Programmable Canonical Correlation Analysis: A Flexible Framework for Blind Adaptive Spatial Filtering

Stephan V. Schell, *Member, IEEE,* and William A. Gardner, *Fellow, IEEE*

Abstract— **We present a new framework known as the programmable canonical correlation analysis (PCCA) for the design of blind adaptive spatial filtering algorithms that attempt to separate one or more signals of interest from unknown cochannel interference and noise. Unlike many alternatives, PCCA does not require knowledge of the calibration data for the array, directions of arrival, training signals, or spatial autocorrelation matrices of the noise or interferers. A novel aspect of PCCA is the ease with which new algorithms, targeted at capturing all signals from particular classes of interest, can be developed within this framework. In this paper, several existing algorithms are unified within the PCCA framework, and new algorithms are derived as examples. Analysis for the infinite-collect case and simulation for the finite-collect case illustrate the operation of specific algorithms within the PCCA framework.**

I. Introduction

In applications such as signal interception, and potentially in wireless communication systems of the future, it can be valuable or even essential for a receiver to adapt an array of sensors to separate one or more signals of interest (SOI's) from signals not of interest (SNOI's) and noise without the aid of a training or preamble signal, without information about directions of arrival and sensor array calibration, and without knowledge of the spatial autocorrelation matrix of the SNOI's and noise. In signals intelligence and interception, the unintended receiver is not likely to have some of this prior knowledge, and the remainder can be prohibitively expensive. In cooperative communication systems, it has been argued (e.g., see [1], [2]) that incorporating techniques for blind adaptive equalization and blind adaptive spatial filtering into the system as an integral part of the design can lead to improvements in both the quality of communication and the capacity of the system. Although some promising work has been done on this problem of blindly adapting a spatial filter, the various techniques (e.g., SCORE [3], AMUSE [4], CMA [5]) developed to date have been largely ad hoc and specific to a particular signal class.

Manuscript received December 7, 1993; revised May 29, 1995. This work was supported by the Office of Naval Research under Contract N00014-92-J-1218 and by E-Systems, Inc. (Greenville Division). The associate editor coordinating review of this paper and approving it for publication was Dr. R. D. Preuss.

S. V. Schell is with the Department of Electrical Engineering, The Pennsylvania State University, University Park, PA 16802 USA.

W. A. Gardner is with the Department of Electrical and Computer Engineering, University of California, Davis, CA 95616 USA.

IEEE Log Number 9415285.

In response to this lack of a *framework* for blind adaptive spatial filtering (as contrasted with a specific algorithm of more restricted utility), we have introduced the notion of programmable canonical correlation analysis (PCCA) [1], [6]–[8]. Using this framework to design an algorithm for a signal class of interest consists of specifying (or programming) a transformation that is used to distinguish each SOI from the others and from the SNOI's and noise. This technique is less direct than the conventional method that uses a training signal, but is potentially more powerful in the sense that less detailed prior knowledge is required.

In this paper we develop the PCCA framework by starting with the conventional approach that uses a known training signal, showing (theoretically and by example) how an alternative (albeit noisy) training signal can be derived from the received data, and then generalizing the technique to accommodate multiple SOI's simultaneously. Although adaptive techniques using data-derived training signals have been explored previously, the techniques described in [9] accommodate only a single SOI of a narrowly prescribed type, and the technique described in [3] does accommodate multiple SOI's but exploits only one statistical property, namely cyclostationarity. Also, the family of constant modulus techniques (cf. [10], [11] and references therein) can be interpreted as generating a training signal by performing memoryless modulus normalization operations on the equalizer output signal. In contrast, we show in this paper how PCCA easily accommodates multiple SOI's in the presence of multiple unknown SNOI's and can be programmed in a wide variety of ways to distinguish among various classes of signals (i.e., to determine what constitutes a SOI and what does not). General conditions on the programmable transformation are derived, and several specific transformations are considered to illustrate the utility of PCCA and the breadth of signal properties that can be used to distinguish each SOI from among the remaining signals and noise in the received data.

Several new algorithms are derived simply by specifying appropriate data transformations. Specifically, we derive new algorithms that separate signals on the basis of their differing temporal correlation properties, differing spectral densities (even when the signals are completely spectrally overlapping), differing temporal activity profiles (even when at least two signals are "on" at any time), and differing symbol-clock phases of bauded digital communication signals. Additionally, an algorithm is designed that can mitigate the effects of multipath

1053-587X/95$04.00 © 1995 IEEE

using prior approximate knowledge of the time-delay between the multiple paths. The class of Cross-SCORE and Conjugate Cross-SCORE algorithms [3] and their accelerated versions [1], [6], [7], which discriminate on the basis of differing spectral correlation properties (including, but not limited to, differing cycle frequencies, such as baud rates, doubled carrier frequencies, baseband tone frequencies, etc.), are shown to be specific algorithms within the PCCA framework.

II. Adaptation Without Conventional Training

In order to motivate the development of blind adaptive algorithms, we first argue briefly how a training signal can be obtained directly from the noisy, interference-corrupted received data, and why this noisy, substantially corrupted training signal can be used to adapt a spatial filter to finally obtain a high-quality signal estimate.

To this end, let the zero-mean sampled complex-envelope of the received data (a vector-valued time series) be modeled as

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

where $\mathbf{a}$ is the array response vector of the signal of interest $s(n)$ and $\mathbf{i}(n)$ is the interference plus noise, for which the spatial autocorrelation matrix $\mathbf{R_{ii}}$ is unknown.

Assume that knowledge of neither $s(n)$ nor $\mathbf{a}$ is available, but that we have somehow formed a signal $\hat{t}(n)$ having the property that it is correlated with $s(n)$ and uncorrelated with everything else (i.e., $\mathbf{i}(n)$) in the data

$$\hat{\mathbf{R}}_{\mathbf{x}\hat{t}} = \mathbf{a}\hat{R}_{s\hat{t}} + \hat{\mathbf{R}}_{\mathbf{i}\hat{t}} \rightarrow \mathbf{a}R_{s\hat{t}}. \tag{1}$$

Here, the symbol "$\rightarrow$" denotes "approaches in the limit as collection time N approaches ∞," and the estimated cross-correlation $\hat{\mathbf{R}}_{\mathbf{uv}}$ is defined as

$$\hat{\mathbf{R}}_{\mathbf{uv}} \triangleq \frac{1}{N}\sum_{n=1}^{N} \mathbf{u}(n)\mathbf{v}(n)^H \triangleq \langle \mathbf{u}(n)\mathbf{v}(n)^H\rangle_N$$

for any vector-valued signals $\mathbf{u}(n)$ and $\mathbf{v}(n)$, with superscript H denoting conjugate-transposition. Significantly, $\hat{t}(n)$ need not be proportional to $s(n)$. In fact, $\hat{t}(n)$ could be a substantially corrupted version of $s(n)$, as long as it is correlated with $s(n)$ and uncorrelated with $\mathbf{i}(n)$. Even though $\hat{t}(n)$ is a poor-quality estimate of $s(n)$, we can use $\hat{t}(n)$ as a training signal to compute a spatial filter

$$\mathbf{w}_x = \hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{x}\hat{t}} \rightarrow g\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{a} = \mathbf{w}_{\mathrm{SINR}} \tag{2}$$

for some scalar g, where $\mathbf{w}_{\mathrm{SINR}}$ is the weight vector that yields the maximum signal to interference and noise ratio (SINR).

We propose that this data-derived training signal $\hat{t}(n)$ be obtained according to

$$\hat{t}(n) = \mathbf{w}_y^H\mathbf{y}(n) \quad \text{where} \quad \mathbf{y}(n) \triangleq T[\mathbf{x}(n)]$$

for any appropriate transformation $T[\cdot]$. This transformation $T[\cdot]$ is called the *reference-path transformation* because it transforms the received data to provide the input to the reference-path spatial filter (having weights $\mathbf{w}_y$) that supplies the reference or training signal eventually used to train the primary spatial filter weights $\mathbf{w}_x$. In general, any transformation $\mathbf{y}(n)$ of $\mathbf{x}(n)$ such that

$$\hat{\mathbf{R}}_{\mathbf{xy}} \rightarrow \mathbf{a}\mathbf{b}^H \tag{3}$$

where $\mathbf{b}$ is not orthogonal to $\mathbf{w}_y$, will work because $\mathbf{w}_x$ is asymptotically (as $N \rightarrow \infty$) proportional to $\mathbf{w}_{\mathrm{SINR}}$

$$\mathbf{w}_x = \hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{xy}}\,\mathbf{w}_y \rightarrow \mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{a}(\mathbf{b}^H\mathbf{w}_y) = g\,\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{a} = \mathbf{w}_{\mathrm{SINR}}. \tag{4}$$

The sufficiency of the condition (3) holds regardless of whether the vector $\mathbf{y}(n)$ has a different number M_y of elements than the number M_x of elements in vector $\mathbf{x}(n)$, and regardless of whether $T[\cdot]$ has memory or is memoryless, is time-invariant or time-variant[1], or is linear or nonlinear.

We note here that an alternative approach to the problem of extracting a **single** SOI is to make use of insights from the signal-subspace perspective. In particular, under the condition $\mathbf{R_{iy}} = 0$, the dominant left singular vector of the matrix $\hat{\mathbf{R}}_{\mathbf{xy}}$ is a high-quality estimator of $\mathbf{a}$ (because $\hat{\mathbf{R}}_{\mathbf{xy}}$ converges to a rank-one matrix with column-space spanned by $\mathbf{a}$). If we call this estimator $\hat{\mathbf{a}}$, then we can form $\mathbf{w}_x = \hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{a}}$ and expect (correctly) to obtain a spatial filter that converges to the maximum SINR estimator of $s(n)$. However, unless some additional knowledge (such as the array manifold) is introduced into this approach, there is no way to separate multiple SOI's in this fashion if they all exhibit the desired property.

In the next section we discuss how our approach of using data-derived training signals can be generalized to handle this problem of separating multiple SOI's successfully.

III. Canonical Correlation Analysis

With the motivation provided by the previous section, we derive a general adaptation framework (which can be identified as being based on canonical correlation analysis) in which both the primary and reference-path spatial filters are jointly adapted. Simultaneously, we increase the number of signals of interest that can be accommodated to an arbitrary number L less than the number M_x of sensors in the array. In this section we derive the new framework and show that it is identical in form with the canonical correlation analysis problem from multivariate statistics. In the next section, we re-derive the new framework from a completely different point of view, one based on the maximum likelihood criterion. Subsequently, we derive the conditions needed for the new framework to yield algorithms that separate multiple SOI's.

Following the previous section, we propose a simple least-squares algorithm for jointly adapting the two spatial filters,

$$\min_{\mathbf{w}_x,\mathbf{w}_y} \langle|\hat{t}(n) - \hat{s}(n)|^2\rangle_N \tag{5}$$

subject to constraints ($\hat{\mathbf{R}}_{\hat{s}\hat{s}} = 1$, $\hat{\mathbf{R}}_{\hat{t}\hat{t}} = 1$), or equivalently (to within an arbitrary complex scalar)

$$\max_{\mathbf{w}_x,\mathbf{w}_y} \frac{|\hat{R}_{\hat{s}\hat{t}}|^2}{\hat{R}_{\hat{s}\hat{s}}\,\hat{R}_{\hat{t}\hat{t}}}. \tag{6}$$

[1] There is some abuse of notation in $\mathbf{y}(n) = T[\mathbf{x}(n)]$ when $T[\cdot]$ has memory.

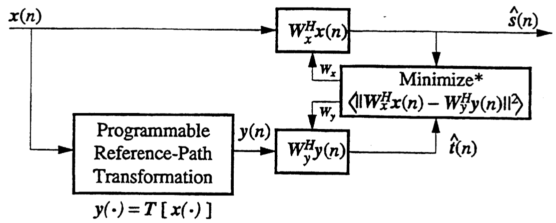

Fig. 1. Block diagram of generic PCCA processor showing the primary (upper) path and the reference (lower) path. *The minimization is constrained as in (7).

To accommodate a number L $(1 \leq L < M_x)$ of SOI's, denoted by the $L \times 1$ vector $\mathrm{s}(n)$, where the received data is now modeled by

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

in which $\mathbf{A}$ is an $M \times L$ matrix composed of L array response vectors, we jointly adapt L primary spatial filters and L reference-path spatial filters, denoted by the $M_x \times L$ matrix $\mathbf{W}_x$ and the $M_y \times L$ matrix $\mathbf{W}_y$, respectively. The corresponding least-squares optimization problem (generalized from (5) with $\hat{\mathrm{s}}(n) = \mathbf{W}_x^H \mathbf{x}(n)$ and $\hat{\mathrm{t}}(n) = \mathbf{W}_y^H \mathbf{y}(n)$) is then given by

$$\min_{\mathbf{W}_x, \mathbf{W}_y} \langle \|\mathbf{W}_y^H \mathbf{y}(n) - \mathbf{W}_x^H \mathbf{x}(n)\|^2 \rangle_N \tag{7}$$

subject to the constraints $\hat{\mathbf{R}}_{\hat{s}\hat{s}} = \mathbf{W}_x^H \hat{\mathbf{R}}_{\mathbf{xx}} \mathbf{W}_x = \mathbf{I}$ and similarly for $\mathbf{W}_y$, as depicted in Fig. 1. The constraints perform two functions here: they prevent the trivial solutions in which $\mathbf{W}_x$ and $\mathbf{W}_y$ equal zero, and they force the estimated SOI's to be uncorrelated. Identifying (7) as the canonical correlation analysis problem from multivariate statistics (with the apparently novel modification that one of the two data sets is actually a transformed version of the other), we move directly to the solution. The sets of spatial filters are given by the L dominant eigenvectors that satisfy the following equations

$$\hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \hat{\mathbf{R}}_{\mathbf{xy}} \hat{\mathbf{R}}_{\mathbf{yy}}^{-1} \hat{\mathbf{R}}_{\mathbf{xy}}^H \mathbf{w}_{x,i} = \lambda_i \mathbf{w}_{x,i} \tag{8}$$

$$\hat{\mathbf{R}}_{\mathbf{yy}}^{-1} \hat{\mathbf{R}}_{\mathbf{xy}}^H \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \hat{\mathbf{R}}_{\mathbf{xy}} \mathbf{w}_{y,i} = \lambda_i \mathbf{w}_{y,i} \tag{9}$$

where $\lambda_1 \geq \cdots \geq \lambda_L > \lambda_{L+1} \geq \cdots \geq \lambda_M$ is required for a unique solution, with $M = \min(M_x, M_y)$ (we ignore the eigenvalues $\lambda_{M+1}, \cdots, \lambda_{\max(M_x, M_y)}$ of the larger of the two systems (8)–(9) since rank $\{\hat{\mathbf{R}}_{\mathbf{xy}}\} \leq M$ implies that these are identically zero). In practice, only (8) need be solved, since $\mathbf{w}_y$ and $t(n)$ are really just intermediate quantities that are useful to us conceptually. The signal estimates of interest are given by $\hat{\mathrm{s}}(n) = \mathbf{W}_x^H \mathbf{x}(n)$.

Alternatively (but equivalently), the columns of $\mathbf{W}_x$ and $\mathbf{W}_y$ can be found from L of the M stationary points of (6). Straightforward complex matrix calculus (or careful use of the Cauchy–Schwartz inequality) leads to the solutions (8).

IV. Constrained Maximum Likelihood

Here we discuss an alternative (but, as we show, equivalent) approach to using canonical correlation analysis for extending the single-SOI technique of (5) to the case of multiple SOI's.

We consider the following two-step process: (1) develop a constrained maximum likelihood (CML) estimator $\hat{\mathrm{t}}(n)$ of $\mathrm{s}(n)$, and (2) use $\hat{\mathrm{t}}(n)$ in the conventional least-squares solution $\mathbf{W}_x = \hat{\mathbf{R}}_{\mathbf{xx}} \hat{\mathbf{R}}_{\mathbf{x}\hat{\mathbf{t}}}$ for the weight vector that will extract the final estimates of the SOI's, $\hat{s}(n) = \mathbf{W}_x^H \mathbf{x}(n)$. We emphasize that step 1 is not sufficient by itself because the transformation $T[\cdot]$ can substantially distort the SOI's (and thus $\hat{\mathrm{t}}(n)$ will contain distorted versions of the SOI's), so step two is necessary to obtain a final estimate of $\mathrm{s}(n)$.

Regarding our terminology, the CML estimator is *constrained* because $\hat{\mathrm{t}}(n)$ is constrained to be a linear transformation of $\mathbf{y}(n)$. The estimator is ML under the assumption that the interference plus noise, $\mathbf{i}(n)$, is a stationary zero-mean temporally-white complex Gaussian process with unknown spatial autocorrelation matrix $\mathbf{R}_{\mathbf{ii}}$, and that the SOI's $\mathrm{s}(n)$ are unknown signals.

A. Step 1: CML Data-Derived Training Signal

The log of the constrained likelihood function for the received data $\mathbf{x}(n)$, $1 \leq n \leq N$, in terms of the estimate $\hat{\mathbf{A}}$ of the unknown array response matrix, the estimate $\hat{\mathrm{t}}(n)$ of the unknown SOI waveforms $\mathrm{s}(n)$, and the estimate $\hat{\mathbf{R}}_{\mathbf{ii}}$ of the unknown spatial autocorrelation of the interference and noise, is

$$\phi(\hat{\mathbf{A}}, \{\hat{\mathrm{t}}(n)\}, \hat{\mathbf{R}}_{\mathbf{ii}}) = c_1 + N \ln|\hat{\mathbf{R}}_{\mathbf{ii}}^{-1}| - N\mathrm{tr}\{\hat{\mathbf{R}}_{\mathbf{ii}}^{-1} \mathbf{Q}(\hat{\mathbf{A}}, \{\hat{\mathrm{t}}(n)\})\} \tag{10}$$

where $\mathrm{tr}\{\cdot\}$ denotes the matrix trace operation, $c_1 = -NM\ln\pi$, and

$$\begin{aligned}\mathbf{Q}(\hat{\mathbf{A}}, \{\hat{\mathrm{t}}(n)\}) &\triangleq \langle(\mathbf{x}(n) - \hat{\mathbf{A}}\hat{\mathrm{t}}(n))(\mathbf{x}(n) - \hat{\mathbf{A}}\hat{\mathrm{t}}(n))^H\rangle_N \\ &= \hat{\mathbf{R}}_{\mathbf{xx}} - \hat{\mathbf{A}}\hat{\mathbf{R}}_{\mathbf{x}\hat{\mathbf{t}}}^H - \hat{\mathbf{R}}_{\mathbf{x}\hat{\mathbf{t}}}\hat{\mathbf{A}}^H + \hat{\mathbf{A}}\hat{\mathbf{R}}_{\hat{\mathbf{t}}\hat{\mathbf{t}}}\hat{\mathbf{A}}^H.\end{aligned}$$

Based on the derivation in Appendix A, we can show that the CML estimators for $\mathbf{R}_{\mathbf{ii}}$ and $\mathbf{A}$ are given by

$$\hat{\mathbf{R}}_{\mathbf{ii}}^{(\mathrm{ML})} = \mathbf{Q}(\hat{\mathbf{A}}^{(\mathrm{ML})}, \{\hat{\mathrm{t}}(n)\})$$
$$\hat{\mathbf{A}}^{(\mathrm{ML})} = \hat{\mathbf{R}}_{\mathbf{x}\hat{\mathbf{t}}} \hat{\mathbf{R}}_{\hat{\mathbf{t}}\hat{\mathbf{t}}}^{-1},$$

and that the remaining task of maximizing $\phi(\hat{\mathbf{A}}^{(\mathrm{ML})}, \{\hat{\mathrm{t}}(n)\}, \hat{\mathbf{R}}_{\mathbf{ii}}^{(\mathrm{ML})})$ with respect to $\mathbf{W}_y$ (in the constraint $\hat{\mathrm{t}}(n) = \mathbf{W}_y^H \mathbf{y}(n)$) is equivalent to minimizing the function

$$f(\mathbf{W}_y) \triangleq \frac{|\mathbf{W}_y^H \hat{\mathbf{P}} \mathbf{W}_y|}{|\mathbf{W}_y^H \hat{\mathbf{R}}_{\mathbf{yy}} \mathbf{W}_y|}$$

where

$$\hat{\mathbf{P}} = \hat{\mathbf{R}}_{\mathbf{yy}} - \hat{\mathbf{R}}_{\mathbf{xy}}^H \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \hat{\mathbf{R}}_{\mathbf{xy}}.$$

By the result proven in Appendix B, the solution $\mathbf{W}_y$ is any full-rank linear combination of the L least dominant generalized eigenvectors of the pair $(\hat{\mathbf{P}}, \hat{\mathbf{R}}_{\mathbf{yy}})$, provided that the eigenvalues satisfy $\lambda_1 \geq \cdots \geq \lambda_{M_y - L} > \lambda_{M_y - L + 1} \geq \cdots \lambda_M$. Finally, we note (omitting the straightforward matrix algebra) that these eigenvectors are identical to the L most dominant generalized eigenvectors of the pair $(\hat{\mathbf{R}}_{\mathbf{xy}}^H \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \hat{\mathbf{R}}_{\mathbf{xy}}, \hat{\mathbf{R}}_{\mathbf{yy}})$, which completes the derivation for step one.

If 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 $(\hat{\mathbf{R}}_{\mathbf{xy}}^H \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \hat{\mathbf{R}}_{\mathbf{xy}}, \hat{\mathbf{R}}_{\mathbf{yy}})$ instead of any linear combination thereof. This is equivalent to the CCA solution (9). In this case, a related result was proposed in [12] and proven in [13]. An alternate proof of their result can be obtained by noting that maximizing $\phi(\hat{\mathbf{A}}^{(\mathrm{ML})}, \{\hat{\mathbf{t}}(n)\}, \hat{\mathbf{R}}_{\mathbf{ii}}^{(\mathrm{ML})})$ is equivalent to maximizing

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

where $\lambda_m(\cdot,\cdot)$ denotes the mth generalized eigenvalue of the matrix pair $(\cdot,\cdot)$, and then applying the Poincaré Separation Theorem for generalized eigenvalues of a pair of Hermitian matrices (e.g., [14]). Also, the result of this section includes that in [15] as the special case in which one signal is present $(L = 1)$ and $\mathbf{y}(n) = \mathbf{x}^*(n)$.

B. Step 2: Using the Data-Derived Training Signal

In this step, $\hat{\mathbf{t}}(n)$ is used as a training signal to minimize the mean-squared-error between $\hat{\mathbf{t}}(n)$ and $\hat{\mathbf{s}}(n) = \mathbf{W}_x^H \mathbf{x}(n)$. Direct substitution of $\hat{\mathbf{t}}(n) = \mathbf{W}_y^H \mathbf{y}(n)$, with $\mathbf{W}_y$ obtained according to step one, into

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

and algebraic manipulation reveal that $\mathbf{W}_x$ can be any full-rank linear combination of the L most dominant eigenvectors of the pair $(\hat{\mathbf{R}}_{\mathbf{xy}} \hat{\mathbf{R}}_{\mathbf{yy}}^{-1} \hat{\mathbf{R}}_{\mathbf{xy}}^H, \hat{\mathbf{R}}_{\mathbf{xx}})$. As in step one, if the elements of $\hat{\mathbf{s}}(n) = \mathbf{W}_x^H \mathbf{x}(n)$ are constrained 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 the pair $(\hat{\mathbf{R}}_{\mathbf{xy}} \hat{\mathbf{R}}_{\mathbf{yy}}^{-1} \hat{\mathbf{R}}_{\mathbf{xy}}^H, \hat{\mathbf{R}}_{\mathbf{xx}})$, instead of any full-rank linear combination thereof. This is equivalent to the CCA solution (8).

V. Separation of Multiple SOIs from SNOIs and Noise

Having derived the new PCCA framework from two different perspectives (CCA and CML), we now discuss the conditions under which PCCA can separate multiple SOI's from each other, multiple SNOI's, and noise. In so doing, we show by example how some existing algorithms are unified under the new framework, and we derive several new algorithms, all by merely choosing appropriate reference-path transformations $T[\cdot]$.

For simplicity of analysis, we assume that the transformation is linear or linear-conjugate-linear (i.e., a linear combination of linear transformations of the signal and the conjugate of the signal). Although we have not analyzed the behavior of PCCA when operating with a nonlinear transformation, we see no fundamental barrier to using such a transformation; this point is left for future investigation. Also we consider only the algorithm behavior for $N \to \infty$; that is, all correlation matrices are assumed to be equal to their limit time-average values (or their ensemble average values, with appropriate assumptions of ergodicity). A full analytical performance evaluation is beyond the scope of this paper and is left for future work. In our discussion, we use the following notation

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{i}(n) \quad \text{and} \quad \mathbf{y}(n) = \mathbf{B}\mathbf{t}(n) + \mathbf{j}(n)$$

where

$$\mathbf{y}(n) \triangleq \mathbf{T}[\mathbf{x}(n)], \quad \mathbf{t}(n) \triangleq \mathbf{T}[\mathbf{s}(n)], \quad \mathbf{j}(n) \triangleq \mathbf{T}[\mathbf{i}(n)]$$

and $\mathbf{T}[\cdot]$ is the user-programmable transformation.

Before considering the case of practical interest, we develop intuition by explaining the behavior of PCCA in the absence of noise and SNOI's $(\mathbf{i}(n) = 0)$. We consider first the presence of uncorrelated SOI's, and second the presence of correlated SOI's. Finally, we consider correlated SOI's, multiple SNOI's, and noise.

A. Case 1: Uncorrelated SOI's, No SNOI's, and No Noise

In this case, $\mathbf{i}(n) = \mathbf{j}(n) = 0$. Denote $\mathbf{A}^H \mathbf{w}_x = \mathbf{g}_x$ (i.e., if $\mathbf{g}_x$ has only one nonzero entry, then $\mathbf{w}_x$ perfectly nulls all of the SOI's except one). Using this definition, the assumption that the number of SOI's is less than the number of sensors, and straightforward algebraic manipulation, we can show that (8) is equivalent to an eigenequation in terms of $\mathbf{g}_x$:

$$\mathbf{R}_{\mathbf{ss}}^{-1} \mathbf{R}_{\mathbf{st}} \mathbf{R}_{\mathbf{tt}}^{-1} \mathbf{R}_{\mathbf{ts}} \mathbf{g}_x = \lambda \mathbf{g}_x. \tag{11}$$

That is, for each $\mathbf{g}_x$ found in this manner, there is a vector $\mathbf{w}_x$ found from (8) having the property $\mathbf{A}^H \mathbf{w}_x = \mathbf{g}_x$. (One such vector, the minimum-norm solution, is $\mathbf{w}_x = \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{g}_x$.)

In light of the meaning of $\mathbf{g}_x$, we would like to find conditions under which these eigenvectors are proportional to the Euclidean basis vectors (i.e., vectors having only one nonzero element). Since a diagonal matrix having distinct diagonal elements is the only full-rank matrix with all eigenvectors proportional to the Euclidean basis vectors, we conclude the following:

A necessary and sufficient condition for PCCA to achieve perfect signal separation in the absence of SNOI's and noise is that the matrix $\mathbf{R}_{\mathbf{ss}}^{-1} \mathbf{R}_{\mathbf{st}} \mathbf{R}_{\mathbf{tt}}^{-1} \mathbf{R}_{\mathbf{ts}}$ be diagonal and have distinct diagonal elements.

We can also state a simpler sufficient condition, along with a useful implication:

A sufficient condition for PCCA to achieve perfect signal separation for uncorrelated SOI's in the absence of SNOI's and noise is that the correlation coefficients defined by

$$\rho_{s_k t_m} \triangleq R_{s_k t_m} / \sqrt{R_{s_k s_k} R_{t_m t_m}} \tag{12}$$

have the properties

$$|\rho_{s_k t_m}|^2 = 0 \qquad \text{for all } k \neq m, \tag{13}$$

$$|\rho_{s_k t_k}|^2 \neq |\rho_{s_n t_n}|^2 \qquad \text{for all } k \neq n. \tag{14}$$

Furthermore, assume that the SOI's are ordered in such a way that

$$|\rho_{s_1 t_1}|^2 > \cdots > |\rho_{s_L t_L}|^2. \tag{15}$$

Then the kth most dominant eigenvector of (8) *perfectly extracts the kth SOI.*

Some examples will help to illustrate the sufficient condition while simultaneously describing some of the reference-path transformations proposed to date. In all of these examples, we assume that there are L uncorrelated SOI's present.

1) Time-Shift Transformation: Assume that $\mathbf{R}_{\mathbf{ss}}(\tau)$ is diagonal for a particular time-shift τ. Let $\mathbf{y}(n) = \mathbf{x}(n-\tau)$. Then the matrix $\mathbf{R}_{\mathbf{ss}}^{-1}\,\mathbf{R}_{\mathbf{st}}\,\mathbf{R}_{\mathbf{tt}}^{-1}\,\mathbf{R}_{\mathbf{ts}}$ is diagonal with kth diagonal element equal to $|\rho_{s_k s_k}(\tau)|^2$. Provided that these correlation coefficient magnitudes are distinct, all SOI's can be perfectly separated by the L most-dominant PCCA eigenvectors. Here, we can interpret the condition that the diagonal elements be distinct to mean that the degree to which each signal is correlated with a time-shifted version of itself is distinct.

In a later section we discuss the behavior of PCCA using the time-shift transformation in the presence of correlated SOI's (e.g., multipath).

2) Frequency-Shift and Time-Shift Transformation: Assume that the L SOI's exhibit cyclostationarity (cf. [16], [17] for tutorial treatments in continuous time and discrete time, respectively) and share a common cycle frequency α. Assume also that the cyclic autocorrelation $\mathbf{R}_{\mathbf{ss}}^{\alpha}(\tau)$ defined by

$$\mathbf{R}_{\mathbf{ss}}^{\alpha}(\tau) \triangleq \langle \mathbf{s}(n)\,\mathbf{s}(n-\tau)^H\, e^{-j2\pi\alpha n}\rangle_{\infty}$$

is diagonal for some fixed time-shift τ (this could be so if the elements of $\mathbf{s}(n)$ are uncorrelated). Let $\mathbf{y}(n) = \mathbf{x}(n-\tau)\, e^{j2\pi\alpha n}$. Then the matrix $\mathbf{R}_{\mathbf{ss}}^{-1}\,\mathbf{R}_{\mathbf{st}}\,\mathbf{R}_{\mathbf{tt}}^{-1}\,\mathbf{R}_{\mathbf{ts}}$ is diagonal with kth diagonal element equal to the cyclic correlation coefficient $|\rho_{s_k s_k}^{\alpha}(\tau)|^2$ defined by

$$\rho_{s_k s_k}^{\alpha}(\tau) = R_{s_k s_k}^{\alpha}(\tau)/R_{ss}(0).$$

Provided that these diagonal elements are distinct, all SOI's can be perfectly separated by the L most-dominant PCCA eigenvectors. We can interpret the condition that the diagonal elements be distinct to mean that the degree to which each signal is correlated with a time-shifted and frequency-shifted version of itself is distinct. Following a convention in the literature on cyclostationarity, we refer to $|\rho_{s_k s_k}^{\alpha}(\tau)|^2$ as the feature strength of $s_k(n)$ for cycle frequency α and lag τ (derived from the interpretation of the quadratically regenerated sine wave at frequency α that appears in the signal $s_k(n)\, s_k^*(n-\tau)$ as a feature of the signal). Thus, PCCA can separate the SOI's if they have distinct feature strengths. This capability is identical to the sorting property exhibited by the Cross-SCORE algorithm, as discussed in [3]. This transformation is appropriate for SOI's created by ASK, FSK, PSK, and QAM modulators.

3) Conjugation, Frequency-Shift, and Time-Shift Transformation: Assume that the L SOI's exhibit conjugate cyclostationarity and share a common cycle frequency α. Let $\mathbf{y}(n) = \mathbf{x}^*(n-\tau)\, e^{j2\pi\alpha n}$. Then we can show by a similar analysis that PCCA can separate the SOI's if they have distinct conjugate feature strengths, where we define the conjugate feature strength to be $|\rho_{s_k s_k^*}^{\alpha}(\tau)|^2$, where $\rho_{s_k s_k^*}^{\alpha}(\tau)$ is the cyclic conjugate correlation coefficient defined by

$$\rho_{s_k s_k^*}^{\alpha}(\tau) = R_{s_k s_k^*}^{\alpha}(\tau)/R_{ss}(0).$$

This transformation is appropriate for SOI's created through double (or vestigial) sideband amplitude modulation with suppressed carrier, BPSK, ASK, real PAM, unbalanced QPSK or QAM, SQPSK, and so forth.

4) Multiple Frequency-Shifts and Time-Shifts and Conjugations: The two cases above can be generalized by defining the reference-path transformation to yield

$$\mathbf{y}(n) = \begin{bmatrix} \mathbf{x}^{(*)}(n-\tau_1)\, e^{-j2\pi\alpha_1 n} \\ \vdots \\ \mathbf{x}^{(*)}(n-\tau_K)\, e^{-j2\pi\alpha_K n} \end{bmatrix}$$

for any desired set of time-shifts $\tau_1, \ldots, \tau_K$ and frequency shifts $\alpha_1, \ldots, \alpha_K$ and optional conjugations $(*)$ on each appearance of $\mathbf{x}$. For example, for digital PSK or QAM SOI's with baud period $T = 4$ and bandwidth efficient pulses, the set of pairs

$$\{(\alpha_i, \tau_i)\} = \{(1/4, 0), (1/4, 1), (1/4, -1), (-1/4, 0), (-1/4, 1), (-1/4, -1)\}$$

allows PCCA to exploit the six significant samples of the cyclic autocorrelation at the baud rate. Simulation results showing a substantial reduction in convergence time, relative to the conventional Cross-SCORE algorithm, appear in [1], [6]–[8]. Hence the resulting algorithm is referred to as the accelerated SCORE algorithm.

5) Frequency Gating and Windowing: We can interpret the example of the time-shift transformation as a special case of the more general transformation in which $\mathbf{y}(n) = \mathbf{x}(n) \star h(n)$, where $h(n)$ is the impulse response of a linear time-invariant (LTI) filter, and $\star$ denotes convolution. Such filters can provide frequency gating (by passing some bands and rejecting others) or, more generally, frequency windowing or shaping.

We can consider the special case in which the filter $h(n)$ nulls out part of the band while passing the remainder with unity gain and zero phase. In this case, it is easy to show that $|\rho_{s_k t_k}|^2$ is simply the degree to which the average power of the kth SOI is contained within the passband of $h(n)$. If this newly defined feature strength is different for each SOI, then PCCA can separate them perfectly. We emphasize that these conditions can be satisfied even when the SOI's are spectrally overlapping in such a way that there is no band to which only a single SOI contributes.

Also, since the passband of $h(n)$ need not be a single contiguous band, potentially intricate schemes of discriminating among SOI's on the basis of their differing spectral densities can be created at the leisure of the transformation designer/programmer.

Finally, in the general case in which $h(n)$ can be any impulse response, the operative condition for separability is still the simply stated one in the sufficient condition. We can interpret it to mean that the degrees to which the $s_k(n)$ are correlated with their corresponding $t_k(n) = s_k(n)\star h(n)$ must be distinct over $k = 1, \ldots, L$.

6) Time Gating and Windowing: Let $\mathbf{y}(n) = \mathbf{x}(n)\, g(n)$, where $g(n)$ is a gating function that is equal to unity for some times and zero for others. Then, analogously to the interpretation of the LTI filtering transformation, we can define the feature strength $|\rho_{s_k t_k}|^2$ to be the degree to which the average power of the kth SOI is due to signal activity during the times when $g(n)$ is "on". We can then interpret the sufficiency condition to mean that PCCA can separate the multiple SOI's provided that each SOI has a distinct degree to which its average power is due to signal activity during the times when $g(n)$ is "on".

The gating function can be periodic (e.g., to match the frequency of a common duty cycle shared by the SOI's), although this is not a requirement. More generally, the gating function need not be binary-valued. That is, it can be a shaped window.

Analogously to our general interpretation of the class of LTI filtering operations, this class of transformations is capable of distinguishing among the SOI's on the basis of their different temporal activity profiles. We emphasize that these activity profiles need not be disjoint, and that there need not be any time periods during which only a single SOI is "on".

For example, for $k = 1, 2$ let $s_k(n)$ be the complex envelope of a PAM or QAM signal

$$s_k(n) = \sum_{m=-\infty}^{\infty} b_{km}\, p(n - mT - n_k) \tag{16}$$

where b_{km} is the mth baud of the kth SOI, $p(n)$ and T are the impulse response of the pulse-shaping filter and the baud period, respectively, that are common to the two SOI's, and n_k is the timing offset of the kth SOI (i.e., if $n_1 \neq n_2$, then the symbol clocks of the SOI's are not phase-synchronized). Also, let the gating function be defined by

$$f(n) = \sum_{m=-\infty}^{\infty} \delta(n - mT) \tag{17}$$

where $\delta(n)$ is the Kronecker delta equal to one for $n = 0$ and zero for $n \neq 0$. Although we could derive a general condition on the pulses and the timing offsets that would be sufficient for separability by PCCA, we choose instead to illustrate this numerically for brevity. As a function of the timing offset n_k, the feature strength $|\rho_{s_k t_k}|^2$ is shown in Table I for the choice $T = 8$ and $p(n)$ equal to a Nyquist-shaped pulse (raised cosine in the frequency domain) with 100% excess bandwidth. With these choices, both signals would be "on" during the entire data collection interval.

From the table, we can conclude that the feature strengths of the two SOI's are distinct (and thus that PCCA can separate them) provided that their timing offsets simultaneously satisfy the two conditions $n_1 \neq n_2$ and $n_1 \neq T - n_2$. Thus, we have shown that a PCCA processor can (theoretically) separate two digital communication signals on the basis of their differing symbol clock phases. Historically, only the Phase-SCORE algorithm [18] and its enhanced-convergence version [19], [8] had been shown to be able to accomplish this feat, although neither ad hoc algorithm has been shown to solve any optimization problem. The task of optimally choosing the windowing function $f(n)$ is left as an open problem. For our purpose of illustrating the potential capabilities of PCCA, our simple choice suffices.

TABLE I
FEATURE STRENGTH VERSUS TIMING OFFSET

n_k	0	1 or 7	2 or 6	3 or 5	4
$\lvert\rho_{s_k t_k}\rvert^2$	0.166	0.154	0.125	0.096	0.084

B. Case 2: Correlated SOI's, No SNOI's, No Noise

Here we generalize our intuition from uncorrelated SOI's to correlated SOI's. Provided that the sufficient conditions (13)–(14) are satisfied, PCCA can separate the uncorrelated SOI's. However, in the presence of correlated SOI's, the capability of PCCA to separate the SOI's from each other is more complicated, as we explain in the following.

To proceed, we denote by $\mathrm{s}_i(n)$ the ith block of signals in $\mathrm{s}(n)$ such that $\mathbf{R}_{\mathrm{s}_i\mathrm{s}_j} = 0$ for $i \neq j$. That is, the blocks are uncorrelated with each other. In addition, assume without loss of generality that the signals within $\mathrm{s}(n)$ are ordered in such a way that the signals belonging to the ith block occupy consecutive row positions in $\mathrm{s}(n)$

$$\mathrm{s}(n) = [\mathrm{s}_1^T(n) \cdots \mathrm{s}_K^T(n)]^T. \tag{18}$$

Thus, the matrix $\mathbf{R}_{\mathrm{ss}}^{-1}\mathbf{R}_{\mathrm{st}}\mathbf{R}_{\mathrm{tt}}^{-1}\mathbf{R}_{\mathrm{ts}}$ is block diagonal, and we denote by $\mathbf{R}_i$ the ith block, which corresponds to the ith block of signals.

Analogously to the uncorrelated-SOI's case, separation of the ith block of signals from the other blocks requires that $\mathbf{R}_i$ and $\mathbf{R}_j$ have no eigenvalues in common,

$$\lambda(\mathbf{R}_i) \cap \lambda(\mathbf{R}_j) = \phi \tag{19}$$

for all j such that $j \neq i$, where $\lambda(\cdot)$ denotes the set of all eigenvalues of its argument and ϕ denotes the empty set. This generalizes the necessary and sufficient condition of the previous subsection to the problem of separating blocks of correlated signals from each other.

In addition, we would like in some applications (e.g., blind adaptive arrays used to mitigate the effects of multipath) to separate the correlated SOI's within a given block. Specifically, we wish to separate the kth signal from all other signals in the ith block. As before, we can accomplish this only if there is associated with the kth signal in the ith block an eigenvalue of multiplicity one and an associated eigenvector belonging to the set of Euclidean basis vectors. These conditions are met if the condition (19) is satisfied and either the kth row or the kth column of $\mathbf{R}_i$ contains only one nonzero element, and this element differs from all of the eigenvalues of the matrix formed from $\mathbf{R}_i$ after the kth row and kth column are deleted.

Thus, although it is possible for PCCA to separate uncorrelated blocks of correlated SOI's from each other, more stringent conditions are needed to separate the correlated SOI's within each block.

We offer a single example to illustrate the point.

1) Time-Shift Transformation: Assume that $\mathrm{s}(n) = [s(n)\; s(n-d)]^T$ for some excess path delay d, and consider the performance of PCCA when $\mathbf{y}(n) = \mathbf{x}(n-d)$; that is, assume that the PCCA has prior knowledge of the multipath delay parameter d. Denote by $\rho(\tau)$ the autocorrelation coefficient of $s(n)$ at log τ. In the absence of noise, using straightforward algebraic manipulations we can show that

$$\mathbf{R}_{\mathbf{ss}}^{-1}\mathbf{R}_{\mathbf{st}}\mathbf{R}_{\mathbf{tt}}^{-1}\mathbf{R}_{\mathbf{ts}} = \frac{1}{b^2}\begin{bmatrix} |\rho(2d) - \rho^2(d)|^2 & 0 \\ r & b^2 \end{bmatrix}$$

where $b = 1 - |\rho(d)|^2$, and $r = bc^* + (\rho(2d) - \rho^2(d) - \rho^2(d))^* c$, and $c = \rho(d) - \rho^*(d)\rho(2d)$. It can be shown that if

$$\frac{|\rho(2d) - \rho^2(d)|^2}{(1 - |\rho(d)|^2)^2} \neq 1$$

then the two distinct eigenvalues are equal to this quotient and unity, respectively, and that the eigenvector corresponding to $\lambda = 1$ is proportional to $[0\ 1]^T$. Furthermore, since the eigenvalues are canonical correlation coefficients, each must satisfy $0 \leq \lambda \leq 1$. Thus, the dominant eigenvector perfectly separates $s(n-d)$ from $\mathrm{s}(n)$. Similarly, if the reference-path transformation $\mathbf{y}(n)$ is defined to be $\mathbf{y}(n) = \mathbf{x}(n+d)$, then it can be shown that the dominant eigenvector perfectly separates $s(n)$ from $\mathrm{s}(n)$.

This example is investigated further in a more realistic environment in Section VI, where additive noise is present and estimated correlation matrices are used.

C. Case 3: Correlated SOI's, Multiple SNOI's, Noise

In this case, neither $\mathbf{i}(n)$ nor $\mathbf{j}(n)$ is zero, which is typically true in practical applications. However, we are still assuming that $\mathbf{R}_{\mathbf{ij}} = 0$. That is, the reference-path transformation decorrelates the transformed noise and the received noise.

Unfortunately for our task of understanding the principle of operation of PCCA in this case, the relatively simple conditions developed in the previous subsection are not directly applicable. However, the intuition is still substantially sound, provided that some additional conditions are met.

The collection of MMSE weight vectors for all L SOI's is

$$\mathbf{W}_{\text{MMSE}} = \mathbf{R}_{\mathbf{xx}}^{-1}\,\mathbf{A}\,\mathbf{R}_{\mathbf{ss}} \tag{20}$$

and the collection of corresponding effective gains and phases of the adapted array for the L SOI's is

$$\mathbf{G}_{\text{MMSE}} = \mathbf{A}^H\mathbf{W}_{\text{MMSE}}. \tag{21}$$

Low MMSE and high SINR occur when

$$\langle\|\mathbf{W}_{\text{MMSE}}^H\mathbf{i}(n)\|^2\rangle \ll \langle\|\mathbf{G}_{\text{MMSE}}^H\mathrm{s}(n)\|^2\rangle \tag{22}$$

and $\mathbf{G}_{\text{MMSE}}$ is approximately diagonal. Using this observation, we can understand the behavior of PCCA in two steps: 1) we show that the interference rejection properties of PCCA are similar to those of the MMSE processor (this part concerns the left hand side of (22)), and 2) we show that the capability of PCCA to separate SOI's from each other is similar to that of the MMSE processor (this part concerns the right hand side of (22)).

To understand the interference rejection properties, we need simply to see that the PCCA eigenequation can be re-expressed as

$$\mathbf{W}_{\text{MMSE}}[\mathbf{R}_{\mathbf{ss}}^{-1}\mathbf{R}_{\mathbf{st}}\mathbf{B}^H\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{B}\mathbf{R}_{\mathbf{ts}}]\mathbf{A}^H\mathbf{w}_x = \lambda\mathbf{w}_x. \tag{23}$$

That is, the L dominant eigenvectors are various linear combinations of the MMSE weight vectors, so their interference rejection properties are similar.

To understand the SOI separation properties, we re-express the PCCA eigenequation as

$$\mathbf{G}_{\text{MMSE}}\mathbf{R}_{\mathbf{ss}}^{-1}\mathbf{R}_{\mathbf{st}}\mathbf{H}_{\text{MMSE}}\mathbf{R}_{\mathbf{tt}}^{-1}\mathbf{R}_{\mathbf{ts}}\mathbf{g}_x = \lambda\mathbf{g}_x \tag{24}$$

where $\mathbf{H}_{\text{MMSE}} \triangleq \mathbf{B}^H\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{B}\mathbf{R}_{\mathbf{tt}}$ is analogous to $\mathbf{G}_{\text{MMSE}}$. To proceed, assume that $\mathbf{G}_{\text{MMSE}}$ is approximately diagonal in the sense that

$$\mathbf{G}_{\text{MMSE}} = \tilde{\mathbf{G}}_{\text{MMSE}} + \mathbf{E}_G \tag{25}$$

where $\tilde{\mathbf{G}}_{\text{MMSE}}$ is diagonal and the residual $\mathbf{E}_G$ is small,

$$\sigma_{\min}\{\tilde{\mathbf{G}}_{\text{MMSE}}\} \gg \sigma_{\max}\{\mathbf{E}_G\} \tag{26}$$

where $\sigma_{\min}$ and $\sigma_{\max}$ denote the smallest and largest singular value, respectively. Assume a similar relation for $\mathbf{H}_{\text{MMSE}}$. Now we can simplify (24) to obtain

$$[\tilde{\mathbf{G}}_{\text{MMSE}}\mathbf{R}_{\mathbf{ss}}^{-1}\mathbf{R}_{\mathbf{st}}\tilde{\mathbf{H}}_{\text{MMSE}}\mathbf{R}_{\mathbf{tt}}^{-1}\mathbf{R}_{\mathbf{ts}} + \Delta]\mathbf{g}_x = \lambda\mathbf{g}_x \tag{27}$$

where Δ is the sum of three terms

$$\begin{aligned}\Delta = {} & \tilde{\mathbf{G}}_{\text{MMSE}}\mathbf{R}_{\mathbf{ss}}^{-1}\mathbf{R}_{\mathbf{st}}\mathbf{E}_H\mathbf{R}_{\mathbf{tt}}^{-1}\mathbf{R}_{\mathbf{ts}} \\ & + \mathbf{E}_G\mathbf{R}_{\mathbf{ss}}^{-1}\mathbf{R}_{\mathbf{st}}\,\tilde{\mathbf{H}}_{\text{MMSE}}\mathbf{R}_{\mathbf{tt}}^{-1}\mathbf{R}_{\mathbf{ts}} \\ & + \mathbf{E}_G\mathbf{R}_{\mathbf{ss}}^{-1}\mathbf{R}_{\mathbf{st}}\mathbf{E}_H\mathbf{R}_{\mathbf{tt}}^{-1}\mathbf{R}_{\mathbf{ts}}.\end{aligned} \tag{28}$$

Since Δ can be interpreted as a small perturbation on the left hand side of (27), we can use a result from perturbation theory for eigenvalue problems [20] to obtain the following expression

$$\mathbf{g}_i = \tilde{\mathbf{g}}_i + \tilde{\mathbf{G}}(\lambda_i\mathbf{I} - \Lambda)^+\,\mathbf{Q}^T\,\Delta\,\tilde{\mathbf{g}}_i \tag{29}$$

where $\mathbf{g}_i$ and $\tilde{\mathbf{g}}_i$ are the ith eigenvectors of the perturbed and unperturbed eigenequations, respectively, $\tilde{\mathbf{G}} = [\tilde{\mathbf{g}}_1 \cdots \tilde{\mathbf{g}}_L]$, λ_i is the ith eigenvalue of the unperturbed eigenequation, superscript + denotes the pseudo-inverse, and $\mathbf{Q}$ contains the right eigenvectors of the unperturbed eigenequation.

For a heuristic understanding, it may be sufficient to note simply that this result implies that the eigenvectors of the PCCA eigenequation have SOI separation properties (which can be inferred from the distance between $\mathbf{A}^H\mathbf{W}_x$ and the closest, possibly permuted, diagonal matrix) that are slightly perturbed from the SOI separation properties of the MMSE weight vectors, provided that the performance of the MMSE processor is sufficiently high (low MMSE, high SINR) and the necessary and sufficient condition (13)–(14) is satisfied.

However, we hasten to remind the reader that the argument leading to this conclusion is valid only asymptotically as $N \to \infty$. Nonetheless, simulation results show that the conclusion is also valid for moderate values of N. This in turn suggests that the results of this section provide a valid starting point for a more detailed analysis of the convergence properties (which is beyond the scope of this paper).

VI. Computer Simulations

An existing set of computer simulations [1], [6]–[8] presents the performance of the accelerated versions of Cross-SCORE as obtained from the PCCA framework, so these are not repeated here.

We present four sets of simulations to investigate the special cases of PCCA described in Section V: PCCA with $\mathbf{y}(n) = \mathbf{x}(n-\tau)$ to separate two signals on the basis of their differing temporal correlation properties, PCCA with $\mathbf{y}(n) = \mathbf{x}(n)*h(n)$ to separate two signals on the basis of their differing spectral densities, PCCA with $\mathbf{y}(n) = \mathbf{x}(n)\,g(n)$ to separate two signals on the basis of their differing symbol-clock phases, and PCCA with $\mathbf{y}(n) = \mathbf{x}(n-\tau)$ to separate a direct-path signal from a multipath reflection.

We use two types of signals in these simulations. The BPSK signals all have square-root-Nyquist (square-root of raised cosine in the frequency domain) pulse shaping with rolloff factor equal to unity (100% excess bandwidth) and zero carrier offset relative to the center of the analysis band. The Gaussian interferer is simply bandpass stationary Gaussian noise created by passing white Gaussian noise through a filter having unity gain in the passband $0 \leq f \leq 0.3$ and zero elsewhere. Simulations of PCCA involving other signal types, including narrowband FM, QPSK, and AMPS FM, appear in [21], [22], [8].

We specify the power γ_i of the ith signal relative to the total noise power in the analysis band, in dB. We measure the quality of the spatial filters found by PCCA in two ways: signal to interference and noise ratio (SINR) and bit error rate (BER). We denote the mean and standard deviation of the SINR of the ith signal by m_i and σ_i, respectively, and the mean and standard deviation of the BER of a BPSK signal by m_b and σ_b, respectively. We compute the BER in each trial by applying a square-root Nyquist filter to the output of the PCCA spatial filter, estimating the symbol timing (by estimating the phase of the quadratically regenerated spectral line at the symbol rate), sampling, making bit decisions, and comparing with the transmitted bit sequence.

In all cases, we perform 100 trials for each collection interval and each pair of signal powers. The array contains four elements, arranged linearly, and uniformly spaced by half the carrier-frequency wavelength.

A. Time-Shift Transformation for Signal Separation

A BPSK signal (signal 1) having baud period equal to 2 samples per baud and a Gaussian interferer (signal 2) arrive at the array. These two signals have different temporal correlation coefficients $|\rho_{s_k s_k}(\tau)|^2$ for $\tau = 1$ sample, with the narrowband signal having the stronger one. Thus, we specify our reference-path transformation to be $\mathbf{y}(n) = \mathbf{x}(n-1)$. According to our analysis in the previous sections, the most dominant PCCA eigenvector should extract the Gaussian interferer, and the next most dominant eigenvector should extract the BPSK signal. For a range of signal powers and collect times, the results are summarized in Table II. Despite the low signal power of the BPSK signal, PCCA converges in a reasonable number of samples to the maximum attainable SINR solutions (which are as follows: $(\gamma_1, \gamma_2) = (-5, 0)$ dB gives maximum attainable SINR's of 0.4 dB and 5.6 dB, respectively; $(\gamma_1, \gamma_2) = (-5, 10)$ dB gives maximum attainable SINR's of 0.25 dB and 15.6 dB, respectively). Other results (not shown here) indicate that this behavior continues for other input power pairs, including $(\gamma_1, \gamma_2) = (-5, 20)$, $(0, 20)$, and $(10, 20)$, and others.

TABLE II
Results Obtained Using Reference Data $\mathbf{y}(n) = \mathbf{x}(n-1)$ for Separating a BPSK Signal (Signal 1) from a Gaussian Signal (Signal 2). The BER Refers to the BPSK SOI (Signal 1)

γ_1	γ_2	T	m_1	σ_1	m_2	σ_2	m_b	σ_b
-5	0	16	-6.3	2.7	0.6	2.0	0.30	0.17
		64	-3.7	1.8	4.3	0.7	0.16	0.15
		256	-1.1	0.8	5.2	0.2	0.044	0.02
		1024	-0.2	0.2	5.4	0.1	0.028	0.005
		4096	0.03	0.07	5.5	0.04	0.024	0.004
-5	10	16	-5.9	2.5	7.9	2.4	0.27	0.17
		64	-3.4	1.8	13.0	1.1	0.15	0.15
		256	-0.87	0.7	14.8	0.4	0.04	0.02
		1024	-0.05	0.18	15.3	0.13	0.02	0.004
		4096	0.15	0.03	15.5	0.05	0.02	0.003

TABLE III
Results Obtained Using Reference Data $\mathbf{y}(n) = \mathbf{x}(n) \star h(n)$, Where $h(n)$ Is a Bandstop Filter that Rejects Only Half of the Band Occupied by the Gaussian Interferer (Signal 2). The BER Refers to the BPSK SOI (Signal 1).

γ_1	γ_2	T	m_1	σ_1	m_2	σ_2	m_b	σ_b
-5	0	16	-7.0	2.7	-0.4	2.5	0.33	0.17
		64	-4.6	2.2	2.8	1.4	0.21	0.16
		256	-2.4	1.4	4.7	0.5	0.09	0.09
		1024	-0.6	0.5	5.3	0.2	0.03	0.01
		4096	-0.1	0.2	5.5	0.1	0.025	0.004
-5	10	16	-6.6	2.7	4.5	3.5	0.31	0.17
		64	-4.2	2.1	8.9	2.2	0.20	0.16
		256	-2.2	1.3	12.5	1.2	0.09	0.10
		1024	-0.5	0.4	14.5	0.6	0.03	0.01
		4096	0.0	0.1	15.3	0.2	0.02	0.004

B. Bandstop Filtering Transformation

In this simulation the signal environment is identical to that in the previous simulation, but the reference-path transformation is different, being a bandstop filter. Two cases are considered: in the first, the bandstop filter rejects the lower half of the Gaussian interferer and part of the BPSK SOI; in the second, the bandstop filter rejects all of the Gaussian interferer and part of the BPSK SOI. In neither case could the bandstop filter suffice for interference rejection since it destroys part of the BPSK signal, but it does provide a reference signal for use by PCCA, which adapts two spatial filters that separate the two signals.

The results are shown in Tables III and IV. The maximum attainable SINR's are the same as in the previous simulation, and PCCA again converges to within a fraction of a dB of these optimal solutions. The strong similarity between the results obtained from the two different bandstop filters in the reference path show that this algorithm is relatively insensitive to the exact characteristics of the bandstop filter (e.g., if the spectral band of the narrowband interferer were known only roughly, the algorithm could still obtain good results). Other results (not shown here) indicate that this behavior continues for other input power pairs, including $(\gamma_1, \gamma_2) = (-5, 20)$, $(0, 20)$, and $(10, 20)$, and others.

TABLE IV
RESULTS OBTAINED USING REFERENCE DATA $\mathbf{y}(n) = \mathbf{x}(n) \star h(n)$, WHERE $h(n)$ IS A BANDSTOP FILTER THAT REJECTS ALL OF THE BAND OCCUPIED BY THE GAUSSIAN INTERFERER (SIGNAL 2). THE BER REFERS TO THE BPSK SOI (SIGNAL 1)

γ_1	γ_2	T	m_1	σ_1	m_2	σ_2	m_b	σ_b
-5	0	16	-6.1	2.4	3.2	1.1	0.27	0.17
		64	-3.6	1.9	5.1	0.3	0.15	0.13
		256	-2.0	1.2	5.5	0.1	0.08	0.08
		1024	-0.5	0.5	5.6	0.02	0.03	0.01
		4096	0.0	0.1	5.6	0.01	0.02	0.004
-5	10	16	-5.9	2.3	11.0	1.9	0.25	0.17
		64	-3.6	1.9	14.8	0.6	0.15	0.13
		256	-1.9	1.2	15.5	0.08	0.07	0.08
		1024	-0.4	0.5	15.6	0.01	0.03	0.01
		4096	0.04	0.1	15.6	0.005	0.02	0.004

TABLE V
RESULTS OBTAINED USING THE PERIODIC TIME-GATING TRANSFORMATION TO SEPARATE TWO SIGNALS HAVING THE SAME BAUD RATE BUT DIFFERENT SYMBOL-CLOCK PHASES

γ_1	γ_2	T	m_1	σ_1	m_2	σ_2	m_b	σ_b
-5	0	16	-3.9	2.1	0.6	2.1	0.2	0.16
		64	-1.6	1.2	3.7	1.0	0.06	0.06
		256	-0.2	0.3	5.0	0.3	0.03	0.006
		1024	0.2	0.1	5.4	0.1	0.02	0.004
		4096	0.3	0.03	5.5	0.03	0.02	0.003
-5	10	16	-4.0	2.1	6.0	2.9	0.2	0.17
		64	-1.6	1.2	10.9	1.7	0.07	0.07
		256	-0.3	0.3	13.8	0.8	0.03	0.006
		1024	0.09	0.09	15.0	0.3	0.02	0.004
		4096	0.2	0.02	15.4	0.1	0.02	0.003

TABLE VI
RESULTS OBTAINED USING PHASE SCORE TO SEPARATE TWO SIGNALS HAVING THE SAME BAUD RATE BUT DIFFERENT SYMBOL-CLOCK PHASES

γ_1	γ_2	T	m_1	σ_1	m_2	σ_2	m_b	σ_b
-5	0	16	-3.3	1.6	1.2	1.7	0.16	0.14
		64	-1.5	1.0	3.7	1.0	0.05	0.03
		256	-0.2	0.3	5.0	0.28	0.03	0.006
		1024	0.2	0.1	5.4	0.09	0.02	0.004
		4096	0.3	0.03	5.5	0.03	0.02	0.003
-5	10	16	-3.3	1.5	6.1	2.8	0.15	0.14
		64	-1.5	1.0	10.9	1.7	0.06	0.05
		256	-0.3	0.3	13.8	0.8	0.03	0.006
		1024	0.1	0.09	15.0	0.3	0.02	0.004
		4096	0.2	0.02	15.4	0.1	0.02	0.003

C. Time Gating Transformation

In this simulation, two BPSK SOI's arrive at the array. They have identical baud periods (2 samples per baud) and identical carriers. However, their symbol-clock phases are offset from each other by 180° (one time sample). Using the time gating transformation described in Section V, PCCA is able to separate these two signals according to their differing symbol-clock phases. The Phase SCORE algorithm, which is able to do this using a different approach, is also simulated, and performs nearly the same as PCCA.

As shown in Tables V and VI, the two approaches have nearly identical performance. However, we note that the performance of PCCA might be improved by modifying the time-gating function; the simple one is used here primarily to demonstrate the concept.

D. Delay Transformation for Multipath Mitigation

In this simulation, we consider the performance of PCCA using a delay transformation to spatially separate a direct path and a delayed path (delayed by one sample) of the same BPSK signal. Based on our discussion in Section V, we choose the reference path transformation to be $\mathbf{y}(n) = \mathbf{x}(n-1)$. PCCA then finds two weight vectors corresponding to two significant eigenvalues: the dominant one extracts the delayed path, and the second one extracts some arbitrary linear combination of the direct path and the delayed path. As shown in Table VII, PCCA converges quickly to the maximum attainable SINR solution (0.6 dB), and the correspondingly low BER shows that the multipath has indeed been mitigated.

TABLE VII
RESULTS OBTAINED USING REFERENCE DATA $\mathbf{y}(n) = \mathbf{x}(n-1)$ FOR MULTIPATH MITIGATION. SIGNAL 1 IS THE DELAYED PATH AND SIGNAL 2 IS THE DIRECT PATH

γ_1	γ_2	T	m_1	σ_1	m_2	σ_2	m_b	σ_b
-5	-5	16	-3.8	2.0	-7.8	2.8	0.18	0.14
		64	-0.4	0.5	-6.1	2.7	0.04	0.02
		256	0.32	0.15	-6.1	2.5	0.02	0.006
		1024	0.5	0.03	-5.7	2.6	0.02	0.005
		4096	0.6	0.001	-5.8	2.6	0.02	0.004

VII. CONCLUSION

We have derived a new framework for blind adaptive spatial filtering that unifies several existing algorithms and facilitates the development of new ones. By analysis of the infinite-collect case and simulation of the finite-collect case, we have illustrated the conditions under which algorithms within the new framework can approach the maximum attainable SINR solutions. In the process of illustrating these properties of PCCA, we derived new algorithms that can separate signals on the basis of their differing temporal correlation properties, differing spectral densities, or differing symbol-clock phases.

In related work not reported here, the PCCA framework has been used to accelerate the convergence of the Cross-SCORE algorithm [1], [6]–[8], and modifications to the PCCA framework, motivated by the work of Biedka [19] on using rank-reduction techniques to accelerate the Cross-SCORE and Phase-SCORE algorithms, have been shown in simulations [21] to accelerate the convergence of numerous algorithms within the PCCA framework.

Several open problems remain to be investigated, including analytical performance evaluation in the finite-collect case, optimization of the time-gating transformation used to separate signals on the basis of their differing symbol-clock phases, and investigation of other reference-path transformations, including nonlinear ones. In [22], the PCCA framework is generalized to incorporate recursive processing; the constant modulus algorithm is then derived from this general recursive PCCA framework by choosing a nonlinear transformation $T[\cdot]$ that normalizes the modulus of the reference-path data.

APPENDIX A
CML DERIVATION: STEP 1

A. Reduction to a Determinantial Form

Using the identities $\nabla_{\mathbf{A}} \text{tr}\{\mathbf{AB}\} = \mathbf{B}^T$ and $\nabla_{\mathbf{A}} \ln|\mathbf{A}| = (\mathbf{A}^{-1})^T$, the complex gradient (defined as in [23]) of $\phi(\hat{\mathbf{A}},$

$\{\hat{\mathbf{t}}(n)\}, \hat{\mathbf{R}}_{\mathrm{ii}})$ with respect to $\hat{\mathbf{R}}_{\mathrm{ii}}^{-1}$ can be shown to be given by

$$\nabla_{\hat{\mathbf{R}}_{\mathrm{ii}}^{-1}}\phi(\hat{\mathbf{A}}, \{\hat{\mathbf{t}}(n)\}, \hat{\mathbf{R}}_{\mathrm{ii}}) = N\hat{\mathbf{R}}_{\mathrm{ii}}^{T} - N\mathbf{Q}(\hat{\mathbf{A}}, \{\hat{\mathbf{t}}(n)\})^{T}.$$

Equating the gradient to zero and solving for $\hat{\mathbf{R}}_{\mathrm{ii}}$ yields the ML estimate $\hat{\mathbf{R}}_{\mathrm{ii}}^{(\mathrm{ML})}$ given by

$$\hat{\mathbf{R}}_{\mathrm{ii}}^{(\mathrm{ML})} = \mathbf{Q}(\hat{\mathbf{A}}, \{\hat{\mathbf{t}}(n)\}).$$

The partially maximized log-likelihood is then given by

$$\phi(\hat{\mathbf{A}}, \{\hat{\mathbf{t}}(n)\}, \hat{\mathbf{R}}_{\mathrm{ii}}^{(\mathrm{ML})}) = c_2 - N\ln|\mathbf{Q}(\hat{\mathbf{A}}, \{\hat{\mathbf{t}}(n)\})|$$

where $c_2 = c_1 - NM$. Using the identity

$$\begin{aligned}\nabla_{\mathbf{A}}\ln|\mathbf{D} - \mathbf{A}\mathbf{B}^H - \mathbf{B}\mathbf{A}^H + \mathbf{A}\mathbf{C}\mathbf{A}^H| \\ = [(-\mathbf{B}^H + \mathbf{C}\mathbf{A}^H)\mathbf{E}^{-1}]^T\end{aligned} \quad (30)$$

where

$$\mathbf{E} = \mathbf{D} - \mathbf{A}\mathbf{B}^H - \mathbf{B}\mathbf{A}^H + \mathbf{A}\mathbf{C}\mathbf{A}^H,$$

the gradient of $\phi(\hat{\mathbf{A}}, \{\hat{\mathbf{t}}(n)\}, \hat{\mathbf{R}}_{\mathrm{ii}}^{(\mathrm{ML})})$ with respect to $\hat{\mathbf{A}}$ can be shown to be given by

$$\begin{aligned}&\nabla_{\hat{\mathbf{A}}}\phi(\hat{\mathbf{A}}, \{\hat{\mathbf{t}}(n)\}, \hat{\mathbf{R}}_{\mathrm{ii}}^{(\mathrm{ML})}) \\ &= N[(\hat{\mathbf{R}}_{\mathbf{x}\hat{\mathbf{t}}} - \hat{\mathbf{A}}\hat{\mathbf{R}}_{\hat{\mathbf{t}}\hat{\mathbf{t}}})^H\mathbf{Q}(\hat{\mathbf{A}}, \{\hat{\mathbf{t}}(n)\})^{-1}]^T.\end{aligned}$$

Equating this gradient to zero and solving for $\hat{\mathbf{A}}$ yields the ML estimate $\hat{\mathbf{A}}^{(\mathrm{ML})}$ given by

$$\hat{\mathbf{A}}^{(\mathrm{ML})} = \hat{\mathbf{R}}_{\mathbf{x}\hat{\mathbf{t}}}\hat{\mathbf{R}}_{\hat{\mathbf{t}}\hat{\mathbf{t}}}^{-1}.$$

Thus, after substituting $\hat{\mathbf{A}}^{(\mathrm{ML})}$ into $\phi(\hat{\mathbf{A}}^{(\mathrm{ML})}, \{\hat{\mathbf{t}}(n)\}, \hat{\mathbf{R}}_{\mathrm{ii}}^{(\mathrm{ML})})$ and using the identity

$$|\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^H| = |\mathbf{A}||\mathbf{C} - \mathbf{B}^H\mathbf{A}^{-1}\mathbf{B}|/|\mathbf{C}|$$

the partially maximized log-likelihood function can be expressed as

$$\begin{aligned}&\phi(\hat{\mathbf{A}}^{(\mathrm{ML})}, \{\hat{\mathbf{t}}(n)\}, \hat{\mathbf{R}}_{\mathrm{ii}}^{(\mathrm{ML})}) \\ &= c_2 - N\ln|\hat{\mathbf{R}}_{\mathbf{xx}} - \hat{\mathbf{R}}_{\mathbf{x}\hat{\mathbf{t}}}\hat{\mathbf{R}}_{\hat{\mathbf{t}}\hat{\mathbf{t}}}^{-1}\hat{\mathbf{R}}_{\mathbf{x}\hat{\mathbf{t}}}^H| \\ &= c_3 - N\ln|\hat{\mathbf{R}}_{\hat{\mathbf{t}}\hat{\mathbf{t}}} - \hat{\mathbf{R}}_{\mathbf{x}\hat{\mathbf{t}}}^H\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{x}\hat{\mathbf{t}}}|/|\hat{\mathbf{R}}_{\hat{\mathbf{t}}\hat{\mathbf{t}}}|\end{aligned} \quad (31)$$

where $c_3 = c_2 - N\ln|\hat{\mathbf{R}}_{\mathbf{xx}}|$.

Maximizing (31) with respect to $\mathbf{W}_y$ (in $\{\hat{\mathbf{t}}(n)\} = \{\mathbf{W}_y^H\mathbf{y}(n)\}$) is equivalent to minimizing the function

$$f(\mathbf{W}_y) \triangleq \frac{|\mathbf{W}_y^H\hat{\mathbf{P}}\mathbf{W}_y|}{|\mathbf{W}_y^H\hat{\mathbf{R}}_{\mathbf{yy}}\mathbf{W}_y|}$$

where

$$\hat{\mathbf{P}} = \hat{\mathbf{R}}_{\mathbf{yy}} - \hat{\mathbf{R}}_{\mathbf{xy}}^H\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{xy}}.$$

This concludes the derivation of the first step of the CML solution.

APPENDIX B
CML DERIVATION: STEP 2

A. Minimization of a Determinantial Form

Result: The function $f(\mathbf{W}) \triangleq |\mathbf{W}^H\mathbf{A}\mathbf{W}|/|\mathbf{W}^H\mathbf{B}\mathbf{W}|$, where $\mathbf{A}$ and $\mathbf{B}$ are $M \times M$ Hermitian matrices with $\mathbf{A} \geq 0$ and $\mathbf{B} > 0$ and $\mathbf{W}$ is an $M \times L$ matrix, is minimized with respect to $\mathbf{W}$ by any full-rank linear combination of the L least dominant generalized eigenvectors $\mathbf{U}$ of the pair $(\mathbf{A}, \mathbf{B})$.

Proof: First, we note that $\mathbf{U}^H\mathbf{A}\mathbf{U} = \mathbf{D}_a$ and $\mathbf{U}^H\mathbf{B}\mathbf{U} = \mathbf{D}_b$, where $\mathbf{D}_a \geq 0$ and $\mathbf{D}_b > 0$ are diagonal matrices, with $\mathbf{D}_a\mathbf{D}_b^{-1} = \Lambda$ being the generalized eigenvalues. Then any $\mathbf{W}$ can be expressed as $\mathbf{W} = \mathbf{U}\mathbf{D}_b^{-1/2}\mathbf{C}$ for some $\mathbf{C}$. Substituting, we have (after straightforward simplification).

$$g(\mathbf{C}) \triangleq f(\mathbf{U}\mathbf{D}_b^{-1/2}\mathbf{C}) = |\mathbf{C}^H\Lambda\mathbf{C}|/|\mathbf{C}^H\mathbf{C}|.$$

For $g(\mathbf{C})$ to be defined, the condition $|\mathbf{C}^H\mathbf{C}| \neq 0$ implies that the $M \times L$ matrix $\mathbf{C}$ has full rank. Since $M > L$, there exists a subset of L linearly independent rows. Denote these rows by the invertible matrix $\mathbf{C}_L$. Without loss of generality (since Λ is not ordered, except with respect to $\mathbf{U}$ which is also unordered), let these rows be the first L rows of $\mathbf{C}$. Thus, we form the partitions

$$\mathbf{C} = \begin{bmatrix}\mathbf{C}_L \\ \mathbf{b}\end{bmatrix}, \quad \text{and} \quad \Lambda = \begin{bmatrix}\Lambda_L & 0 \\ 0 & \Lambda_b\end{bmatrix}$$

where $\mathbf{b}$ has dimension $(M-L) \times L$. Re-expressing $g(\mathbf{C})$, we obtain

$$\begin{aligned}g(\mathbf{C}) &= \frac{|\mathbf{C}_L^H\Lambda_L\mathbf{C}_L + \mathbf{b}^H\Lambda_b\mathbf{b}|}{|\mathbf{C}_L^H\mathbf{C}_L + \mathbf{b}^H\mathbf{b}|} \\ &= |\Lambda_L|\frac{|\mathbf{I} + (\Lambda_b^{1/2}\mathbf{d}\Lambda_L^{-1/2})^H(\Lambda_b^{1/2}\mathbf{d}\Lambda_L^{-1/2})|}{|\mathbf{I} + \mathbf{d}^H\mathbf{d}|}\end{aligned}$$

where we obtain the simplification by using $\mathbf{d} \triangleq \mathbf{b}\mathbf{C}_L^{-1}$ and the identity $|\mathbf{X}\mathbf{Y}| = |\mathbf{X}||\mathbf{Y}|$ for any square matrices $\mathbf{X}$ and $\mathbf{Y}$.

Now, a consequence of the Minkowski determinant inequality (cf. [24], pp. 115–117) is that both $|\mathbf{I} + (\Lambda_b^{1/2}\mathbf{d}\Lambda_L^{-1/2})^H(\Lambda_b^{1/2}\mathbf{d}\Lambda_L^{-1/2})|$ and $|\mathbf{I} + \mathbf{d}^H\mathbf{d}|$ are nondecreasing (concave) functions of $\mathbf{d}$. To aid our analysis of the ratio of these two functions, we note that

$$\begin{aligned}&\nabla_{\mathbf{d}}|\mathbf{I} + (\Lambda_b^{1/2}\mathbf{d}\Lambda_L^{-1/2})^H(\Lambda_b^{1/2}\mathbf{d}\Lambda_L^{-1/2})| \\ &= \Lambda_b^{1/2}\left[(\nabla_{\mathbf{d}}|\mathbf{I} + \mathbf{d}^H\mathbf{d}|)|_{\mathbf{d}\Lambda_b^{1/2}\mathbf{d}\Lambda_L^{-1/2}}\right]\Lambda_L^{-1/2}.\end{aligned}$$

That is, if $[\Lambda_b]_{ii}/[\Lambda_L]_{jj} > 1$ then the numerator increases *more* rapidly as a function of d_{ij} than does the denominator. On the other hand, if $[\Lambda_b]_{ii}/[\Lambda_L]_{jj} < 1$ then the numerator increases *less* rapidly as a function of d_{ij} than does the denominator. We treat the ratio in two exhaustive cases.

Case 1: There exists at least one pair i, j such that $[\Lambda_b]_{ii}/[\Lambda_L]_{jj} < 1$. Thus, since the numerator of $g(\mathbf{C})$ increases less rapidly as a function of d_{ij} than does the denominator, $g(\mathbf{C})$ is a strictly decreasing function of d_{ij}. Conversely, if $[\Lambda_b]_{ii}/[\Lambda_L]_{jj} > 1$, then $g(\mathbf{C})$ is a strictly increasing function of d_{ij}. That is, $g(\mathbf{C})$ is minimized only by increasing without bound the magnitudes of all d_{ij} for which $[\Lambda_b]_{ii}/[\Lambda_L]_{jj} < 1$. In this limit, $g(\mathbf{C})$ approaches the product of the L smallest diagonal elements of Λ. However, since we are not interested in infinite-norm solutions, we pass on to Case 2.

Case 2: For all pairs i, j we have $[\Lambda_b]_{ii}/[\Lambda_L]_{jj} > 1$. Thus, $g(\mathbf{C})$ is a strictly increasing function of $\mathbf{d}$ and has its global minimum at $\mathbf{d} = 0$, at which $g(\mathbf{C}) = |\Lambda_L|$, where Λ_L are the L smallest elements of Λ. Thus, with $\mathbf{C} = [\mathbf{C}_L^H,\ 0]^H$ for any invertible $\mathbf{C}_L$, we see that the corresponding $\mathbf{W} = \mathbf{U}\mathbf{D}_2^{-1/2}\mathbf{C}$ is any full-rank linear combination of the least dominant generalized eigenvectors of $(\mathbf{A}, \mathbf{B})$, provided that the Lth smallest eigenvalue is less than the $(L+1)$th.

This completes the proof of this result and with it the second step of the derivation of the CML solution.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their comments that led to substantial improvements to this paper.

REFERENCES

[1] S. V. Schell, "An overview of sensor array processing for cyclostationary signals," in *Cyclostationarity in Communications and Signal Processing*, W. A. Gardner, Ed. New York: IEEE, 1994, ch. 3, pp. 168–239.

[2] S. V. Schell and W. A. Gardner, "Blind adaptive antenna arrays for increased capacity in cellular communications," in *Wireless Personal Communications: Trends and Challenges*, T. S. Rappaport, B. D. Woerner, and J. H. Reed, Eds. Boston: Kluwer, 1994, pp. 15–24.

[3] 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, pp. 753–767, Apr. 1990.

[4] L. Tong, R. Liu, V. C. Soon, and Y.-F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circ. Syst.*, vol. 38, pp. 499–509, May 1991.

[5] R. Gooch and J. Lundell, "The CM array: An adaptive beamformer for constant modulus signals," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, Tokyo, Japan, 1986, pp. 2523–2526.

[6] S. V. Schell and W. A. Gardner, "Spatio-temporal filtering and equalization for cyclostationary signals," in *Control and Dynamic Systems, Volume 64: Stochastic Techniques in Digital Signal Processing Systems, Part 1 of 2*, C. T. Leondes, Ed. New York: Academic, ch. 1, pp. 1–68, 1994.

[7] ——, "Programmable canonical correlation analysis: A flexible framework for blind adaptive spatial filtering," in *Proc. 27th Asilomar Conf. Sig. Syst. Comp.*, Nov. 1993, pp. 647–652.

[8] W. A. Gardner, J. Schenck, and S. V. Schell, "Programmable blind adaptive spatial filtering," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, Adelaide, Australia, 1994, pp. IV:53–56.

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

[10] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 459–472, Apr. 1993.

[11] J. J. Shynk and R. P. Gooch, "Convergence properties of the multistage CMA adaptive beamformer," in *Proc. Asilomar Conf. Signals Syst. Comput.*, 1993, pp. 622–626.

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

[13] M. Siotani, T. Hayakawa, and Y. Fujikoshi, *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook.* Columbus, OH: Amer. Sci., 1985.

[14] 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.

[15] B. G. Agee, "Maximum-likelihood approaches to blind adaptive signal extraction using narrowband antenna arrays," in *Proc. 25th Asilomar Conf. Signals Syst. Comput.*, 1991.

[16] W. A. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Processing Mag.*, vol. 8, pp. 14–37, Apr. 1991.

[17] W. A. Gardner, "An introduction to cyclostationary signals," in *Cyclostationarity in Communications and Signal Processing*, W. A. Gardner, Ed. New York: IEEE, 1994, ch. 1, pp. 1–90.

[18] 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. 22nd Ann. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, Nov. 1988, pp. 274–278.

[19] T. E. Biedka, "Subspace-constrained score algorithms," in *Proc. 27th Asilomar Conf. Signals Syst. Comput.*, Nov. 1993, pp. 716–720.

[20] A. Y. T. Leung, "Perturbed general eigensolutions," *Commun. Appl. Numer. Methods*, vol. 6, pp. 401–409, 1990.

[21] J. L. Schenck and W. A. Gardner, "Blind adaptive spatial processing in a mobile radio environment," in *Proc. 2nd Workshop Cyclostationary Signals*, S. V. Schell and C. M. Spooner, Eds., Monterey, CA, Aug. 1994, pp. 17.1–17.7.

[22] M. F. Kahn, M. A. Mow, W. A. Gardner, and T. E. Biedka, "A recursive programmable canonical correlation analyzer," in *Proc. 2nd Workshop on Cyclostationary Signals*, S. V. Schell and C. M. Spooner, Eds., Monterey, CA, Aug. 1994, pp. 18.1–18.12.

[23] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *Proc. Inst. Elect. Eng.*, vol. 130, pts. F and H, no. 1, pp. 11–16, Feb. 1983.

[24] M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities.* Boston, MA: Allyn and Bacon, 1964.

Stephan V. Schell (S'84–M'91) was born in Livermore, CA in 1964. He received the B.S. (highest honors), M.S., and Ph.D. degrees from the University of California, Davis in 1986, 1987, and 1990, respectively.

From 1986 to 1990, he worked as a Research and Teaching Assistant in the Department of Electrical and Computer Engineering at the University of California, Davis (UCD). From 1990 to 1992, he pursued postdoctoral research and served as an instructor at UCD. He is currently taking a one-year leave of absence from the Department of Electrical Engineering at The Pennsylvania State University, where he has been an Assistant Professor since 1992, to focus on research as a Visiting Scholar at UCD. He is also an independent consultant for Statistical Signal Processing, Inc., Mission Research Corporation, and Booz-Allen and Hamilton, Inc. His research interests include blind adaptive filtering, digital communication, system identification, detection and estimation theory, the theory and application of cyclostationarity, signal processing in entomology, and identification of ionospheric propagation modes.

Dr. Schell has been chair of the Central Pennsylvania chapter of the IEEE Signal Processing Society (1993–1995) and was co-organizer (with C. M. Spooner) of the 1994 Workshop on Cyclostationary Signals. He holds two patents in the area of sensor array processing. While at UCD, he was awarded a UCD MICRO Fellowship, UCD Distinguished Scholar Research Award, UCD Graduate Fellowship, UCD College of Engineering TOPS Research Award, UCD Regents Fellowship, and the 1991 Anil K. Jain Memorial Prize for the Best Dissertation in the Department of Electrical and Computer Engineering. He is a member of the honor societies Pi Mu Epsilon, Tau Beta Pi, Phi Kappa Phi, and Sigma Xi.

William A. Gardner (F'91) was born in Palo Alto, CA, USA, in 1942. He received the M.S. degree from Stanford University, Stanford, CA, USA, in 1967 and the Ph.D. degree from the University of Massachusetts, Amherst, USA, in 1972, both in electrical engineering.

He was a member of the Technical Staff at Bell Laboratories in Massachusetts from 1967 to 1969. He has been a faculty member at the University of California at Davis, USA, since 1972, where he is Professor of Electrical and Computer Engineering. His research interests are in the general area of statistical signal processing, with primary emphasis on the theories of time-series analysis, stochastic processes, and signal detection and estimation and applications to communications and signals intelligence. He is the author of *Introduction to Random Processes* (McGraw-Hill, 1990) and *Statistical Spectral Analysis* (Prentice-Hall, 1987). He is the Editor of *Cyclostationarity in Communications and Signal Processing* (IEEE, 1994). He holds five patents and is the author of over 100 journal research papers.

Dr. Gardner received the Best Paper of the Year Award from the European Association for Signal Processing in 1986, the 1987 Distinguished Engineering Alumnus Award from the University of Massachusetts, and the Stephen O. Rice Prize Paper Award in the Field of Communication Theory from the IEEE in 1988. He organized and chaired the NSF/ONR/ARO/AFOSR-sponsored workshop on Cyclostationary Signals in 1992. He was elected Fellow of the IEEE in 1991 "for contributions to the development of time-series analysis and stochastic processes with applications to statistical signal processing and communication and for contributions to engineering education."