Spatio-Temporal Filtering and Equalization for Cyclostationary Signals

Stephan V. Schell[1]

William A. Gardner[2]

[1]Dept. of Electrical Engineering
Pennsylvania State University
University Park, PA 16802

[2]Dept. of Elec. & Comp. Eng.
University of California
Davis, CA 95616

I Introduction

In such application areas as communication systems, signals intelligence, radar, sonar, commercial communications monitoring, biomedical signal processing, and geophysical exploration, signals of interest (SOIs) are often corrupted by channel distortion, interfering signals, and noise. To mitigate these sources of corruption and thus enable the receiver to obtain high-quality estimates of the SOI, it is often necessary to use adaptive spatial, temporal, or spatio-temporal filtering. Conventional methods of adaptive filtering typically require prior knowledge of the SOI and/or of the corruption, such as a training signal, channel transfer function, or interference covariance matrix. However, this prior knowledge can be difficult or impossible to obtain in some applications.

For example, in cellular communication systems that use time division multiple access (TDMA) (i.e., each SOI is active during only a short periodically occurring time slot) and must operate in the presence of rapidly changing multipath propagation, the characteristics of the corruption are time-varying and unknown, and periodic retransmission of a sufficiently

CONTROL AND DYNAMIC SYSTEMS, VOL. 64

long training signal during each time slot would decrease prohibitively the time that remains to communicate the message. Furthermore, many conventional adaptive methods are derived without regard for the statistical structure that uniquely identifies the SOIs.

The primary goal of this chapter is to explain how the statistical structure of *cyclostationary* signals (cf. [1, 2, 3] and references therein), and of man-made communication signals in particular, can be used to address some of the drawbacks of conventional methods. Almost all man-made communication signals exhibit *cyclostationarity* as the result of periodic gating, keying, and mixing operations in modulators and transmitters. A key point is that this property can be exploited in signal processors to favor desired signals and to discriminate against undesired signals, interference, and noise. This signal selectivity offers appealing benefits: post-processing that determines which signals and their parameters are of interest can sometimes be reduced or eliminated, and applicability to some difficult environments is enhanced. For example, direction-finding methods that exploit cyclostationarity can operate properly even when the total number of signals arriving at the sensor array exceeds the number of sensors, provided that the number of signals having the specified cyclostationarity property is less than the number of sensors (cf. [2, 4, 5, 6]). In addition, many cyclostationarity-exploiting algorithms offer these benefits while using much less prior knowledge of signal characteristics than that required by conventional methods. In particular, some methods of adaptive spatial filtering that exploit cyclostationarity require only knowledge of the baud rate or carrier frequency, or other frequency that characterizes the underlying periodicity exhibited by the desired signals. It is explained in this chapter how this knowledge is used to avoid the need for training signals, estimates of directions of arrival, and array calibration data in both adaptive sensor arrays (spatial filtering) and fractionally-spaced equalization for digital communication systems (temporal filtering).

In addition to explaining how cyclostationarity can be exploited by blind adaptive algorithms, this chapter also illustrates how other statistical properties can be exploited in a unifying framework referred to as Programmable Canonical Correlation Analysis.

This chapter is organized as follows. In Section II, the mathematical models for the sensor array data are described. In Section III, notation and basic results on the measurement of spatial characteristics of cyclostationary signals are summarized, and their implications for the basic problems of spatial filtering and fractionally-spaced equalization are briefly discussed. In Section IV, algorithms for blindly adapting (i.e., without using a training signal) a spatial or spatio-temporal filter to extract high-quality estimates of signal waveforms are summarized. In Section V, blind adaptive fractionally-spaced equalization (using one or more sensors) is addressed. Finally, in Section VI a few brief conclusions are drawn and the chapter is briefly summarized.

II Modeling Data from Sensor Arrays

In this section the mathematical models for the signals at the output of a sensor array are developed from basic physical considerations. The general model for wideband data is derived by assuming that a single sinusoid arrives at the array and then applying superposition to build up the expressions for multiple nonsinusoidal signals. An extremely useful simplifying approximation is then justified and applied to yield a description referred to in the sensor-array signal processing literature as the *narrowband model*. Extensive use is made of this model in this chapter.

II.A General Wideband Model

Consider the analytic signal $\exp(j2\pi ft)$ corresponding to a real sine wave having frequency f 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, and assume that the wavefronts impinging on the array are planar. 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 locations 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 mth 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. Since the signal is sinusoidal, the propagation delay t_m is equivalent to a phase shift by the amount $\psi_m = -2\pi f t_m$, which is in turn equivalent to multiplication by $\exp(j\psi_m)$. Thus, the signal received by the array can be expressed in the vector form

$$\mathbf{x}(t) \triangleq \begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \begin{bmatrix} \exp(j\psi_1(\theta, f)) \\ \vdots \\ \exp(j\psi_M(\theta, f)) \end{bmatrix} e^{j2\pi f t}, \tag{1}$$

where

$$\psi_m(\theta, f) = [q_m \sin(\theta) + r_m \cos(\theta)]2\pi f/c. \tag{2}$$

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 (1). Denoting the gain and phase of the mth sensor by $g_m(\theta, f)$ and $\phi_m(\theta, f)$, respectively, the analytic signal at the outputs of the sensors can be expressed as

$$\mathbf{x}(t) = \mathbf{a}(\theta, f)e^{j2\pi f t} \tag{3}$$

where $a_m(\theta, f) = g_m(\theta, f) \exp[j(\psi_m(\theta, f) + \phi_m(\theta, f))]$ is the mth element of the vector $\mathbf{a}(\theta, f)$, which is referred to in this chapter as the *array response vector*, although 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 f of interest is referred to as the *array manifold.*

In the more general (and interesting) case in which multiple non-sinusoidal signals arrive at the array, the data can be modeled by decomposing it in

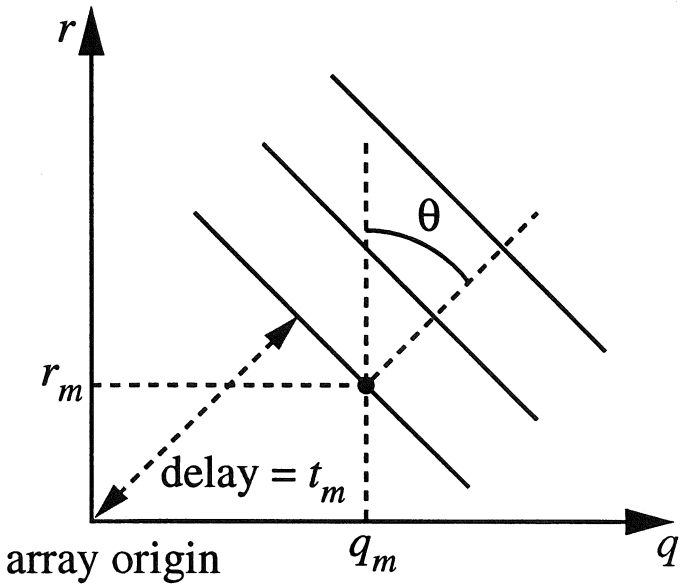

Figure 1: Plane waves propagating from angle θ toward the array origin.

the frequency domain (temporarily assuming that the signals are Fourier-transformable) and using linear superposition:

$$\begin{aligned}\tilde{\mathbf{x}}(f) &= \sum_{l=1}^{L} \mathbf{a}(\theta_l, f)\tilde{s}_l(f) + \tilde{\mathbf{i}}(f) \\ &= [\mathbf{a}(\theta_1, f) \cdots \mathbf{a}(\theta_L, f)]\,[\tilde{s}_1(f) \cdots \tilde{s}_L(f)]^T + \tilde{\mathbf{i}}(f) \\ &= \mathbf{A}(\Theta, f)\tilde{\mathbf{s}}(f) + \tilde{\mathbf{i}}(f), \end{aligned} \tag{4}$$

where L signals having Fourier transforms $\tilde{s}_1(f)$, ..., $\tilde{s}_L(f)$ arrive from angles θ_1, ..., θ_L and $\tilde{\mathbf{i}}(f)$ 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 $\mathbf{a}(\theta, f)$.

II.B Narrowband Model

The general wideband model is needlessly complex 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). For example, in some applications the data may be channelized into very narrow bands which are then processed individually. Alternatively, if it is known that the signals of interest occupy a certain frequency band then it is advantageous to reject interference components and noise that lie outside this band. If this band is sufficiently narrow that the array response vector $\mathbf{a}(\theta, f)$ is approximately constant with respect to f over the band of interest for all angles θ (e.g., if the reciprocal of the bandwidth of the signal is much greater than the time required for the signal to propagate across the array, and if the sensor characteristics do not vary significantly across this bandwidth), then the dependence on f can be dropped and the array data can be modeled in the time domain as the analytic signal

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

where $\mathbf{s}(t)$ and $\mathbf{i}(t)$ are analytic signals. That is, these are the complex representations obtained by suppressing the negative frequency portions of the 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. This observation is the essence of the narrowband model.

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 [7]. A brief justification is offered here. Consider a single signal $s(t)$ having flat power spectral density over the band $[f_0 - B/2, f_0 + B/2]$, and arriving at a uniform linear array (ULA) for which $(q_m, r_m) = (dm, 0)$ where d is the sensor spacing. The spectral density of $\mathbf{x}(t)$ is $\mathbf{S}_{\mathbf{xx}}(f) = \mathbf{a}(\theta, f)\mathbf{a}^H(\theta, f)$ for $|f - f_0| \leq B/2$, and the

autocorrelation at lag $\tau = 0$ can be expressed as

$$[\mathbf{R_{xx}}(0)]_{m,n} = B \exp\left(-j2\pi\frac{f_0 d}{c}(m-n)\sin\theta\right) \operatorname{sinc}\left(\pi\frac{Bd}{c}(m-n)\sin\theta\right).$$

The total average power P_{tot} of $\mathbf{x}(t)$ is $P_{tot} = \operatorname{tr}\{\mathbf{R_{xx}}(0)\} = MB$. Also, the average power P_{f_0} in $\mathbf{x}(t)$ that can be represented by the narrowband model $\mathbf{a}(\theta, f_0)\, s(t)$ can be found (e.g., conceptually by projecting $\mathbf{x}(t)$ onto the space spanned by $\mathbf{a}(\theta, f_0)$ and then computing the power of the projection); that is,

$$\begin{aligned} P_{f_0} &= \operatorname{tr}\{\mathbf{P}_{\mathbf{a}(\theta,f_0)}\mathbf{R_{xx}}(0)\} \\ &= \frac{B}{M}\sum_{k,m=1}^{M} \operatorname{sinc}\left(\frac{\pi\eta}{2}(k-m)\sin\theta\right) \end{aligned}$$

where $\eta \triangleq B/f_0$ is the relative bandwidth, f_0 is the frequency chosen for use in the narrowband model (it can be any frequency in the reception band), and $\mathbf{P_v}$ denotes the orthogonal projection matrix for vector $\mathbf{v}$. Thus, the ratio $\gamma \triangleq P_{f_0}/P_{tot}$ is a measure of the degree to which $\mathbf{x}(t)$ admits a narrowband model:

$$\gamma \triangleq \frac{P_{f_0}}{P_{tot}} = \frac{1}{M^2}\sum_{k,m=1}^{M} \operatorname{sinc}\left(\frac{\pi\eta}{2}(k-m)\sin\theta\right).$$

Clearly, the narrowband model is exact ($P_{f_0} = P_{tot}$) for $\eta = 0$ but degrades as the relative bandwidth increases, as shown in Figure 2. For example, at the worst-case value of θ ($\theta = 90$ degrees), the narrowband model for the 4-element ULA accounts for 99% of the received power for $\eta = 10\%$, and 90% of the received power for $\eta = 30\%$.

In this chapter and in much of the literature on sensor-array signal processing, 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 $\exp(-j2\pi ft)$ for some appropriate f), the corresponding model for the sampled complex envelope is essentially the same as in (5),

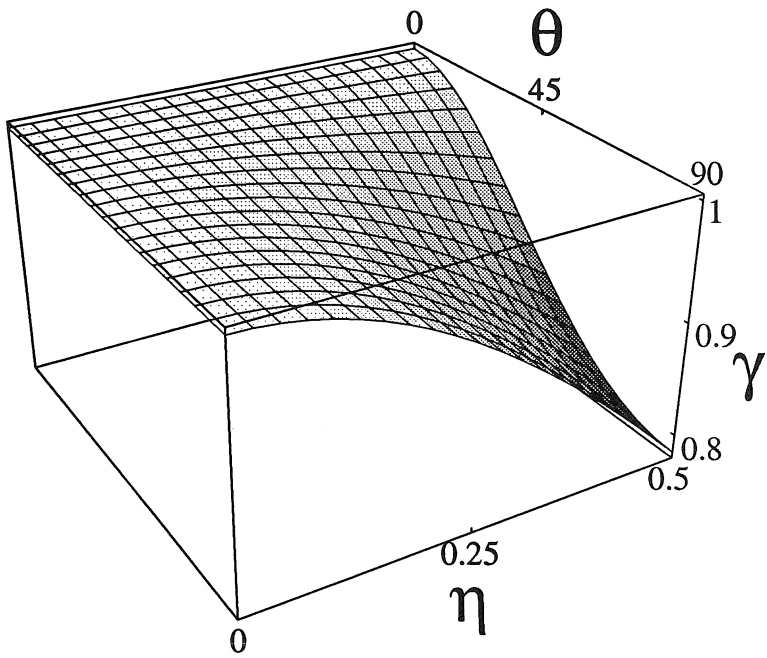

Figure 2: Accuracy of narrowband model versus relative bandwidth η and angle of arrival θ for a 4-element ULA.

except that $\mathbf{x}(t)$, $\mathbf{s}(t)$, and $\mathbf{i}(t)$ denote the complex envelopes of the array data, the signals, and the noise, respectively, and t is replaced with n:

$$\mathbf{x}(n) = \sum_{l=1}^{L} \mathbf{a}(\theta_l) s_l(n) + \mathbf{i}(n) = \mathbf{A}(\Theta)\mathbf{s}(n) + \mathbf{i}(n). \tag{6}$$

Almost all of the algorithms discussed in this chapter are based on this model.

III Cyclostationarity

In this section the relevant results from the theory of cyclostationary signals are briefly summarized. Cyclostationary signals are also referred to as being periodically correlated or almost periodically correlated, among other terms (cf. [8, 3]). A useful partial characterization of a cyclostationary signal is its

set of *cycle frequencies*, where a signal is said to exhibit cyclostationarity with cycle frequency α if and only if its cyclic autocorrelation function $R_{xx}^{\alpha}(\tau)$ (defined in Section III.A) is not identically zero. This requirement can be interpreted as meaning that a finite-amplitude additive sine wave at frequency α is present in the output $x(n)\,x^*(n-\tau)$ of the delay-and-multiply quadratic processor, or equivalently that some spectral components of $x(n)$ that are separated by α are correlated.

The emphasis in this chapter is on man-made communication signals, which typically have cycle frequencies equal to harmonics of their respective symbol rates (e.g., for PSK, QAM, FSK, and spread spectrum signals), doubled-carrier frequencies (e.g., for DSB-AM, narrowband FM, and BPSK signals), and sums and differences of these (e.g., for BPSK, staggered QPSK, and MSK signals). In many applications, one or more of these cycle frequencies can be known by the receiver (e.g., especially in commercial communication systems). Alternatively, they can be estimated (cf. [9, 10, 11, 12, 13] and references therein), although much work remains to be done in this area.

First, notation used in the rest of this chapter is summarized. Then, the implications of cyclostationarity for sensor array processing (especially interference rejection) are summarized by discussing certain temporal correlation statistics of cyclostationarity. Finally, the implications for equalization of channel distortion are discussed by making use of an alternative but completely equivalent interpretation of cyclostationarity in terms of spectral redundancy. Detailed tutorial treatments of cyclostationarity can be found in [2, 3, 14, 12].

III.A Notation

The following abbreviations are used throughout this chapter:

SOI: Signal of interest

SNOI: Signal not of interest

STF: Spatio-temporal filter

DOA: Direction of arrival

LCL-PTV: Linear-conjugate-linear polyperiodic time varying

SNR: Signal to noise ratio

SINR: Signal to interference and noise ratio

MSE: Mean squared error

PSK: Phase-shift-keyed

QAM: Quadrature amplitude modulated

With reference to the abbreviation LCL-PTV, it is noted that the term *linear-conjugate-linear multiply-periodic time-varying* (LCL-MPTV) is used in earlier work [15] to denote the same concept. The new term is chosen for consistency with the terminology used in [3]. The following definitions are used throughout this chapter:

D-1 The notation $\langle f(n)\rangle_N$ denotes the time average of $f(n)$:

$$\langle f(n)\rangle_N \triangleq \frac{1}{N}\sum_{n=0}^{N-1} f(n).$$

D-2 The function $Z_N(f,\tau)$ is defined as

$$\begin{aligned} Z_N(f,\tau) &\triangleq \frac{1}{N}\sum_{n=0}^{N-1-\tau} e^{-j2\pi fn} \\ &= \left(1-\frac{\tau}{N}\right)\frac{\operatorname{sinc}(\pi f(N-\tau))}{\operatorname{sinc}(\pi f)}e^{-j\pi f(N-\tau-1)} \end{aligned}$$

and is $O\left(\frac{1}{N}\right)$ for fixed $f \neq 0$ and τ, where $\operatorname{sinc}(x) \triangleq \sin x/x$.

D-3 The superscripts $*$, T, H, $+$, and $\perp$ denote conjugation, matrix transposition, matrix conjugate transposition, pseudo-inversion, and the orthogonal complement operation, respectively. Furthermore, $\star$ and $\otimes$ denote the convolution and Kronecker (tensor) product (cf. [16]) operators, respectively.

D-4 The expectation operation is denoted by $E\{\}$ and can denote either the ensemble-averaging operation for which the argument is typically a stochastic process, or the sinewave (or polyperiodic component)

extraction operation based on time-averaging. In this chapter, this operation can be interpreted in either way with complete equivalence of the results, provided that underlying assumptions of cycloergodicity and so forth are properly attended to. Extensive discussion of these two analytical frameworks can be found in [3].

D-5 The time-variant cross-correlation $\mathbf{R}_{\mathbf{xy}}(m,n)$ between $\mathbf{x}(m)$ and $\mathbf{y}(n)$ is defined by

$$\mathbf{R}_{\mathbf{xy}}(m,n) \triangleq E\left\{\mathbf{x}(m)\mathbf{y}^H(n)\right\}.$$

D-6 If $\mathbf{R}_{\mathbf{xy}}(n+\tau,n)$ has a Fourier series representation in n, it is given by

$$\mathbf{R}_{\mathbf{xy}}(n+\tau,n) = \sum_{\beta} \mathbf{R}^{\beta}_{\mathbf{xy}}(\tau) e^{j2\pi\beta n}$$

where the sum is taken over all β for which the cyclic cross correlation $\mathbf{R}^{\beta}_{\mathbf{xy}}(\tau)$, defined by

$$\mathbf{R}^{\beta}_{\mathbf{xy}}(\tau) \triangleq \left\langle \mathbf{R}_{\mathbf{xy}}(n+\tau,n) e^{-j2\pi\beta n} \right\rangle_{\infty},$$

is not identically zero as a function of τ.

D-7 The symbol $\hat{\mathbf{R}}^{\alpha}_{\mathbf{xy}}(\tau)$ denotes the cyclic cross-correlogram of $\mathbf{x}(n)$ and $\mathbf{y}(n)$ at cycle frequency α and lag τ:

$$\hat{\mathbf{R}}^{\alpha}_{\mathbf{xy}}(\tau) \triangleq \begin{cases} \frac{1}{N}\sum_{n=0}^{N-1-\tau} \mathbf{x}(n+\tau)\mathbf{y}^H(n)\, e^{-j2\pi\alpha n}, & \text{for } \tau \geq 0, \\ \frac{1}{N}\sum_{n=-\tau}^{N-1} \mathbf{x}(n+\tau)\mathbf{y}^H(n)\, e^{-j2\pi\alpha n}, & \text{for } \tau < 0, \end{cases}$$

and it is useful to note that $\hat{\mathbf{R}}^{\alpha}_{\mathbf{xy}}(\tau) \equiv \hat{\mathbf{R}}^{-\alpha H}_{\mathbf{yx}}(-\tau) e^{j2\pi\alpha\tau}$.

D-8 The cyclic cross correlation coefficient between $u(n)$ and $v(n)$ for cycle frequency α and lag τ is defined as

$$\rho^{\alpha}_{uv}(\tau) \triangleq R^{\alpha}_{uv}(\tau) \Big/ \sqrt{R_{uu}(0) R_{vv}(0)},$$

where it is noted that $|\rho^{\alpha}_{uv}(\tau)| \leq 1$ for all α and τ (e.g., [14]). The magnitude of $\rho^{\alpha}_{ss}(\tau)$ is sometimes referred to as the feature strength (at cycle frequency α and lag τ) of the signal $s(n)$. The feature strength is a normalized measure that can be interpreted as indicating the degree to which a signal is correlated with a time- and frequency-shifted copy of itself).

III.B Implications for Spatial Filtering

Under the assumption that the narrowband model holds well and L_α signals of interest (SOI) $s_1(n), \ldots, s_{L_\alpha}(n)$ having cycle frequency α arrive at the array, and that the remaining signals not of interest (SNOIs) and noise $\mathbf{i}(n)$ do not have cycle frequency α (i.e., $\mathbf{R}_{\mathbf{ii}}^{\alpha}(\tau) \equiv 0$ and $\mathbf{R}_{\mathbf{ii}^*}^{\alpha}(\tau) \equiv 0$), where

$$\mathbf{x}(n) = \sum_{l=1}^{L_\alpha} \mathbf{a}(\theta_l)\, s_l(n) + \mathbf{i}(n),$$

then it can be shown (see [8, 6]) that the estimated cyclic autocorrelation matrix converges to the ideal cyclic autocorrelation matrix,

$$\begin{aligned} \hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha}(\tau) &\longrightarrow \mathbf{A}(\Theta)\mathbf{R}_{\mathbf{ss}}^{\alpha}(\tau)\mathbf{A}^H(\Theta) \\ \hat{\mathbf{R}}_{\mathbf{xx}^*}^{\alpha}(\tau) &\longrightarrow \mathbf{A}(\Theta)\mathbf{R}_{\mathbf{ss}^*}^{\alpha}(\tau)\mathbf{A}^T(\Theta) \end{aligned}$$

with bias and covariance $O\left(\frac{1}{N}\right)$. This in turn implies that a measurement of the spatial characteristics (the vector space spanned by the SOIs' array response vectors) of the L_α SOIs can be made, even if the SOI waveforms and directions of arrival are unknown, and even if $\mathbf{i}(n)$ has completely arbitrary and unknown spatial characteristics. However, note that restrictions on the ambiguity of the manifold and on the rank of $\mathbf{R}_{\mathbf{ss}}^{\alpha}(\tau)$ are needed to determine Θ and/or $\mathbf{A}(\Theta)$ from $\mathbf{R}_{\mathbf{xx}}^{\alpha}(\tau)$.

In Section IV it is shown how this signal selectivity can be exploited to solve waveform and parameter estimation problems in sensor array processing.

An implication of the convergence results obtained in [6] is that error in the knowledge of the cycle frequency manifests itself as cycle leakage in the sense that the regenerated spectral line at the true cycle frequency α_0 leaks into the estimate at α with weight $Z_N(\alpha_0 - \alpha)$. For example, it can be shown under reasonable assumptions that the mean of $\hat{\mathbf{R}}_{\mathbf{xy}}^{\alpha}(\tau)$ for $\tau > 0$ can be expressed as

$$E\left\{\hat{\mathbf{R}}_{\mathbf{xy}}^{\alpha}(\tau)\right\} = \left(1 - \frac{\tau}{N}\right)\mathbf{R}_{\mathbf{xy}}^{\alpha}(\tau) + \sum_{\beta \neq \alpha} \mathbf{R}_{\mathbf{xy}}^{\beta}(\tau)\, Z_N(\alpha - \beta, \tau),$$

where the summation is performed over all cycle frequencies of $\mathbf{x}(n)$ and $\mathbf{y}(n)$. Thus, the allowable error in the estimated cycle frequency should be

less than $1/2N$. More extensive discussion of cycle leakage can be found in [12, 17, 11]. To reduce the error, methods for estimating the cycle frequencies of a signal can be used, although they themselves are somewhat susceptible to error and require a potentially large number of data samples [9, 10], or they are computationally intensive (requiring computation of the cyclic spectrum over a potentially large portion of the (f, α) plane as in [13]) (cf. [17, 11] for efficiently implemented cyclic spectrum analyzers). However, it has been observed in some cases [18] that the CRLB on the variance of cycle frequency estimates decreases as $1/N^3$ (in contrast to $1/N$ for estimates of direction of arrival, carrier phases, and so forth), so it may be possible to develop extremely reliable cycle frequency estimators. Even in the case of scalar data this remains an open problem, although the literature on synchronization in digital communication systems (cf. [19] and references therein) offers useful starting points.

III.C Implications for Equalization

In the equalization of digital QAM communication signals, the statistical structure of the SOI is restricted even further. Let $x(n)$ be a scalar-valued digital communication signal corrupted by channel distortion and interference,

$$x(n) = \sum_{l=-\infty}^{\infty} h(n - lT)s(l) + i(n)$$

where the unknown channel has impulse response $h(n)$, T is the baud period, $s(k)$ is the kth symbol (unknown), and $i(n)$ is unknown noise. It can be shown (cf. [20]) that $R_{xx}(m + k, m)$ is periodic with period T and thus that the cyclic autocorrelation evaluated at α equal to integer multiples of the symbol rate $1/T$ is not identically zero.

Although this interpretation of the cyclostationarity of $x(n)$ in terms of cyclic correlations could be pursued in a manner similar to that done in Section III.B, an alternative but completely equivalent interpretation in the frequency domain is given here. In particular, $x(n)$ can also be expressed as

$$x(n) = h(n) * t(n) + i(n)$$

where

$$t(n) = \sum_{l=-\infty}^{\infty} \delta(n - lT)s(l),$$

and thus the power spectral density (PSD) of $t(n)$ is given by

$$S_t(f) = \sum_{l} S_s(f + l/T).$$

Under the reasonable assumption that $s(l)$ and $i(n)$ are white sequences, the power spectral densities of $s(l)$, $t(n)$, and $h(n)*t(n)$ are depicted in Figure 3 for a typical example of $h(n)$ having 100% excess bandwidth beyond the minimum (Nyquist) bandwidth needed for transmission of $s(l)$ without intersymbol interference. Clearly, $t(n)$ and $h(n) * t(n)$ exhibit spectral redundancy (which is reflected in nonzero spectral correlation), which in this case means that the same components of the message stream $s(l)$ are transmitted and received over multiple spectral bands. Specifically, the signal component in the band $[-1/T, -1/(2T)]$ is a filtered and frequency-shifted copy of the component in the band $[0, 1/(2T)]$, and similarly for the bands $[-1/(2T), 0]$ and $[1/(2T), T]$.

This property of digital QAM communication signals implies that a receiver can use information from one spectral band to repair or replace information that is degraded or destroyed in a different spectral band. Fractionally spaced equalizers, which are special cases of the more general filters called frequency shift filters (or LCL-PTV filters), exploit this property to advantage, as discussed in Section V. Also, more extensive discussion of frequency-shift filters can be found in [21, 22].

Also, this property of these signals implies that information on the channel phase is available in the second-order statistics of cyclostationarity (i.e., $\hat{R}^{\alpha}_{xx}(\tau)$ and its Fourier transform, the cyclic spectrum $\hat{S}^{\alpha}_{xx}(f)$). Methods that implicitly or explicitly exploit this fact can be found in [23] and in Section V of this chapter.

IV Spatial Filtering

In this section the problem of using a sensor array to spatially filter the received signals without knowing a training signal or direction of arrival

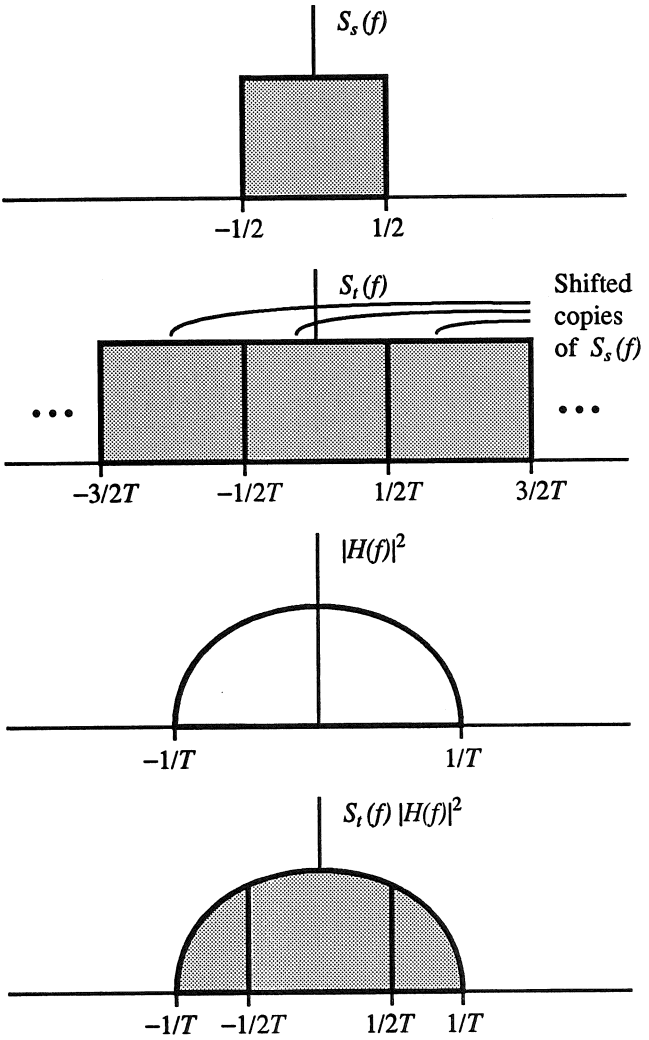

Figure 3: PSDs of $s(l)$, $t(n)$, and $t(n) * h(n)$.

is addressed. First, structures for spatial and spatio-temporal filtering, performance measures, and conventional adaptive methods are reviewed. Then, the use of data-derived training signals is discussed and a flexible general framework, called Programmable Canonical Correlation Analysis (PCCA), is introduced. The utility of some specific realizations within this framework is then illustrated through computer simulations, and potential applications are discussed briefly.

IV.A Structures for Spatial Filtering

Spatial filtering is used for purposes similar to those of temporal filtering: to enhance desired signal components, to attenuate undesired signal components, and to minimize noise.

The simplest spatial filter considered here linearly combines the signals from the sensors to yield an output signal $\hat{s}(n)$:

$$\hat{s}(n) = \mathbf{w}^H \mathbf{x}(n).$$

If the narrowband model (6) holds well then, to a close approximation,

$$\hat{s}(n) = \sum_{l=1}^{L_\alpha} \left[\mathbf{w}^H \mathbf{a}(\theta_l)\right] s_l(n) + \mathbf{w}^H \mathbf{i}(n),$$

which clearly shows that the gain applied to a signal arriving from angle θ is $\mathbf{w}^H \mathbf{a}(\theta)$. Thus, $\mathbf{w}^H \mathbf{a}(\theta)$ is analogous to the transfer function (Fourier-transformed impulse response) of a linear time-invariant (LTI) temporal filter and is referred to as the spatial transfer function or antenna pattern of the spatial filter. When multiple signals arrive from different directions, a carefully chosen spatial filter can extract one of these signals while rejecting the others, as depicted by the antenna pattern in Figure 4, where $\mathbf{w}$ has been chosen to extract the signal arriving from 30 degrees and to reject (or null out) the signals arriving from -20 degrees and 0 degrees.

More generally, a spatial filter can include an LTI filter on each sensor to perform spatio-temporal filtering (STF)

$$\hat{s}(n) = \sum_{m=1}^{M} w_m^*(n) \star x_m(n) = \mathbf{w}^H(n) \star \mathbf{x}(n),$$

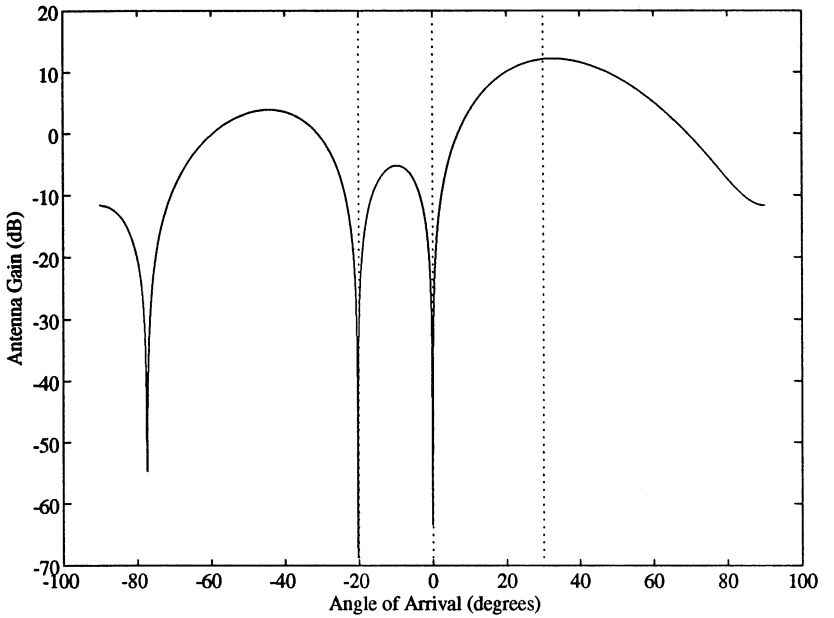

Figure 4: Antenna pattern of a spatial filter to extract a SOI arriving from 30 degrees in the presence of signals at -20 and 0 degrees and white noise that is uncorrelated from sensor to sensor.

for which the antenna pattern is now a function of both DOA θ and frequency f (i.e., $\mathbf{w}^H(-f)\mathbf{a}(\theta, f)$), and the frequency dependence of $\mathbf{a}(\theta)$ is included to emphasize that the STF is appropriate even if the narrowband model does not hold.

Yet more generally, a spatial filter can include a linear-conjugate-linear polyperiodic time-variant (LCL-PTV) filter on each sensor to perform LCL-PTV STF, which is the generalization to multiple inputs of the scalar LCL-PTV temporal filter. An extensive discussion of LCL-PTV filters is given in [21] where they are referred to as polyperiodic linear filters. As discussed there and particularly in [22], the LCL-PTV filter structure can implement the Cyclic Wiener filter, which is the generalization of the Wiener filter from stationary to cyclostationary signals. The fractionally-spaced equal-

izer structure that is ubiquitous in digital communication systems is also a restricted implementation of an LCL-PTV filter [22, 24]. Since a properly chosen LCL-PTV filter can separate two cyclostationary signals even if they are completely overlapping both temporally and spectrally, it is reasonable to expect there to be substantial benefits of LCL-PTV STFs over LTI STFs in some cases.

In addition to the preceding, LCL-PTV STFs can be motivated simply by observing that they make it possible to simultaneously exploit multiple cyclostationarity features. For example, a digital communication SOI having baud rate f_{baud} exhibits useful cyclostationarity at multiple cycle frequencies, such as $\pm f_{baud}$ (and harmonics thereof if the SOI is not bandwidth-efficient), and the cyclic autocorrelations at these cycle frequencies contain useful information at multiple values of the lag parameter τ. Thus, the LCL-PTV STF can be interpreted as linearly combining multiple frequency-shifted and filtered and possibly conjugated versions of the received data, or as simply providing a signal, which when correlated with the original data, allows multiple cyclostationarity properties to be manifested simultaneously in a single measurement. That is, the output of the LCL-PTV STF is given by

$$\hat{s}(n) = \mathbf{w}^H \mathbf{y}(n) \tag{7}$$

where

$$\mathbf{y}(n) = \begin{bmatrix} (\mathbf{x}(n) \star h_1(n))\, e^{j2\pi\alpha_1 n} \\ \vdots \\ (\mathbf{x}(n) \star h_J(n))\, e^{j2\pi\alpha_J n} \\ (\mathbf{x}(n)^* \star h_{J+1}(n))\, e^{j2\pi\alpha_{J+1} n} \\ \vdots \\ (\mathbf{x}(n)^* \star h_K(n))\, e^{j2\pi\alpha_K n} \end{bmatrix} \tag{8}$$

and $h_k(n)$ for $k = 1, \ldots, K$ are the impulse responses of arbitrary LTI filters, and α_k for $k = 1, \ldots, K$ are typically related to the cycle frequencies of the desired cyclostationary signals (and possibly undesired interfering signals) $\mathbf{s}(n)$ (e.g., doubled carrier frequencies, baud rates and their harmonics, and sums and differences of these). Figure 5 depicts an implementation of (7), without the branches that use conjugation.

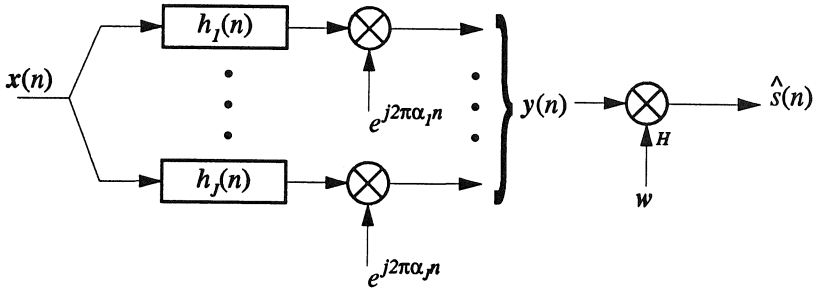

Figure 5: Block diagram of an LCL-PTV STF. The conjugate branches are omitted for simplicity.

It is noted here that several open problems relate to LCL-PTV STFs. In particular, practical methods of adapting them, their convergence time and SINR at convergence, methods for *blindly* adapting them (i.e., without the use of training signals), and their use in such applications as equalization and/or demodulation of severely corrupted digital communication signals (e.g., see Section V) and interception (blind de-spreading) of spread spectrum communication signals are of significant interest. As described in Section IV.E, the primary use of LCL-PTV STF in this chapter is as a means of deriving training signals (for use by an algorithm that adapts a memoryless linear spatial filter) directly from the data.

IV.B Performance Criteria

The performance measures most commonly used to evaluate waveform estimators such as spatial filters are summarized here.

Perhaps the most obvious of these measures is the mean squared error (MSE) of the estimated waveform relative to the desired waveform,

$$MSE(\hat{s}, s) = \left\langle \|\hat{s}(n) - s(n)\|^2 \right\rangle_N . \tag{9}$$

A closely related measure is the signal to interference and noise ratio (SINR) which can be expressed as

$$SINR(\hat{s}, s) \triangleq \frac{\text{power of desired signal components in } \hat{s}}{\text{power of everything else in } \hat{s}}$$

$$= \frac{\left|\hat{R}_{\hat{s}s}\right|^2 / \hat{R}_{ss}}{\hat{R}_{\hat{s}\hat{s}} - \left|\hat{R}_{\hat{s}s}\right|^2 / \hat{R}_{ss}} = \frac{|\hat{\rho}_{\hat{s}s}|^2}{1 - |\hat{\rho}_{\hat{s}s}|^2}. \tag{10}$$

The SINR increases without bound as $\hat{s}$ becomes arbitrarily highly correlated with s (i.e., as $|\hat{\rho}_{\hat{s}s}| \to 1$). All three of these measures ($MSE(\hat{s}, s)$, $SINR(\hat{s}, s)$, and $|\hat{\rho}_{\hat{s}s}|^2$) are applicable to single time-series, but are often averaged over multiple realizations (e.g., sample paths of a stochastic process) or multiple data segments of length N comprising a much longer time series.

It should be noted that minimizing $MSE(g\hat{s}, s)$ with respect to the complex scalar g and estimated waveform $\hat{s}$ is exactly equivalent to maximizing $SINR(\hat{s}, s)$ and then computing $g = \hat{R}_{s\hat{s}}/R_{\hat{s}\hat{s}}$.

It should also be noted that in this chapter SINR is distinct from SNR, which is used here to specify the power levels of received signals relative to the power of Gaussian noise.

Finally, in the special, though increasingly important, case in which the desired signal is a digital communication signal, the ultimate goal is typically to obtain a good estimate of the underlying bit or symbol stream. Thus, the relevant performance measure is the bit error rate (BER), which is strongly dependent both on the type of interference, noise, and channel distortion that corrupts the signal at the input to the demodulator and on the demodulator itself. Simple expressions for BER as a function of SNR (relative to white Gaussian noise) in the absence of channel distortion are well-known and can be found in any standard text on digital communications (e.g., [25]).

IV.C Conventional Methods

Conventional approaches to adapting a spatial filter have been well understood for more than two decades, and a tutorial that includes them and more recent advances can be found in [26]. Comprehensive treatments can also be found in [27] and [28]. For the purposes of this chapter, three of the most popular methods are summarized here.

The most direct method simply minimizes $MSE(\hat{s}, s)$ to obtain the

minimum MSE (MMSE) solution

$$\mathbf{w}_{MMSE} = \hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{x}s}, \tag{11}$$

for which the corresponding SINR (referred to in this chapter as the maximum SINR SINR_{max} attainable by a memoryless time-invariant spatial filter) is given by (10) with

$$|\hat{\rho}_{\hat{s}s}|^2 = \hat{\mathbf{R}}_{\mathbf{x}s}^H\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{x}s}/\hat{R}_{ss}.$$

As can be inferred from (10), a spatial filter attaining this maximum SINR can also be found by maximizing $|\hat{\rho}_{\hat{s}s}|^2$. A general depiction of this adaptation scheme is shown in Figure 6. For applications in which it is undesirable, expensive, or impossible for a known training signal to be sent by the transmitter and a local copy of it produced for use by the receiver for adaptation, this method is unsuitable.

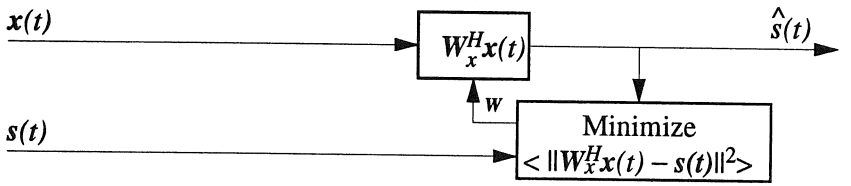

Figure 6: Block diagram of MMSE adaptive filter.

An alternative to the MMSE method makes use of prior knowledge of the direction of arrival (DOA) of the signal of interest (SOI) to form a beam on the SOI and to null any interferers. This method is referred to in the literature as minimum variance distortionless response (MVDR) beamforming and is a special case of a more general framework called linearly constrained minimum variance beamforming. Yet more general variants exist which accommodate quadratic, derivative, and eigenvector constraints (cf. [26]). The spatial filter weights are chosen so as to minimize the average output power while maintaining unity array gain in the direction θ_0 of the SOI, since doing so necessarily minimizes the contributions of interferers that arrive from other directions:

$$\max_{\mathbf{w}\ \text{s.t.}\ \mathbf{w}^H\mathbf{a}(\theta_0)=1} \hat{\mathbf{R}}_{\hat{s}\hat{s}} \iff \max_{\mathbf{w}\ \text{s.t.}\ \mathbf{w}^H\mathbf{a}(\theta_0)=1} \mathbf{w}^H\hat{R}_{\mathbf{xx}}\mathbf{w}.$$

It can be shown (e.g., via the method of Lagrange multipliers) that the solution is given by

$$\mathbf{w}_{MVDR} = \left[\mathbf{a}^H(\theta_0)\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\mathbf{a}(\theta_0)\right]^{-1} \hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\mathbf{a}(\theta_0). \tag{12}$$

If the SOI arrives only from angle θ_0, then $\hat{\mathbf{R}}_{\mathbf{x}s}$ converges to $\mathbf{a}(\theta_0)R_{ss}$, which with (11) and (12) implies that the output SINR of the MVDR spatial filter converges to SINR_{max}. For applications in which it is impossible to know θ_0 or $\mathbf{a}(\theta_0)$ and difficult to estimate them (e.g., using one of the direction finding methods described in [29]), this method is unsuitable.

As yet another alternative, when neither a known training signal nor a known direction of arrival is available, a direction-finding algorithm can be applied. For example, given known array calibration data (the array manifold), the method of [30] estimates the directions of arrival of all signals and then uses these to compute the weights of the spatial filter for each signal. However, in some applications, the computational load of estimating all of the directions of arrival and post-processing the spatially filtered signals to select those of interest may be prohibitive. Also, accurate array calibration data can be difficult or impossible to obtain in some applications.

IV.D Data-Derived Training Signals

In this section, one possible motivation for the framework of blind adaptive spatial filtering to be discussed in Section IV.E is presented.

The challenge of blind adaptive spatial filtering is to minimize $MSE(\hat{s}, s)$ without knowing s, without knowing $\mathbf{a}(\theta_0)$, and without resorting to methods based on direction finding which require knowledge of the array calibration data $\mathbf{a}(\cdot)$. At first this challenge may seem insurmountable. However, it is noted that if s arrives from only one direction (i.e., no multipath or smart jamming is present), then $\hat{\mathbf{R}}_{\mathbf{x}s}$ converges to $\mathbf{a}(\theta_0)R_{ss}$; indeed, any vector that converges to a vector that is proportional to $\mathbf{a}(\theta_0)$ can reasonably serve as a substitute for $\hat{\mathbf{R}}_{\mathbf{x}s}$ in (11). In particular, if $d(n)$ is any signal that is correlated with $s(n)$ but uncorrelated with the remaining signals comprising $\mathbf{x}(n)$, then $\hat{\mathbf{R}}_{\mathbf{x}d} = \mathbf{a}(\theta_0)\hat{R}_{sd} + \hat{\mathbf{R}}_{\mathbf{i}d}$ converges to $\mathbf{a}(\theta_0)R_{sd}$, which is clearly proportional to $\mathbf{a}(\theta_0)$ as desired. This implies that $\mathbf{w} = \hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{x}d}$

might be a suitable adaptive spatial filtering algorithm. If possible, the convergence time should be reduced by choosing the signal $d(n)$ so as to naximize the ratio of the desired contributions $\mathbf{a}(\theta_0)\hat{R}_{sd}$ to the residual contributions $\hat{\mathbf{R}}_{id}$.

Of particular interest here is that the alternative training signal $d(n)$ can be derived directly from the data. Several specific examples of this technique can be found in Chapter 7 of [28], where the adaptation is implemented using the LMS algorithm. However, the approach taken in Section IV.E is more general and flexible, accommodating multiple signals of interest simultaneously and user-programmability. This approach relies on an auxiliary spatial filter $\mathbf{w}_y$ to generate $\hat{d}(n)$:

$$\hat{d}(n) = \mathbf{w}_y^H \mathbf{y}(n),$$

where $\mathbf{y}(n)$ is the output of a user-programmable transformation applied to $\mathbf{x}(n)$. A key observation is that a carefully selected transformation can have the following properties:

1. $s(n)$ and its transformed version are correlated
2. $\mathbf{i}(n)$ and its transformed version are uncorrelated (although it will be seen in Section IV.E.6 that this property is desirable but not necessary).

Thus, provided that the auxiliary spatial filter $\mathbf{w}_y$ does not reject the ,ransformed version of a $s(n)$, then $\hat{d}(n)$ can be a useful data-derived training signal:

$$\min_{\mathbf{w}_x} MSE(\hat{s}, \hat{d}).$$

Since $\hat{d}(n)$ serves as a training or reference signal, the transformation that yields $\mathbf{y}(n)$ from $\mathbf{x}(n)$ is referred to as the *reference-path transformation.*

An example of this approach applied to cyclostationary signals illustrates the main ideas. If it is known that $s(n)$ is cyclostationary with cycle frequency α, but $\mathbf{i}(n)$ is not, then a reference signal $\hat{d}(n)$ can be derived directly from the received data,

$$\hat{d}(n) = \mathbf{w}_y^H \mathbf{y}(n) \text{ with } \mathbf{y}(n) = \mathbf{x}(n-\tau)\, e^{j2\pi\alpha n}, \tag{13}$$

where $\mathbf{w}_y$ is any vector of spatial filter weights such that $\mathbf{w}_y^H \mathbf{a}(\theta_0) \neq 0$ (e.g., $\mathbf{w}_y = [1, 0, \ldots, 0]^T$ is reasonable for use with omnidirectional sensors or identically-oriented directional sensors), and τ is chosen so as to maximize (if possible) $R_{ss}^{\alpha}(\tau)$. The key observation is that $\hat{d}(n)$ is correlated with $s(n)$ but uncorrelated with $\mathbf{i}(n)$. This observation leads directly to the *Least-Squares SCORE* (LS-SCORE) optimization problem

$$\min_{\mathbf{w}_x} MSE(\hat{s}, \hat{d}) \Longleftrightarrow \max_{\mathbf{w}_x} \left|\hat{\rho}_{\hat{s}\hat{d}}\right|^2 \tag{14}$$

which has solution

$$\mathbf{w}_x = \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \hat{\mathbf{R}}_{\mathbf{x}\hat{d}} = \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha}(\tau) \mathbf{w}_y \tag{15}$$

referred to as the LS-SCORE spatial filter. Alternatively, if $s(n)$ exhibits conjugate cyclostationarity, then $\hat{d}(n)$ can be given by

$$\hat{d}(n) = \mathbf{w}_y^H \mathbf{y}(n) \text{ with } \mathbf{y}(n) = \mathbf{x}^*(n - \tau)\, e^{j2\pi\alpha n}, \tag{16}$$

which yields

$$\mathbf{w}_x = \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \hat{\mathbf{R}}_{\mathbf{x}\hat{d}} = \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \hat{\mathbf{R}}_{\mathbf{xx}^*}^{\alpha}(\tau) \mathbf{w}_y.$$

The performance of LS-SCORE is discussed in [31], where it is shown to provide high-quality estimates of a single SOI in the presence of interference and noise by using only the prior knowledge of the SOI's baud rate (e.g., for a PSK, QAM, or FSK signal) or carrier frequency (e.g., for a BPSK, DSB-AM, or NBFM signal).

IV.E Programmable Canonical Correlation Analysis

Two generalizations immediately follow, from the simple approach just discussed, by minimizing the MSE with respect to both $\mathbf{w}_x$ *and* $\mathbf{w}_y$, and by generalizing the spatial filter vectors to be matrices to accommodate multiple SOIs. In particular, the latter generalization requires a notational change in which $s(n)$, $\hat{s}(n)$, $\mathbf{w}_x$, and $\mathbf{w}_y$ become $\mathbf{s}(n)$, $\hat{\mathbf{s}}(n)$, $\mathbf{W}_x$, and $\mathbf{W}_y$, respectively. An additional benefit is a substantial reduction in convergence time in some cases, as demonstrated in Section IV.E.6. These generalizations are accommodated naturally by the two alternative (but equivalent) approaches to the problem, namely canonical correlation analysis and

constrained conditional maximum likelihood, discussed in Sections IV.E.1 and IV.E.2, respectively. These two approaches are completely equivalent except for one detail that has implications for processing signals in the presence of multipath or correlated interference, as discussed in Section IV.E.3. The two approaches allow substantial programmability in choosing the reference-path transformations, hence the name Programmable Canonical Correlation Analysis. Several possible transformations are discussed in Section IV.E.4. It is then shown in Section IV.E.5 that the Cross-SCORE algorithms (cf. [31, 15, 32] and references therein) are special cases of PCCA that exploit a single cyclostationarity property of the SOIs. The flexibility and utility of PCCA are demonstrated in Section IV.E.6 where it is shown how to simultaneously exploit multiple cyclostationarity properties to accelerate the convergence of Cross-SCORE, how to exploit differences in spectral support of spectrally overlapping signals to separate them, and how to exploit the temporal shift (delay) structure of multipath propagation to mitigate the effects of multipath channel distortion.

IV.E.1 Canonical Correlation Analysis

In the canonical correlation analysis (CCA) (cf. [33, 34, 35]) of two data sets $\mathbf{x}(n)$ and $\mathbf{y}(n)$ that are believed to share some number L of additive components (e.g., signals) jointly denoted by $\mathbf{s}(n)$, it is desired to minimize the mean-squared error between the estimates of $\mathbf{s}(n)$ linearly obtained from each of $\mathbf{x}(n)$ and $\mathbf{y}(n)$. Denoting $\hat{\mathbf{s}}(n) = \mathbf{W}_x^H\mathbf{x}(n)$ and $\hat{\mathbf{d}}(n) = \mathbf{W}_y^H\mathbf{y}(n)$ and constraining $\mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}(0) = \mathbf{I}$ and $\mathbf{R}_{\hat{\mathbf{d}}\hat{\mathbf{d}}}(0) = \mathbf{I}$, then this can be accomplished by minimizing

$$MSE(\hat{\mathbf{s}}, \hat{\mathbf{d}}) = \left\langle \left\| \mathbf{W}_x^H\mathbf{x}(n) - \mathbf{W}_y^H\mathbf{y}(n) \right\|^2 \right\rangle_N \tag{17}$$

subject to the constraints that $\mathbf{W}_x^H\hat{\mathbf{R}}_{\mathbf{xx}}\mathbf{W}_x = \mathbf{I}$ and $\mathbf{W}_y^H\hat{\mathbf{R}}_{\mathbf{yy}}\mathbf{W}_y = \mathbf{I}$.

Equivalently, this can be accomplished (cf. [34]) by first maximizing the magnitude of the cross-correlation coefficient between $\hat{s}_1(n)$ and $\hat{d}_1(n)$, and then by successively maximizing the magnitude of the cross-correlation coefficient between $\hat{s}_m(n)$ and $\hat{d}_m(n)$, subject to the constraint that $\hat{s}_m(n)$

be uncorrelated with $\hat{s}_1(n), \cdots, \hat{s}_{m-1}(n)$ and similarly for $\hat{d}_m(n)$:

$$\max_{\mathbf{w}_{x,m},\mathbf{w}_{y,m}} \frac{\left|\mathbf{w}_{x,m}^H \hat{\mathbf{R}}_{\mathbf{xy}} \mathbf{w}_{y,m}\right|^2}{\left[\mathbf{w}_{x,m}^H \hat{\mathbf{R}}_{\mathbf{xx}} \mathbf{w}_{x,m}\right]\left[\mathbf{w}_{y,m}^H \hat{\mathbf{R}}_{\mathbf{yy}} \mathbf{w}_{y,m}\right]}$$

subject to the orthogonality constraints

$$\mathbf{w}_{x,m}^H \hat{\mathbf{R}}_{\mathbf{xx}} \mathbf{w}_{x,k} = \delta_{m-k} \text{ and } \mathbf{w}_{y,m}^H \hat{\mathbf{R}}_{\mathbf{yy}} \mathbf{w}_{y,k} = \delta_{m-k} \tag{18}$$

for all $k = 1, \ldots, m-1$, where δ_{m-k} is the Kronecker delta function.

Either way, the resulting weight matrices $\mathbf{W}_x$ and $\mathbf{W}_y$ are given by the L most dominant eigenvectors of

$$\mathbf{T}_{xy} = \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \hat{\mathbf{R}}_{\mathbf{xy}} \hat{\mathbf{R}}_{\mathbf{yy}}^{-1} \hat{\mathbf{R}}_{\mathbf{yx}} \tag{19}$$

and

$$\mathbf{T}_{yx} = \hat{\mathbf{R}}_{\mathbf{yy}}^{-1} \hat{\mathbf{R}}_{\mathbf{yx}} \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \hat{\mathbf{R}}_{\mathbf{xy}}, \tag{20}$$

respectively.

In this chapter, since $\mathbf{y}(n)$ is a user-programmable transformation of $\mathbf{x}(n)$, rather than being simply another measured data set, this approach to blind adaptation is referred to as Programmable CCA (PCCA). A general block diagram of the processor is shown in Figure 7.

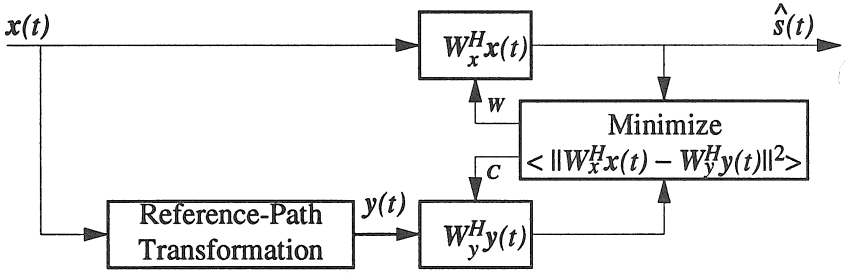

Figure 7: Generic block diagram of the PCCA adaptive processor.

IV.E.2 Constrained Conditional Maximum Likelihood

An alternative interpretation of canonical correlation analysis is discussed here to provide additional insight and justification for the approach.

In [15] a constrained conditional maximum-likelihood (CCML) STF is found for unknown cyclostationary signals in stationary zero-mean complex white Gaussian noise having arbitrary unknown spatial covariance matrix. Since the publication of [15], a more complete interpretation, described here, has been developed.

The derivation of the algorithm consists of two steps: 1) the CCML STF is developed to obtain a set of optimal (in the CCML sense) data-derived training signals for the SOIs; and 2) the resulting estimates of the training signals replace the ideal training signal in (11) to yield the spatial filter weights used to estimate the SOI waveforms. The two-step procedure is adopted because it will be seen that the data-derived training signals cannot be high quality estimates of the SOIs, but they *can* be effective training signals.

Step 1: CCML Data-Derived Training Signal

Before becoming embroiled in the details of the algorithm, the justifications for the qualifiers *constrained* and *conditional* are presented.

To exploit cyclostationarity, the reference-path transformation which is used to generate the data-derived training signals must be constrained to be an LCL-PTV STF. However, it is noted here (and discussed in more detail in Section IV.E.4) that other signal properties could also be exploited by simply choosing $\mathbf{y}(n)$ to be some other transformation of $\mathbf{x}(n)$. For example, the Kronecker products $\mathbf{x}(n)\otimes\mathbf{x}(n-\tau)$, $\mathbf{x}(n)\otimes\mathbf{x}(n-\tau_1)\otimes\mathbf{x}(n-\tau_2)$, variations of these involving conjugation of some terms, and further variations in which frequency shifts are included, could be useful transformations of $\mathbf{x}(n)$ that yield data-derived training signals. These examples could be appropriate for exploitation of higher-order stationarity or higher-order cyclostationarity. These statistical properties are discussed in depth in [36].

The ML problem is conditional (on the unknown waveforms of the cyclostationary signals) because noise can often be accurately modeled as being stationary and Gaussian, whereas communication signals of interest are almost never Gaussian nor stationary, and their probability distribution functions are typically unknown or virtually intractable to work with. In particular, in some signal interception and signal classification applications, almost nothing might be known about the signals of interest. Also, the con-

ditional ML problem easily admits the user-programmable constraints just described.

The vector of sampled complex envelopes at the output of an M-element sensor array is denoted by $\mathbf{x}(n)$, which can be modeled under the narrowband assumption by

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{i}(n)$$

where $\mathbf{s}(n)$ denotes the vector of L unknown signals, $\mathbf{A}$ denotes the $M \times L$ matrix of unknown array response vectors of the signals, and $\mathbf{i}(n)$ denotes the stationary temporally-white Gaussian noise. The estimates of the data-derived training signals are constrained to be the linear combinations

$$\hat{\mathbf{d}}(n) = \mathbf{W}_y^H \mathbf{y}(n) \tag{21}$$

where $\mathbf{y}(n)$ is almost any specific realization of the general LCL-PTV form in (8) or another suitable transformation as discussed above. It is necessary to constrain the processor structure to something other than the usual linear processor $\hat{\mathbf{d}}(n) = \mathbf{W}^H \mathbf{x}(n)$ since the resulting solution for $\mathbf{W}$ in this case could be any arbitrary matrix having column-rank equal to L, as will be seen in Section IV.E.4. Furthermore, for the same reason, $\mathbf{x}(n)$ cannot appear in unmodified form in $\mathbf{y}(n)$. Thus, $\hat{\mathbf{d}}(n)$ cannot in general be a high quality estimate of $\mathbf{s}(n)$, but it will be seen to be a useful training signal.

The constrained conditional likelihood function for the estimates $\hat{\mathbf{A}}$, $\hat{\mathbf{D}}$, and $\hat{\mathbf{R}}_{\mathbf{ii}}$ of $\mathbf{A}$, $\mathbf{S} \triangleq \{\mathbf{s}(n) : 1 \le n \le N\}$, and $\mathbf{R}_{\mathbf{ii}} \triangleq \langle \mathbf{i}(n)\mathbf{i}(n)^H \rangle_\infty$, respectively, is given by

$$L(\hat{\mathbf{A}}, \hat{\mathbf{D}}, \hat{\mathbf{R}}_{\mathbf{ii}}) = \left|\pi \hat{\mathbf{R}}_{\mathbf{ii}}\right|^{-N} \exp\left\{-N \left\langle \mathbf{e}(n)^H \hat{\mathbf{R}}_{\mathbf{ii}}^{-1} \mathbf{e}(n) \right\rangle_N\right\}$$

where $\mathbf{e}(n) = \mathbf{x}(n) - \hat{\mathbf{A}}\hat{\mathbf{d}}(n)$.

By application of results from matrix calculus [16] and complex gradients of non-analytic functions [37], it can be shown [38] that the CCML estimates are given by

$$\begin{aligned}
\hat{\mathbf{R}}_{\mathbf{ii}}^{(ML)} &= \hat{\mathbf{R}}_{\mathbf{xx}} - \hat{\mathbf{A}}^{(ML)} \hat{\mathbf{R}}_{\mathbf{x}\hat{\mathbf{d}}}^H - \hat{\mathbf{R}}_{\mathbf{x}\hat{\mathbf{d}}} \left[\hat{\mathbf{A}}^{(ML)}\right]^H + \hat{\mathbf{A}}^{(ML)} \hat{\mathbf{R}}_{\hat{\mathbf{d}}\hat{\mathbf{d}}} \left[\hat{\mathbf{A}}^{(ML)}\right]^H, \\
\hat{\mathbf{A}}^{(ML)} &= \hat{\mathbf{R}}_{\mathbf{x}\hat{\mathbf{d}}} \hat{\mathbf{R}}_{\hat{\mathbf{d}}\hat{\mathbf{d}}}^{-1},
\end{aligned}$$

and $\hat{\mathbf{d}}(n)$ is given by (21). The matrix $\mathbf{W}_y$ has L columns given by any full-column-rank linear combination of the L most dominant eigenvectors $\mathbf{E}$ of the matrix $\mathbf{T}_{yx}$ defined by

$$\mathbf{T}_{yx} = \hat{\mathbf{R}}_{\mathbf{yy}}^{-1}\hat{\mathbf{R}}_{\mathbf{yx}}\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{xy}}. \tag{22}$$

That is, $\mathbf{W}_y = \mathbf{EQ}$ for any full-rank $L \times L$ matrix $\mathbf{Q}$.

If, in addition to the constraint on the processor structure, it is required that the estimated signals be uncorrelated, $\mathbf{W}_y^H\hat{\mathbf{R}}_{\mathbf{yy}}\mathbf{W}_y = \mathbf{I}$, then the likelihood is maximized *only* by the L most dominant eigenvectors of $\mathbf{T}_{yx}$, instead of any linear combination thereof; that is, $\mathbf{W}_y = \mathbf{E}$. A mathematical result related to the solution (22) was proposed in [39] and proven in [33]. Also, the result of this section includes that in [40] as the special case for which one signal is present ($L = 1$) and $\mathbf{y}(n) = \mathbf{x}^*(n)$.

An alternate proof of (22) can be obtained by noting that maximizing $L(\hat{\mathbf{A}}^{(ML)}, \hat{\mathbf{D}}, \hat{\mathbf{R}}_{\mathbf{ii}}^{(ML)})$ is equivalent to maximizing

$$f(\mathbf{W}_y) = \prod_{m=1}^{M} \lambda_m(\mathbf{W}_y^H\hat{\mathbf{R}}_{\mathbf{yx}}\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{xy}}\mathbf{W}_y, \mathbf{W}_y^H\hat{\mathbf{R}}_{\mathbf{yy}}\mathbf{W}_y),$$

where $\lambda_m(\cdot,*)$ denotes the mth generalized eigenvalue of the matrix pair $(\cdot,*)$, and then applying the Poincaré Separation Theorem for generalized eigenvalues of a pair of Hermitian matrices (e.g., [41]). In any case, the additional constraint that the data-derived training signals be uncorrelated yields a unique solution $\mathbf{W}_y$ that can also be obtained from the CCA perspective as discussed in Section IV.E.1.

Step 2: Using the Data-Derived Training Signal

In this step, $\hat{\mathbf{d}}(n)$ is used as a training signal to minimize $MSE(\hat{\mathbf{s}}, \hat{\mathbf{d}})$.

Direct substitution of $\hat{\mathbf{d}} = (\mathbf{EQ})^H\mathbf{y}(n)$ into

$$\mathbf{W}_x = \hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{x}\hat{\mathbf{d}}}$$

and algebraic manipulation reveal that $\mathbf{W}_x$ can be any full-rank linear combination of the L most dominant eigenvectors of $\mathbf{T}_{xy}$,

$$\mathbf{T}_{xy} = \hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{xy}}\hat{\mathbf{R}}_{\mathbf{yy}}^{-1}\hat{\mathbf{R}}_{\mathbf{yx}}. \tag{23}$$

As before, if the elements of $\hat{\mathbf{s}}(n) = \mathbf{W}_x^H \mathbf{x}(n)$ are costrained to be uncorrelated, then the additional constraint $\mathbf{W}_x^H \hat{\mathbf{R}}_{\mathbf{xx}} \mathbf{W}_x = \mathbf{I}$ implies that $\mathbf{W}_x$ is exactly the L most dominant eigenvectors of $\mathbf{T}_{xy}$.

IV.E.3 Contrast Between CCML and CCA

Clearly, $\mathbf{T}_{yx}$ in (20) and $\mathbf{T}_{yx}$ in (22) are identical, and similarly for $\mathbf{T}_{xy}$, which shows that the CCML and CCA frameworks yield identical algorithms when the signal estimates are constrained to be uncorrelated. Thus, a primary utility of having these two approaches is simply a conceptual one, providing the engineer with some flexibility in how the application of interest is conceptualized.

The CCML conceptual framework is obviously applicable when the spatial covariance of the interference and noise $\mathbf{i}(n)$ is unknown, but it requires that $\mathbf{i}(n)$ be stationary, white, and Gaussian. In contrast, the CCA conceptual framework is not obviously optimum when $\mathbf{R}_{\mathbf{ii}}(0)$ is unknown (although from its equivalence with CCML we know it is), but its simple least-squares formulation does not require that $\mathbf{i}(n)$ be stationary, white, or Gaussian.

The other contrast of interest here is that CCML can be formulated without the constraint that the signal estimates be uncorrelated, whereas CCA cannot so easily admit this flexibility. In many applications, uncorrelated signal estimates are required. However, in the presence of multipath it is conjectured (based on results for Cross-SCORE in [31]) that multiple signal estimates, one for each propagation path of a SOI, are obtained by CCML/CCA. In this case, the desired action of the processor is first to reject other SOIs, interference, and noise from the multiple signal estimates of a single SOI, and then to linearly combine these multiple estimates to mitigate the multipath. Thus, a conceptual benefit of the CCML framework, without the constraint that signal estimates be uncorrelated, is that linearly combining the multipath estimates is the optimal type of processing (in the CCML sense). Note, however, that the CCML framework does not suggest *how* to choose the weights in the linear combiner. Another adaptive processor (e.g., one of the blind equalization algorithms reviewed in [42], or either of the algorithms in Section V) could be configured to perform this task.

IV.E.4 Reference-Path Transformations for PCCA

Here the restriction on the choice of $\mathbf{y}(n)$ mentioned earlier is discussed, and it is noted that the CCML/CCA framework can exploit a wider variety of signal properties than just those related to cyclostationarity.

From (17) it can be seen that $\mathbf{y}(n)$ should not contain $\mathbf{x}(n)$ as a literal element, since any solution of the form

$$\mathbf{W}_y = \begin{bmatrix} \mathbf{W}_x \\ \mathbf{W} \end{bmatrix}$$

would minimize (17); this observation implies that this CCML/CCA framework cannot directly yield a blind adaptive Cyclic Wiener spatio-temporal filter.

As alluded to earlier, the CCML/CCA framework need not use an LCL-PTV transformation to obtain $\mathbf{y}(n)$. Indeed, the transformation is entirely up to the user, provided that the restriction just discussed is met. To emphasize this flexibility, the *programmable canonical correlation analyzer* (PCCA) is proposed, wherein the transformation used to obtain $\mathbf{y}(n)$ is completely programmable by the user. Thus, the PCCA can use many types of signal properties to distinguish between desired signals and interference. A non-exhaustive list of transformations is proposed here:

1. $\mathbf{y}(n)$ is a frequency-shifted (by α) and delayed (by τ) version of $\mathbf{x}(n)$ or $\mathbf{x}^*(n)$, which yield the Cross-SCORE and conjugate Cross-SCORE algorithms, respectively [31] (cf. Section IV.E.5); this defines as SOIs those signals that exhibit cyclostationarity or conjugate cyclostationarity with cycle frequency α, and can be generalized to multiple frequency shifts, multiple delays, and pre-filtering.

2. $\mathbf{y}(n)$ is the output of a band-stop (or band-pass) LTI filter applied to $\mathbf{x}(n)$; this defines as SOIs those signals that have spectral support outside (or inside) the stop-band (or pass-band), and can be generalized to more complicated regions of spectral support.

3. $\mathbf{y}(n)$ is a delayed version of $\mathbf{x}(n)$; this defines as SOIs those signals for which the coherence time is greater than or equal to τ.

4. $\mathbf{y}(n)$ is the output of a temporal interval-stop (gating) device applied to $\mathbf{x}(n)$; this defines as SOIs those signals that are active outside the stop intervals.

5. $\mathbf{y}(n)$ is the narrowband (or wideband) output of an adaptive spectral-line enhancer applied to $\mathbf{x}(n)$; this defines as SOIs those signals that are relatively narrowband (or wideband).

6. $\mathbf{y}(n)$ is the enhanced (or degraded) output of a spectral-correlation enhancer (a blind adaptive LCL-PTV filter) applied to $\mathbf{x}(n)$; this defines as SOIs those signals that exhibit (or don't exhibit) cyclostationarity at a specified cycle frequency α.

7. $\mathbf{y}(n)$ is the constant modulus (or non-constant modulus) output of an LTI filter (or LTI canceller) adapted by the constant modulus (CM) algorithm (CMA); this defines as SOIs those signals that have (or do not have) constant modulus.

8. $\mathbf{y}(n)$ is the output of a demodulation/remodulation device that is applied to $\mathbf{x}(n)$ and is structured to select FM, PM, FSK, or PSK signals.

9. $\mathbf{y}(n)$ is the output of a nonlinear transformation such as $\mathbf{x}(n) \odot \mathbf{x}(n) \odot \mathbf{x}(n)$, $\mathbf{x}(n) \odot \mathbf{x}(n) \odot \mathbf{x}^*(n)$, $\mathbf{x}(n) \otimes \mathbf{x}(n) \otimes \mathbf{x}(n)$, $\mathbf{x}(n) \otimes \mathbf{x}(n) \otimes \mathbf{x}^*(n)$, or time-variant non-memoryless generalizations thereof, where $\odot$ denotes the elementwise product and $\otimes$ denotes the Kronecker product; this defines as SOIs those signals that have the higher-order stationarity or higher-order cyclostationarity properties selected for by the chosen transformation.

Examples of the first three transformations are used in the computer simulations of the PCCA described in Section IV.E.6.

IV.E.5 SCORE Algorithms

It is noted here that the cross-SCORE and conjugate cross-SCORE [31] blind adaptive beamformers are specific realizations of the CCML/CCA beamformer. Two special choices of $\mathbf{y}(n)$ merit consideration. The first

choice is $\mathbf{y}(n) = \mathbf{x}(n-\tau)e^{j2\pi\alpha n}$ for which each of $\mathbf{W}_x^H\mathbf{x}(n)$ and $\mathbf{W}_y^H\mathbf{y}(n)$ is a CCML and CCA beamformer. Also, $\mathbf{W}_x^H\mathbf{x}(n)$ in this case is exactly the Cross-SCORE blind adaptive STF that is found by solving

$$\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha}(\tau)\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha}(\tau)^H\mathbf{w}_m = \lambda_m\mathbf{w}_m \tag{24}$$

for $m = 1, \ldots, L$, where

$$\hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha}(\tau) \triangleq \left\langle \mathbf{x}(n)\mathbf{x}(n-\tau)^H e^{-j2\pi\alpha n} \right\rangle$$

is the cyclic autocorrelation matrix of $\mathbf{x}(n)$ for cycle frequency α and lag τ. The block diagram of the signal processor for this case is shown in Figure 8, although the general CCML/CCA processor can be obtained by replacing the dash-boxed section with the diagram for the appropriate LCL-PTV transformation of $\mathbf{x}(t)$ (i.e., the relevant part of Figure 5).

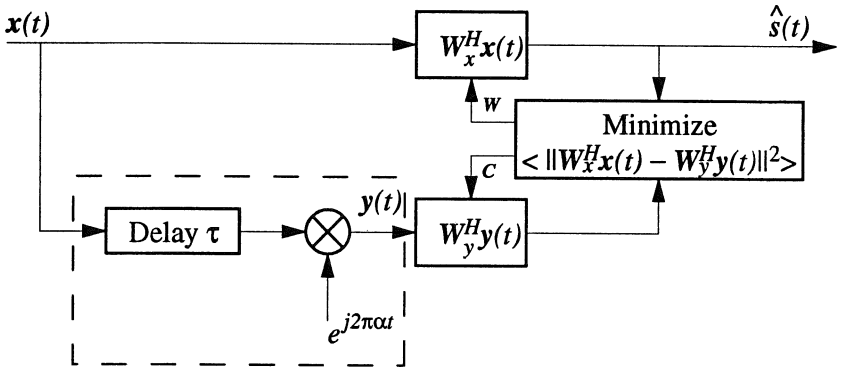

Figure 8: Block diagram of Cross-SCORE processor.

Similarly, the second choice is $\mathbf{y}(n) = \mathbf{x}(n-\tau)^* e^{j2\pi\alpha n}$ for which each of $\mathbf{W}_x^H\mathbf{x}(n)$ and $\mathbf{W}_y^H\mathbf{y}(n)$ is a CCML and CCA beamformer. Also, $\mathbf{W}_x^H\mathbf{x}(n)$ in this case is exactly the *conjugate* Cross-SCORE blind adaptive beamformer that is found by solving

$$\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{xx}^*}^{\alpha}(\tau)\hat{\mathbf{R}}_{\mathbf{x}^*\mathbf{x}^*}^{-1}\hat{\mathbf{R}}_{\mathbf{xx}^*}^{\alpha}(\tau)^H\mathbf{w}_m = \lambda_m\mathbf{w}_m \tag{25}$$

for $m = 1, \ldots, L.$, where

$$\hat{\mathbf{R}}_{\mathbf{xx}^*}^{\alpha}(\tau) \triangleq \left\langle \mathbf{x}(n)\mathbf{x}(n-\tau)^T e^{-j2\pi\alpha n} \right\rangle$$

is the cyclic conjugate correlation matrix of $\mathbf{x}(n)$ for cycle frequency α and lag τ.

The observation that cross-SCORE and conjugate cross-SCORE are CCA beamformers was first made in [43] where it was used only to motivate the Cyclic Correlation Significance Test for estimating the number of signals having a specified cycle frequency.

Performance Attributes

In the following, assume that L_α SOIs have cycle frequency α. In [31] it is shown that the L_α most dominant eigenvectors $\mathbf{w}_1, \cdots, \mathbf{w}_{L_\alpha}$ in (24), which correspond to the L_α highest stationary points of $\left|\hat{\rho}_{\hat{s}d}\right|^2$ (i.e., locations at which the gradients with respect to $\mathbf{w}_x$ and $\mathbf{w}_y$ are both equal to zero), are spatial filter weights that can be used to obtain estimates of the L_α signals having cycle frequency α and lag τ. It is also shown that each of these signal estimates has nearly the same SINR as the corresponding MMSE signal estimate, provided that a sufficient number of data samples are available for adaptation, that input SNR is not too low, that fewer signals than sensors arrive at the array, the array response vectors are linearly independent, no coherent multipath is present, and that each signal has a different feature strength. The condition on the feature strengths can be easily derived in the absence of noise and effects due to finite-averaging by simplifying (24) to obtain

$$\mathbf{R}_{\mathbf{ss}}^{\alpha}(\tau)\mathbf{R}_{\mathbf{ss}}^{-1}\mathbf{R}_{\mathbf{ss}}^{\alpha H}(\tau)\mathbf{g} = \lambda\mathbf{R}_{\mathbf{ss}}\mathbf{g} \tag{26}$$

where $\mathbf{g} = \mathbf{A}(\Theta)^H\mathbf{w}$ and it is assumed that the L_α columns of $\mathbf{A}(\Theta)$ are linearly independent, and noticing that perfect signal separation implies that the eigenvectors $[\mathbf{g}_1 \cdots \mathbf{g}_{L_\alpha}]$ are a permutation of the identity matrix $\mathbf{I}$, which occurs only when $\mathbf{R}_{\mathbf{ss}}$ and $\mathbf{R}_{\mathbf{ss}}^{\alpha}(\tau)$ are diagonal and $\left|\rho_{s_n s_n}^{\alpha}(\tau)\right| \neq \left|\rho_{s_m s_m}^{\alpha}(\tau)\right|$ for all $m \neq n$. Thus, the SOIs must be uncorrelated and have distinct feature strengths for perfect signal separation in Cross-SCORE. In practice, noise is present (so (26) does not apply), and thus signal separation is less than perfect. Nonetheless, signal separation has been observed in simulations to be adequate, provided that the input SNR (but not necessarily the input SINR) is greater than 0 dB.

In practical terms, Cross-SCORE avoids the need for a known training signal, albeit at the expense of increased convergence time relative to the

MMSE method. The benefits of this to cellular communication system design are discussed in [44, 6]. For example, using only the knowledge of the baud rate of desired bauded digital communications signals (e.g., PSK, QAM, etc.) or only the doubled carrier of communications signals that exhibit cyclic conjugate correlation at the doubled carrier frequency (e.g., BPSK, DSB-AM, NBFM), the cross-SCORE and conjugate cross-SCORE algorithms, respectively, can achieve nearly the maximum SINR attainable by an LTI beamformer [31].

Although empirical results on the performance of cross-SCORE for finite numbers of time samples are available in [31, 45, 46, 47, 48, 15], no analytical results are available, such as the mean and variance of the output SINR as a function of the number N of time samples. It is likely that existing results in the multivariate statistics literature on the behavior of canonical correlation analyzers could be applicable to the performance analysis of Cross-SCORE and the CCML/CCA generalizations thereof.

Simulation Results

A brief illustration of the performance of LS-SCORE and Cross-SCORE is given here. In the computer simulations, 100 independent trials are performed and the average SINR at the output of the SCORE processor is computed as a function of the number N of time samples used for adaptation. In each trial, independent BPSK signals having baud rates 0.25 and 0.2, carrier frequencies 0 and 0.1, and angles of arrival 0 and 20 degrees, respectively, arrive at a 4-element ULA in the presence of additive white Gaussian noise that is uncorrelated from sensor to sensor. Both signals use Nyquist-shaped pulses with 100% excess bandwidth and have SNR equal to 10 dB. LS-SCORE and Cross-SCORE are both applied to the data, first with $\alpha = 0.25$ and then using the conjugate form with $\alpha = 0$; that is, first the baud rate feature and then the doubled-carrier feature of the first signal is exploited. Similar results (not shown here) are obtained using the cycle frequencies of the second BPSK signal. As shown in Figure 9, Cross-SCORE adapts more quickly than LS-SCORE because Cross-SCORE adapts both $\mathbf{w}_x$ and $\mathbf{w}_y$. Convergence is much quicker for doubled-carrier features than for baud rate features because the feature strength at the doubled carrier is much stronger than at the baud rate; a stronger feature enables more re-

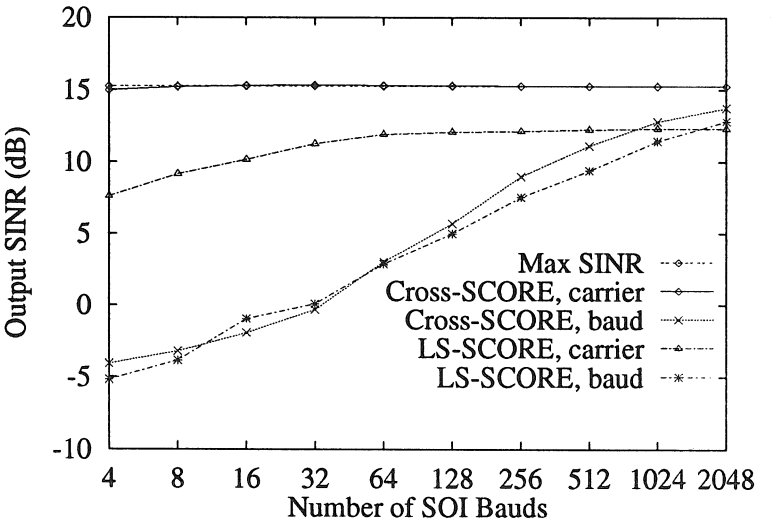

Figure 9: Average output SINR of LS-SCORE and Cross-SCORE for different feature choices.

liable discrimination between the desired signal and the interference, thus reducing convergence time. This is analogous to the fact that the MMSE method adapted with a slightly noisy training signal converges more quickly than the same method adapted with a very noisy training signal. Indeed, the results in Figure 9 show that Cross-SCORE, using the doubled carrier feature, converges at nearly the same rate as the MMSE method, which uses perfect knowledge of the signal waveform.

Other performance results are given in Section IV.E.6, and in [45, 46, 47, 31], where it is also shown that, in certain environments in which a baud-rate feature is being used, Cross-SCORE can converge much more quickly than LS-SCORE, in contrast to the results shown in Figure 9.

IV.E.6 Performance of Example Realizations of PCCA

Here the performance attributes of four different realizations of the PCCA beamformer are briefly illustrated via computer simulations. In the first

three examples, the PCCA is applied to problems in which two independent signals arrive at the sensor array, and the objective is to separate them from each other. In the first example, the PCCA structure is used to accelerate the convergence of the Cross-SCORE algorithm as it extracts an estimate of one BPSK signal in the presence of an interfering BPSK signal having a different baud rate. In the second example, one of the signals is replaced by a narrowband Gaussian interferer, and the reference-path transformation is simply a bandstop filter, where the stop band coincides with the spectral support of the Gaussian signal. In the third example, the reference-path transformation from the second example is replaced by a unit-delay. In the fourth example, a single SOI arrives at the array and a delayed version of the SOI arrives from a different direction; in this example, the objective is to separate the signals corresponding to the two paths, and the reference-path transformation is again simply a delay. In all of the examples, the average output SINR is obtained by averaging the output SINRs from one hundred independent trials.

Cyclostationarity Exploitation

In the simulated environment, a 4-element ULA having half-wavelength sensor-spacing receives 2 BPSK signals in the presence of stationary complex white Gaussian noise. The two signals are spectrally overlapping and have equal signal to noise ratios (SNRs) but arrive from different directions (0 and 20 degrees) and have different baud rates (0.25 and 0.33 times the sampling rate). Two different input SNRs (0 dB and 10 dB) are considered. In each case, the average output SINR is evaluated for different numbers of cycle frequencies α and filters $h_k(n) = \delta_{n-\tau_k}$ in (8). In particular, 6 combinations of one or two values of α (0.25 and -0.25) and 3 values of τ (0, 1, and -1) are considered; since these values of α correspond to cycle frequencies of the BPSK signal with baud rate 0.25 (because the cyclic autocorrelations at these cycle frequencies and lags τ are nonzero), this signal is extracted by the spatial filter. If α were chosen to be 0.33 and/or -0.33, then the other BPSK signal would be chosen. If only one value of α and one value of τ are used, then the CCA beamformer is equivalent to the cross-SCORE beamformer. As seen in Figures 10 and 11, the use of multiple values of α

and τ substantially improves convergence, with between only one-half and one-eighth as many data samples required to achieve the same performance as cross-SCORE.

Bandstop Filtering

In this simulation, the array receives a BPSK signal and a narrowband Gaussian signal. The BPSK signal has a baud rate of 0.25, zero carrier offset, 10 dB SNR, and direction of arrival of 0 degrees. The narrowband Gaussian signal consists of white Gaussian noise passed through a filter with passband $[0, 0.1]$, and arrives from 20 degrees. In the PCCA, the reference-path transformation is simply a bandstop filter that rejects the passband of the Gaussian signal. Since this also causes irreparable damage to the BPSK signal, the bandstop filter is unsuitable as the sole interference rejection device. However, it does allow the PCCA to distinguish between the signals, and thereby to reject either signal by spatial filtering alone, as demonstrated in Figures 12 and 13, which show the output SINR obtained by PCCA as a function of the number of data samples and the SNR of the Gaussian signal (which ranges from 10 dB to 50 dB). For both signals, the output SINR converges to the maximum attainable during the adaptation period considered. Of the four spatial filters found by PCCA for this array (i.e., one per antenna), the two obtained from the most dominant eigenvectors rejected both signals, and the least dominant eigenvector extracted the Gaussian signal (which is predictable from the fact that the reference-path transformation rejects this signal), and the next-to-least dominant eigenvector extracts the BPSK signal. As of this writing, this ordering of the solutions is not completely understood. In some applications, relatively simple classification algorithms could be applied to the multiple output signals to determine which are of interest.

Delay: Signal Separation

In this simulation, the signals are exactly the same as in the previous simulation, although the reference-path transformation is simply a unit-sample delay instead of a bandstop filter. The corresponding results are shown in Figures 14 and 15. In this simulation, the most dominant eigenvector

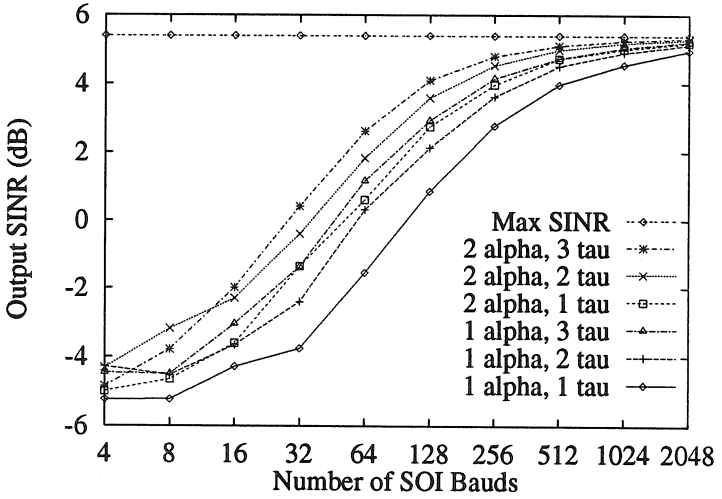

Figure 10: Average output SINR for different numbers of cycle frequencies and lags. The input SNR is 0 dB.

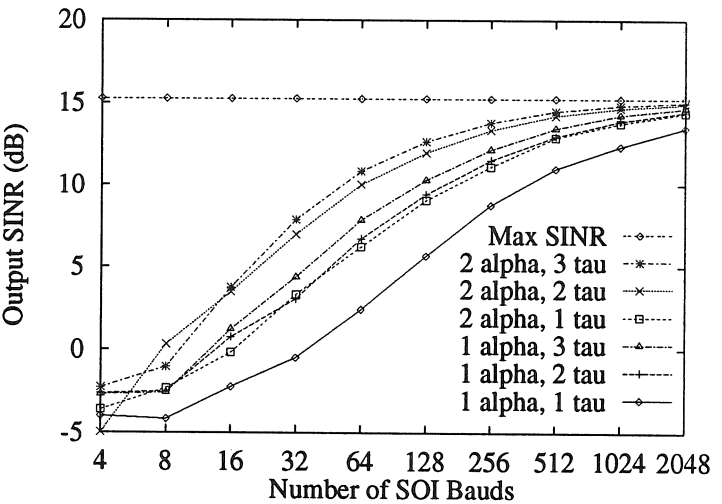

Figure 11: Average output SINR for different numbers of cycle frequencies and lags. The input SNR is 10 dB.

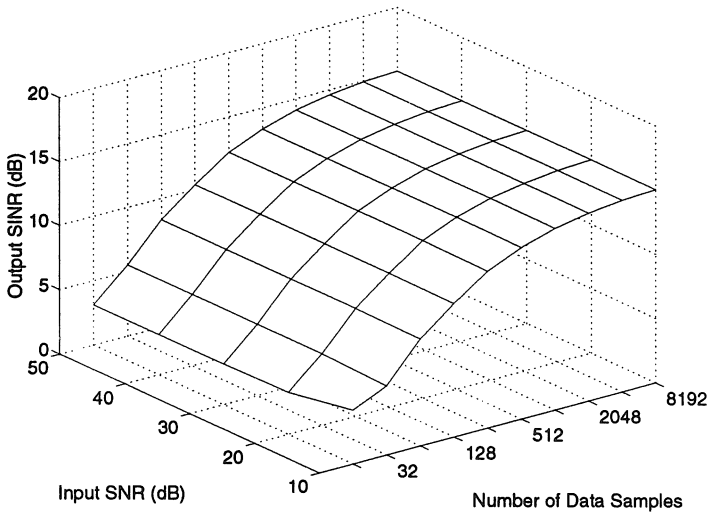

Figure 12: Output SINR for BPSK signal obtained by PCCA using bandstop reference-path transformation.

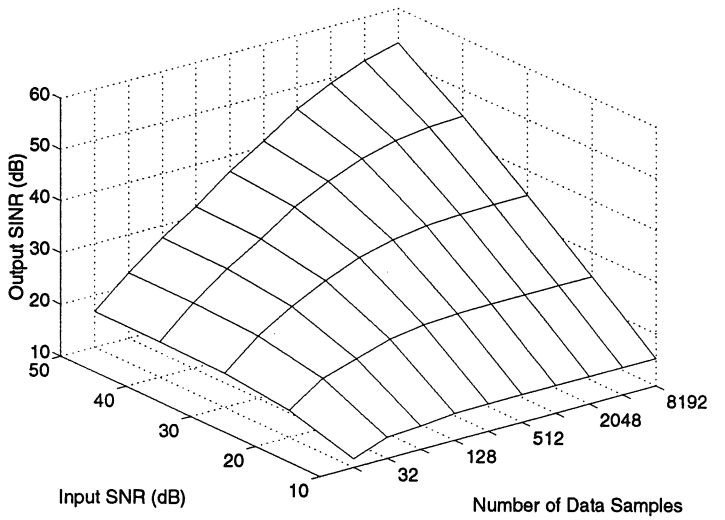

Figure 13: Output SINR for Gaussian interference obtained by PCCA using bandstop reference-path transformation.

found by PCCA extracts the Gaussian signal (because it has the longest coherence time) , the next-most dominant eigenvector extracts the BPSK signal, and the least dominant eigenvectors reject the signals in favor of the noise (which, being white, is uncorrelated with the delayed version of itself, resulting in a feature strength of zero).

Delay: Multipath Mitigation

In this simulation, a single BPSK SOI having baud rate 0.25 arrives at the array from two different directions, simulating a two-ray multipath propagation environment. Each arrival is given a random carrier phase that is uniformly distributed on $[0, 2\pi)$ radians. One arrival is designated as the direct path and has 10 dB SNR; the other arrival is designated as the reflected path; it has an SNR that is randomly chosen at each trial from the range 5 to 15 dB, and it is delayed by a fixed positive amount relative to the direct path. The range of delays is 0.5 to 3 samples in increments of one-half sample. The reference-path transformation used by PCCA is simply a unit-sample advance, which causes the most dominant eigenvector to select the direct path, whereas a unit-sample delay causes the most dominant eigenvector to select the reflected path. The output SINR for the most dominant eigenvector is shown in Figure 16, and a typical antenna pattern (obtained for multipath delay equal to one sample) shown in Figure 17 confirms that the reflected signal is being rejected, rather than coherently combined with the direct path (which would result in multipath distortion and thus require an adaptive filter to compensate). As the multipath delay increases beyond the advance value used in the reference-path transformation, the output SINR decreases, which suggests that a reasonable estimate of the range of multipath delays is needed for this method to work well.

IV.F Related Algorithms

Two additional blind adaptive methods are summarized here. The first is intended for use where the narrowband approximation does not hold, and it can be interpreted as a frequency-domain version of the PCCA. The second is not a PCCA method but the algorithm has a similar structure to

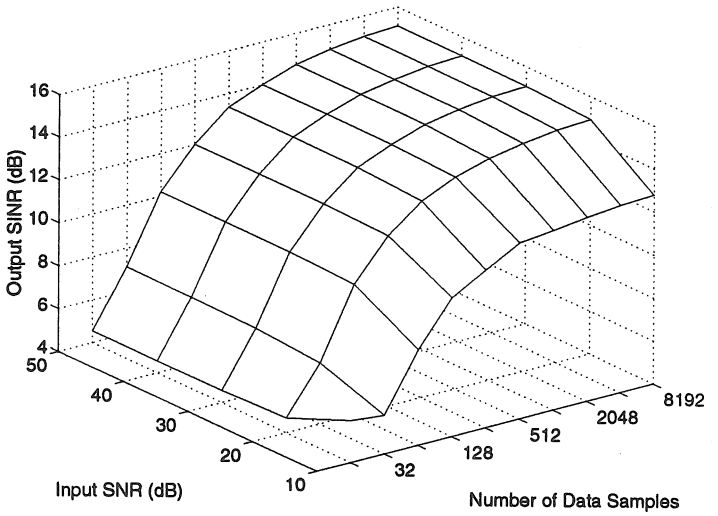

Figure 14: Output SINR for BPSK signal obtained by PCCA using unit-delay reference-path transformation.

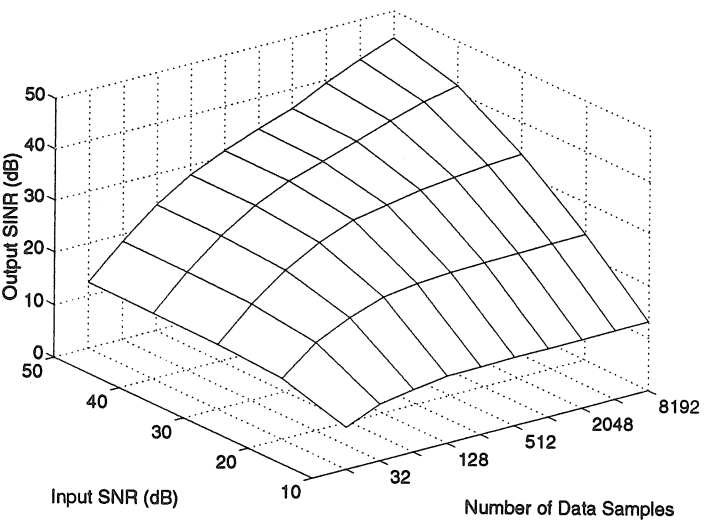

Figure 15: Output SINR for Gaussian signal obtained by PCCA using unit-delay reference-path transformation.

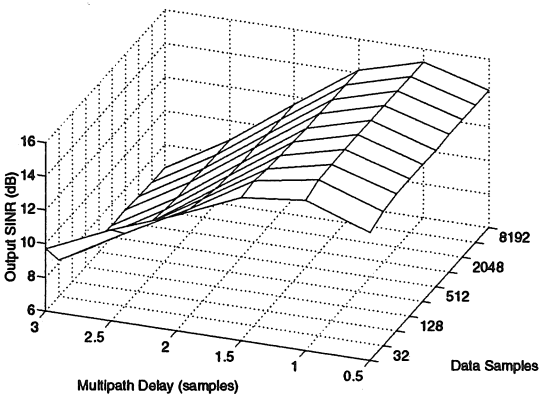

Figure 16: Output SINR for most dominant eigenvector of PCCA using a unit-sample advance to reject a multipath reflection from 20 degrees.

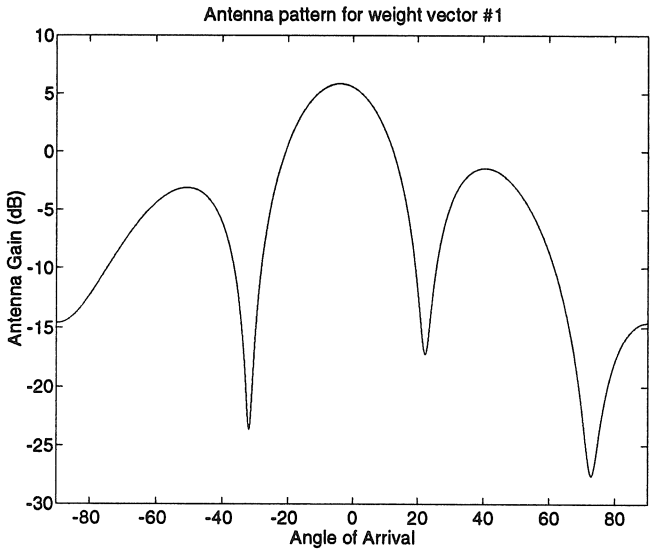

Figure 17: Typical antenna pattern showing the rejection of the multipath reflection that arrives from 20 degrees.

the PCCA eigenvalue problem, and the performance attributes are simlar, with the exception that SOIs are separated on the basis of both feature strength *and* phase, rather than feature strength only as in PCCA.

IV.F.1 Wideband SCORE

SCORE can also be applied to wideband environments, in which the received data $\mathbf{x}(t)$ can be modeled in terms of finite-time Fourier transforms (FTFT):

$$\begin{aligned} \tilde{\mathbf{x}}(t,f) &\approx \sum_{l=1}^{L_\alpha} \tilde{\mathbf{a}}(\theta_l,f)\tilde{s}_l(t,f) + \tilde{\mathbf{i}}(t,f) \\ &= \tilde{\mathbf{A}}(\Theta,f)\tilde{\mathbf{s}}(t,f) + \tilde{\mathbf{i}}(t,f), \end{aligned} \tag{27}$$

where $\tilde{\mathbf{x}}(t,f) \triangleq \left(1/\sqrt{\Delta}\right) \int_{t-\Delta/2}^{t+\Delta/2} \mathbf{x}(t)\, e^{-j2\pi ft}\, dt$ is the FTFT of the received data, $\tilde{\mathbf{a}}(\theta,f)$ is the transfer function of the array for a signal arriving from angle θ, $\tilde{\mathbf{s}}(t,f)$ is the FTFT of the L_α cyclostationary signals impinging on the array from angles $\theta_1, \ldots, \theta_{L_\alpha}$, respectively, and having cycle frequency α, and $\tilde{\mathbf{i}}(t,f)$ is the FTFT of all other signals and noise that do not have cycle frequency α. The approximation in (27) holds well if the FTFT integration time Δ is much greater than the duration of the impulse response of the array. Similarly, the FTFT of the lth extracted signal is given by $\tilde{s}_l(t,f) = \tilde{\mathbf{w}}_{x,l}(t,f)^H \tilde{\mathbf{x}}(t,f)$ for $l = 1, \ldots, L$.

Application of the SCORE concept to this problem can be pursued by re-expressing it as a CCA problem. Notice that the FTFT $\tilde{\mathbf{y}}(t,f)$ of the auxiliary signal defined by $\mathbf{y}(t) \triangleq \mathbf{x}(t)\, e^{j2\pi\alpha t}$ is given by

$$\begin{aligned} \tilde{\mathbf{y}}(t,f) &= \tilde{\mathbf{x}}(t,f-\alpha) \\ &= \tilde{\mathbf{A}}(\Theta,f-\alpha)\,\tilde{\mathbf{s}}(t,f-\alpha) + \tilde{\mathbf{i}}(t,f-\alpha). \end{aligned} \tag{28}$$

Since $\mathbf{s}(t)$ is cyclostationary with cycle frequency α, $\tilde{\mathbf{s}}(t,f-\alpha)$ is correlated (shares a common component) with $\tilde{\mathbf{s}}(t,f)$. Consequently, a solution based on canonical correlation analysis can be found for each value of f:

$$\hat{\mathbf{S}}_{\mathbf{xy}}(f)\,\hat{\mathbf{S}}_{\mathbf{yy}}(f)^{-1}\,\hat{\mathbf{S}}_{\mathbf{yx}}(f)\tilde{\mathbf{w}}_{x,l}(f) = \lambda_l(f)\hat{\mathbf{S}}_{\mathbf{xx}}(f)\tilde{\mathbf{w}}_{x,l}(f) \tag{29}$$

$$\hat{\mathbf{S}}_{\mathbf{yx}}(f)\,\hat{\mathbf{S}}_{\mathbf{xx}}(f)^{-1}\,\hat{\mathbf{S}}_{\mathbf{xy}}(f)\tilde{\mathbf{w}}_{y,l}(f) = \lambda_l(f)\hat{\mathbf{S}}_{\mathbf{yy}}(f)\tilde{\mathbf{w}}_{y,l}(f), \tag{30}$$

where $\hat{\mathbf{S}}_{\mathbf{xy}}(f)$ is the estimated cross-correlation matrix of $\tilde{\mathbf{x}}(t, f)$ and $\tilde{\mathbf{y}}(t, f)$ (and can also be interpreted as the cross-periodogram matrix of $\mathbf{x}(t)$ and $\mathbf{y}(t)$).

Other details of the method, including the regions over which $\tilde{\mathbf{w}}_{x,l}(f)$ and $\tilde{\mathbf{w}}_{y,l}(f)$ are defined, are given in [32]. Also, as discussed in [32], Wide-band SCORE differs from the frequency-dependent SCORE method presented in [49], which typically requires that a complicated set of coupled matrix equations be solved, which hinders practical application.

IV.F.2 Phase SCORE

One very important occurrence of multiple spectrally overlapping SOIs having the same cycle frequency α is in modern communication networks based on code-division multiple-access (CDMA). In CDMA, multiple signals can share the same spectral band because the unique code used by the transmitter to spread a message signal across the wide spectral band is also used by the receiver to despread it. At the receiver, this code despreads the corresponding SOI but does not despread the signals of other users. Depending upon the choice of spreading codes, the feature strengths of the CDMA signals of the multiple users can be identical. Thus, in this application, cross-SCORE's requirement that each SOI has a different feature strength (to enable signal separation) is especially restrictive. However, unless the network is synchronous (i.e., all users are synchronized to a single symbol clock), the cyclic correlation coefficients $\rho^{\alpha}_{s_m s_m}(\tau)$ will have different phases even if the magnitudes are the same. This result follows because the phase of $\rho^{\alpha}_{s_m s_m}(\tau)$ is linearly dependent on the phase of the symbol clock when α is a harmonic of the symbol rate. Thus, the task of signal separation may be viewed as one of designing a SCORE-like algorithm that preserves this phase information.

An ad hoc solution can be obtained from the noiseless infinite-time analysis (26) by simply deleting $\mathbf{R}^{-1}_{\mathbf{ss}}\mathbf{R}^{\alpha H}_{\mathbf{ss}}(\tau)$ from the left-hand side to obtain

$$\mathbf{R}^{\alpha}_{\mathbf{ss}}(\tau)\mathbf{g}_l = \lambda_l \mathbf{R}_{\mathbf{ss}}\mathbf{g}_l \quad \text{for} \quad l = 1, \ldots, L_\alpha. \tag{31}$$

Here, perfect signal separation (i.e., $[\mathbf{g}_1, \cdots, \mathbf{g}_{L_\alpha}]$ is a permutation of $\mathbf{I}$) occurs only when the SOIs are uncorrelated and $\rho^{\alpha}_{s_m s_m}(\tau) \neq \rho^{\alpha}_{s_n s_n}(\tau)$ for

all $m \neq n$. That is, the phase information of the feature at α is preserved and is used to distinguish each signal from the others. Working backwards from (31) yields the *Phase-SCORE* algorithm

$$\hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha}(\tau)\mathbf{w} = \lambda \hat{\mathbf{R}}_{\mathbf{xx}}\mathbf{w}. \tag{32}$$

As shown in [46] and [31] via computer simulations, phase-SCORE can separate SOIs having the same feature strengths but different feature phases (e.g., caused by the signals having different timing of their symbol clocks).

As with Cross-SCORE, analytical results for finite-time performance of Phase-SCORE are desirable but currently unavailable. Unlike Cross-SCORE, which has been reinterpreted as a particular instance of the canonical correlation analyzer, or as the solution to a particular constrained conditional maximum likelihood problem, no link between Phase-SCORE and existing methodology in multivariate statistics has been found. Furthermore, it is unclear (at least to the authors as of this writing) whether Phase-SCORE is the solution to an optimization problem or it is a purely ad hoc algorithm. Also, it is not clear if Phase-SCORE can be extended to accommodate multiple cycle frequencies, or if it can exploit features associated with the cyclic conjugate correlation (e.g., features at the doubled carrier frequency).

IV.G Application Examples

In this section two specific applications of blind adaptive spatial filtering to digital communication systems are summarized very briefly.

In the first application, the Cross-SCORE algorithm forms the basis for a completely new architecture for a land mobile digital cellular radio system that accommodates a substantially greater number of users than existing or proposed systems. This system is discussed in [48, 44] where it is demonstrated in computer simulations to achieve up to 5 times the capacity of an optimistically evaluated CDMA system. The new scheme achieves this capacity by using spatial filtering during both transmission and reception at the base station to separate spectrally overlapping signals of multiple users. Adaptation of the spatial filter for each user's signal is performed by the conjugate Cross-SCORE algorithm, which exploits a novel

carrier allocation scheme. Each user is assigned a distinct carrier frequency (which is used by conjugate Cross-SCORE to extract that particular user without wasting message capacity on a training signal), but the separation between adjacent carrier frequencies can be much less than the bandwidths of the signals, thus providing a large increase in spectral efficiency.

In the second application, the wideband SCORE algorithm is applied to a direct-sequence spread spectrum communication system to reduce the amount of processing gain required from spreading alone, thereby allowing an increase in channel capacity and/or a decrease in bit error rate. In particular, in so-called overlay systems one or more spread spectrum signals use the same spectral band as pre-existing conventional fixed-point microwave signals or other relatively narrowband signals. Provided that the average power of the spread spectrum signals is low, the pre-existing users suffer a tolerable amount of added interference. A potentially large spreading factor is then needed in the spread spectrum system to overcome the interference caused by the pre-existing users. The solution proposed in [32] is to use the Wideband SCORE algorithm to adapt an antenna array at each spread spectrum receiver to spatially reject the pre-existing users prior to the despreading operation. This implies that the spreading factor can be smaller (thus allowing the data rate of the message signal to be larger for the same receiver bandwidth) or that the number of CDMA users can be greater.

IV.H Summary

Several blind adaptive spatial filtering algorithms are available, although their performance is not thoroughly understood and none appears to be statistically optimum and devoid of undesirable behavior. Nonetheless, their performance and capabilities suggest that blind adaptive spatial filtering and blind adaptive LCL-PTV STF may significantly improve the performance (i.e., signal quality and capacity) of communication systems relative to that obtainable with conventional methods. Also, the flexibility of the PCCA framework and its performance attributes (illustrated by simulations) suggest that interference and multipath can be rejected (or at least mitigated) and multiple SOIs separated even when little is known about

them at the receiver.

V Fractionally-Spaced Equalization

In this section the problem of blindly adapting a fractionally spaced equalizer (for either single-sensor or multi-sensor data) for bauded digital communication signals (such as PAM, QAM, ASK, and PSK) is addressed. In a loose sense, this problem is similar in character to that addressed in Section IV because the objective in both sections is to blindly adapt a spatio-temporal filter to obtain a high-quality estimate of the SOI. However, the emphasis here is on using spatio-temporal degrees of freedom to compensate for channel distortion rather than using them to separate multiple SOIs and to reject SNOIs and noise.

Fractionally spaced equalizers (FSEs) (cf. Section 7.4 of [50], or any standard text on digital communications) for single-sensor data are widely used in digital communication systems to compensate for channel distortion and to mitigate the effects of timing jitter. An FSE filters the received data (typically sampled at two or four times the baud rate) and then samples the filtered output at the baud rate to obtain a stream of estimated symbols. Provided that the LTI filter in the FSE is chosen properly, it can be shown that this structure is the optimum linear receiver for QAM signals corrupted by linear channel distortion and additive stationary noise [24, 51]. The FSE structure can also be understood in terms of LCL-PTV filtering (also called LCL frequency-shift filtering or LCL-FRESH filtering), as discussed in [3, 22, 52].

Tutorial overviews of equalization in digital communication systems are given in [53, 54], and additional discussion of blind equalization are given in [42] and references therein. Most of the blind channel identification and equalization methods studied to date have exploited the fact that well-equalized baud-sampled data should have samples drawn from a finite alphabet, or that the higher-order statistics of the baud-sampled data can be used to infer the effective channel seen by the symbol stream as it propagates through the modulator, transmitter, propagation channel, and receiver. Many of these methods require the iterative optimization of non-

linear objective functions, which can cause misconvergence or noise capture, as discussed in [55, 56, 57]. Furthermore, the methods typically converge at a rate that is unacceptably slow for applications in which the channel varies rapidly (e.g., land mobile cellular communications).

In constrast, the methods considered in [42, 58] and the new method [6] discussed in Section V.B exploit only the second-order cyclostationarity of the fractionally-sampled data, a property which is destroyed by the baud-sampling typically performed prior to application of other blind equalization methods. Unlike the second-order statistics of stationary signals, the second-order statistics of cyclostationary signals have been shown to be sufficient to perform identification and equalization of most nonminimum-phase channels [42, 58].

In this section, the blind FSE presented in [59, 58] is rederived using least-squares minimization instead of observations about the algebraic structure of the ideal autocorrelation function of the received data, and a new optimum method proposed in [6] is explained. Brief comparative examples of performance are also given.

V.A The Blind FSE of Tong, Xu, and Kailath

In this section the blind channel identification and equalization method of Tong, Xu, and Kailath, referred to here as the TXK method for the sake of brevity, is reinterpreted as the solution to a least-squares problem. The method and an algebraic derivation is presented in [59] and is generalized in a straightforward way in [58] to accommodate multiple sensors. It accommodates arbitrary time-limited pulse shapes and channel distortion but requires that the noise be spatially and temporally white (i.e., it has no interference rejection capability).

The TXK method performs blind adaptive temporal equalization of a noisy distorted QAM signal,

$$x(n) = \sum_{k=-\infty}^{\infty} g(n - kT)s(k) + i(n) \tag{33}$$

received at a single sensor, where $\{s(k)\}$ is the stream of independent symbols, T is the symbol period, and $g(n)$ is the unknown distorted pulse. In

[59, 58], the TXK method is derived by exploiting various signal subspace properties of the ideal correlation matrices of the received data to yield an algebraically motivated solution. In [58], the straightforward extension to sensor array or multichannel data is proposed by replacing $x(\cdot)$ in (33) by the received data $\mathbf{x}(\cdot)$ at the output of the sensor array, and replacing $h(\cdot)$ by the multi-sensor impulse response $\mathbf{g}(\cdot)$. This general case is used in this section.

The method operates on the vector-stationarized data

$$\mathbf{z}(k) = \begin{bmatrix} \mathbf{x}(kT) \\ \vdots \\ \mathbf{x}(kT + KT - 1) \end{bmatrix} = \mathbf{H}\mathbf{s}(k) + \mathbf{i}(k) \tag{34}$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{g}(-(k_0 + 0)T) & \cdots & \mathbf{g}(-(k_0 + d - 1)T) \\ \vdots & & \vdots \\ \mathbf{g}(KT - 1 - (k_0 + 0)T) & \cdots & \mathbf{g}(KT - 1 - (k_0 + d - 1)T) \end{bmatrix}$$

and

$$\mathbf{s}(k) = \begin{bmatrix} s(k_0 + k) \\ \vdots \\ s(k_0 + k + d - 1) \end{bmatrix},$$

with K equal to the desired length (in symbol periods) of the equalizer, k_0 equal to an arbitrary integer (e.g., zero for simplicity), and d equal to a best guess of the number of symbols that contribute significantly to any given sample of the signal (i.e., the estimated number of symbol periods in the distorted pulse $\mathbf{h}(n)$). In addition to re-expressing the data as the stationary output of a multidimensional LTI system, this representation is useful because the autocorrelations $\mathbf{R}_{\mathbf{zz}}(0)$ and $\mathbf{R}_{\mathbf{zz}}(1)$ exhibit a special structure, namely $\mathbf{R}_{\mathbf{zz}}(0) = \mathbf{H}\mathbf{H}^H + \sigma^2\mathbf{I}$ and $\mathbf{R}_{\mathbf{zz}}(1) = \mathbf{H}\mathbf{J}\mathbf{H}^H + \sigma^2\mathbf{J}^T$, where the superscript T here temporarily denotes exponentiation, not transposition. This structure is exploited by the TXK method as described in [58].

The objective is to identify the channel $\mathbf{h}(n)$ through which the symbols are sent, equalize the channel, and estimate the symbol values. The equalization structure is constrained to be linear, with the estimated symbols

$\hat{\mathbf{s}}(k)$ given by

$$\hat{\mathbf{s}}(k) = \mathbf{W}^H \mathbf{z}(k).$$

Thus, not only must d be long enough that the estimated channel can be a reasonable approximation of the true channel, but K must be large enough that the equalization filter $\mathbf{W}$ can adequately compensate for the distortion.

Here, the TXK method for blind adaptive spatio-temporal equalization is re-derived as the solution to a least-squares problem for a finite data set:

$$\min_{\mathbf{W},\hat{\mathbf{H}}} \left\langle \left\| \mathbf{z}(k) - \hat{\mathbf{H}}\hat{\mathbf{s}}(k) \right\|^2 \right\rangle_{N/T}, \tag{35}$$

where the time average $\langle \cdot \rangle$ is taken over the symbol index k, the estimated symbols are $\hat{\mathbf{s}}(k) = \mathbf{W}^H \mathbf{z}(k)$, and the autocorrelations of $\hat{\mathbf{s}}$ are restricted to obey $\hat{\mathbf{R}}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}(0) = \mathbf{I}$ and $\hat{\mathbf{R}}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}(1) = \mathbf{J}$, with

$$\mathbf{J} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ & & & 0 & 1 \\ 0 & & \cdots & & 0 \end{bmatrix}.$$

Re-expressing the cost function (35) in terms of estimated correlations and minimizing with respect to $\mathbf{W}$ yields the solution

$$\mathbf{W} = \hat{\mathbf{H}} \left(\hat{\mathbf{H}}^H \hat{\mathbf{H}} \right)^{-1}.$$

Substituting this solution into (35) and minimizing with respect to $\hat{\mathbf{H}}$ can be shown to lead to the maximization problem

$$\max_{\hat{\mathbf{H}}} \operatorname{tr} \left\{ \mathbf{P}_{\hat{\mathbf{H}}} \hat{\mathbf{R}}_{\mathbf{zz}}(0) \right\}$$

where $\mathbf{P}_{\hat{\mathbf{H}}} = \hat{\mathbf{H}} \left(\hat{\mathbf{H}}^H \hat{\mathbf{H}} \right)^{-1} \hat{\mathbf{H}}^H$, which in turn is equivalent to

$$\max_{\mathbf{W}} \operatorname{tr} \left\{ \mathbf{P}_{\mathbf{W}} \hat{\mathbf{R}}_{\mathbf{zz}}(0) \right\}.$$

Expressing the problem in terms of the SVD of $\mathbf{W} = \mathbf{U}\Sigma\mathbf{V}^H$ yields the constrained optimization problem

$$\max_{\mathbf{W}} \operatorname{tr} \left\{ \mathbf{U}\mathbf{U}^H \hat{\mathbf{R}}_{\mathbf{zz}}(0) \right\}$$

subject to

$$\begin{aligned} \mathbf{V}\Sigma\mathbf{U}^H\hat{\mathbf{R}}_{\mathbf{zz}}(0)\mathbf{U}\Sigma\mathbf{V}^H &= \mathbf{I} \\ \mathbf{V}\Sigma\mathbf{U}^H\hat{\mathbf{R}}_{\mathbf{zz}}(1)\mathbf{U}\Sigma\mathbf{V}^H &= \mathbf{J}. \end{aligned}$$

Straightforward manipulations reveal that maximization subject to the first constraint requires that $\mathbf{U}$ equal the d most dominant eigenvectors of $\hat{\mathbf{R}}_{\mathbf{zz}}$ and Σ equal the reciprocal of the square root of the corresponding eigenvalues. The second constraint then defines $\mathbf{V}$.

This derivation leads to the same algorithm as obtained in [59, 58] by a different, algebraically motivated argument for the single-channel case. However, since the second constraint can be satisfied only if the matrix $\Sigma\mathbf{U}^H\hat{\mathbf{R}}_{\mathbf{zz}}(1)\mathbf{U}\Sigma$ has Jordan form $\mathbf{J}$, which occurs with probability zero due to finite averaging time, the TXK method is seen to lead to an ill-posed optimization problem. A possible solution to this dilemma is given without proof (but motivated by standard practices in the sensor array processing literature) in [59, 58].

V.B Another Blind FSE

Partially motivated by the least-squares derivation of the TXK method, another algorithm, first proposed in [6], that blindly adapts an FSE is discussed here. Unlike the TXK method, this method can be interpreted as a maximum likelihood blind adaptive FSE under some reasonable assumptions, described in Section V.B.2.

V.B.1 Notation

Let $\mathbf{x}(kT+n)$, for $0 \leq n < T$, be the received data given by

$$\mathbf{x}(kT+n) = \sum_{l=-\infty}^{0} \mathbf{g}(n-lT)s(l+k) + \mathbf{i}(kT+n),$$

which is a noisy bauded communication signal distorted by an unknown multi-sensor channel having impulse response $\mathbf{g}(\cdot)$, where T is the baud period, $s(k)$ is the kth symbol (unknown), and $\mathbf{i}(\cdot)$ is unknown noise. Note

that $s(k)$ is not necessarily the dominant symbol within the sampling interval $\{kT, \ldots, kT+T-1\}$, and that the time interval $kT \leq kT+n < kT+T$ need not be centered on an optimal instant for making a symbol decision. Provided that $\mathbf{g}(t)$ decays sufficiently fast as $t \to \infty$, $\mathbf{x}(kT+n)$ can be approximated by

$$\mathbf{x}(kT+n) \approx \sum_{l=-L}^{0} \mathbf{g}(n-lT)s(l+k) + \mathbf{i}(kT+n). \tag{36}$$

For most physical channels of interest, this simply means that L is sufficiently large, or equivalently that only $L+1$ symbols contribute significantly to any particular baud period of the signal. Defining

$$\begin{aligned}
\mathbf{y}(k) &= [\mathbf{x}(kT),\ \mathbf{x}(kT+1),\ \cdots,\ \mathbf{x}(kT+T-1)]^T, \\
\mathbf{h}(k) &= [\mathbf{g}(-kT),\ \mathbf{g}(-kT+1),\ \cdots,\ \mathbf{g}(-kT+T-1)]^T, \\
\mathbf{H} &= [\mathbf{h}(-L)\ \mathbf{h}(-L+1)\ \cdots\ \mathbf{h}(0)], \\
\mathbf{j}(k) &= [\mathbf{i}(kT),\ \mathbf{i}(kT+1),\ \cdots,\ \mathbf{i}(kT+T-1)]^T, \\
\mathbf{s}(k) &= [s(-L+k),\ s(-L+1+k),\ \cdots,\ s(k)]^T,
\end{aligned}$$

(36) can be expressed as

$$\mathbf{y}(k) \approx \mathbf{H}\mathbf{s}(k) + \mathbf{j}(k), \tag{37}$$

which is a stationary complex vector-valued signal. It is noted that $\mathbf{H}$ as defined here is unstructured because $\mathbf{y}(k)$ contains only a single baud period of the signal, whereas $\mathbf{H}$ as defined in the TXK method is a block Toeplitz matrix in most cases.

The signal is to be filtered by an FIR filter of length $(K+1)T$ samples to yield an estimate of the symbol stream,

$$\hat{s}(k) = \begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_K \end{bmatrix}^H \begin{bmatrix} \mathbf{y}(k-K) \\ \mathbf{y}(k-K+1) \\ \vdots \\ \mathbf{y}(k) \end{bmatrix} = \mathbf{w}^H \begin{bmatrix} \mathbf{y}(k-K) \\ \mathbf{y}(k-K+1) \\ \vdots \\ \mathbf{y}(k) \end{bmatrix}, \tag{38}$$

and thus the vector $\hat{\mathbf{s}}(k)$ of symbol estimates is given by

$$\hat{\mathbf{s}}(k) = \mathbf{W}^H \begin{bmatrix} \mathbf{y}(k-K-L) \\ \mathbf{y}(k-K-L+1) \\ \vdots \\ \mathbf{y}(k) \end{bmatrix} = \mathbf{W}^H \mathbf{z}(k), \tag{39}$$

where $\mathbf{z}(k)$ is defined in the obvious way and $\mathbf{W}$ is a block Toeplitz matrix of dimension $(K+L+1) \times (L+1)$ with block size $T \times 1$ (or $MT \times 1$ if $\mathbf{x}(n)$ comes from an M-sensor array) and first column equal to $[\mathbf{w}_0^H \ \mathbf{w}_1^H \cdots \mathbf{w}_K \ 0 \cdots 0]^H$ and first row equal to $[\mathbf{w}_0 \ 0 \cdots \ 0]$. This formulation insures that $\hat{s}_l(k) = \hat{s}_{l-m}(k+m)$, where as the TXK method attempts to satisfy this indirectly through the correlation constraints, although as noted in Section V.A these constraints are satisfied with zero probability.

V.B.2 Derivation

In a vein similar to (35) but having different constraints, a least-squares minimization problem can be posed,

$$\min_{\mathbf{W},\hat{\mathbf{H}}} \left\langle \left\| \mathbf{y}(k) - \hat{\mathbf{H}}\hat{\mathbf{s}}(k) \right\|^2 \right\rangle_{N/T}, \tag{40}$$

subject only to the constraint that $\mathbf{W}$ is block Toeplitz. Note that there is no requirement that the symbols be uncorrelated with each other or that the noise be spatio-temporally white (although these assumptions might be needed to insure consistency; this is an open problem). Note also that there can be phase ambiguities and temporal-shift ambiguities in $\hat{\mathbf{H}}$ and $\hat{\mathbf{s}}(k)$, since only their product must match the signal component of $\mathbf{y}(k)$ to minimize the cost function. Equivalently,

$$\min_{\mathbf{W},\hat{\mathbf{H}}} \operatorname{tr} \left\{ \hat{\mathbf{R}}_{\mathbf{yy}} - \hat{\mathbf{H}}\hat{\mathbf{R}}_{\mathbf{y}\hat{\mathbf{s}}}^H - \hat{\mathbf{R}}_{\mathbf{y}\hat{\mathbf{s}}}\hat{\mathbf{H}}^H + \hat{\mathbf{H}}\hat{\mathbf{R}}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\hat{\mathbf{H}}^H \right\}. \tag{41}$$

Alternatively, if $\mathbf{i}(n)$ is zero-mean stationary spatio-temporally white Gaussian noise, then it is easily shown that (40)-(41) maximize the conditional likelihood of the received data, with $\mathbf{H}$ being an unknown parameter matrix to be estimated and the conditioning being performed over the unknown

symbol values, whose estimates are then constrained to be given by (38)-(39).

Minimization of (41) with respect to $\hat{\mathbf{H}}$ is a standard problem whose solution is well-known. However, in preparation for the subsequent optimization with respect to $\mathbf{W}$ which does not seem to be a standard problem, the basics of complex matrix calculus are introduced here. It is shown in [37] that an appropriate definition of the complex gradient of a real-valued function with respect to a complex matrix and its conjugate can be found simply by treating the matrix and its conjugate as independent variables. This step is necessary because the usual definition of differentiation with respect to a complex variable is not valid unless the function is analytic. Consequently, it is also shown in [37] that the stationary point of a real-valued function of a complex matrix can be found by setting the complex gradient with respect to the conjugate of the matrix equal to zero (equivalently, the gradient with respect to the matrix could be equated with zero). Then, it is easily shown that

$$\nabla_{\hat{\mathbf{H}}^*} \mathrm{tr}\left\{\hat{\mathbf{H}}^H \mathbf{A}\right\} = \mathbf{A} \quad \text{and} \quad \nabla_{\hat{\mathbf{H}}^*} \mathrm{tr}\left\{\hat{\mathbf{H}}^H \mathbf{A} \hat{\mathbf{H}}\right\} = \mathbf{A}\hat{\mathbf{H}}.$$

Consequently, the gradient of the cost function in (41) is

$$\nabla_{\hat{\mathbf{H}}^*}(\text{cost func}) = -\hat{\mathbf{R}}_{\mathbf{y}\hat{\mathbf{s}}} + \hat{\mathbf{H}}\hat{\mathbf{R}}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}.$$

Equating to zero and solving yields the standard solution,

$$\hat{\mathbf{H}} = \hat{\mathbf{R}}_{\mathbf{y}\hat{\mathbf{s}}}\hat{\mathbf{R}}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}^{-1}. \tag{42}$$

Substituting (42) into (41) yields

$$\min_{\mathbf{W}} \mathrm{tr}\left\{\hat{\mathbf{R}}_{\mathbf{yy}} - \hat{\mathbf{R}}_{\mathbf{y}\hat{\mathbf{s}}}\hat{\mathbf{R}}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}^{-1}\hat{\mathbf{R}}_{\mathbf{y}\hat{\mathbf{s}}}^H\right\}, \tag{43}$$

which can be re-expressed using (39) as

$$\max_{\mathbf{W}} \mathrm{tr}\left\{\mathbf{W}^H\hat{\mathbf{R}}_{\mathbf{yz}}^H\hat{\mathbf{R}}_{\mathbf{yz}}\mathbf{W}\left(\mathbf{W}^H\hat{\mathbf{R}}_{\mathbf{zz}}\mathbf{W}\right)^{-1}\right\}. \tag{44}$$

It is crucial to note that $\mathbf{W}$ is a block Toeplitz matrix, so care is required in the application of conventional optimization techniques. However, (44)

can be re-expressed in terms of the unstructured vector $\mathbf{w}$ as follows. Note that the (m,n)th element of a quadratic form in $\mathbf{W}$ can be expressed as

$$\left[\mathbf{W}^H\mathbf{R}\mathbf{W}\right]_{m,n} = \sum_{l=0}^{K}\sum_{k=0}^{K}\mathbf{w}_l^H\mathbf{R}_{l+m,k+n}\mathbf{w}_k \tag{45}$$

where $\mathbf{R}_{l+m,k+n}$ is the block element (matrix) at block-row $l+m$ and block-column $k+n$ in $\mathbf{R}$, and m and n each range from 0 to L. Denoting

$$\mathbf{R}^{m,n} = \begin{bmatrix} \mathbf{R}_{m,n} & \cdots & \mathbf{R}_{m,n+K} \\ \vdots & & \vdots \\ \mathbf{R}_{m+K,n} & \cdots & \mathbf{R}_{m+K,n+K} \end{bmatrix},$$

(45) can be expressed as

$$\left[\mathbf{W}^H\mathbf{R}\mathbf{W}\right]_{m,n} = \mathbf{w}^H\mathbf{R}^{m,n}\mathbf{w}. \tag{46}$$

Thus, (44) is equivalent to

$$\max_{\mathbf{w}} \operatorname{tr}\left\{\mathbf{A}(\mathbf{w})\left[\mathbf{B}(\mathbf{w})\right]^{-1}\right\} \tag{47}$$

where

$$\left[\mathbf{A}(\mathbf{w})\right]_{m,n} = \mathbf{w}^H\left[\hat{\mathbf{R}}_{\mathbf{yz}}^H\hat{\mathbf{R}}_{\mathbf{yz}}\right]^{m,n}\mathbf{w} \quad \text{and} \quad \left[\mathbf{B}(\mathbf{w})\right]_{m,n} = \mathbf{w}^H\left[\hat{\mathbf{R}}_{\mathbf{zz}}\right]^{m,n}\mathbf{w}$$

with m and n each ranging from 0 to L. As of this writing, no further simplification of (47) has been found, except in the special case where $L=0$ (i.e., the impulse response of the channel is assumed to be negligible outside of one symbol period). In this special case, $\mathbf{A}(\mathbf{w})$ and $\mathbf{B}(\mathbf{w})$ are scalars and the standard solution for $\mathbf{w}$ is simply the most dominant eigenvector $\mathbf{w}_{max}$ of the system

$$\hat{\mathbf{R}}_{\mathbf{yz}}^H\hat{\mathbf{R}}_{\mathbf{yz}}\mathbf{w}_{max} = \lambda_{max}\hat{\mathbf{R}}_{\mathbf{zz}}\mathbf{w}_{max}. \tag{48}$$

This special case provides some hope that (47) can be solved without resort to a gradient-based search method. However, this remains an open problem.

V.B.3 Discussion

It is interesting to note that if $K=L=0$ is chosen then $\mathbf{y}\equiv\mathbf{z}$ and (48) reduces to

$$\hat{\mathbf{R}}_{\mathbf{yy}}\mathbf{w}_{max} = \lambda_{max}\mathbf{w}_{max}. \tag{49}$$

That is, the equalizer weights are obtained from the most dominant eigenvector used in the Karhunen-Loéve (KL) expansion of $\mathbf{y}$ over one symbol period. Thus, the equalizer output for each symbol interval is the dominant coefficient in the KL expansion over that interval. Furthermore, the TXK FSE also degenerates to (49) when its parameters are chosen as $K = 1$ and $d = 1$ (which have the same meaning as $K = 0$ and $L = 0$, respectively in the new method of this section), although this degenerate case was not specifically considered in [59, 58].

This simplified algorithm for equalization of a scalar channel,

$$\hat{s}(k) = \mathbf{w}_{max}^H \mathbf{y}(k)$$

with $\mathbf{w}_{max}$ having length equal to one symbol interval, can be derived directly from consideration of the properties of the KL expansion of $\mathbf{y}(k)$ over the kth symbol interval:

$$\mathbf{y}(k) = \sum_{l=0}^{T-1} \sigma_l(k)\boldsymbol{\eta}_l \tag{50}$$

where $[\boldsymbol{\eta}_0 \quad \cdots \quad \boldsymbol{\eta}_{T-1}]$ is the unitary matrix of eigenvectors of $\mathbf{R}_{yy} = E\{\mathbf{y}(k)\mathbf{y}^H(k)\}$ with corresponding eigenvalues $\lambda_0 \geq \cdots \geq \lambda_{T-1}$, and

$$\sigma_l(k) = \boldsymbol{\eta}_l^H \mathbf{y}(k) \quad \text{with} \quad E\{\sigma_l(k)\sigma_m^*(k)\} = \lambda_l \delta_{l-m}.$$

The fact that $\boldsymbol{\eta}_l$ is independent of k for $l = 0, \ldots, T-1$, follows because $y(k)$ is stationary. Note that $\mathbf{w}_{max} = \boldsymbol{\eta}_0$.

Now, let us compare the KL expansion (50) with the noise-free signal at the channel output,

$$\mathbf{y}(k) = \sum_{l=0}^{\infty} s(k-l)\mathbf{h}(-l) = \sum_{l=0}^{\infty} \tilde{s}_l(k)\tilde{\mathbf{h}}(l) \tag{51}$$

where

$$\begin{aligned} \tilde{s}_l(k) &= s(k-l)\ \|\mathbf{h}(-l)\| \\ \tilde{\mathbf{h}}(l) &= \mathbf{h}(-l)/\|\mathbf{h}(-l)\|. \end{aligned}$$

In the typical case in which the symbols $s(k)$ are independent and identically distributed, $\tilde{s}_l(k)$ can be modeled as orthogonal random variables, similar

to $\sigma_l(k)$:

$$E\{\tilde{s}_l(k)\tilde{s}_m^*(k)\} = E\left\{|s(k)|^2\right\}\|\mathbf{h}(-l)\|^2\,\delta_{l-m} = \gamma_l\delta_{l-m}.$$

In (51), denote by d the index of the dominant symbol $\tilde{s}_d(k)$ within the symbol interval k; that is,

$$\gamma_d = \max_l \gamma_l.$$

Then the intersymbol interference (ISI) in the kth symbol interval corresponds to all of the terms in (51) for which $l \neq d$. To remove this ISI, the weight vector $\mathbf{w}$ would have to satisfy

$$\mathbf{w}^H\tilde{\mathbf{h}}(l) = 0, \quad l \neq d, \quad l = 0, 1, \ldots. \tag{52}$$

We see from the preceding that T of the terms in (50) and (51) would be identical if

$$\tilde{s}_{n(l)}(k) = \sigma_l(k), \quad \tilde{\mathbf{h}}(n(l)) = \eta_l, \quad \mathbf{w} = \eta_0, \tag{53}$$

where $n(l)$ is an index mapping such that $\gamma_{n(0)} \geq \cdots \geq \gamma_{n(T-1)}$ and in this case we would obtain perfect equalization because (52) would be satisfied. However, in general, the set of vectors $\tilde{\mathbf{h}}(l)$ is not orthogonal, and the $T \times 1$ vector $\mathbf{w}$ can be orthogonal to no more than $T-1$ vectors $\tilde{\mathbf{h}}(l)$; thus, except for certain cases, (53) cannot be satisfied.

Nonetheless, (53) is sufficient but not necessary for the zero ISI condition (52) to be satisfied. For practical implementation, given N/T symbol intervals of data, we can replace $\mathbf{R}_{\mathbf{yy}}$ with its time-averaged estimator $\hat{\mathbf{R}}_{\mathbf{yy}}$ and use $\mathbf{w}$ equal to the dominant eigenvector $\mathbf{w}_{max}$ of $\hat{\mathbf{R}}_{\mathbf{yy}}$. Some tradeoff can be seen to exist in the choice of T, namely that larger T is needed for $\mathbf{w}$ to be orthogonal to more ISI components but that larger T entails an increased sampling rate. If the orthogonality (52) is sufficiently closely approximated, then the signal can be very effectively equalized even though η_0 might be a poor estimate of the channel (because the estimate spans only one symbol interval). Thus, the KL method should be considered an equalization method, not a channel identification method.

This derivation/explanation of the KL algorithm can be easily generalized for vector channels. The solution is still the dominant eigenvector of

$\hat{\mathbf{R}}_{\mathbf{yy}}$ where $\mathbf{y}(k)$ is the vector-stationarized version of the vector data as in (37).

Although this derivation omits the noise from the model of the received data, it is easily shown that the eigenvectors of the ideal autocorrelation matrix $\mathbf{R}_{\mathbf{yy}}$ are invariant to additive stationary white noise. Consequently, the same argument that leads to the KL algorithm in the absence of noise applies when noise is taken into consideration.

Curiously, the eigenvectors of the ideal correlation matrix $\mathbf{R}_{\mathbf{yy}}$ have been studied previously (cf. [60] and references therein) without the observation being made that the dominant eigenvector of $\hat{\mathbf{R}}_{\mathbf{yy}}$ might serve usefully as a blind adaptive FSE.

V.B.4 Algorithm Summary

In general, the algorithm can be summarized in the following steps:

1. Estimate $\hat{\mathbf{R}}_{\mathbf{yz}}$ and $\hat{\mathbf{R}}_{\mathbf{zz}}$.
2. If $L = 0$, then solve (48) or (49), and exit.
3. If $L \neq 0$, initialize the weight vector $\mathbf{w}$ with a Kronecker delta function, or prior knowledge of a suitable equalizer for the multichannel impulse response (e.g., in single-channel communication systems for which the transmitter filter has transfer function equal to the square-root of the raised-cosine function, the initial guess for $\mathbf{w}$ could be the impulse response of this filter), or a centered version of the weight vector found from running the algorithm with $K = L = 0$.
4. Iteratively update the guess $\mathbf{w}^{(k)}$ at the kth iteration using the simple gradient method

$$\mathbf{w}^{(k)} = \mathbf{w}^{(k-1)} + \mu \nabla_{\mathbf{w}^*} \tag{54}$$

where $\nabla_{\mathbf{w}^*}$ is the gradient with respect to $\mathbf{w}^*$ of the objective function in (47) evaluated at $\mathbf{w}^{(k-1)}$, and can be shown using results in [37, 16] to be equal to the following,

$$\nabla_{\mathbf{w}^*} = \sum_{m,n} \left\{ C_{n,m} \ \tilde{\mathbf{A}}_{m,n}\mathbf{w} - \left(\mathbf{w}^H \tilde{\mathbf{A}}_{m,n}\mathbf{w}\right) \sum_{p,q} C_{n,p} \ \tilde{\mathbf{B}}_{p,q}\mathbf{w} \ C_{q,m} \right\} \tag{55}$$

where each summation index ranges from 0 to L and

$$\tilde{\mathbf{A}}_{m,n} = \left[\hat{\mathbf{R}}_{\mathbf{yz}}^{H}\hat{\mathbf{R}}_{\mathbf{yz}}\right]^{m,n}, \quad \tilde{\mathbf{B}}_{m,n} = \left[\hat{\mathbf{R}}_{\mathbf{zz}}\right]^{m,n}, \quad \text{and} \quad \mathbf{C} = [\mathbf{B}(\mathbf{w})]^{-1}. \tag{56}$$

Cauchy's method (cf. [61]) could be an alternative to the simple gradient method (54), and consists simply of (54) with μ replaced by a scalar variable over which maximization occurs at each iteration. Another alternative is a limited-step Newton's method, although this requires that the Hessian matrix be computed and inverted at each iteration, which significantly increases the computational load at each iteration. These issues are left as open problems, as are those concerning the the choice of the values for the parameters L and K.

V.C A Simulation Example

Here the TXK method and the method of Section V.B are simulated in the example environment considered in [58]. That is, the transmitter filter and the propagation channel together have impulse response $(c(n) - 0.8c(n - 0.5) + 0.4c(n-3))u_{6T}(n)$, where $c(n)$ is a Nyquist-shaped pulse having 11% excess bandwidth and $u_{6T}(n)$ is a rectangular window of duration 6 symbol intervals. The symbol stream is binary, and the sampling rate is 4 samples per symbol. The TXK method is implemented according to the description in [58], with the parameter settings $d = 10$ and $K = 5$. The new method is implemented as described at the end of Section V.B, with $K = L = 0$ chosen to enable the simple optimization based on the eigenvalue problem (48) (i.e., the KL method is implemented). One hundred independent trials of each method at each SNR were performed. In each trial, 100 symbols were collected and used to adapt the equalizer, and the performance of this equalizer was then measured by processing 1000 symbols. As shown in Figure 18, the BER of the KL method is significantly less than the BER of the TXK method over a wide range of SNRs in this environment.

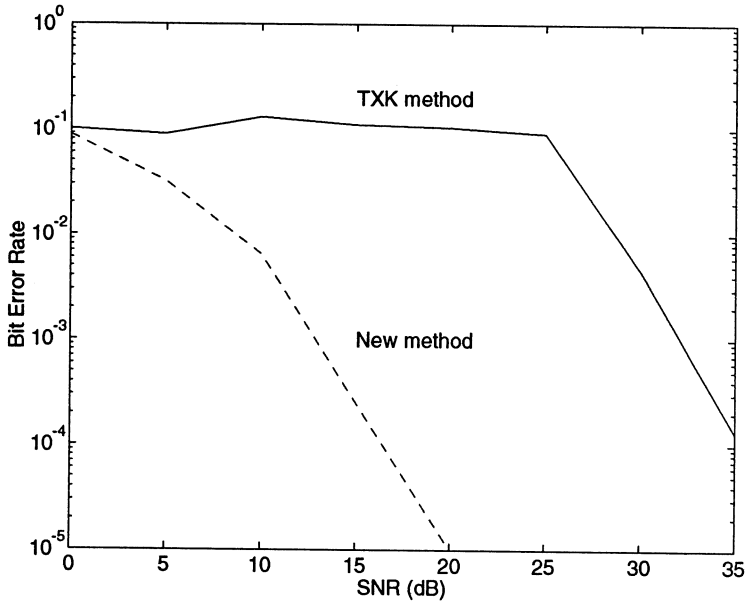

Figure 18: Bit error rate after blind equalization by the TXK method and the KL method.

V.D Summary

A comprehensive evaluation of these second-order cyclostationarity-exploiti methods and the higher-order/finite-alphabet methods referred to at the beginning of this section has yet to be performed, so no firm conclusions can be drawn as to their relative merits. Issues such as relative convergence time, robustness to deviation from underlying assumptions, and computational complexity must all be considered. However, it is clear that neither approach by itself can be optimum since the signals in question exhibit both properties, collectively described by their nth-order cyclostationarity properties for $n = 2, 3, \ldots$.

Many open problems exist in blind spatio-temporal equalization, even if the scope is restricted to a single, distorted, digital communication signal in noise. For example, the consistency properties of the method of Section V.B remain to be determined. Finite-time analyses and complete simulation studies of the two methods considered in this section remain to be done. Also, more computationally efficient methods of solving (47) need to be found. Both the TXK method and the new method of Section V.B could be easily extended to constrain the estimated symbols to be real, which is appropriate for signals such as BPSK and ASK having real constellations. On the speculative side, extension of these algorithms to the new capability of simultaneously separating and equalizing multiple spectrally overlapping signals could be used to increase communication capacity in wireless cellular networks. Except for the temporal equalization step, such a scheme based on Cross-SCORE has already been proposed in [44, 6].

VI Conclusions

Work on algorithms for spatial filtering and equalization of cyclostationary signals has shown that exploitation of cyclostationarity can offer substantially improved performance over that of conventional algorithms in some signal environments and can offer reasonably good performance in other environments where conventional algorithms fail completely. However, some of these signal-selective algorithms are suboptimal and exhibit some un-

desirable performance attributes. Thus, numerous problems remain to be investigated.

VII Acknowledgements

The authors gratefully acknowledge assistance provided by Mr. Dale L. Smith with the computer simulations of the blind adaptive FSEs.

The work surveyed in this chapter was supported in part by the Office of Naval Research under contract N00014-92-J-1218 (PIs: W.A. Gardner and S.V. Schell) and E-Systems, Inc. (Greenville Division).

References

[1] W. R. Bennett, "Statistics of regenerative digital transmission," *Bell Syst. Tech. J.*, vol. 37, pp. 1501–1542, 1958.

[2] W. A. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Processing Mag.*, vol. 8, no. 2, pp. 14–37, Apr. 1991.

[3] W. A. Gardner, "An introduction to cyclostationary signals," in *Cyclostationarity in Communications and Signal Processing* (W. A. Gardner, ed.), IEEE Press, 1993.

[4] G. Xu and T. Kailath, "Direction of arrival estimation via exploitation of cyclostationarity — a combination of temporal and spatial processing," *IEEE Trans. Signal Processing*, vol. SP-40, no. 7, pp. 1775–1786, July 1992.

[5] L. Izzo, L. Paura, and G. Poggi, "An interference-tolerant algorithm for localization of cyclostationary-signal sources," *IEEE Trans. Signal Processing*, vol. SP-40, no. 7, pp. 1682–1686, July 1992.

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

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

[8] H. L. Hurd, "Nonparametric time series analysis for periodically correlated processes," *IEEE Trans. Inform. Theory*, vol. IT-35, no. 2, pp. 350–359, Mar. 1989.

[9] S. V. Schell and W. A. Gardner, "Progress on signal–selective direction finding," in *Proc. of the Fifth ASSP Workshop on Spectrum Estimation and Modeling*, (Rochester, NY), pp. 144–148, Oct. 1990.

[10] T. E. Biedka and B. G. Agee, "Subinterval cyclic MUSIC – robust DF with error in cyclic frequency knowledge," in *Proc. Twenty–Fifth Asilomar Conf. on Signals, Systems and Computers*, (Pacific Grove, CA), pp. 262–266, Nov. 1991.

[11] R. S. Roberts, W. A. Brown, and H. H. Loomis, Jr., "A review of digital spectral correlation analysis: Theory and implementation," in *Cyclostationarity in Communications and Signal Processing* (W. A. Gardner, ed.), IEEE Press, 1993.

[12] W. A. Gardner, *Statistical Spectral Analysis: A Nonprobabilistic Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[13] H. L. Hurd and N. L. Gerr, "Graphical methods for determining the presence of periodic correlation," *J. of Time Series Anal.*, vol. 12, no. 4, pp. 337–350, 1991.

[14] W. A. Gardner, *Introduction to Random Processes with Applications to Signals and Systems,* Second Edition. New York, NY: McGraw–Hill, 1989.

[15] S. V. Schell and W. A. Gardner, "Maximum likelihood and common factor analysis-based blind adaptive spatial filtering for cyclostationary signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (Minneapolis, Minnesota), pp. IV:292–295, Apr. 1993.

[16] A. Graham, *Kronecker Products and Matrix Calculus: with Applications.* England: Ellis Horwood, Ltd., 1981.

[17] R. S. Roberts, W. A. Brown, and H. H. Loomis, Jr., "Computationally efficient algorithms for cyclic spectral analysis," *IEEE Signal Processing Mag.*, vol. 8, no. 2, pp. 38–49, Apr. 1991.

[18] S. V. Schell, *Exploitation of Spectral Correlation for Signal–Selective Direction Finding.* Ph.D. dissertation, Dept. of Electrical Engineering and Computer Science, University of California, Davis, CA, 1990.

[19] L. E. Franks, "Carrier and bit synchronization in data communication – a tutorial review," *IEEE Trans. Comm.*, vol. COM-28, no. 8, pp. 1107–1121, Aug. 1980.

[20] W. A. Gardner, W. A. Brown, III, and C.-K. Chen, "Spectral correlation of modulated signals: Part II — digital modulation," *IEEE Trans. Comm.*, vol. COM–35, no. 6, pp. 595–601, June 1987.

[21] L. E. Franks, "Polyperiodic linear filtering," in *Cyclostationarity in Communications and Signal Processing* (W. A. Gardner, ed.), IEEE Press, 1993.

[22] W. A. Gardner, "Cyclic Wiener filtering: Theory and method," *IEEE Trans. Comm.*, vol. COM-41, no. 1, pp. 151–163, Jan. 1993.

[23] W. A. Gardner, "A new method of channel identification," *IEEE Trans. Comm.*, vol. COM-39, no. 6, pp. 813–817, June 1991.

[24] W. A. Gardner, "The structure of least-mean-square linear estimators for synchronous m-ary signals," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 2, pp. 240–243, Mar. 1973.

[25] J. G. Proakis, *Digital Communications.* New York: McGraw–Hill, second ed., 1989.

[26] B. D. van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[27] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays.* NY: John Wiley and Sons, 1980.

[28] R. T. Compton, Jr., *Adaptive Antennas.* Englewood Cliffs, NJ: Prentice–Hall, 1988.

[29] S. V. Schell and W. A. Gardner, "High–resolution direction finding," in *Handbook of Statistics, Vol. 10* (N. K. Bose and C. R. Rao, eds.), North-Holland, 1993. In press.

[30] B. Ottersten, R. Roy, and T. Kailath, "Signal waveform estimation in sensor array processing," in *Proc. Twenty–Third Asilomar Conf. on Signals, Systems and Computers*, (Pacific Grove, CA), pp. 787–791, 1989.

[31] B. G. Agee, S. V. Schell, and W. A. Gardner, "Spectral self–coherence restoral: A new approach to blind adaptive signal extraction," *Proc. IEEE*, vol. 78, no. 4, pp. 753–767, Apr. 1990.

[32] S. V. Schell and W. A. Gardner, "Blind adaptive spatio-temporal filtering for wideband cyclostationary signals," *IEEE Trans. Signal Processing*, vol. SP-41, no. 5, pp. 1961–1964, May 1993.

[33] M. Siotani, T. Hayakawa, and Y. Fujikoshi, *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook.* Columbus, Ohio: American Sciences Press, 1985.

[34] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis.* Englewood Cliffs, NJ: Prentice Hall, second ed., 1988.

[35] D. R. Brillinger, *Time Series: Data Analysis and Theory.* San Francisco, CA: Holden-Day, expanded edition ed., 1981.

[36] C. M. Spooner, "Higher-order statistics for nonlinear processing of cyclostationary signals," in *Cyclostationarity in Communications and Signal Processing* (W. A. Gardner, ed.), IEEE Press, 1993.

[37] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proc.*, vol. 130, no. Pts. F and H, No. 1, pp. 11–16, Feb. 1983.

[38] S. V. Schell and W. A. Gardner, "Improved blind adaptive spatio-temporal filters for cyclostationary signals," *IEEE Trans. Signal Processing*, 1993. In preparation.

[39] V. J. Yohai and M. S. G. Ben, "Canonical variables as optimal predictors," *Ann. Statist.*, vol. 8, pp. 865–869, 1980.

[40] B. G. Agee, "Maximum-likelihood approaches to blind adaptive signal extraction using narrowband antenna arrays," in *Proc. of 25th Asilomar Conf. on Signals, Systems, and Computers*, IEEE Computer Society Press, 1991.

[41] C. R. Rao, "Separation theorems for singular values of matrices and their applications in multivariate analysis," *J. Multiv. Anal.*, vol. 9, pp. 362–377, 1979.

[42] Z. Ding, "Blind channel identification and equalization using spectral correlation measurements, part I: Frequency domain analysis," in *Cyclostationarity in Communications and Signal Processing* (W. A. Gardner, ed.), IEEE Press, 1993.

[43] S. V. Schell and W. A. Gardner, "Detection of the number of cyclostationary signals in unknown interference and noise," in *Proc. of the Twenty–Fourth Annual Asilomar Conf. on Signals, Systems, and Computers*, (Pacific Grove, CA), pp. 473–477, Nov. 1990.

[44] S. V. Schell, W. A. Gardner, and P. A. Murphy, "Blind adaptive antenna arrays in cellular communications for increased capacity," in *Proc. of 3rd Virginia Tech. Symp. on Wireless Personal Comm.*, June 1993.

[45] B. G. Agee, S. V. Schell, and W. A. Gardner, "Self–coherence restoral: A new approach to blind adaptation of antenna arrays," in *Proc. of the Twenty–First Asilomar Conf. on Signals, Systems, and Computers*, (Pacific Grove, CA), pp. 589–593, Nov. 1987.

[46] S. V. Schell and B. G. Agee, "Application of the SCORE algorithm and SCORE extensions to sorting in the rank-l self–coherence environment," in *Proc. of the Twenty–Second Annual Asilomar Conf. on Signals, Systems, and Computers*, (Pacific Grove, CA), pp. 274–278, Nov. 1988.

[47] B. G. Agee, S. V. Schell, and W. A. Gardner, "The SCORE approach to blind adaptive signal extraction: An application of the theory of spectral correlation," in *Proc. of the Fourth ASSP Workshop on Spectrum Estimation and Modelling*, (Minneapolis, MN), pp. 277–282, Aug. 1988.

[48] W. A. Gardner, S. V. Schell, and P. A. Murphy, "Multiplication of cellular radio capacity by blind adaptive spatial filtering," in *Proc. of IEEE Int'l. Conf. on Selected Topics in Wireless Comm.*, (Vancouver, B.C., Canada), pp. 102–106, June 1992.

[49] B. G. Agee, *The Property–Restoral Approach to Blind Adaptive Signal Extraction.* Ph.D. dissertation, Dept. of Electrical Engineering and Computer Science, University of California, Davis, CA, 1989.

[50] R. D. Gitlin, J. F. Hayes, and S. B. Weinstein, *Data Communication Principles.* New York: Plenum, 1992.

[51] R. D. Gitlin and S. B. Weinstein, "Fractionally-spaced equalization: An improved digital transversal equalizer," *Bell Syst. Tech. J.*, vol. 60, no. 2, pp. 275–296, Feb. 1981.

[52] W. A. Gardner and W. A. Brown, "Frequency–shift filtering theory for adaptive co–channel interference removal," in *Proc. of the Twenty–Third Annual Asilomar Conf. on Signals, Systems, and Computers*, (Pacific Grove, CA), pp. 562–567, Nov. 1989.

[53] J. G. Proakis, "Advances in equalization for intersymbol interference," in *Advances in Communication Systems* (A. J. Viterbi, ed.), New York: Academic, 1975. vol. 4.

[54] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 53, pp. 1349–1387, Sept. 1985.

[55] J. R. Treichler, V. Wolff, and C. R. Johnson, Jr., "Observed misconvergence in the constant modulus algorithm," in *Proc. Twenty–Fifth Asilomar Conf. on Signals, Systems and Computers*, (Pacific Grove, CA), pp. 663–667, 1991.

[56] Z. Ding, R. A. Kennedy, B. D. O. Anderson, and C. R. Johnson, Jr., "Ill-convergence of Godard blind equalizers in data communication systems," *IEEE Trans. Comm.*, vol. COM-39, no. 9, pp. 1313–1327, Sept. 1991.

[57] Z. Ding, C. R. Johnson, Jr., and R. A. Kennedy, "On the (non)existence of undesirable equilibria of Godard blind equalizers," *IEEE Trans. Signal Processing*, vol. SP-40, no. 10, pp. 2425–2432, Oct. 1992.

[58] L. Tong, G. Xu, and T. Kailath, "Blind channel identification and equalization using spectral correlation measurements, part II: A time-domain approach," in *Cyclostationarity in Communications and Signal Processing* (W. A. Gardner, ed.), IEEE Press, 1993.

[59] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Proc. Twenty–Fifth Asilomar Conf. on Signals, Systems and Computers*, (Pacific Grove, CA), pp. 856–860, Nov. 1991.

[60] G. Long, F. Ling, and J. G. Proakis, "Fractionally-spaced equalizers based on singular value decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (New York), pp. 1514–1517, 1988.

[61] M. Avriel, *Nonlinear Programming: Analysis and Methods.* Englewood Cliffs, N.J.: Prentice-Hall, 1976.