



Gradient Adaptive Paraunitary Filter Banks for Spatio-Temporal Subspace Analysis and Multichannel Blind Deconvolution*

SCOTT C. DOUGLAS

Department of Electrical Engineering, Southern Methodist University, Dallas, Texas 75275, USA

SHUN-ICHI AMARI

Laboratory for Mathematical Neuroscience, RIKEN Brain Science Institute, Wako-shi, Saitama 351-0198, Japan

S.-Y. KUNG

Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, USA

Abstract. Paraunitary filter banks are important for several signal processing tasks, including coding, multichannel deconvolution and equalization, adaptive beamforming, and subspace processing. In this paper, we consider the task of adapting the impulse response of a multichannel paraunitary filter bank via gradient ascent or descent on a chosen cost function. Our methods are spatio-temporal generalizations of gradient techniques on the Grassmann and Stiefel manifolds, and we prove that they inherently maintain the paraunitariness of the multichannel adaptive system over time. We then discuss the necessary practical approximations, modifications, and simplifications of the methods for solving two relevant signal processing tasks: (i) spatio-temporal subspace analysis and (ii) multichannel blind deconvolution. Simulations indicate that our methods can provide simple, useful solutions to these important problems.

Keywords: adaptive algorithms, gradient descent, Grassmann manifold, multichannel blind deconvolution, paraunitary filter banks, Stiefel manifold, subspace analysis

1. Introduction

Consider the following problem: Given a cost function $\mathcal{J}(\{\mathbf{W}_p\}_{-\infty}^{\infty})$ for the sequence of $(m \times n)$ matrices $\mathbf{W}_p = [\mathbf{w}_{1p} \dots \mathbf{w}_{mp}]^T$ with $\mathbf{w}_{ip} = [w_{i1p} \dots w_{inp}]^T$ and $m \leq n$,

$$\text{maximize } \mathcal{J}(\{\mathbf{W}_p\}_{-\infty}^{\infty}) \quad (1)$$

$$\text{such that } \sum_{p=-\infty}^{\infty} \mathbf{W}_p \mathbf{W}_{p+l}^T = \mathbf{I} \delta_l, \quad (2)$$

where \mathbf{I} is the identity matrix and δ_l is the Kronecker impulse function. Defining the z -transform

of \mathbf{W}_p as

$$\mathbf{W}[z] = \sum_{p=-\infty}^{\infty} \mathbf{W}_p z^{-p}, \quad (3)$$

the constraint in (2) can be expressed in the frequency domain as

$$\mathbf{W}[z] \mathbf{W}^T [z^{-1}]|_{z=e^{j\omega}} = \mathbf{I}. \quad (4)$$

Multichannel linear systems that satisfy (2) or (4) are called *paraunitary filter banks*. They are useful for designing perfect-reconstruction filter banks in coding and image processing tasks [1–7] as well as for determining the eigenstructure of multichannel time

*This work was supported in part by the Office of Research and Development under Contract No. 98F135700-000.

series in direction-of-arrival estimation and wideband array processing tasks [8–12]. When $m = n = 1$, (4) guarantees that the single-input, single-output linear system $W[z]$ is an *all-pass filter*, and thus solutions to (1)–(4) are useful for single-channel deconvolution, equalization, and adaptive control tasks [13–20].

To our knowledge, simple on-line adaptive methods for solving FIR approximations to (1)–(4) have not been extensively explored in the signal processing literature. Works in FIR perfect reconstruction filter bank design have largely focused on constructive methods using classic mean-square or least-squares approximation theory with constraints [3–6]. Realizations of paraunitary systems have largely been parametric or intrinsic parametrizations that employ lattice parameterizations in a linear or tree-based structure [7, 10–12]. Direct-form adaptive all-pass filters have also been developed [14, 15], but these are based on rational transfer functions rather than direct impulse response adjustment and thus require stability-monitoring in practice. Methodologies for adapting FIR approximations to paraunitary filter banks would likely increase the popularity of such systems for solving practical problems while providing useful solutions to new problems.

In this paper, we provide gradient-based adaptive algorithms for solving (1)–(4) iteratively over time. Although other adaptive methods could be used, gradient methods represent the simplest to implement and thus form the baseline to which others are often compared. Our algorithms are spatio-temporal extensions of gradient techniques on the Grassmann and Stiefel manifolds [21–30] and can be applied to any appropriately-defined cost function. We prove that our algorithms in differential form preserve (2) or (4) over time, and thus they adjust \mathbf{W}_p in the impulse response space of paraunitary systems. As these idealized systems are non-causal and doubly-infinite in length, we explore the necessary modifications of these methods that produce practical and implementable solutions to both the spatio-temporal subspace analysis and the multichannel blind deconvolution tasks. The resulting algorithms are computationally-simple, can be easily implemented in digital signal processing hardware, and are shown via simulations to be effective solutions for their respective tasks.

The organization of this paper is as follows. In the next section, gradient algorithms for adaptation of orthogonal matrices are briefly reviewed, extensions of these methods to adaptive paraunitary systems

are provided, and general implementation issues are discussed. In Section 3, we show how the proposed methods can be applied to the task of spatio-temporal subspace analysis task, and Section 4 describes applications of these methods to multichannel blind deconvolution. Both of these sections include simulations that verify the algorithms' usefulness in these tasks. Section 5 contains our conclusions.

2. Gradient Adaptive Paraunitary Systems

2.1. Grassmann and Stiefel Manifolds

To develop gradient-based methods for solving (1)–(4), we first consider the situation where $\mathbf{W}_p = \mathbf{W}\delta_p$ defines a memoryless system. In this case, (1)–(4) reduces to

$$\text{maximize } \mathcal{J}(\mathbf{W}) \quad (5)$$

$$\text{such that } \mathbf{W}\mathbf{W}^T = \mathbf{I}. \quad (6)$$

Note that (6) restricts \mathbf{W} to the space of $(m \times n)$ unitary matrices. If in addition the cost function obeys $\mathcal{J}(\mathbf{W}) = \mathcal{J}(\mathbf{Q}\mathbf{W})$ for any $(m \times m)$ unitary matrix \mathbf{Q} , then the solutions to (5) and (6) for the subspace spanned by the rows of \mathbf{W} lie within the *Grassmann manifold* of orthonormal subspaces [21]. If $\mathcal{J}(\mathbf{W}) \neq \mathcal{J}(\mathbf{Q}\mathbf{W})$, then the solutions to (6) for \mathbf{W} lie within the *Stiefel manifold* of orthonormal matrices [22]. For a general discussion of the geometric structures of the Grassmann and Stiefel manifolds, see [30].

Adaptive solutions to (5) and (6) are useful in many problems. For example, if we define

$$\mathcal{J}(\mathbf{W}) = \pm \text{tr}[\mathbf{W}\mathbf{R}_{\mathbf{x}\mathbf{x}}\mathbf{W}^T] \quad (7)$$

where $\mathbf{R}_{\mathbf{x}\mathbf{x}}$ is a symmetric positive-definite matrix, then (5) and (6) forms the framework for principal subspace analysis (PSA) or minor subspace analysis (MSA) [23–32]. Alternatively, if $\mathcal{J}(\mathbf{W})$ is a contrast function, then (5) and (6) forms the framework for contrast-based blind source separation of prewhitened instantaneous mixtures [33–36].

2.2. Gradient Adaptation on the Grassmann and Stiefel Manifolds

We now discuss the forms of several methods for solving (5) and (6) via gradient techniques. Perhaps

the most-obvious gradient-based method for solving (5) and (6) employs unconstrained adaptation using

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mathbf{G}(k), \quad (8)$$

where

$$\mathbf{G}(k) = \mu(k) \frac{\partial \mathcal{J}(\mathbf{W}(k))}{\partial \mathbf{W}} \quad (9)$$

is the scaled (Euclidean) gradient of the cost function with respect to \mathbf{W} evaluated at $\mathbf{W}(k)$ and $\mu(k)$ is a chosen step size sequence. Then, $\mathbf{W}(k)$ is periodically projected back to the constraint space defined by (6) using a Gram-Schmidt, singular-value-decomposition (SVD), or equivalent orthonormalization procedure. This crude solution is likely to be computationally-difficult to extend to the case of paraunitary systems, however, due to complications in the projection process. In addition, the complexity of the projection step often dominates the overall complexity of the procedure. For example, in principal and minor subspace analysis the Gram-Schmidt orthonormalization procedure is of $O(m^2n)$ complexity, whereas the calculation of $\mathbf{G}(k)$ is of $O(mn)$ complexity.

An alternative procedure to the above method is to parameterize the form of $\mathbf{W}(k)$ using products of $mn - m(m+1)/2$ Givens rotation matrices and adjust each rotational angle using unconstrained gradient adaptation [11, 34]. Since this procedure maintains an *intrinsic parameterization* of $\mathbf{W}(k)$, the constraint in (6) is automatically satisfied, and no modifications of standard gradient procedures are required. The gradient calculations inevitably involve trigonometric calculations, however, hampering their implementation on simple DSP platforms that are optimized for arithmetic (multiply and add) operations.

In this paper, we shall focus on spatio-temporal extensions of methods that use the *extrinsic parameterization* given by the rows of $\mathbf{W}(k)$ and that attempt to adapt $\mathbf{W}(k)$ to maintain the constraint in (6) iteratively over time. Such methods inherently make use of the multiply/accumulate operation in practical applications and thus are ideally suited for DSP platform implementations. Algorithms for adjusting $\mathbf{W}(k)$ within the Grassmann and Stiefel manifolds have received much attention recently due to the elegant adaptation properties that such methods possess [24, 28, 29]. The simplest gradient-based algorithms attempt to move

$\mathbf{W}(k)$ via differential changes along a geodesic in the associated parameter space to maximize $\mathcal{J}(\mathbf{W})$. These algorithms modify the Euclidean gradient using the curvature of the parameter manifold as contained in the associated Riemannian metric for the manifold, and as such they are known as *natural gradient* methods. See [37, 38] for further details on natural gradient adaptation.

We now describe the differential forms of these methods. The natural gradient algorithm for adapting the subspace defined by the linear span of \mathbf{W} within the Grassmann manifold is

$$\frac{d\mathbf{W}}{dt} = \mathbf{W}\mathbf{W}^T\mathbf{G} - \mathbf{G}\mathbf{W}^T\mathbf{W}. \quad (10)$$

The natural gradient algorithm for adapting the rows of \mathbf{W} within the Stiefel manifold is

$$\frac{d\mathbf{W}}{dt} = \mathbf{W}\mathbf{W}^T\mathbf{G} - \mathbf{W}\mathbf{G}^T\mathbf{W}. \quad (11)$$

It can be easily shown for (10) and (11) that, if $\mathbf{W}(t_0)\mathbf{W}^T(t_0) = \mathbf{I}$ at some time instant t_0 , then [28]

$$\frac{d\mathbf{W}\mathbf{W}^T}{dt} = \mathbf{0}, \quad \text{for all } t \geq t_0. \quad (12)$$

Moreover, since no other processing is required, (10) and (11) are ideal for extension to the case of adaptive paraunitary systems.

2.3. Extensions to Spatio-Temporal Systems

To develop procedures similar to (10) and (11) to solve (1)–(4), we rely on recent works relating instantaneous blind source separation and blind deconvolution [39–42]. These works have indicated that spatial-only adaptive algorithms can be extended to adaptive single- and multichannel linear systems if the following three rules are followed:

1. Multiplication of two matrices in the spatial-only case is equivalent to convolution of their associated sequences in the spatio-temporal case.
2. Addition of two matrices in the spatial-only case is equivalent to element-by-element addition of their associated sequences in the spatio-temporal case.
3. Transposition of a matrix in the spatial-only case is equivalent to element-by-element transposition

and time-reversal of its associated sequence in the spatio-temporal case.

Using these rules, we can extend (10) and (11) to solve (1)–(4). These algorithms are given in continuous-time update form for analysis; discrete-time versions are developed for specific applications in later sections.

Extension of (10) [Grassmann]:

$$\begin{aligned} \frac{d\mathbf{W}_p}{dt} &= \mathbf{W}_p * \mathbf{W}_{-p}^T * \mathbf{G}_p - \mathbf{G}_p * \mathbf{W}_{-p}^T * \mathbf{W}_p \\ &= \sum_{q,r} \mathbf{W}_{p-r} \mathbf{W}_{q-r}^T \mathbf{G}_p - \mathbf{G}_{p-r} \mathbf{W}_{q-r}^T \mathbf{W}_q. \end{aligned} \quad (13)$$

Extension of (11) [Stiefel]:

$$\begin{aligned} \frac{d\mathbf{W}_p}{dt} &= \mathbf{W}_p * \mathbf{W}_{-p}^T * \mathbf{G}_p - \mathbf{W}_p * \mathbf{G}_{-p}^T * \mathbf{W}_p \\ &= \sum_{q,r} \mathbf{W}_{p-r} \mathbf{W}_{q-r}^T \mathbf{G}_p - \mathbf{W}_{p-r} \mathbf{G}_{q-r}^T \mathbf{W}_q. \end{aligned} \quad (14)$$

The following theorem indicates the desirable properties possessed by (13) and (14) for adaptive paraunitary systems.

Theorem 1. *The updates in (13) and (14) maintain the constraint in (2) for all $t \geq t_0$ if $\mathbf{W}_p(t_0)$ is paraunitary.*

Proof: For brevity, we only show this property for (13); the proof for (14) is similar. Taking derivatives of both sides of (2) gives

$$\frac{\sum_p \mathbf{W}_p \mathbf{W}_{p+l}^T}{dt} = \sum_p \frac{d\mathbf{W}_p}{dt} \mathbf{W}_{p+l}^T + \mathbf{W}_p \frac{d\mathbf{W}_{p+l}^T}{dt}. \quad (15)$$

Substituting the relation in (13) into the RHS of (15) gives

$$\begin{aligned} &\frac{\sum_p \mathbf{W}_p \mathbf{W}_{p+l}^T}{dt} \\ &= \sum_{p,q,r} \mathbf{W}_{p-r} \mathbf{W}_{q-r}^T \mathbf{G}_q \mathbf{W}_{p+l}^T - \mathbf{G}_{p-r} \mathbf{W}_{q-r}^T \mathbf{W}_q \mathbf{W}_{p+l}^T \\ &\quad + \mathbf{W}_p \mathbf{G}_q^T \mathbf{W}_{q-r} \mathbf{W}_{p+l-r}^T - \mathbf{W}_p \mathbf{W}_q^T \mathbf{W}_{q-r} \mathbf{G}_{p+l-r}^T. \end{aligned} \quad (16)$$

Let $s = p - r$ and $t = q - r$, respectively, so that (16) is

$$\begin{aligned} &\frac{\sum_p \mathbf{W}_p \mathbf{W}_{p+l}^T}{dt} \\ &= \sum_{p,q} \left[\sum_r \mathbf{W}_{p-r} \mathbf{W}_{q-r}^T \right] \mathbf{G}_q \mathbf{W}_{p+l}^T \\ &\quad - \sum_{s,t} \mathbf{G}_s \mathbf{W}_t^T \left[\sum_r \mathbf{W}_{t+r} \mathbf{W}_{s+r+l}^T \right] \\ &\quad + \sum_{p,q} \mathbf{W}_p \mathbf{G}_q^T \left[\sum_r \mathbf{W}_{q-r} \mathbf{W}_{p-r+l}^T \right] \\ &\quad - \sum_{s,t} \left[\sum_r \mathbf{W}_{s+r} \mathbf{W}_{t+r}^T \right] \mathbf{W}_t \mathbf{G}_{s+l}^T. \end{aligned} \quad (17)$$

Assume that (2) holds at some time instant t_0 . Then, the four summations within brackets on the RHS of (17) are $\delta_{p-q} \mathbf{I}$, $\delta_{s-t+l} \mathbf{I}$, $\delta_{p-q+l} \mathbf{I}$, and $\delta_{s-t} \mathbf{I}$, respectively. Substituting these relations and simplifying gives

$$\begin{aligned} \frac{\sum_p \mathbf{W}_p \mathbf{W}_{p+l}^T}{dt} &= \sum_p \mathbf{G}_p \mathbf{W}_{p+l}^T - \sum_s \mathbf{G}_s \mathbf{W}_{s+l}^T \\ &\quad + \sum_p \mathbf{W}_p \mathbf{G}_{p+l}^T - \sum_s \mathbf{W}_s \mathbf{G}_{s+l}^T \\ &= 0. \end{aligned} \quad (18)$$

Thus, the impulse response sequence \mathbf{W}_p never deviates from paraunitarity, as desired. \square

Remark 1. The choice of whether to use (13) or (14) to adjust the impulse response of the paraunitary system depends on the type of cost function being optimized. In analogy with the results for the spatial-only case, we provide the following algorithm selection guidelines:

- For cost functions in which

$$\mathcal{J}(\{\mathbf{W}_p\}_{-\infty}^{\infty}) = \mathcal{J}(\{\mathbf{Q}_p * \mathbf{W}_p\}_{-\infty}^{\infty}) \quad (19)$$

for any arbitrary $(m \times m)$ paraunitary impulse response $\{\mathbf{Q}_p\}$, $-\infty < p < \infty$, the algorithm in (13) is the appropriate choice.

- For cost functions in which

$$\mathcal{J}(\{\mathbf{W}_p\}_{-\infty}^{\infty}) \neq \mathcal{J}(\{\mathbf{Q}_p * \mathbf{W}_p\}_{-\infty}^{\infty}) \quad (20)$$

for any arbitrary $(m \times m)$ paraunitary impulse response $\{\mathbf{Q}_p\}$, $-\infty < p < \infty$, the algorithm in (14) is the appropriate choice.

The reasons for these choices are seen if one considers the gradient of $\mathcal{J}(\{\mathbf{Q}_p * \mathbf{W}_p\}_{-\infty}^{\infty})$ with respect to the sequence \mathbf{W}_p , which assuming (19) is

$$\frac{\partial \mathcal{J}(\{\mathbf{Q}_p * \mathbf{W}_p\}_{-\infty}^{\infty})}{\partial \mathbf{W}_p} = \mathbf{Q}_{-p}^T * \mathbf{G}_p. \quad (21)$$

Defining the sequence $\mathbf{T}_p(t_0) = \mathbf{Q}_p * \mathbf{W}_p(t_0)$ at some time $t = t_0$, it is straightforward to show for the update in (13) that

$$\frac{d\mathbf{T}_p}{dt} = \mathbf{T}_p * \mathbf{T}_{-p}^T * \mathbf{G}_p - \mathbf{G}_p * \mathbf{T}_{-p}^T * \mathbf{T}_p, \quad (22)$$

and thus

$$\mathbf{T}_p = \mathbf{Q}_p * \mathbf{W}_p \quad (23)$$

for all $t \geq t_0$. In other words, the coefficient trajectories provided by (13) are invariant to the orientations of the rows of $\mathbf{W}[z]$ in the space of paraunitary filters, which is the spatio-temporal extension of the subspace invariance property possessed by the algorithm in (10). Thus, (13) is the appropriate algorithm form for solutions to spatio-temporal subspace analysis tasks. As (14) does not possess this subspace invariance property, it is the appropriate algorithm form for solutions to multichannel blind deconvolution tasks.

Remark 2. The procedures in (13) and (14) are given in terms of the doubly-infinite impulse response of the paraunitary system $\mathbf{W}[z]$. If additional constraints on $\mathbf{W}[z]$ are specified, then the procedures would have to be changed. For example, if $\mathbf{W}[z]$ belongs to the class of finite impulse response (FIR) paraunitary filters in which

$$\mathbf{W}_l = 0, \quad l < -\frac{L}{2} \quad \text{or} \quad l > \frac{L}{2}, \quad (24)$$

then the above updates may not exactly maintain the corresponding paraunitary constraints given by

$$\sum_{p=-L/2}^{L/2} \mathbf{W}_p \mathbf{W}_{p+l}^T = \mathbf{I} \delta_l. \quad (25)$$

Algorithms for adapting multichannel FIR filters to exactly maintain (25) at each time instant are outside the scope of this paper. These facts do not preclude, however, the use of appropriate *approximations* of the generic procedures in (13) and (14) using truncated

multichannel FIR filter models, to be discussed shortly. Technically, truncating $\mathbf{W}[z]$ to FIR form within our algorithms yields procedures that cannot exactly maintain the constraints in (2), and thus there is not a one-to-one correspondence between the impulse responses produced by our adaptive procedures and the intrinsic parameters of an FIR paraunitary filter. As we shall show, however, the resulting adaptive filters' deviations from paraunitariness are extremely small—typically less than 2% of the total filter coefficient energy—throughout the adaptive process.

2.4. General Implementation Issues

In practice, discrete-time versions of (13) and (14) are required for real-time digital implementations. Substituting finite differences for differentials in these algorithms, however, can create numerical problems due to error accumulation. For example, when $\mathbf{W}_p = \mathbf{W} \delta_p$, the algorithm

$$\begin{aligned} \mathbf{W}(k+1) &= \mathbf{W}(k) + \mathbf{W}(k) \mathbf{W}^T(k) \mathbf{G}(k) \\ &\quad - \mathbf{G}(k) \mathbf{W}^T(k) \mathbf{W}(k) \end{aligned} \quad (26)$$

$$\mathbf{G}(k) = \mu(k) \mathbf{W}(k) \mathbf{R}_{\mathbf{xx}} \quad (27)$$

is numerically-unstable for PSA and MSA, such that the rows of $\mathbf{W}(k)$ slowly diverge from orthonormality over time [29, 32]. For this reason, modifications of the updates are required. For PSA, an approximation to (26) and (27) given by

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \hat{\mathbf{G}}(k) - \hat{\mathbf{G}}(k) \mathbf{W}^T(k) \mathbf{W}(k) \quad (28)$$

$$\hat{\mathbf{G}}(k) = \mu(k) \mathbf{y}(k) \mathbf{x}^T(k) \quad (29)$$

and $\mu(k) > 0$ yields the principal subspace estimate in a numerically-stable manner [23]. For MSA,

$$\begin{aligned} \mathbf{W}(k+1) &= \mathbf{W}(k) + \mathbf{W}(k) \mathbf{W}^T(k) \mathbf{W}(k) \mathbf{W}^T(k) \hat{\mathbf{G}}(k) \\ &\quad - \hat{\mathbf{G}}(k) \mathbf{W}^T(k) \mathbf{W}(k) \end{aligned} \quad (30)$$

with $\hat{\mathbf{G}}(k)$ defined in (29) and $\mu(k) < 0$ yields asymptotically-stable behavior [29]. Extensions of these self-stabilized methods to the spatio-temporal subspace analysis and multichannel blind deconvolution tasks are provided in the next two sections.

In addition to the above issues, one also must address three others in practice: (i) the infinite memory, (ii) the non-causality, and (iii) the computational complexity,

respectively, of the adaptive system. The first two issues typically can be addressed by truncating the filter model to finite length and by employing delayed updates within the algorithm, respectively. The system's overall computational complexity usually can then be reduced by assuming that the system adapts slowly, so that delayed versions of previously-computed signals are used in place of similar signals appearing within the parameter updates. Similar approximations have been used in other natural gradient methods for single- and multichannel deconvolution tasks [43, 45, 50]. Details of these approximations are given in the specific applications that follow.

3. Spatio-Temporal Subspace Analysis

3.1. Problem Description

Let $\mathbf{x}(k)$ be a sequence of n -dimensional vectors from a wide-sense stationary vector random process with autocorrelation function

$$\mathbf{R}_{\mathbf{xx}}(l) = E\{\mathbf{x}(k)\mathbf{x}^T(k-l)\}. \quad (31)$$

In spatio-temporal principal subspace analysis, an $(m \times n)$, $m < n$ paraunitary system is used to compute a sequence of m -dimensional vectors $\mathbf{y}(k)$ as

$$\mathbf{y}(k) = \sum_{p=-\infty}^{\infty} \mathbf{W}_p \mathbf{x}(k-p). \quad (32)$$

The goal is to calculate a (non-unique) paraunitary system impulse response \mathbf{W}_p such that the cost function

$$\mathcal{J}(\{\mathbf{W}_p\}_{-\infty}^{\infty}) = E\{\|\mathbf{y}(k)\|^2\} \quad (33)$$

is maximized, where $\|\mathbf{y}(k)\|$ denotes the L_2 or Euclidean norm of $\mathbf{y}(k)$. The resulting family of solutions obtain optimal energy compaction in m -dimensional signal space, such that the sequence

$$\mathbf{u}(k) = \sum_{q=-\infty}^{\infty} \mathbf{W}_{-q}^T \mathbf{y}(k-q) \quad (34)$$

is the optimal rank- m linear filtered approximation to the vector sequence $\mathbf{x}(k)$ in a mean-square-error sense. One possible application of such a method is the efficient coding and storage of multisensor signals such as n -microphone audio recordings of m talkers in a

room environment. Similarly, minimizing (33) under paraunitary constraints is the spatio-temporal extension of minor subspace analysis and is potentially-useful for direction-of-arrival in wideband array processing systems [8–12].

3.2. An Algorithm for Spatio-Temporal PSA

We now apply the gradient methods described in Section 2 to the spatio-temporal PSA task. Since (33) is the appropriate averaged cost function for this task, we consider its instantaneous version

$$\hat{\mathcal{J}}(\{\mathbf{W}_p(k)\}_{-\infty}^{\infty}) = \|\mathbf{y}(k)\|^2 \quad (35)$$

for use within a stochastic gradient procedure that inherently averages across time samples in its operation. A straightforward application of (13) to this task yields

$$\begin{aligned} \mathbf{W}_p(k+1) &= \mathbf{W}_p(k) + \mu(k) \sum_{r=-\infty}^{\infty} \mathbf{W}_{p-r}(k) \\ &\quad \times \sum_{q=-\infty}^{\infty} \mathbf{W}_{q-r}^T(k) \mathbf{y}(k) \mathbf{x}^T(k-q) - \mu(k) \mathbf{y}(k) \\ &\quad \times \sum_{q=-\infty}^{\infty} \left[\sum_{r=-\infty}^{\infty} \mathbf{x}^T(k-p+r) \mathbf{W}_{q-r}^T(k) \right] \mathbf{W}_q(k) \end{aligned} \quad (36)$$

This system can be expected to be numerically-unstable, however, as (26) is numerically-unstable. To obtain a potentially-useful algorithm, we instead begin with a version that is analogous in form to (28) as given by

$$\begin{aligned} \mathbf{W}_p(k+1) &= \mathbf{W}_p(k) + \mu(k) \mathbf{y}(k) \left\{ \mathbf{x}^T(k-p) \right. \\ &\quad \left. - \sum_{q=-\infty}^{\infty} \left[\sum_{r=-\infty}^{\infty} \mathbf{x}^T(k-p+r) \mathbf{W}_{q-r}^T(k) \right] \mathbf{W}_q(k) \right\}. \end{aligned} \quad (37)$$

Such a system cannot be implemented, however, due to the doubly-infinite impulse response model. For practical systems, we truncate the paraunitary filter model by

setting $\mathbf{W}_p(k) = \mathbf{0}$ for $p < 0$ and $p > L$, respectively, which yields

$$\mathbf{y}(k) = \sum_{l=0}^L \mathbf{W}_l(k) \mathbf{x}(k-l) \quad (38)$$

in place of (32) and

$$\begin{aligned} \mathbf{W}_p(k+1) &= \mathbf{W}_p(k) + \mu(k) \mathbf{y}(k) \left\{ \mathbf{x}^T(k-p) \right. \\ &\quad \left. - \sum_{q=0}^L \left[\sum_{r=q-L}^q \mathbf{x}^T(k-p+r) \mathbf{W}_{q-r}^T(k) \right] \mathbf{W}_q(k) \right\} \end{aligned} \quad (39)$$

in place of (37) for $0 \leq p \leq L$.

Equation (39) is both computationally-complex and non-causal. In [43–45, 50], a procedure is described and used to simplify related algorithms for multichannel blind deconvolution. This procedure assumes that the adaptive system is slowly varying, such that

$$\mathbf{W}_p(k) \approx \mathbf{W}_p(k-1) \approx \dots \approx \mathbf{W}_p(k-2L), \quad (40)$$

such that delayed coefficient values can be used in place of more-recent coefficient values within the filtered gradient calculations. We follow a similar procedure in what follows. Letting $l = q - r$, we can make the approximation

$$\begin{aligned} &\sum_{r=q-L}^q \mathbf{W}_{q-r}(k) \mathbf{x}(k-p+r) \\ &\approx \sum_{l=0}^L \mathbf{W}_l(k-p+q) \mathbf{x}(k-p+q-l) \\ &= \mathbf{y}(k-p+q), \end{aligned} \quad (41)$$

and thus

$$\begin{aligned} &\sum_{q=0}^L \mathbf{W}_q^T(k) \sum_{r=q-L}^q \mathbf{W}_{q-r}(k) \mathbf{x}(k-p+r) \\ &\approx \sum_{q=0}^L \mathbf{W}_{L-q}^T(k) \mathbf{y}(k-p+L-q) \\ &\approx \sum_{q=0}^L \mathbf{W}_{L-q}^T(k-p+L) \mathbf{y}(k-p+L-q). \end{aligned} \quad (42)$$

Finally, substituting the above results into the RHS of (39) and delaying these terms by L time samples, we obtain a causal delayed-update version of (39) as

$$\begin{aligned} \mathbf{W}_p(k+1) &= \mathbf{W}_p(k) + \mu(k) \mathbf{y}(k-L) \\ &\quad \times [\mathbf{x}^T(k-L-p) - \mathbf{u}^T(k-p)], \end{aligned} \quad (43)$$

where

$$\mathbf{u}(k) = \sum_{q=0}^L \mathbf{W}_{L-q}^T(k) \mathbf{f}(\mathbf{y}(k-q)) \quad (44)$$

and $\mathbf{f}(\mathbf{y}) = \mathbf{y}$ is linear in this case.

Equations (38), (43) and (44) describe the proposed spatio-temporal extension of the principal subspace rule in (26). This algorithm requires $(3mn + n)(L + 1) + m$ multiply/accumulates (MACs) per iteration to implement, and its average complexity per adaptive parameter is about the same as that in the scalar case ($L = 0$).

3.3. An Algorithm for Spatio-Temporal MSA

Using similar principles, we can extend the MSA algorithm in (30) to the spatio-temporal case. For brevity, the derivations are omitted, and only the final form of the algorithm is given: for $0 \leq p \leq L$,

$$\begin{aligned} \mathbf{W}_p(k+1) &= \mathbf{W}_p(k) - \mu(k) [\underline{\zeta}(k) \mathbf{x}^T(k-2L-p) \\ &\quad - \mathbf{y}(k-2L) \mathbf{u}^T(k-L-p)], \end{aligned} \quad (45)$$

where $\mathbf{y}(k)$ and $\mathbf{u}(k)$ are computed as in (38) and (44),

$$\underline{\zeta}(k) = \sum_{l=0}^L \mathbf{W}_l(k) \mathbf{z}(k-l) \quad (46)$$

$$\mathbf{z}(k) = \sum_{q=0}^L \mathbf{W}_{L-q}^T(k) \underline{\psi}(k-q) \quad (47)$$

$$\underline{\psi}(k) = \sum_{l=0}^L \mathbf{W}_l(k) \mathbf{u}(k-l). \quad (48)$$

Note that this algorithm requires $(7nm + 2m)(L + 1)$ MACs to implement, such that its average complexity per adaptive parameter is about the same as that of (30).

3.4. Simulations

We now explore the performance of the spatio-temporal PSA algorithm in (38), (43) and (44) via simulations. Consider a two-input, four-output system in which

$\mathbf{s}(k) = [s_1(k) \ s_2(k)]^T$, $s_i(k)$, $i \in \{1, 2\}$ are independent zero-mean Gaussian sequences with autocorrelations $r_{ss,i}(l) = \delta_l$,

$$\mathbf{x}(k) = \hat{\mathbf{x}}(k) + \underline{\eta}(k) \quad (49)$$

$$\hat{\mathbf{x}}(k) = \sum_{i=1}^2 \mathbf{A}_i \hat{\mathbf{x}}(k-i) + \sum_{j=0}^1 \mathbf{B}_j \mathbf{s}(k-j) \quad (50)$$

$$\mathbf{A}_1 = \begin{bmatrix} 0.38 & 0.39 & -0.22 & 0.08 \\ 0.24 & -0.30 & -0.03 & -0.08 \\ -0.36 & -0.20 & -0.44 & 0.02 \\ -0.49 & 0.16 & 0.49 & -0.17 \end{bmatrix} \quad (51)$$

$$\mathbf{A}_2 = \begin{bmatrix} -0.01 & 0.01 & 0.06 & 0.06 \\ -0.05 & 0.03 & 0.04 & -0.09 \\ 0.02 & -0.06 & -0.01 & 0.02 \\ 0.05 & -0.02 & 0.01 & -0.09 \end{bmatrix} \quad (52)$$

$$\mathbf{B}_0^T = \begin{bmatrix} -0.02 & -0.04 & 0.07 & -0.10 \\ 0.05 & 0.09 & 0.10 & 0.06 \end{bmatrix} \quad (53)$$

$$\mathbf{B}_1^T = \begin{bmatrix} -0.1 & 0.0 & -0.6 & 0.3 \\ -0.04 & 0.9 & 0.5 & -0.2 \end{bmatrix}, \quad (54)$$

$\underline{\eta}(k) = [\eta_1(k)\eta_2(k)\eta_3(k)\eta_4(k)]^T$, and $\eta_i(k)$, $i \in \{1, 2, 3, 4\}$ are independent zero-mean Gaussian sequences with $r_{\eta\eta,i}(l) = \sigma_\eta^2 \delta_l$ and $\sigma_\eta^2 = 10^{-4}$. We apply (38), (43) and (44) to $\mathbf{x}(k)$, where $m = 2$, $n = 4$, $L = 14$, $\mu(k) = 0.001$, and $w_{ijp}(0)$ is unity if $i = j$ and $p = L/2$ and is zero otherwise. Shown in Fig. 1(a) and (b) are the evolutions of

$$\rho_{PSA}(k) = \|\mathbf{x}(k-L) - \mathbf{u}(k)\|^2 \quad (55)$$

and

$$\gamma(k) = \sum_{l=-L}^L \left\| \mathbf{I}\delta_l - \sum_{p=0}^L \mathbf{W}_p(k) \mathbf{W}_{p+l}^T(k) \right\|_F^2, \quad (56)$$

respectively, as averaged over twenty different simulation runs. As can be seen, the proposed algorithm effectively adapts to the signal subspace while approximately maintaining the system's paraunitariness through its adaptive characteristics. The steady-state value of $\rho_{PSA}(k)$ is approximately 2.10×10^{-4} , close to the minimum value of $\min E\{\rho_{PSA}(k)\} = 2\sigma_\eta^2 = 2 \times 10^{-4}$ theoretically-obtainable from the data model.

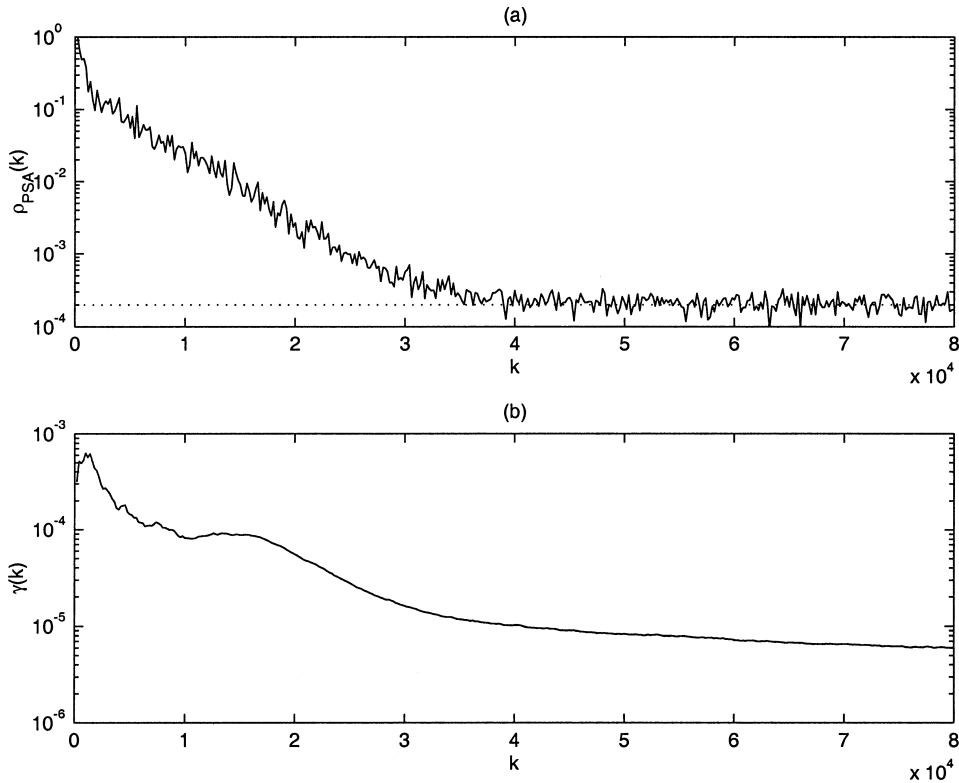


Figure 1. Evolutions of (a) $E\{\rho_{PSA}(k)\}$ and (b) $E\{\gamma(k)\}$ for the spatio-temporal PSA algorithm in the first simulation example.

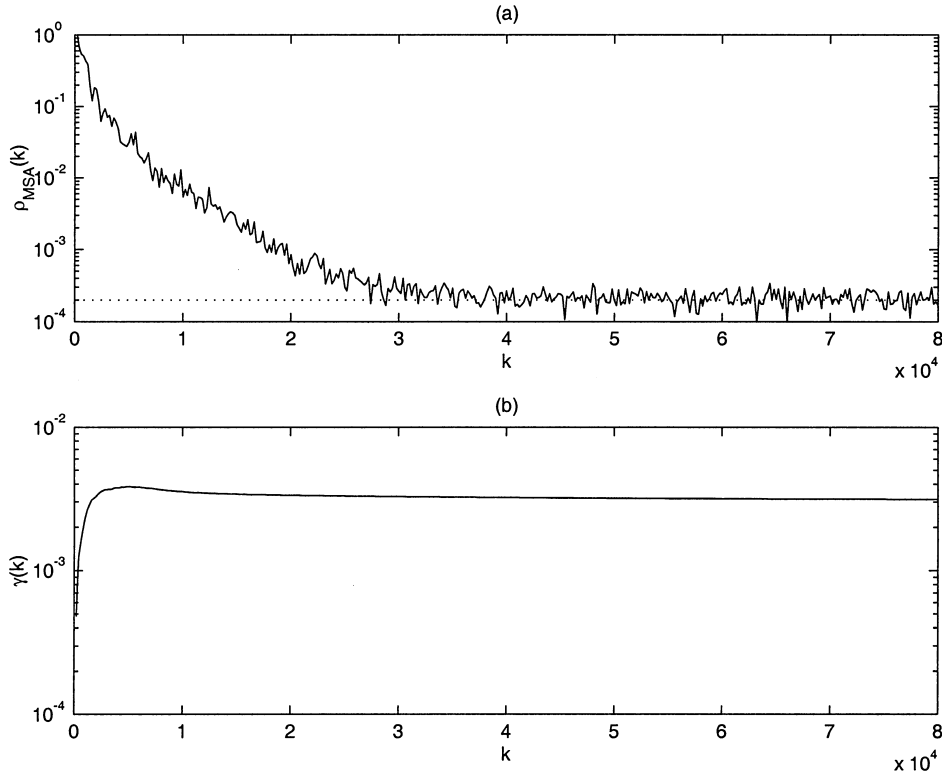


Figure 2. Evolutions of (a) $E\{\rho_{MSA}(k)\}$ and (b) $E\{\gamma(k)\}$ for the spatio-temporal MSA algorithm in the second simulation example.

To verify the performance of (38), (44)–(48) in the spatio-temporal MSA task, Fig. 2(a) and (b) show the evolutions of the averaged values of the performance factors

$$\rho_{MSA}(k) = \|\mathbf{y}(k)\|^2 \quad (57)$$

and $\gamma(k)$ as computed by this algorithm on data generated according to (49)–(54) over twenty different simulation runs, where $\mu(k) = 0.001$. In this case, $\rho_{MSA}(k)$ converges to a steady-state value of approximately 2.14×10^{-4} , close to the theoretical minimum value of 2×10^{-4} for this data. Moreover, the algorithm approximately maintains the paraunitariness of $\mathbf{W}_p(k)$ over time.

4. Multichannel Blind Deconvolution

4.1. Problem Description

Let $\mathbf{x}(k)$ be a sequence of n -dimensional vectors generated from the linear model

$$\mathbf{x}(k) = \sum_{j=-\infty}^{\infty} \mathbf{A}_j \mathbf{s}(k-j), \quad (58)$$

where each vector $\mathbf{s}(k) = [s_1(k) \dots s_m(k)]^T$ contains m samples from m different i.i.d. source signals and the sequence of $(n \times m)$ matrices \mathbf{A}_j , $-\infty < j < \infty$ constitute the impulse response of an unknown linear channel. The goal is to process these signals with a second linear system such that estimates of each of the source signals $s_i(k)$, $1 \leq i \leq m$ are obtained.

Suppose an $(m \times n)$ multichannel system with impulse response \mathbf{P}_j , $-\infty < j < \infty$ is designed such that the m -dimensional vector sequence

$$\mathbf{v}(k) = \sum_{j=-\infty}^{\infty} \mathbf{P}_j \mathbf{x}(k-j) \quad (59)$$

is uncorrelated in space and time; i.e.

$$\mathbf{R}_{\mathbf{v}\mathbf{v}}(l) = E\{\mathbf{v}(k)\mathbf{v}^T(k-l)\} = \mathbf{I}\delta_l. \quad (60)$$

An adaptive algorithm for performing this spatio-temporal decorrelation task can be found in [54]. Then, it can be shown that the paraunitary system given by

$$\mathbf{y}(k) = \sum_{p=-\infty}^{\infty} \mathbf{W}_p \mathbf{v}(k-p) \quad (61)$$

is sufficient for the multichannel blind deconvolution task, such that

$$y_i(k) = c_{j_i} s_{j_i}(k - \Delta_{j_i}), \quad (62)$$

for some non-unique, non-replicative assignment $i \rightarrow j_i$, $\{i, j_i\} \in \{1, 2, \dots, m\}$ can be obtained, where Δ_{j_i} denotes the integer delay of the j_i th source at the i th output. Moreover, a simple contrast function that can be used within the framework in (1)–(4) to solve this task is [46–48]

$$\mathcal{J}(\{\mathbf{W}_p\}_{-\infty}^{\infty}) = \sum_{i=1}^M |\kappa[y_i(k)]| \quad (63)$$

$$\kappa[y_i(k)] = E\{|y_i(k)|^4\} - 3, \quad (64)$$

where $\kappa[y]$ is the (normalized) kurtosis of the unit-variance random variable y . Solutions to the above task are useful for separating digital communications signals in an unknown multiuser environment [49, 50] as well as separating acoustic signals from multisensor recordings in reverberant room environments [51–53].

We now develop algorithms for multichannel blind deconvolution for prewhitened signal mixtures. These algorithms are spatio-temporal generalizations of methods for blind source separation of instantaneous prewhitened signal mixtures, in which $\mathbf{A}_l = \mathbf{A}\delta_l$, $\mathbf{P}_l = \mathbf{P}\delta_l$, and $\mathbf{W}_l = \mathbf{W}\delta_l$ in (58), (61) and (59), respectively. It can be shown that two simple self-stabilized algorithms for adjusting an $(m \times n)$ demixing matrix $\mathbf{W}(k)$ to separate the source signals in the instantaneous blind source separation task are [36]

$$\begin{aligned} \mathbf{W}(k+1) &= \mathbf{W}(k) + \mu(k)[\mathbf{f}(\mathbf{y}(k))\mathbf{v}^T(k) \\ &\quad - \mathbf{v}(k)\mathbf{f}^T(\mathbf{y}(k))\mathbf{W}(k)] \end{aligned} \quad (65)$$

and

$$\begin{aligned} \mathbf{W}(k+1) &= \mathbf{W}(k) - \mu(k)[\mathbf{W}(k)\mathbf{W}^T(k)\mathbf{W}(k)\mathbf{W}^T(k)\mathbf{f}(\mathbf{y}(k))\mathbf{v}^T(k) \\ &\quad - \mathbf{v}(k)\mathbf{f}^T(\mathbf{y}(k))\mathbf{W}(k)], \end{aligned} \quad (66)$$

respectively, where $\mathbf{f}(\mathbf{y}) = [f_1(y_1) \dots f_m(y_m)]^T$, $f_i(y) = \partial g_i(y)/\partial y$, and $g_i(y)$ is one term of the instantaneous generalized contrast function

$$\hat{\mathcal{J}}(\mathbf{W}) = \sum_{i=1}^m g_i(y_i(k)). \quad (67)$$

These algorithms are similar in form to (28) and (30), respectively. When $\mathbf{W}(k)$ is square ($m = n$), we can also consider the multiplicative equivariant forms of (65) and (66), given by

$$\begin{aligned} \mathbf{W}(k+1) &= \mathbf{W}(k) + \mu(k)[\mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k) \\ &\quad - \mathbf{y}(k)\mathbf{f}^T(\mathbf{y}(k))\mathbf{W}(k)\mathbf{W}^T(k)]\mathbf{W}(k), \end{aligned} \quad (68)$$

and

$$\begin{aligned} \mathbf{W}(k+1) &= \mathbf{W}(k) - \mu(k)[\mathbf{W}(k)\mathbf{W}^T(k)\mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k) \\ &\quad - \mathbf{y}(k)\mathbf{f}^T(\mathbf{y}(k))]\mathbf{W}(k), \end{aligned} \quad (69)$$

respectively. These algorithms can be shown to have similar behaviors to their non-equivariant counterparts [36]. The main reason for their inclusion is computational complexity; in particular, (69) is simpler to implement than (66). For this reason, we shall only consider extensions of (65) and (69) in what follows.

4.2. Algorithms for Multichannel Blind Deconvolution

To extend (65) and (69) to the multichannel blind deconvolution task, we consider a generalized spatio-temporal version of the joint contrast function in (67), given by

$$\hat{\mathcal{J}}(\{\mathbf{W}_p(k)\}_{-\infty}^{\infty}) = \sum_{i=1}^m g_i(y_i(k)). \quad (70)$$

Noting that

$$\mathbf{G}_p(k) = \mu(k)\mathbf{f}(\mathbf{y}(k))\mathbf{v}^T(k-p), \quad (71)$$

the spatio-temporal equivalent of (65) is found from (14) to be

$$\begin{aligned} \mathbf{W}_p(k+1) &= \mathbf{W}_p(k) + \mu(k) \left[\mathbf{f}(\mathbf{y}(k))\mathbf{v}^T(k-p) - \sum_{r=-\infty}^{\infty} \mathbf{W}_{p-r}(k) \right. \\ &\quad \left. \times \sum_{q=-\infty}^{\infty} \mathbf{v}(k-q+r)\mathbf{f}^T(\mathbf{y}(k))\mathbf{W}_q(k) \right]. \end{aligned} \quad (72)$$

As (72) cannot be implemented as written, we employ similar approximations and simplifications that were used to derive (43) from (37). The details of these

manipulations are omitted for brevity, and only the final algorithm form is given: for $0 \leq p \leq L$,

$$\mathbf{W}_p(k+1) = \mathbf{W}_p(k) + \mu(k)[\mathbf{f}(\mathbf{y}(k-L))\mathbf{v}^T(k-L-p) - \mathbf{y}(k-L)\mathbf{u}^T(k-p)] \quad (73)$$

$$\mathbf{y}(k) = \sum_{l=0}^L \mathbf{W}_l(k)\mathbf{v}(k-l), \quad (74)$$

and $\mathbf{u}(k)$ is defined in (44). This algorithm requires $(4m^2 + 2m)(L+1)$ MACs at each iteration, not including the evaluation of the elements of $\mathbf{f}(\mathbf{y}(k))$ at each iteration.

Similarly, we can develop an approximate spatio-temporal extension of the algorithm in (69). This update for $0 \leq p \leq L$ is

$$\mathbf{W}_p(k+1) = \mathbf{W}_p(k) - \mu(k)[\underline{\psi}(k)\underline{\xi}^T(k-p) - \mathbf{y}(k-L)\mathbf{u}^T(k-p)], \quad (75)$$

where $\mathbf{u}(k)$ and $\underline{\psi}(k)$ are as defined in (44) and (48), respectively, and

$$\underline{\xi}(k) = \sum_{q=0}^L \mathbf{W}_{L-q}^T(k)\mathbf{y}(k-q). \quad (76)$$

Taken together, the algorithm in (44), (48), (74), (75) and (76) requires $(6m^2 + 2m)(L+1)$ MACs at each iteration, not including the evaluation of the elements of $\mathbf{f}(\mathbf{y}(k))$ at each iteration.

4.3. Simulations

We now verify the usefulness of the multichannel blind deconvolution algorithms via simulations. In all cases, we have generated the received signal vector sequence $\mathbf{x}(k)$ as

$$\mathbf{x}(k) = \mathbf{Q}\mathbf{t}(k), \quad (77)$$

where \mathbf{Q} is the eigenvector matrix of the symmetric matrix

$$\mathbf{R} = \begin{bmatrix} 5 & 4 & 5 \\ 4 & 1 & 7 \\ 5 & 7 & 9 \end{bmatrix}, \quad (78)$$

the elements of $\mathbf{t}(k) = [t_1(k) \dots t_3(k)]^T$ are given by

$$t_1(k) = s_1(k-2) + 0.5\{t_1(k-1) - s_1(k-1)\} - 0.1\{t_1(k-2) - s_1(k)\} \quad (79)$$

$$t_2(k) = s_2(k-2) + 0.3\{t_2(k-1) - s_2(k-1)\} - 0.2\{t_2(k-2) - s_2(k)\} \quad (80)$$

$$t_3(k) = s_3(k-2) + 0.4\{t_3(k-1) - s_3(k-1)\} + 0.1\{t_3(k-2) - s_3(k)\}, \quad (81)$$

and each $s_i(k)$, $i \in \{1, 2, 3\}$ is an i.i.d. signal with a specified distribution. It can be easily shown that the relationship between $\mathbf{x}(k)$ and $\mathbf{s}(k)$ is paraunitary, such that we may assign $\mathbf{x}(k) = \mathbf{v}(k)$ without spatio-temporal prewhitening. Both of the algorithms in (44), (73), (74) and (44), (48), (74)–(76) were then applied to this data, where $m = n = 3$, $L = 14$, and $w_{ijp}(0)$ is unity if $i = j$ and $p = L/2$ and is zero otherwise. For each algorithm, the nonlinearities $f_i(y)$ have been chosen in each case to guarantee the local stability of the algorithms about a separating solution [36]. Averaged values of the performance factors

$$\rho_{BSS}(k) = \sum_{i=1}^m |s_i(k) - c_{j_i}y_i(k - \Delta_{j_i})|^2 \quad (82)$$

and $\gamma(k)$ were then computed using twenty different simulation runs, where fixed values of c_{j_i} , Δ_{j_i} , and the assignments $\{i \rightarrow j_i\}$ were chosen to minimize the value of $\rho_{BSS}(k)$ in steady-state.

Figure 3(a) shows the averaged values of $\rho_{BSS}(k)$ for the algorithm in (44), (73) and (74) with $f_i(y) = y/|y|$ and $\mu(k) = 0.0007$ as well as the algorithm in (44), (48), (74)–(76) with $f_i(y) = y^3$ and $\mu(k) = 0.0005$ when $s_1(k)$, $s_2(k)$, and $s_3(k)$ are zero-mean, unit variance i.i.d. signals with binary, binary, and uniform probability density functions (p.d.f.'s), respectively. In each case, we have chosen step sizes that yield similar steady-state values of $E\{\rho_{BSS}(k)\}$ for each algorithm on this data. As can be seen, both algorithms deconvolve and separate the source signals from the prewhitened mixtures, and the low values of $E\{\gamma(k)\}$ for both algorithms in Fig. 3(b) indicate that they adaptively maintain the paraunitariness of $\mathbf{W}_p(k)$ over time.

Shown in Fig. 4(a) and (b) are the corresponding averaged values of $\rho_{BSS}(k)$ and $\gamma(k)$ for the algorithm in (44), (73), (74) with $f_i(y) = y^3$ and $\mu(k) = 0.00005$ as well as the algorithm in (44), (48), (74)–(76) with $f_i(y) = y/|y|$ and $\mu(k) = 0.0009$ when each $s_i(k)$ is a zero-mean, unit variance i.i.d. signal with Laplacian p.d.f. $p_s(s) = 2^{-1/2} \exp(-\sqrt{2}|s|)$. Again, both algorithms perform multichannel deconvolution with these nonlinearity choices on this data, and they adaptively maintain the paraunitariness of $\mathbf{W}_p(k)$ over time.

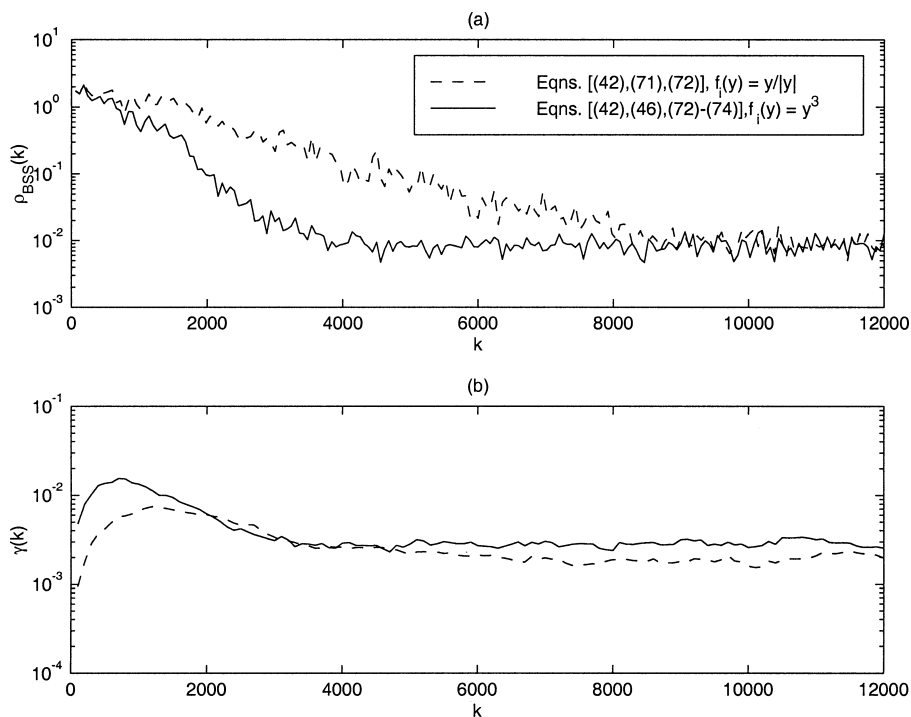


Figure 3. Evolutions of (a) $E\{\rho_{BSS}(k)\}$ and (b) $E\{\gamma(k)\}$ for the two multichannel blind deconvolution algorithms in the third simulation example.

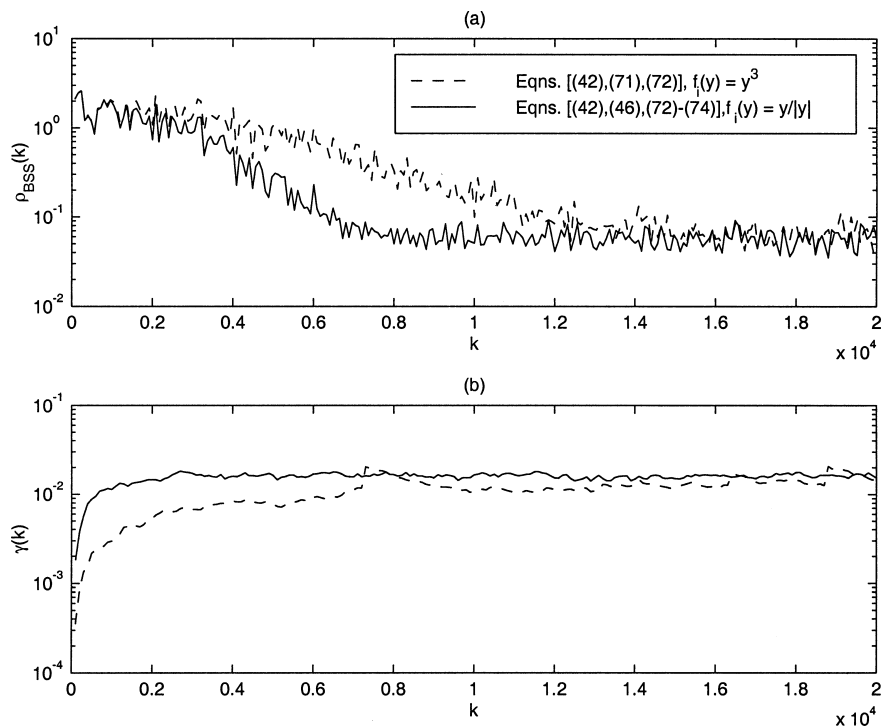


Figure 4. Evolutions of (a) $E\{\rho_{BSS}(k)\}$ and (b) $E\{\gamma(k)\}$ for the two multichannel blind deconvolution algorithms in the fourth simulation example.

5. Conclusions

In this paper, we have proposed gradient adaptive methods for paraunitary filter banks that are extensions of gradient methods on the Grassmann and Stiefel manifolds of orthogonal matrices. When applied to spatio-temporal subspace analysis and blind deconvolution, the proposed algorithms are computationally simple and effective at their respective tasks. The techniques are expected to be useful for signal-adaptive filter bank design, coding, and equalization tasks.

References

- P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Englewood Cliffs, NJ: Prentice-Hall, 1993.
- D.T.M. Slock, "Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks, and Multichannel Linear Prediction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Adelaide, Australia, vol. 4, 1994, pp. 585–588.
- M.K. Tsatsanis and G.B. Giannakis, "Principal Component Filter Banks for Optimum Multiresolution Analysis," *IEEE Trans. Signal Processing*, vol. 43, 1995, pp. 1766–1777.
- A. Kiraç and P.P. Vaidyanathan, "Theory and Design of Optimum FIR Compaction Filters," *IEEE Trans. Signal Processing*, vol. 46, 1998, pp. 903–919.
- P. Moulin and M.K. Mthçak, "Theory and Design of Signal-Adapted FIR Paraunitary Filter Banks," *IEEE Trans. Signal Processing*, vol. 46, 1998, pp. 920–929.
- B. Xuan and R.H. Bamberger, "FIR Principal Component Filter Banks," *IEEE Trans. Signal Processing*, vol. 46, 1998, pp. 930–940.
- O.S. Jahromi, M.A. Masnadi-Shirazi, and M. Fu, "A Fast $O(N)$ Algorithm for Adaptive Filter Bank Design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Seattle, WA, vol. 3, 1998, pp. 1325–1328.
- B. Porat and B. Friedlander, "Estimation of Spatial and Spectral Parameters of Multiple Sources," *IEEE Trans. Inform. Theory*, vol. 29, 1983, pp. 412–425.
- B. Ottersten and T. Kailath, "Direction-of-Arrival Estimation for Wideband Sources Using the ESPRIT Algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, 1990, pp. 317–327.
- P. Loubaton and P.A. Regalia, "Blind Deconvolution of Multivariate Signals by Using Adaptive FIR Lossless Filters," in *Proc. EUSIPCO92*, Brussels, Belgium, 1992, pp. 317–327.
- P.A. Regalia and P. Loubaton, "Rational Subspace Estimation Using Adaptive Lossless Filters," *IEEE Trans. Signal Processing*, vol. 40, 1992, pp. 2392–2405.
- P.A. Regalia and D.-Y. Huang, "Attainable Error Bounds in Multirate Adaptive Lossless FIR Filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 2, Detroit, MI, 1995, pp. 1460–1463.
- A. Benveniste, M. Goursat, and G. Ruget, "Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications," *IEEE Trans. Automat. Contr.*, vol. 25, 1980, pp. 385–399.
- B. Farhang-Boroujeny and S. Nooshfar, "Adaptive Phase Equalization Using All-Pass Filters," in *Proc. IEEE Int. Conf. Commun.*, vol. 3, Denver, CO, 1991, pp. 1403–1407.
- T. J. Lim and M.D. Macleod, "Adaptive Allpass Filtering for Nonminimum-Phase System Identification," *IEE Proc.—Vision, Image, Signal Processing*, vol. 141, 1994, pp. 373–379.
- S. Hakyin (Ed.), *Blind Deconvolution*, Englewood Cliffs, NJ: Prentice-Hall, 1994.
- P.A. McEwen and J.G. Kenney, "Allpass Forward Equalizer for Decision Feedback Equalization," *IEEE Trans. Magnetics*, vol. 31, 1995, pp. 3045–3047.
- E. Abreu, S.K. Mitra, and R. Marchesani, "Nonminimum Phase Channel Equalization Using Noncausal Filters," *IEEE Trans. Signal Processing*, vol. 45, 1997, pp. 1–13.
- S.C. Douglas and S.-Y. Kung, "Gradient Adaptive Algorithms for Contrast-Based Blind Deconvolution," *J. VLSI Signal Processing Syst.*, vol. 26, nos. 1/2, 2000, pp. 47–60.
- B. Widrow and E. Walach, *Adaptive Inverse Control*, Upper Saddle River, NJ: Prentice-Hall, 1996.
- H.G. Grassmann, *Die Ausdehnungslehre*, Berlin: Enslin, 1862.
- E. Stiefel, "Richtungsfelder und Fernparallelismus in n -Dimensionalem Manning Faltigkeiten," *Commentarii Math. Helvetici*, vol. 8, 1935/1936, pp. 305–353.
- E. Oja and J. Karhunen, "On Stochastic Approximation of the Eigenvectors and Eigenvalues of the Expectation of a Random Matrix," *J. Math. Anal. Appl.*, vol. 106, no. 1, 1985, pp. 69–84.
- L. Xu, "Least-Mean-Square Error Recognition Principle for Self-Organizing Neural Nets," *Neural Networks*, vol. 6, 1993, pp. 627–648.
- S.T. Smith, "Geometric Optimization Methods for Adaptive Filtering," Ph.D. thesis, Harvard Univ., Cambridge, MA, 1993.
- U. Helmke and J.B. Moore, *Optimization and Dynamical Systems*, New York: Springer-Verlag, 1994.
- D.R. Fuhrmann, "A Geometric Approach To Subspace Tracking," in *Proc. 31st Asilomar Conf. Signals, Syst., Comput.*, vol. 1, Pacific Grove, CA, 1997, pp. 783–787.
- T.-P. Chen, S. Amari, and Q. Lin, "A Unified Algorithm for Principal and Minor Components Extraction," *Neural Networks*, vol. 11, 1998, pp. 385–390.
- S.C. Douglas, S.-Y. Kung, and S. Amari, "A Self-Stabilized Minor Subspace Rule," *IEEE Signal Processing Lett.*, vol. 5, 1998, pp. 328–330.
- A. Edelman, T. Arias, and S.T. Smith, "The Geometry of Algorithms with Orthogonality Constraints," *Siam J. Matrix Anal. Appl.*, vol. 20, 1998, pp. 303–353.
- K.I. Diamantaras and S.-Y. Kung, *Principal Component Neural Networks*, New York: Wiley, 1996.
- S.C. Douglas, S. Amari, and S.Y. Kung, "On Gradient Adaptation with Unit-Norm Constraints," *IEEE Trans. Signal Processing*, vol. 48, no. 6, 2000, pp. 1843–1847.
- P. Comon, "Independent Component Analysis: A New Concept?" *Signal Processing*, vol. 36, 1994, pp. 287–314.
- N. Delfosse and P. Loubaton, "Adaptive Blind Separation of Independent Sources: A Deflation Approach," *Signal Processing*, vol. 45, 1995, pp. 59–83.

35. S.-Y. Kung and C. Mejuto, "Extraction of Independent Components from Hybrid Mixture: KuicNet Learning Algorithm and Applications," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 2, Seattle, WA, 1998, pp. 1209–1212.
36. S.C. Douglas, "Self-Stabilized Gradient Algorithms for Blind Source Separation with Orthogonality Constraints," *IEEE Trans. Neural Networks*, vol. 11, no. 6, 2000, pp. 1490–1497.
37. S. Amari, "Natural Gradient Works Efficiently in Learning," *Neural Computation*, vol. 10, 1998, pp. 251–276.
38. S.C. Douglas and S. Amari, "Natural Gradient Adaptation," in *Unsupervised Adaptive Filtering*, vol. I: *Blind Source Separation*, S. Haykin (Ed.), New York: Wiley, 2000, pp. 13–61.
39. R.H. Lambert, "Multichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath Mixtures," Ph.D. thesis, University of Southern California, Los Angeles, CA, May 1996.
40. S.C. Douglas and S. Haykin, "On the Relationship Between Blind Deconvolution and Blind Source Separation," in *Proc. 31st Ann. Asilomar Conf. Signals, Syst., Comput.*, vol. 2, Pacific Grove, CA, 1997, pp. 1591–1595.
41. I. Sabala, A. Cichocki, and S. Amari, "Relationships Between Instantaneous Blind Source Separation and Multichannel Blind Deconvolution," in *Proc. IEEE Int. Joint Conf. Neural Networks*, vol. 1, Anchorage, AK, 1998, pp. 39–44.
42. S.C. Douglas and S. Haykin, "Relationships Between Blind Deconvolution and Blind Source Separation," in *Unsupervised Adaptive Filtering*, vol. II: *Blind Deconvolution*, S. Haykin (Ed.), New York: Wiley, 2000, pp. 113–145.
43. S.C. Douglas, A. Cichocki, and S. Amari, "Fast-Convergence Filtered Regressor Algorithms for Blind Equalisation," *Electron. Lett.*, vol. 32, no. 23, 1996, pp. 2114–2115.
44. S.C. Douglas, A. Cichocki, and S. Amari, "Quasi-Newton Filtered-Error and Filtered-Regressor Algorithms for Adaptive Equalization and Deconvolution," in *Proc. 1st IEEE Workshop Signal Proc. Adv. Wireless Commun.*, Paris, France, 1997, pp. 109–112.
45. S.C. Douglas, A. Cichocki, and S. Amari, "Self-Whitening Algorithms for Adaptive Equalization and Deconvolution," *IEEE Trans. Signal Processing*, vol. 47, 1999, pp. 1161–1165.
46. P. Comon, "Contrasts for Multichannel Blind Deconvolution," *IEEE Signal Processing Lett.*, vol. 3, 1996, pp. 209–211.
47. E. Moreau and J.-C. Pesquet, "Generalized Contrasts for Multichannel Blind Deconvolution of Linear Systems," *IEEE Signal Processing Lett.*, vol. 4, 1997, pp. 182–183.
48. R.-W. Liu and Y. Inouye, "Direct Blind Signal Separation of Convolutional Mixtures of White Non-Gaussian Signals," in *Proc. 1st Int. Conf. Indep. Compon. Anal. Signal Sep.*, Aussois, France, 1999, pp. 233–238.
49. A. Gorokhov and J.-F. Cardoso, "Equivariant Blind Deconvolution of MIMO-FIR Channels," in *Proc. 1st IEEE Workshop Signal Processing Adv. Wireless Commun.*, Paris, France, 1997, pp. 89–92.
50. S. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang, "Multichannel Blind Deconvolution Using the Natural Gradient," in *Proc. 1st IEEE Workshop Signal Proc. Adv. Wireless Commun.*, Paris, France, 1997, pp. 101–104.
51. K. Torkkola, "Blind Separation of Convolved Sources Based on Information Maximization," *IEEE Workshop Neural Networks Signal Processing*, Kyoto, Japan, 1996, pp. 423–432.
52. R.H. Lambert and A.J. Bell, "Blind Separation of Multiple Speakers in a Multipath Environment," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, Munich, Germany, 1997, pp. 423–426.
53. S. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang, "Novel On-Line Adaptive Learning Algorithms for Blind Deconvolution Using the Natural Gradient Approach," in *Proc. 11th Int. Conf. Syst. Ident.*, vol. 3, Kitakyushu, Japan, 1997, pp. 1057–1062.
54. S.C. Douglas and A. Cichocki, "Neural Networks for Blind Decorrelation of Signals," *IEEE Trans. Signal Processing*, vol. 45, 1997, pp. 2829–2842.



Scott C. Douglas received the B.S. (with distinction), M.S., and Ph.D. degrees in Electrical Engineering from Stanford University, Stanford, CA. He is currently an Associate Professor in the Department of Electrical Engineering at Southern Methodist University and the Associate Director of the Institute for Engineering Education at SMU. His research activities include adaptive filtering, active noise control, blind deconvolution and source separation, and VLSI/hardware implementations of digital signal processing systems.

Dr. Douglas is the author or co-author of two books, six book chapters, and more than 120 articles in journals and conference proceedings; he has also been the proceedings editor or co-editor on six international symposia and workshops. He received the NSF CAREER Award in 1995 and the Best Paper Award of the IEEE Signal Processing Society in 2002 for his work on fast algorithms for multichannel active noise and vibration control. He is a senior member of the IEEE, a past associate editor for the IEEE Transactions on Signal Processing (IEEE) and the IEEE Signal Processing Letters (IEEE), and a current associate editor of the Journal of VLSI Signal Processing Systems (Kluwer). He has served in various IEEE organizations and committees, most notably as the Chair of the Neural Networks for Signal Processing Technical Committee and the Secretary of the Signal Processing Education Technical Committee of the IEEE Signal Processing Society. Most recently, he has played an integral role in developing and managing the Infinity Project, an effort among university faculty, high-tech industry, and civic educational leaders to bring an exciting and practical engineering curriculum to all U.S. high school students. He is a frequent public speaker and consultant to industry in the areas of signal processing and engineering education and is a member of both Phi Beta Kappa and Tau Beta Pi.



Shun-Ichi Amari was born in Tokyo, Japan, on January 3, 1936. He received the bachelor's degree in mathematical engineering from the University of Tokyo in 1958 and the Dr. Eng. degree from the University of Tokyo in 1963. He was an Associate Professor at Kyushu University, an Associate and then Full Professor at the Department of Mathematical Engineering and Information Physics, University of Tokyo, and is now Professor-Emeritus at the University of Tokyo. He is currently the Director of the Brain Science Institute at The Institute of Physical and Chemical Research—RIKEN, where he is also the director of the Brain Style Information Systems Group and the team leader of the Mathematical Neuroscience Laboratory. He has been engaged in research in wide areas of mathematical engineering and applied mathematics, such as topological network theory, differential geometry of continuum mechanics, pattern recognition, mathematical foundations of neural networks, and information geometry.

Dr. Amari served as President of the International Neural Network Society, Council Member of Bernoulli Society for Mathematical Statistics and Probability Theory, and Vice President of the Institute of Electrical, Information, and Communication Engineers. He was founding Coeditor-in-Chief of Neural Networks. He has been awarded the Japan Academy Award, the IEEE Neural Networks Pioneer Award, and the IEEE Emanuel R. Piore Award.



Sun-Yuan Kung received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA. Since 1987, he has

been a Professor of electrical engineering at Princeton University, Princeton, NJ. In 1974, he was an Associate Engineer with Amdahl Corporation, Sunnyvale, CA. From 1977 to 1987, he was a Professor of electrical engineering-systems, University of Southern California, Los Angeles. In 1984, he was a Visiting Professor with the Stanford University and the Delft University of Technology, Delft, The Netherlands. In 1994, he was a Toshiba Chair Professor at Waseda University, Waseda, Japan, and a Honorary Professor with the Central China University of Science and Technology, Wuhan, China. In 2001, he was a Distinguished Chair of Multimedia Signal Processing, Hong Kong Polytechnic University, Hong Kong. His research interests include VLSI array processors, image/video/multimedia signal processing, neural networks for biometric and bioinformatic signal processing, and wireless digital communication. He has served as an Editor-In-Chief of the Journal of VLSI Signal Processing Systems since 1990. He has authored more than 300 technical publications, including three books: VLSI Array Processors (Englewood Cliffs, NJ: Prentice-Hall, 1988) (with Russian and Chinese translations); Digital Neural Networks (Englewood Cliffs, NJ: Prentice-Hall, 1993), and Principal Component Neural Networks (New York: Wiley, 1996). He has edited numerous reference books, including VLSI and Modern Signal Processing (Englewood Cliffs, NJ: Prentice-Hall, 1985) (with Russian translation), VLSI Signal Processing, Vol. III (Piscataway, NJ: IEEE Press), Neural Networks for Signal Processing, Vol. I, II, and III (Piscataway, NJ: IEEE Press), Multimedia Signal Processing, Vol. I (Piscataway, NJ: IEEE Press), Systolic Arrays (Menlo Park, CA: IEEE Comput. Soc. Press), and Application-Specific Array Processors (Menlo Park, CA: IEEE Comput. Soc. Press). He has recently co-edited a book on "Multimedia Image and Video Processing" (Boca Raton, FL: CRC, 2001).