathrm{Re}\{[\boldsymbol{D}^{\mathrm{H}}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}}(\Theta))\boldsymbol{D}] \odot \boldsymbol{R}_{ss}^{\mathrm{T}}\}\right]^{-1}, \tag{79}$$

where $\odot$ denotes the element-wise matrix product, and the matrix $\boldsymbol{D}$ (not to be confused with the matrix $\boldsymbol{D}$ in (47)) is given by

$$\boldsymbol{D} = \left[\frac{\mathrm{d}\boldsymbol{a}(\theta_1)}{\mathrm{d}\theta_1} \quad \cdots \quad \frac{\mathrm{d}\boldsymbol{a}(\theta_L)}{\mathrm{d}\theta_L}\right]. \tag{80}$$

Like the expressions for the stochastic CRLB, this expression does not yield much insight into the behavior of the CRLB as a function of angular separation between sources, SNR, or number of sensors. However, for the special case of a uniform linear array, if the number M of sensors and the number N of time samples are sufficiently large, then the CRLB can be expressed simply as

$$\operatorname{cov}\{\hat{\Theta}\} \geqslant \frac{6}{M^3 N} \begin{bmatrix} 1/\mathrm{SNR}_1 & & 0 \\ & \ddots & \\ 0 & & 1/\mathrm{SNR}_L \end{bmatrix}, \tag{81}$$

where SNR_l is the signal-to-noise ratio for the l-th signal. This expression shows that the deterministic CRLB is reduced drastically as the number M of sensors increases and that increasing SNR or collect time N also decreases the CRLB in inverse proportion. Although it seems strange that (81) is independent of the angular separations, it is a direct result of the requirement that M be sufficiently large. However, for practical numbers of sensors, analytical evaluation of the deterministic CRLB is too complex to offer much intuitively pleasing information.

Although the preceding two versions of the CRLB can provide useful lower bounds for the RMSE of estimators operating on stationary data, they neglect the reduction in RMSE that is potentially obtainable when the signals are cyclostationary. In Schell (1990) and Schell and Gardner (1992a) it is shown that the CRLB for the DOAs of cyclostationary signals can be orders of magnitude less than the CRLB for the DOAs of stationary signals, even if the cycle frequencies (and SNRs and other parameters) of the signals are unknown, provided that not all signals have the same cycle frequencies. This potential performance increase is based on the fact (mentioned in Section 5.5.1) that simultaneously resolving in the cycle-frequency domain (even though the signals are spectrally overlapping) and the spatial domain can require much less data than resolving only in the spatial domain alone.

7. Departures from ideality

Given the difficulty in understanding the physical behavior of most DF methods, the bulk of the discussion in the previous sections of this chapter is devoted to the ideal case in which all quantities processed by the methods are known exactly and all operating assumptions are satisfied. As seen from the preceding discussion, there are substantial differences in performance and applicability among the methods without introducing other dependencies. However, in practice, virtually all of the departures from ideality discussed in this section can occur. Effects due to finite averaging-time, calibration errors, unknown number of signals, the presence of more signals than sensors, unknown interference-and-noise spatial autocorrelation matrix, and wideband signals can all conspire to degrade performance. For each of these sources of

error there are both analyses of existing methods in the presence of error and designs of new methods for reducing the effects of error, although much more work remains to be done. In this section, some of the most important departures from ideality and potential methods for accommodating them are briefly surveyed.

7.1. Finite averaging time

The DF methods discussed in this chapter are explained in terms of their behavior when given ideal correlation matrices in order to concentrate on the simplest interpretations of the methods. In practice, the correlation matrices must be measured from the received data. The consequences of using these nonideal values include possible bias (average error) in the estimates, variability in the estimates from trial to trial for the same environment, incorrect detection of the number of signals, failure to resolve the signals even when given the correct number of them, and errors resulting from dynamic or transient environments. Usually it is assumed that the physical model (source locations, array geometry, signal power, etc.) is stationary. Thus, as the number of data samples increases without bound, the measured correlation matrices are assumed to converge to the ideal matrices. However, depending on the DF method, this convergence does not imply convergence of the DOA estimates to the true values, as noted in the numerical evaluations presented in the preceding sections of this chapter.

To investigate performance for a finite number of data samples, most papers on DF methods contain the results of Monte Carlo simulations in which the random number seeds used to generate the signals and noise vary from trial to trial. From these trials are calculated the bias and RMSE of the DOA estimates, but these results are typically applicable only to the specific signal environment being simulated. In contrast, analytical results can be used for performance evaluation in a variety of environments, their explicit functional dependence can reveal relationships between performance and environmental parameters not previously known, and they facilitate the comparison of performance obtained for different DF methods. However, due to the nonlinear dependence of the DOA estimates on both the data and the correlation matrices measured from the data, analysis of bias and RMSE for a finite amount of data can be difficult at best or intractable. For example, analysis of recent methods based on eigenvalue problems and subspace-fitting typically involves asymptotic (as the number of data samples becomes large) probabilistic behavior of both the correlation matrices in question and the eigenvectors or singular vectors of those matrices. Although results on such probabilistic behavior are available, the difficulty in obtaining very specific results (such as the number of data samples needed to resolve two signals) remains. Analytical investigations of more complicated environments (involving more than one or two signals) are typically avoided in favor of Monte Carlo simulations.

Nonetheless, a number of useful analyses are presented in the literature,

including results on the SNR needed for MUSIC and the minimum-norm methods to resolve two signals (Kaveh and Barabell, 1986), and the bias and RMSE of DOA estimates obtained by MUSIC (Porat and Friedlander, 1988; Stoica and Nehorai, 1989), ESPRIT (Ottersten et al., 1991), ML (Stoica and Nehorai, 1989), WSF (Viberg and Ottersten, 1991) and Cyclic MUSIC (Schell, 1993). Other useful performance analyses and results of Monte Carlo simulations are presented in Wang and Kaveh (1986), Barabell et al. (1984), and Jeffries and Farrier (1985).

7.2. *Calibration errors*

Throughout most of the literature on DF it is assumed that the array manifold is known precisely. However, the effective array manifold can differ from the assumed or measured one due to measurement error during calibration, sensor characteristics changing with time and temperature, array geometry being perturbed by vibration and mechanical deformation (e.g., in arrays that are dismantled and reassembled), and more complex effects such as mutual coupling among sensors. If the corresponding perturbations in the array manifold are small, then many DF algorithms can still provide useful results for a single signal or for widely spaced signals, although the estimates will likely be biased, but the ability to resolve two closely spaced sources can be severely reduced. Thus, it is important both to quantify the sensitivity of algorithms to calibration error and to formulate algorithms that either are tolerant to such error or attempt to adaptively self-calibrate the array.

In some respects the analyses of DOA estimation error in the presence of calibration errors and finite-time effects are similar because they can both be interpreted as involving perturbations to $\boldsymbol{R}_{xx}$. In fact, the general perturbation equations from Rao and Hari (1989) on asymptotic analysis of root-MUSIC for finite time are used in Swindlehurst et al. (1989) to obtain results for DOA estimation error of root-MUSIC for calibration error. A more general expression obtained there which is applicable to MUSIC and root-MUSIC is

$$\hat{\theta}_i - \theta_i \approx -\frac{\mathrm{Re}[\mathbf{d}^{\mathrm{H}}(\theta_i)\boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^{\mathrm{H}}\tilde{\boldsymbol{a}}(\theta_i)]}{\mathbf{d}^{\mathrm{H}}(\theta_i)\boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^{\mathrm{H}}\mathbf{d}(\theta_i)}\,, \tag{82}$$

where $\tilde{\boldsymbol{a}}(\theta) = \boldsymbol{a}_{\mathrm{true}}(\theta) - \boldsymbol{a}_{\mathrm{assumed}}(\theta)$ is the perturbation which is assumed to be small, $\mathbf{d}(\theta)$ is the partial derivative of $\boldsymbol{a}(\theta)$, $\mathbf{d}(\theta) = \partial \boldsymbol{a}(\theta)/\partial\theta$, and $\boldsymbol{E}_{\mathrm{n}}$ is the matrix of noise eigenvectors of $\boldsymbol{R}_{xx}$ or the null eigenvectors of $\boldsymbol{R}_{xx}^{\alpha}$. Specific numerical evaluations in Friedlander (1990) and Swindlehurst et al. (1989) show that calibration error can be rather large (e.g., random phase errors on the order of 30 degrees) without causing significant degradation in the performance of MUSIC when only one signal is present, but that even small calibration errors (e.g., a phase error of 2 degrees) can prevent MUSIC from resolving two signals.

In light of this last result it appears particularly important to formulate DF algorithms that can adaptively adjust their calibration. Most such algorithms proposed to date, including those in Weiss and Friedlander (1989), Rockah and Schultheiss (1987a,b), Weiss et al. (1988), and Friedlander and Weiss (1988), are applicable to restricted types of calibration error. One type of error comes from gain and phase errors that are independent of DOA. In this case the effective array response vector $\boldsymbol{a}(\theta) + \tilde{\boldsymbol{a}}(\theta)$ is related to the assumed vector $\boldsymbol{a}(\theta)$ by a diagonal matrix Γ,

$$\boldsymbol{a}(\theta) + \tilde{\boldsymbol{a}}(\theta) = \Gamma \boldsymbol{a}(\theta) , \tag{83}$$

so in addition to estimating the DOAs, the DF algorithm must also estimate Γ. Another type of error results from errors in sensor locations. In this case the sensor locations must be estimated and then substituted into the analytical expression for the array response vectors in order to obtain the estimated array response vectors. For these two types of error, a common approach is to iterate between estimating DOAs using the current corrections for calibration error and estimating corrections for calibration error using the current DOA estimates. For example, an ML version of this approach is shown with Monte Carlo simulations in Weiss and Friedlander (1989) to converge quickly (e.g., 5 to 10 iterations) for sensor location perturbations of 34% of the element spacing in a 6-element uniform circular array. This self-calibration method reduced the DOA errors from 8 degrees to less than 1 degree. Despite the encouraging nature of these results, the more general problem of combined sensor and gain/phase errors and the problem of gain and phase errors that vary with DOA remain.

An alternate approach to addressing the calibration problem is to explore algorithms that are inherently less sensitive to calibration errors. For example, the ESPRIT algorithm (cf. Section 5.2) uses only the distance between the two subarrays, thereby greatly reducing the dependence on proper calibration. However, if the two subarrays are not identical due to physical perturbations of the array geometry or component drift in the sensors, then performance can be degraded. As another example, the methods that exploit cyclostationarity (cf. Section 5.5) can yield better performance in the presence of calibration errors than their conventional counterparts by reducing the number of signals to be processed. For example, consider only one signal present that is cyclostationary with a particular cycle frequency α and other closely-spaced signals that are present and do not have that cycle frequency. If the calibration error is great enough that closely-spaced signals cannot be spatially resolved, then a conventional method will fail. In contrast, the signal-selective DOA estimate obtained by a cyclostationarity-exploiting method will be biased, but the amount of such bias can be small. Since the conventional methods must spatially resolve all of the signals, they are more susceptible to calibration error than are the cyclostationarity-exploiting methods which pre-select one signal (or a relatively

small subset of signals) in the cycle-frequency domain independently of any calibration error. This is not to say that the cyclostationarity-exploiting methods are not affected by calibration error, but that they can be much less affected than their conventional counterparts because they typically operate on fewer signals at a time (cf. Schell, 1992b).

7.3. Unknown number of signals

Throughout this chapter it is assumed that the number of signals impinging on the array is known. Of course, if the ideal spatial autocorrelation matrix $\boldsymbol{R}_{xx}$ of the data is known, the ideal autocorrelation matrix of the signals $\boldsymbol{R}_{ss}$ has full rank, and the interference-and-noise autocorrelation matrix $\boldsymbol{R}_{ii}$ is known, then the number of signals can be found by subtracting the multiplicity of the smallest eigenvalue of $\boldsymbol{R}_{xx}$ from the number of sensors (cf. equation (26)).

In practice, the number of signals is often unknown and only a finite amount of data is available, preventing the use of the preceding technique. Although a sequence of hypothesis tests can be applied to the eigenvalues of the estimated autocorrelation matrix (Bartlett, 1954; Lawley, 1956), this procedure requires that some subjective judgement be made to choose the threshold value used in the tests. More specifically, the log-likelihood of the data (assumed to be Gaussian) is evaluated for increasing trial value L of the number of signals. Then, since this function is nondecreasing, some threshold is needed to determine whether the likelihood is high enough for the trial value to be taken as the final estimate. The subjectivity in choosing this threshold, then, is a weakness of the method.

In contrast, the type A Information Criterion (AIC) of Akaike (1973), the minimum description length (MDL) criterion (Wax and Kailath, 1985), and the efficient detection criterion (EDC) (Zhao et al., 1986) avoid this problem by employing a more sophisticated threshold that is chosen according to a statistical motivation specific to each method. These methods estimate the number of signals as the value of L that minimizes the function

$$f(L, p, N) = N(M - L)\log[a(L)/g(L)] + p(L, N)\,, \tag{84}$$

where $a(L)$ and $g(L)$ are the arithmetic and geometric means, respectively, of the $M - L$ smallest eigenvalues of $\boldsymbol{R}_{xx}$, and $p(L, N)$ is a penalty function that is different for each of the AIC, MDL, and EDC methods. The purpose of the penalty function is to offset the first term (which is recognizable as the negative of the log-likelihood and is a nonincreasing function of L) and thereby induce a minimum in the function. Alternatively, the penalty function can be interpreted as a sophisticated threshold function. A different alternative considered in Chen et al. (1991) involves a statistical method for predicting the best thresholds and has been shown to outperform the AIC and MDL methods.

For the cyclic DF methods such as Cyclic MUSIC and Cyclic Least Squares,

the aforementioned detection techniques do not apply. They do not apply directly because the number L_α of only the signals having cycle frequency α is desired, and they do not apply indirectly because the statistical behavior of the singular values of the cyclic autocorrelation matrix is much different than that of the eigenvalues of the conventional autocorrelation matrix. A new detection scheme that combines a penalty function (so no threshold needs to be chosen) with some results in multivariate statistics regarding common factor analysis (which is concerned with estimating signal components that are common to two different data sets) is presented in Schell and Gardner (1990a) and is shown there to perform satisfactorily. More recent methods are considered in Schell (1993).

7.4. *Number of signals > number of sensors*

In cluttered signal environments such as tactical surveillance, air traffic control, and communication networks, the number of signals impinging on the array can exceed the number of sensors. Given the cost or impracticality of increasing the number of sensors in some applications, an attractive alternative is to use a DF algorithm that can operate properly in such cluttered environments.

The possibility of such an alternative follows from a recent result in Wax (1992), where it is shown that a number $L > M$ of signals can be uniquely localized if the signals satisfy certain constraints. For example, it is shown that up to $2M - 2$ uncorrelated signals having unknown but constant amplitudes can almost always be uniquely localized. A more powerful result is shown under the constraints that $M > 2$ sensors are used and that the signals are uncorrelated complex exponentials (e.g., each signal $s_l(n)$ has the form $s_l(n) = a_l \exp(j2\pi f_l n + \phi_l)$). Then any number L of signals can almost always be uniquely localized provided that L time samples are taken.

Another alternative reduces the inherent redundancy in a ULA without reducing the number of resolvable signals or the aperture (spatial extent) of the array. The primary example of this is the minimum redundancy linear array (Moffet, 1968) which is based on the representation of all pairwise differences among the integers $1, 2, \ldots, M$ by the pairwise differences in a set containing fewer than M integers. For example, the integers $1, 2, \ldots, 24$ yield pairwise differences $0, 1, \ldots, 23$, and many pairwise differences are the same (e.g., $24 - 10 = 23 - 9 = \cdots = 15 - 1 = 14$). However, the set of integers 1, 2, 3, 12, 16, 19, 22, 24 also yields all differences between 0 and 23, but this set contains only 8 elements. This property can be exploited to reconstruct the spatial autocorrelation matrix of the 24-element ULA from the autocorrelation matrix of the minimum redundancy 8-element array because redundant differences in the set of integers $1, \ldots, M$ correspond to redundant cross-correlations between sensors. Since (for infinite time-averaging) $\boldsymbol{R}_{xx}$ for a ULA is a Toeplitz matrix (i.e., $R_{x_i x_j} = R_{x_{i+k} x_{j+k}}$), an entire sub-diagonal or super-diagonal of $\boldsymbol{R}_{xx}$ can be reconstructed from a single element in it. This reconstructed matrix, obtained from only 8 sensors, can then be processed by a DF algorithm such as

MUSIC to resolve up to 23 signals. Larger gains can be achieved by considering larger values of M above.

Yet another alternative is to apply a signal-selective DF algorithm such as Cyclic MUSIC or Cyclic Least Squares (cf. Section 5.5). For these algorithms, the only requirement on the number of signals is that the number L_{α_1} of signals with the cycle frequency α_1 be less than the number of sensors (considering only the case of full rank $\boldsymbol{R}_{ss}^{\alpha_1}$ for simplicity here). Consequently, if there exist $L_{\alpha_2} < M$ signals with cycle frequency α_2 then the DOAs of those signals can also be estimated simply by changing the value of α from α_1 to α_2. This situation can be repeated for a number of distinct cycle frequencies limited only by measurement noise due to finite time-averaging and computational precision effects so that, in principle, the DOAs of a virtually unlimited number of temporally and spectrally overlapping signals can be estimated. Also, when each signal of interest has a unique cycle frequency, only a two-sensor array is needed, although some reduction in RMSE can be obtained by adding a third sensor to substantially increase the aperture beyond that required to avoid ambiguities for the two-element array.

7.5. *Unknown* $\boldsymbol{R}_{ii}$

The need to know the interference-plus-noise autocorrelation matrix $\boldsymbol{R}_{ii}$ can prevent many DF algorithms including MUSIC, ESPRIT, ML, and WSF, from obtaining any useful DOA estimates or can impose extra cost, added complexity (e.g., if an automatic system were somehow added to estimate $\boldsymbol{R}_{ii}$), or hardship on the operator to obtain this information, when it is indeed available. Typically, $\boldsymbol{R}_{ii}$ is assumed to be proportional to the identity matrix. However, if one sensor is particularly noisy, or if significant noise is actually received (rather than being generated by the sensors themselves), then $\boldsymbol{R}_{ii}$ can deviate significantly from the assumed value.

Covariance differencing techniques (Paulraj and Kailath, 1986; Prasad et al., 1988) sidestep this requirement under certain conditions by computing the difference between the matrix and a transformed version, where the transformation is chosen such that the noise correlation is invariant to it, thus removing the contribution of $\boldsymbol{R}_{ii}$. However, these techniques work well only when the noise does indeed satisfy this invariance property. Also, covariance differencing does not remove the need to know $\boldsymbol{R}_{ii}$ when undesired and interfering signal sources are to be included in $\boldsymbol{i}(n)$ as a means of eliminating them from consideration in the DOA estimation process.

However, cyclostationarity-exploiting DF methods such as Cyclic MUSIC and Cyclic Least Squares (cf. Section 5.5) inherently avoid the need to know $\boldsymbol{R}_{ii}$ if the interference and noise do not exhibit cyclostationarity with the cycle frequency of the signals of interest, allowing them to be applied in a broader range of environments. This is particularly important in analysis applications in which – for some cases – virtually no prior knowledge of the signal environment exists.

7.6. *Potential limitations of cyclostationarity-exploiting methods*

An obvious limitation of cyclostationarity-exploiting methods is that not all signals whose directions of arrival are to be found exhibit cyclostationarity. Although some natural signals do (e.g., due to seasonal variations), many natural signals do not. Also, although many man-made signals do exhibit cyclostationarity, some do not (e.g., analog frequency-modulated signals with large modulation index, and analog single-sideband amplitude-modulated signals).

Although the performance of methods that exploit cyclostationarity can be greatly superior to that of conventional methods in difficult environments, the performance can be worse in some environments, and knowledge of the unconventional parameters required can be difficult to obtain in some applications.

Inferior performance of the cyclostationarity-exploiting methods is likely to occur when the conventional methods are operating well within their operating limits, such as when received noise is negligible, internal noise is uncorrelated from sensor to sensor and the noise power at each sensor is known, few signals are present, and the signals are not extremely closely spaced. Examples of this behavior can be found in Schell (1990) and Schell and Gardner (1990b). One explanation for this behavior is that signal selectivity is not essential to acceptable performance in these environments and that the DOA estimates of the cyclostationarity-exploiting methods are based on a weaker (more difficult to estimate accurately) property of the signals than the property used by the conventional methods. Specifically, the additive sine wave component at frequency $\alpha \neq 0$ in the lag-product waveform $x(n)x^{H}(n-\tau)$ that is exploited by the cyclostationarity-exploiting methods is weaker (and for bandwidth-efficient communication signals can be much weaker) than the additive constant component at frequency $\alpha = 0$ exploited by the conventional methods. Consequently, more data samples can be required to obtain a reliable estimate of the amplitude and phase of the sine wave and hence of the DOAs.

Although some parameters required by the conventional methods (e.g., autocorrelation matrix of the noise, total number of signals) are not needed by the cyclostationarity-exploiting methods, knowledge of some unconventional parameters *is* needed. In particular, the cycle frequency α must be known accurately (to within about $100/N$ percent of the width of the receiver band, where N is the number of data samples) to obtain acceptable performance. Some initial work (Schell and Gardner, 1990b) on estimating α from the data indicates that excellent performance can be obtained even when α is essentially unknown. Another parameter that is required to be known is the number of signals present that have the desired cycle frequency. As with the conventional methods, the estimation of this parameter (analogous to the estimation of the total number of signals) requires a statistical test to be applied when only a finite amount of data is available. Unfortunately, using the eigenvalues or singular values of the cyclic autocorrelation matrix in the conventional detec-

tion criteria does not result in useful estimates of the number of signals having the desired cycle frequency. However, a new detection criterion presented in Schell and Gardner (1990a), which is based on existing results in multivariate analysis of testing the significance of the correlation of one data set with another, appears to yield acceptable performance (cf. Schell, 1993).

7.7. *Breakdown of narrowband approximation*

Throughout this chapter it is assumed that the signals are sufficiently narrowband that equation (7) is a close approximation. Thus, the contribution of each signal to the spatial autocorrelation matrix of the received data has rank equal to one, and the rank of the signals-only part of that matrix is equal to the number of signals (in the absence of perfectly correlated signals, pathological signal environments, and ambiguities in the array manifold). However, in many applications the bandwidths of the signals can be comparable to the center frequency of interest and/or tapped delay lines can be attached to the sensors, so the more general model that includes frequency dependence must be used instead. A more detailed discussion of the conditions under which the wideband model is needed can be found in Buckley (1987).

Perhaps the most obvious method for accommodating wideband data is simply to decompose the data into disjoint frequency bands that individually satisfy the narrowband assumption. Then, the DOA estimates obtained from applying a narrowband DF algorithm to each band must somehow be combined to yield one DOA estimate for each wideband signal. This approach is investigated in Wax et al. (1984). However, as in any non-linear estimation problem, as SNR decreases below some threshold, the errors in the individual estimates increase dramatically and prevent the final combination from being effective (cf. Van Trees, 1968, Chapter 4, Section 2). This is so because the data from the multiple narrow bands are combined incoherently in the sense that no combining is performed until after the data from each band is processed by a nonlinear processor. Specific examples of this effect in the wideband DF problem are presented in Wang and Kaveh (1985). Consequently, methods that coherently combine the data from the multiple narrow bands are needed.

One such coherent wideband DF method, broadband signal-subspace spatial-spectrum estimation (BASS-ALE) (Buckley and Griffiths, 1988), is based on a reduced rank representation of wideband signals. This representation is based on the results shown in Buckley (1987) that more than 99.99% of the received average power from a single signal is characterized by the r largest eigenvalues of the autocorrelation matrix, where

$$r = \lceil 2WT(\theta) + 1 \rceil , \tag{85}$$

$\lceil x \rceil$ represents the smallest integer greater than or equal to x, W is the

bandwidth of the received data, and $T(\theta)$ is the propagation time across the array (including time spent traveling through the delay lines on the sensors) for a signal arriving from angle θ. Two sampling criteria are also assumed to hold: (1) the sensors are spaced closely enough that

$$M \geqslant 2f_{\max}\tau(\theta)\,, \tag{86}$$

where M is the number of sensors, $f_{\max}$ is the highest frequency present in the data, and $\tau(\theta)$ is the propagation time across the physical array (not including delay lines); and (2) the delay increment in the delay-line is $1/2f_{\max}$.

Thus, the effective rank of the autocorrelation matrix is given by r in (85). A pure sine wave ($W=0$) has effective rank $r=1$, being perfectly narrowband. For nonzero W and $T(\theta)$, r is always greater than one. However, if $2WT(\theta)+1$ is sufficiently close to one then the second eigenvalue of $\boldsymbol{R}_{xx}$ is negligible, making the data effectively narrowband. And, for wider bandwidths and greater propagation times across the array, the effective rank r can be much greater than one but still be less than the dimension of the received data vector (number of sensors plus the number of delay elements per sensor). This notion of effective rank is exploited for wideband DF in Buckley and Griffiths (1988) to define the notion of signal subspace in the context of wideband signals. Specifically, it is used to show that a signal subspace and an orthogonal complement (noise subspace) do indeed exist. Then, as in narrowband MUSIC, the array response vectors that are orthogonal to that noise subspace are found, and the corresponding angles are taken to be the DOA estimates.

Another coherent wideband method, the coherent signal subspace (CSS) method described in Wang and Kaveh (1985) and Hung and Kaveh (1988), is based on focusing the data from the narrow bands onto a common signal subspace, say, the signal subspace corresponding to the center frequency of the receiver band. Since the signal subspace of each narrow band is different, the subspaces must be rotated onto a common subspace before they can be usefully combined. Then, given the autocorrelation matrix of the focused data, existing signal subspace algorithms such as MUSIC can be applied.

Yet another alternative uses the steered covariance matrix (STCM) (Krolik and Swingler, 1989), which is measured after steering delays are applied to the received data. Under the assumption that the output of the m-th sensor can be modeled as

$$x_m(n) = \sum_{l=1}^{L} s_l(n - \tau_m(\theta_l)) + i_m(n)\,, \tag{87}$$

where $\tau_m(\theta)$ is the signal propagation delay to the m-th sensor (relative to the coordinate origin of the array) for a signal from angle θ, then applying delays $\tau_m(\theta)$, $m=1,\ldots,M$, to the sensor outputs aligns in time the signal component from angle θ. This pre-steering enhances the contribution of the signal from direction θ to the autocorrelation matrix, similar to the way in which the

focusing matrices in the CSS align the data from the multiple narrow bands before the autocorrelation matrices are computed. Notice that the STCM method maps the propagation time across the array to zero, in contrast to the CSS method which maps the bandwidth of the signal to zero. Both methods have the effect of mapping the time-bandwidth product to zero for a particular look direction (or set of directions). Then, for each direction within the analysis region, the autocorrelation matrix of the pre-steered data (the STCM) for that look direction is computed and then processed using modified versions of existing narrowband methods.

8. Summary

Sensor arrays can be used to obtain high-resolution estimates of the directions of arrival of propagating signals. Several recent direction-finding methods are described in terms of how they use spatial filters (linear combiners) to enhance the contribution of some signals and/or attenuate others in the process of estimating the directions of arrival. It is shown that this physically motivated interpretation can be used to derive some of the methods, explain their behavior in different signal environments, and lead smoothly to the more abstract framework of subspace fitting which is prevalent in the research literature. Included in this discussion are descriptions of recent advances in unifying apparently diverse methods and exploiting cyclostationarity properties of signals to obtain better performance. Statistical bounds on the errors of the estimates are briefly described, and several departures from ideality are considered. Future research is likely to continue to focus on accommodating multipath and jamming signals, wideband signals, and array calibration errors, and on achieving even better performance at less computational expense.

References

Agee, B., S. Schell and W. Gardner (1990). Spectral self-coherence restoral: A new approach to blind adaptive signal extraction. *Proc. IEEE* **78**(4), 753–767. Special Issue on Multidimensional Signal Processing.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: *Proc. 2nd Internat. Symp. Inform. Theory*. Suppl., *Problems of Control and Inform. Theory*, 267–281.

Ball, D. (1989). *Soviet Signals Intelligence* (*SIGINT*) **47**. Strategic and Defence Studies Centre of Australian National University, Canberra, Australia.

Bangs, W. (1971). Array processing with generalized beamformers. Ph.D. dissertation, Yale University, New Haven, CT.

Barabell, A., J. Capon, D. DeLong, J. Johnson and K. Senne (1984). Performance comparison of superresolution array processing algorithms. Proj. Rep. TST-72, MIT Lincoln Labs, Lexington, MA.

Bartlett, M. (1954). A note on the multiplying factors for various χ^2 approximations. *J. Roy. Statist. Soc.* **16**, 296–298.

Bienvenu, G. and L. Kopp (1980). Adaptivity to background noise spatial coherence for high resolution passive methods. In: *Proc. IEEE Internat. Conf. Acoust., Speech, Signal Processing*. Denver, Colorado, 307–310.

Bresler, Y. and A. Macovski (1986). Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Trans. Acoust. Speech Signal Process.* **34**(5), 1081–1089.

Buckley, K. (1987). Spatial/spectral filtering with linearly constrained minimum variance beamformers. *IEEE Trans. Acoust. Speech Signal Process.* **35**(3), 249–266.

Buckley, K. and L. Griffiths (1988). Broadband signal-subspace spatial-spectrum (BASS-ALE) estimation for sensor array processing. *IEEE Trans. Acoust. Speech Signal Process.* **36**(7), 953–964.

Burg, J. P. (1967). Maximum entropy spectral analysis. In: *Proc. 37th Ann. Internat. SEG Meeting*, Oklahoma City, OK.

Burg, J. (1972). The relationship between maximum entropy spectra and maximum likelihood spectra. *Geophysics* **37**(2), 375–376.

Capon, J. (1969). High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE* **57**(8), 1408–1418.

Capon, J. (1979). Maximum-likelihood spectral estimation. In: S. Haykin, ed., *Nonlinear Methods of Spectral Analysis*. Springer, New York 155–179.

Carter, G., ed. (1981). Special issue on time delay estimation. *IEEE Trans. Acoust. Speech Signal Process.* **29**(3, Part II).

Chen, C.-K. and W. Gardner (1992). Signal-selective time-difference-of-arrival estimation for passive location of manmade signal sources in highly corruptive environments. Part II: Algorithms and performance. *IEEE Trans. Signal Process.* **40**(5), 1185–1197.

Chen, W., K. Wong and J. Reilly (1991). Detection of the number of signals: A predicted eigen-threshold approach. *IEEE Trans. Signal Process.* **39**(5), 1088–1098.

Compton Jr, R. (1988). *Adaptive Antennas*. Prentice-Hall, Englewood Cliffs, NJ.

Evans, J., J. Johnson and D. Sun (1982). Applications of advanced signal processing techniques to angle of arrival estimation in ATC navigation and surveillance system. Tech. Rep. 582, Lincoln Laboratory, MIT, Lexington, MA.

Friedlander, B. (1990). A sensitivity analysis of the MUSIC algorithm. *IEEE Trans. Acoust. Speech Signal Process.* **38**(10), 1740–1751.

Friedlander, B. and A. Weiss (1988). Eigenstructure methods for direction finding with sensor gain and phase uncertainty. In: *Proc. IEEE Internat. Conf. Acoust., Speech, Signal Processing*, New York, 2681–2684.

Gabriel, W., ed. (1976). Special issue on adaptive antennas. *IEEE Trans. Antennas and Propagation* **24**(5).

Gabriel, W., ed. (1980). Spectral analysis and adaptive array superresolution techniques. *Proc. IEEE* **68**(6), 654–666.

Gabriel, W. (1986). Special issue on adaptive processing antenna systems. *IEEE Trans. Antennas and Propagation* **34**(3).

Gardner, W. (1987a). Spectral correlation of modulated signals: Part I – Analog modulation. *IEEE Trans. Comm.* **35**(6), 584–594.

Gardner, W. (1987b). *Statistical Spectral Analysis: A Nonprobabilistic Theory*. Prentice-Hall, Englewood Cliffs, NJ.

Gardner, W. (1988). Simplification of MUSIC and ESPRIT by exploitation of cyclostationarity. *Proc. IEEE* **76**(7), 845–847.

Gardner, W. (1989). *Introduction to Random Processes with Applications to Signals and Systems*. 2nd ed., McGraw-Hill, New York.

Gardner, W. (1991). Exploitation of spectral redundancy in cyclostationary signals. *IEEE Signal Process. Mag.* **8**(2), 14–37.

Gardner, W., B. Agee and S. Schell (1988). Self-coherence restoring signal extraction apparatus and method. Patent pending.

Gardner, W. and W. Brown (1991). Fraction-of-time probability for time series that exhibit cyclostationarity. *Signal Process. J. EURASIP* **23**(3), 273–292.

Gardner, W., W. Brown III and C.-K. Chen (1987). Spectral correlation of modulated signals: Part II – Digital modulation. *IEEE Trans. Comm.* **35**(6), 595–601.

Gardner, W. and C.-K. Chen (1992). Signal-selective time-difference-of-arrival estimation for passive location of manmade signal sources in highly corruptive environments. Part I: Theory and method. *IEEE Trans. Signal Process.* **40**(5), 1168–1184.

Golub, G. and C. van Loan (1989). *Matrix Computations*. 2nd ed., Johns Hopkins Univ. Press, Baltimore, MD.

Haber, F. and M. Zoltowski (1986). Spatial spectrum estimation in a coherent signal environment using an array in motion. *IEEE Trans. Antennas and Propagation* **34**(3), 301–310.

Hayakawa, M., K. Hattori, S. Shimakura, M. Parrot and F. Lefeuvre (1990). Direction finding of chorus emission in the outer magnetosphere and their generation and propagation. *Planetary Space Sci.* **38**(1), 135.

Haykin, S., ed. (1980). *Array Processing: Applications to Radar*. Benchmark Papers in Electrical Engineering and Computer Science, Vol. 22. Dowden, Hutchinson and Ross, Stroudsburg, PA.

Haykin, S. (1985). Radar array processing for angle of arrival estimation. In: S. Haykin, ed., *Array Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, ch. 4, 194–292.

Haykin, S. and J. Cadzow, eds. (1982). Special issue on spectral estimation. *Proc. IEEE*, **70**(9).

Hung, H. and M. Kaveh (1988). Focussing matrices for coherent signal-subspace processing. *IEEE Trans. Acoust. Speech Signal Process.* **36**(8), 1272–1281.

Jeffries, D. and D. Farrier (1985). Asymptotic results for eigenvector methods. *Proc. IEE-F* **132**(7), 589–594.

Johnson, D. (1982). The application of spectral estimation methods to bearing estimation problems. *Proc. IEEE* **70**(9), 1018–1028.

Johnson, R. and D. Wichern (1988). *Applied Multivariate Statistical Analysis*. 2nd ed., Prentice-Hall, Englewood Cliffs, NJ.

Justice, J. (1985). Array processing in exploration seismology. In: Haykin, S., ed., *Array Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, ch. 2, 6–114.

Kaveh, M. and A. Barabell (1986). The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise. *IEEE Trans. Acoust. Speech Signal Process.* **34**(2), 331–341.

Kesler, S. (1982). Generalized Burg algorithm for beamforming in correlated multipath field. In: *Proc. IEEE Internat. Conf. Acoust., Speech, Signal Process.* 1481–1484.

Kesler, S. and V. Shahmirian (1988). Bias and resolution of the MUSIC and the modified FBLP algorithms in the presence of coherent plane waves. *IEEE Trans. Acoust. Speech Signal Process.* **36**(8), 1351–1352.

Knapp, C. and G. Carter (1976). The generalized correlation method for estimation of time delay. *IEEE Trans. Acoust. Speech Signal Process.* **24**, 320–327.

Krolik, J. and D. Swingler (1989). Multiple broad-band source location using steered covariance matrices. *IEEE Trans. Acoust. Speech Signal Process.* **37**(10), 1481–1494.

Lawley, D. (1956). Tests of significance of the latent roots of the covariance and correlation matrices. *Biometrika* **43**, 128–136.

Marr, J. (1986). A selected bibliography on adaptive antenna arrays. *IEEE Trans. Aerospace Electron. Systems* **22**(6), 781–794.

Matthews, P. and B. Mohebbi (1989). Direction of arrival measurements at UHF. *Electron. Lett.* **25**(16), 1069–1070.

Moffet, A. (1968). Minimum redundancy linear arrays. *IEEE Trans. Antennas and Propagation* **16**(2), 172–175.

Monzingo, R. and T. Miller (1980). *Introduction to Adaptive Arrays*. Wiley, New York.

Oppenheim, A. and R. Schafer (1975). *Digital Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ.

Ottersten, B. and L. Ljung (1989). Asymptotic results for sensor array processing. In: *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.* Glasgow, Scotland, 2266–2269.

Ottersten, B., M. Viberg and T. Kailath (1991). Performance analysis of the total least squares ESPRIT algorithm. *IEEE Trans. Signal Process.* **39**(5), 1122–1135.

Paulraj, A. and T. Kailath (1986). Eigenstructure methods for direction of arrival estimation in the presence of unknown noise fields. *IEEE Trans. Acoust. Speech Signal Process.* **34**(1), 13–20.

Paulraj, A., R. Roy and T. Kailath (1985). Estimation of signal parameters via rotational invariance techniques – ESPRIT. In: *Proc. 19th Asilomar Conf. on Circuits, Systems and Computers*. Pacific Grove, CA, 83–89.

Pillai, S. (1989). *Array Signal Processing*. Springer, New York.

Porat, B. and B. Friedlander (1988). Analysis of the asymptotic relative efficiency of the MUSIC algorithm. *IEEE Trans. Acoust. Speech Signal Process.* **36**(4), 532–544.

Prasad, S., R. Williams, A. Mahalanabis and L. Sibul (1988). A transform-based covariance differencing approach for some classes of parameter estimation problems. *IEEE Trans. Acoust. Speech Signal Process.* **36**(5), 631–641.

Rao, B. and K. Hari (1989). Performance analysis of root-MUSIC. *IEEE Trans. Acoust. Speech Signal Process.* **37**(12), 1939–1941.

Readhead, A. (1982). Radio astronomy by very long baseline interferometry. *Sci. Amer.* **246**, 52–61.

Reddi, S. (1987). On a spatial smoothing technique for multiple sources. *IEEE Trans. Acoust. Speech Signal Process.* **35**(5), 709.

Rockah, Y. and P. Schultheiss (1987a). Array shape calibration using sources in unknown locations – Part I: Far-field sources. *IEEE Trans. Acoust. Speech Signal Process.* **35**(3), 286–299.

Rockah, Y. and P. Schultheiss (1987b). Array shape calibration using sources in unknown locations – Part II: Near-field sources and estimation implementation. *IEEE Trans. Acoust. Speech Signal Process.* **35**(6), 724–735.

Roy, R. (1987). ESPRIT: Estimation of signal parameters via rotational invariance techniques. Ph.D. dissertation, Stanford University, Stanford, CA.

Roy, R. and T. Kailath (1989). ESPRIT – Estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.* **37**(7), 984–995.

Schell, S. (1990). Exploitation of spectral correlation for signal-selective direction finding. Ph.D. dissertation, Dept. of Electrical Engineering and Computer Science, University of California, Davis, CA.

Schell, S. V. (1993). An overview of sensor array processing for cyclostationary signals. In: W. A. Gardner, ed., *Cyclostationarity in Signal Processing and Communications*. IEEE Press.

Schell, S. and B. Agee (1988). Application of the SCORE algorithm and SCORE extensions to sorting in the rank-l self-coherence environment. In: *Proc. 22nd Ann. Asilomar Conf. on Signals, Systems, and Computers*. Pacific Grove, CA, 274–278.

Schell, S., R. Calabretta, W. Gardner and B. Agee (1989). Cyclic MUSIC algorithms for signal-selective DOA estimation. In: *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.* Glasgow, Scotland, 2278–2281.

Schell, S. and W. Gardner (1990a). Detection of the number of cyclostationary signals in unknown interference and noise. In: *Proc. 24th Ann. Asilomar Conf. on Signals, Systems, and Computers*. Pacific Grove, CA, 473–477.

Schell, S. and W. Gardner (1990b). Progress on signal-selective direction finding. In: *Proc. 5th ASSP Workshop on Spectrum Estimation and Modeling*. Rochester, NY, 144–148.

Schell, S. and W. Gardner (1990c). Signal-selective high-resolution direction finding in multipath. In: *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.* Albuquerque, NM, 2667–2670.

Schell, S. V. and W. A. Gardner (1992a). Cramér–Rao lower bound for directions of arrival of Gaussian cyclostationary signals. *IEEE Trans. Inform. Theory* **38**(4), 1418–1422.

Schell, S. V. and W. A. Gardner (1992b). Robustness of direction-finding methods for cyclostationary signals in the presence of array calibration error. In: *Proc. IEEE Sixth SP Workshop on Statistical Signal and Array Processing*. Victoria, BC, Canada, 346–349.

Schmidt, R. (1979). Multiple emitter location and signal parameter estimation. In: *Proc. RADC Estimation Workshop*. Rome Air Development Center, New York.

Schmidt, R. (1981). A signal subspace approach to multiple source location and spectral estimation. Ph.D. dissertation, Stanford Univ., Stanford, CA.

Schmidt, R. (1986). Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas and Propagation* **34**(3), 276–280.

Shan, T.-J., M. Wax and T. Kailath (1985). On spatial smoothing for direction-of-arrival estimation of coherent signals. *IEEE Trans. Acoust. Speech Signal Process.* **33**(4), 806–811.

Stoica, P. and A. Nehorai (1989). MUSIC, maximum likelihood, and Cramér–Rao bound. *IEEE Trans. Acoust. Speech Signal Process.* **37**(5), 720–741.

Swindlehurst, A., B. Ottersten and T. Kailath (1989). An analysis of MUSIC and root-MUSIC in the presence of sensor perturbations. In: *Proc. 23rd Asilomar Conf. on Signals, Systems and Computers*. Pacific Grove, CA, 930–934.

Tufts, D. and R. Kumaresan (1982). Frequency estimation of multiple sinusoids: Making linear prediction like maximum likelihood. *Proc. IEEE* **70**(9), 975–990.

Van Trees, H. (1968). *Detection, Estimation, and Modulation Theory: Part I*. Wiley, New York.

Van Veen, B. and K. Buckley (1988). Beamforming: A versatile approach to spatial filtering. *IEEE Acoust. Speech Signal Process. Mag.* **5**(2), 4–24.

Viberg, M. and B. Ottersten (1991). Sensor array processing based on subspace fitting. *IEEE Trans. Signal Process.* **39**(5), 1110–1121.

Wagstaff, R. and A. Baggeroer, eds. (1985). *High-Resolution Spatial Processing in Underwater Acoustics*. Naval Ocean Research and Development Activity, NSTL, Missouri.

Wang, H. and M. Kaveh (1985). Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources. *IEEE Trans. Acoust. Speech Signal Process.* **33**(4), 823–831.

Wang, H. and M. Kaveh (1986). On the performance of signal-subspace processing – Part I: Narrowband systems. *IEEE Trans. Acoust. Speech Signal Process.* **34**(10) 1201–1209.

Wax, M. (1985). Detection and estimation of superimposed signals. Ph.D. dissertation, Stanford University, Stanford, CA.

Wax, M. (1992). On unique localization of constrained-signals sources. *IEEE Trans. Signal Process.* **40**(6), 1542–1547.

Wax, M. and T. Kailath (1985). Detection of signals by information theoretic criterion. *IEEE Trans. Acoust. Speech Signal Process.* **33**(2), 387–392.

Wax, M. T.-J. Shan and T. Kailath (1984). Spatio-temporal spectral analysis by eigenstructure methods. *IEEE Trans. Acoust. Speech Signal Process.* **32**(4), 817–827.

Wax, M. and I. Ziskind (1989). On unique localization of multiple sources by passive sensor arrays. *IEEE Trans. Acoust. Speech Signal Process.* **37**(7), 996–1000.

Weiss, A. and B. Friedlander (1989). Array shape calibration using sources in unknown locations – A maximum likelihood approach. *IEEE Trans. Acoust. Speech Signal Process.* **37**(12), 1958–1966.

Weiss, A., A. Willsky and B. Levy (1988). Eigenstructure approach for array processing with unknown intensity coefficients. *IEEE Trans. Acoust. Speech Signal Process.* **36**(10), 1613–1617.

Widrow, B. and S. Stearns (1985). *Adaptive Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ.

Wiley, R. (1985). *Electronic Intelligence: The Interception of Radar Signals*. Artech House, Dedham, MA.

Xu, G. and T. Kailath (1989). Array signal processing via exploitation of spectral correlation – A combination of temporal and spatial processing. In: *Proc. 23rd Asilomar Conf. on Signals, Systems and Computers*. Pacific Grove, CA, 945–949.

Xu, G. and T. Kailath (1992). Direction of arrival estimation via exploitation of cyclostationarity – A combination of temporal and spatial processing. *IEEE Trans. Signal Process.* **40**(7), 1775–1786.

Zhao, L., P. Krishnaiah and Z. Bai (1986). On detection of the number of signals in presence of white noise. *J. Multivariate Anal.*, **20**, 1–25.

Ziskind, I. and M. Wax (1988). Maximum likelihood localization of multiple sources by alternating projection. *IEEE Trans. Acoust. Speech Signal Process.* **36**(10), 1553–1560.

Zoltowski, M. and F. Haber (1986). A vector space approach to direction finding in a coherent multipath environment. *IEEE Trans. Antennas and Propagation* **34**(9), 1069–1079.