ADAPTIVE PARAUNITARY FILTER BANKS FOR PRINCIPAL AND MINOR SUBSPACE ANALYSIS

S.C. Douglas[1]*, S. Amari*[2]*, and S.-Y. Kung*[3]

[1]Department of Electrical Engineering, Southern Methodist University, Dallas, TX 75275 USA

[2]Brain-Style Information Systems Group, RIKEN Brain Science Institute, Wako-shi, Saitama 351-0198 JAPAN

[3]Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA

ABSTRACT

Paraunitary filter banks are important for several signal processing tasks. In this paper, we consider the task of adapting the coefficients of a multichannel FIR paraunitary filter bank via gradient ascent or descent on a chosen cost function. The proposed generalized algorithms inherently adapt the system's parameters in the space of paraunitary filters. Modifications and simplifications of the techniques for spatio-temporal principal and minor subspace analysis are described. Simulations verify one algorithm's useful behavior in this task.

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 $\mathbf{W}_p = [\mathbf{w}_{1p} \ \cdots \ \mathbf{w}_{mp}]^T$ with $\mathbf{w}_{ip} = [w_{i1p} \ \cdots \ w_{inp}]^T$ and $m \leq n$,

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

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

where $\mathbf{I}$ is the identity matrix and δ_l is the Kronnecker impulse function. Defining the z-transform of $\mathbf{W}_p$ as $\mathbf{W}(z) = \sum_{p=-\infty}^{\infty} \mathbf{W}_p z^{-p}$, (2) is

$$\mathbf{W}(z)\mathbf{W}^T(z^{-1})\big|_{z=e^{j\omega}} \quad = \quad \mathbf{I}. \tag{3}$$

Multichannel linear systems that satisfy (2) or (3) are called *paraunitary* systems. They are useful for designing perfect-reconstruction filter banks for multirate systems used in coding, multichannel deconvolution and equalization, and image processing [1]–[5]. When $m = n = 1$, (3) guarantees that $W(z)$ is an *all-pass filter*, and thus solutions to (1)–(3) are useful for single-channel equalization and control tasks [6]–[10].

To our knowledge, adaptive methods for solving (1)–(3) have not been extensively explored. Works in perfect reconstruction filter bank design have focused on constructive methods using classic mean-square or least-squares approximation theory [2]–[5]. Direct-form adaptive all-pass filters have been developed [7, 8] but are challenging to extend to higher-order systems. In addition, adaptive solutions to (1)–(3) would enable efficient coding and storage of multisensor signals, such as multiple-microphone recordings of several talkers in a room.

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

In this paper, we provide gradient-based adaptive algorithms for solving (1)–(3) iteratively over time. Our algorithms in differential form preserve (2) or (3), and thus they adjust $\mathbf{W}_p$ in the impulse response space of paraunitary filters. We then explore spatio-temporal extensions of principal and minor subspace analysis algorithms [11]–[14], developing the necessary modifications and simplifications to enable their simple and stable implementation in truncated FIR filter form. Simulations verifying the useful behavior of one algorithm are provided.

2. GRADIENT ADAPTIVE PARAUNITARY FILTERS

To develop gradient-based methods for solving (1)–(3), we consider the situation where $\mathbf{W}_p = \mathbf{W}\delta_p$ defines a memoryless system, in which case (1)–(3) reduces to

$$\text{maximize } \mathcal{J}(\mathbf{W}) \qquad \text{such that } \mathbf{W}\mathbf{W}^T = \mathbf{I}. \tag{4}$$

Adaptive solutions to (4) are useful in many problems. For example, when $\mathcal{J}(\mathbf{W}) = \pm\text{tr}[\mathbf{W}\mathbf{R}_{\mathbf{xx}}\mathbf{W}^T]$ where $\mathbf{R}_{\mathbf{xx}}$ is a symmetric positive-definite matrix, (4) becomes principal component analysis (PCA) or minor component analysis (MCA) [11]–[14]. Since solutions to (4) exist, we shall leverage this knowledge by extending one of these solutions to the multichannel dispersive case.

One simple gradient-based method for solving (4) employs unconstrained adaptation via the update

$$\mathbf{W}(k+1) \quad = \quad \mathbf{W}(k) + \mathbf{G}(k) \tag{5}$$

$$\mathbf{G}(k) \quad = \quad \mu(k)\frac{\partial \mathcal{J}(\mathbf{W}(k))}{\partial \mathbf{W}}, \tag{6}$$

where $\mu(k)$ is the step size, combined with the periodic projection of $\mathbf{W}(k)$ back to the constraint space using a Gram-Schmidt, singular-value-decomposition (SVD), or equivalent procedure. Such a solution is likely to be computationally-difficult to extend to the more-general dispersive case, however, due to complications in the projection process.

Note that (4) restricts $\mathbf{W}$ to the space of $(m \times n)$ orthonormal matrices. If in addition $\mathcal{J}(\mathbf{W}) = \mathcal{J}(\mathbf{Q}\mathbf{W})$ for any $(m \times m)$ Hermitian matrix $\mathbf{Q}$, any $\mathbf{W}$ satisfying (4) lies within the Grassman manifold [15]. If $\mathcal{J}(\mathbf{W}) \neq \mathcal{J}(\mathbf{Q}\mathbf{W})$, then the solution to (4) lies within the Stiefel manifold [16]. Algorithms for adjusting $\mathbf{W}(k)$ directly within the Grassman and Stiefel manifolds have received attention recently [17]–[22]. The simplest gradient-based algorithms attempt to move $\mathbf{W}(k)$ via differential changes along a geodesic in the associated parameter space to maximize $\mathcal{J}(\mathbf{W})$, as calculating exact geodesic motions is known to involve m two-dimensional rotations in the Grassman manifold [21]. For the Grassman and Stiefel manifolds, these gradient algorithms in differential form are

$$\frac{d\mathbf{W}}{dt} = \mathbf{W}\mathbf{W}^T\mathbf{G} - \mathbf{G}\mathbf{W}^T\mathbf{W} \quad (7)$$

and

$$\frac{d\mathbf{W}}{dt} = \mathbf{W}\mathbf{W}^T\mathbf{G} - \mathbf{W}\mathbf{G}^T\mathbf{W}, \quad (8)$$

respectively. It can be proven for (7) and (8) that

$$\frac{d\mathbf{W}\mathbf{W}^T}{dt} = \mathbf{0} \quad (9)$$

for all $t \geq t_0$ if $\mathbf{W}(t_0)\mathbf{W}^T(t_0) = \mathbf{I}$ at some time instant $t = t_0$, such that $\mathbf{W}\mathbf{W}^T = \mathbf{I}$ is maintained [22]. Moreover, since no other processing is required, (7) and (8) are ideal for extension to adaptive paraunitary systems.

To develop procedures similar to (7) and (8) to solve (1)–(3), we rely on recent works relating instantaneous blind source separation and blind deconvolution [23]–[25]. 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 multichannel dispersive case.
2. Addition of two matrices in the spatial-only case is equivalent to element-by-element addition of their associated sequences in the multichannel dispersive 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 multichannel dispersive case.

Using these rules, we now extend (7) and (8) to solve (1)–(3). These algorithms are given in continuous-time form for analysis; discrete-time versions are developed later.

Extension of (7) :

$$\frac{d\mathbf{W}_p}{dt} = \mathbf{W}_p * \mathbf{W}_{-p}^T * \mathbf{G}_p - \mathbf{G}_p * \mathbf{W}_{-p}^T * \mathbf{W}_p = \sum_{q,r} \mathbf{W}_{p-r}\mathbf{W}_{q-r}^T\mathbf{G}_q - \mathbf{G}_{p-r}\mathbf{W}_{q-r}^T\mathbf{W}_q. \quad (10)$$

Extension of (8) :

$$\frac{d\mathbf{W}_p}{dt} = \mathbf{W}_p * \mathbf{W}_{-p}^T * \mathbf{G}_p - \mathbf{W}_p * \mathbf{G}_{-p}^T * \mathbf{W}_p = \sum_{q,r} \mathbf{W}_{p-r}\mathbf{W}_{q-r}^T\mathbf{G}_q - \mathbf{W}_{p-r}\mathbf{G}_{q-r}^T\mathbf{W}_q. \quad (11)$$

It can be proven for both (10) and (11) that (2) is satisfied for all $t \geq t_0$ if $\mathbf{W}_p(t_0)$ is paraunitary. For brevity, we only show this property for (10); the proof for (11) is similar. Taking derivatives of both sides of (2) gives

$$\frac{\sum_p \mathbf{W}_p\mathbf{W}_{p+l}^T}{dt} = \sum_p \frac{d\mathbf{W}_p}{dt}\mathbf{W}_{p+l}^T + \mathbf{W}_p\frac{d\mathbf{W}_{p+l}^T}{dt} \quad (12)$$

Substituting the relation in (10) into the RHS of (12) gives

$$\frac{\sum_p \mathbf{W}_p\mathbf{W}_{p+l}^T}{dt} = \sum_{p,q,r} \mathbf{W}_{p-r}\mathbf{W}_{q-r}^T\mathbf{G}_q\mathbf{W}_{p+l}^T - \mathbf{G}_{p-r}\mathbf{W}_{q-r}^T\mathbf{W}_q\mathbf{W}_{p+l}^T + \mathbf{W}_p\mathbf{G}_q^T\mathbf{W}_{q-r}\mathbf{W}_{p+l-r}^T - \mathbf{W}_p\mathbf{W}_q^T\mathbf{W}_{q-r}\mathbf{G}_{p+l-r}^T \quad (13)$$

Let $s = p - r$ and $t = q - r$, respectively, so that (13) is

$$\frac{\sum_p \mathbf{W}_p\mathbf{W}_{p+l}^T}{dt} = \sum_{p,q}\left[\sum_r \mathbf{W}_{p-r}\mathbf{W}_{q-r}^T\right]\mathbf{G}_q\mathbf{W}_{p+l}^T - \sum_{s,t}\mathbf{G}_s\mathbf{W}_t^T\left[\sum_r \mathbf{W}_{t+r}\mathbf{W}_{s+r+l}^T\right] + \sum_{p,q}\mathbf{W}_p\mathbf{G}_q^T\left[\sum_r \mathbf{W}_{q-r}\mathbf{W}_{p-r+l}^T\right] - \sum_{s,t}\left[\sum_r \mathbf{W}_{s+r}\mathbf{W}_{t+r}^T\right]\mathbf{W}_t\mathbf{G}_{s+l}^T \quad (14)$$

Assume that (2) holds at some time instant t_0. Then, the four summations within brackets on the RHS of (14) 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

$$\frac{\sum_p \mathbf{W}_p\mathbf{W}_{p+l}^T}{dt} = \sum_p \mathbf{G}_p\mathbf{W}_{p+l}^T - \sum_s \mathbf{G}_s\mathbf{W}_{s+l}^T + \sum_p \mathbf{W}_p\mathbf{G}_{p+l}^T - \sum_s \mathbf{W}_s\mathbf{G}_{s+l}^T \quad (15)$$

$$= \mathbf{0}. \quad (16)$$

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

3. IMPLEMENTATION ISSUES

In practice, discrete-time versions of (10) and (11) 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

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mathbf{W}(k)\mathbf{W}^T(k)\mathbf{G}(k) - \mathbf{G}(k)\mathbf{W}^T(k)\mathbf{W}(k) \quad (17)$$

$$\mathbf{G}(k) = \mu(k)\mathbf{W}(k)\mathbf{R}_{\mathbf{xx}} \quad (18)$$

is numerically-unstable for PCA and MCA [13, 14]. For this reason, modifications of the updates are required. For PCA, an approximation to (17)–(18) given by

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mathbf{G}(k) - \mathbf{G}(k)\mathbf{W}^T(k)\mathbf{W}(k) \quad (19)$$

with $\mu(k) > 0$ yields the principal subspace estimate in a numerically-stable manner [11]. For MCA, the update

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mathbf{W}(k)\mathbf{W}^T(k)\mathbf{W}(k)\mathbf{W}^T(k)\mathbf{G}(k) - \mathbf{G}(k)\mathbf{W}^T(k)\mathbf{W}(k) \quad (20)$$

with $\mu(k) < 0$ yields asymptotically-stable behavior [14].

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.

3.1. Spatio-Temporal Principal Subspace Analysis

To better understand the procedure for developing simplified, numerically-stable discrete-time implementations of (10) and (11), we consider the problem of spatio-temporal principal subspace analysis. Let the m-dimensional vector $\mathbf{y}(k)$ be the output at time k of the $(m \times n)$-dimensional multichannel linear system given by

$$\mathbf{y}(k) = \sum_{l=-\infty}^{\infty} \mathbf{W}_l(k)\mathbf{x}(k-l), \quad (21)$$

where $\mathbf{x}(k)$ is an n-dimensional vector sequence and $\mathbf{W}_p(k)$, $-\infty < p < \infty$ is the adaptive impulse response of the multichannel system. We wish to maximize the cost function

$$\mathcal{J}(\{\mathbf{W}_p\}_{-\infty}^{\infty}) = E\{||\mathbf{y}(k)||^2\} \quad (22)$$

with respect to $\{\mathbf{W}_p(k)\}$ subject to (2). A stochastic gradient algorithm that employs the cost function

$$\widehat{\mathcal{J}}(\{\mathbf{W}_p\}_{-\infty}^{\infty}) = ||\mathbf{y}(k)||^2 \quad (23)$$

can be used to adjust $\mathbf{W}_p(k)$, as a slowly-adapting system averages across successive time samples in its operation.

A straightforward application of (10) to this task yields

$$\begin{aligned}\mathbf{W}_p(k+1) &= \mathbf{W}_p(k) \\ &+ \mu(k)\sum_{r=-\infty}^{\infty}\mathbf{W}_{p-r}(k)\sum_{q=-\infty}^{\infty}\mathbf{W}_{q-r}^T\mathbf{y}(k)\mathbf{x}^T(k-q) \\ &- \mu(k)\mathbf{y}(k)\sum_{q=-\infty}^{\infty}\left[\sum_{r=-\infty}^{\infty}\mathbf{x}^T(k-p+r)\mathbf{W}_{q-r}^T(k)\right]\mathbf{W}_q(k) \quad (24)\end{aligned}$$

This system can be expected to be numerically-unstable, however, as (17) is numerically-unstable. To obtain a potentially-useful algorithm, we instead begin with a version that is analogous in form to (19) as given by

$$\begin{aligned}\mathbf{W}_p(k+1) &= \mathbf{W}_p(k) + \mu(k)\mathbf{y}(k)\mathbf{x}^T(k-p) \\ &- \mu(k)\mathbf{y}(k)\sum_{q=-\infty}^{\infty}\left[\sum_{r=-\infty}^{\infty}\mathbf{x}^T(k-p+r)\mathbf{W}_{q-r}^T(k)\right]\mathbf{W}_q(k) \quad (25)\end{aligned}$$

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

$$\mathbf{y}(k) = \sum_{l=0}^{L} \mathbf{W}_l(k)\mathbf{x}(k-l) \quad (26)$$

in place of (21) and

$$\begin{aligned}\mathbf{W}_p(k+1) &= \mathbf{W}_p(k) + \mu(k)\mathbf{y}(k)\mathbf{x}^T(k-p) \\ &- \mu(k)\mathbf{y}(k)\sum_{q=0}^{L}\left[\sum_{r=q-L}^{q}\mathbf{x}^T(k-p+r)\mathbf{W}_{q-r}^T(k)\right]\mathbf{W}_q(k) \quad (27)\end{aligned}$$

in place of (25) for $0 \le p \le L$.

Eqn. (27) is both computationally-complex and noncausal. In [26]–[29], a procedure is described and used to simplify related algorithms for multichannel blind deconvolution. We follow a similar procedure in what follows. Letting $l = q - r$, we can make the approximation

$$\begin{aligned}&\sum_{r=q-L}^{q}\mathbf{W}_{q-r}(k)\mathbf{x}(k-p+r) \\ &\approx \sum_{l=0}^{L}\mathbf{W}_l(k-p+q)\mathbf{x}(k-p+q-l) = \mathbf{y}(k-p+q) \quad (28)\end{aligned}$$

and thus

$$\begin{aligned}&\sum_{q=0}^{L}\mathbf{W}_q^T(k)\sum_{r=q-L}^{q}\mathbf{W}_{q-r}(k)\mathbf{x}(k-p+r) \\ &\approx \sum_{q=0}^{L}\mathbf{W}_{L-q}^T(k)\mathbf{y}(k-p+L-q) \quad (29) \\ &\approx \sum_{q=0}^{L}\mathbf{W}_{L-q}^T(k-p+L)\mathbf{y}(k-p+L-q). \quad (30)\end{aligned}$$

Finally, substituting the above results into the RHS of (27) and delaying these terms by L time samples, we obtain a causal delayed-update version of (27) as

$$\begin{aligned}\mathbf{W}_p(k+1) &= \mathbf{W}_p(k) \\ &+ \mu(k)\mathbf{y}(k-L)\left[\mathbf{x}^T(k-L-p) - \mathbf{u}^T(k-p)\right], \quad (31)\end{aligned}$$

where

$$\mathbf{u}(k) = \sum_{q=0}^{L}\mathbf{W}_{L-q}^T(k)\mathbf{y}(k-q). \quad (32)$$

Eqns. (26), (31), and (32) describe the proposed spatio-temporal extension of the principal subspace rule in (17). 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 ($L = 0$).

3.2. Spatio-Temporal Minor Subspace Analysis

Using similar principles, we can extend the MSA algorithm in (20) to the spatio-temporal case. For brevity, the derivations are omitted, and only the final form of the algorithm is given: for $0 \le p \le L$,

$$\begin{aligned}\mathbf{W}_p(k+1) &= \mathbf{W}_p(k) + \mu(k)\left[\zeta(k)\mathbf{x}^T(k-2L-p)\right. \\ &\left.-\mathbf{y}(k-2L)\mathbf{u}^T(k-L-p)\right], \quad (33)\end{aligned}$$

where $\mathbf{y}(k)$ and $\mathbf{u}(k)$ are computed as in (26) and (32),

$$\zeta(k) = \sum_{l=0}^{L}\mathbf{W}_l(k)\mathbf{z}(k-l) \quad (34)$$

$$\mathbf{z}(k) = \sum_{q=0}^{L}\mathbf{W}_{L-q}^T(k)\mathbf{v}(k-q) \quad (35)$$

$$\mathbf{v}(k) = \sum_{l=0}^{L}\mathbf{W}_l(k)\mathbf{u}(k-l). \quad (36)$$

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 (20).

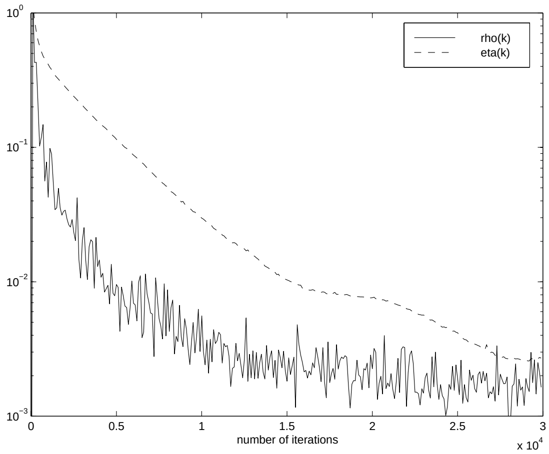

Fig. 1: Evolution of $\rho(k)$ and $\eta(k)$ for the proposed spatio-temporal PSA scheme in the simulation example.

4. SIMULATIONS

We now explore the performance of [(26),(31),(32)] 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 $r_{ss,i}(l) = ((-1)^i 0.5)^{|l|}$, and

$$\mathbf{x}(k) = \sum_{i=1}^{2} \mathbf{A}_i \mathbf{x}(k-i) + \sum_{j=0}^{1} \mathbf{B}_j \mathbf{s}(k-j), \quad (37)$$

$$\mathbf{A}_1 = \begin{bmatrix} 0.38 & 0.39 & -0.22 & 0.08 \\ 0.24 & -0.30 & -0.03 & -0.08 \\ -0.36 & -0.20 & -0.44 & 0.02 \\ -0.49 & 0.16 & 0.49 & -0.17 \end{bmatrix} \quad (38)$$

$$\mathbf{A}_2 = \begin{bmatrix} -0.01 & 0.01 & 0.06 & 0.06 \\ -0.05 & 0.03 & 0.04 & -0.09 \\ 0.02 & -0.06 & -0.01 & 0.02 \\ 0.05 & -0.02 & 0.01 & -0.09 \end{bmatrix} \quad (39)$$

$$\mathbf{B}_0^T = \begin{bmatrix} -0.02 & -0.04 & 0.07 & -0.10 \\ 0.05 & 0.09 & 0.10 & 0.06 \end{bmatrix} \quad (40)$$

$$\mathbf{B}_1^T = \begin{bmatrix} -0.1 & 0.0 & -0.6 & 0.3 \\ -0.4 & 0.9 & 0.5 & -0.2 \end{bmatrix}. \quad (41)$$

We apply [(26),(31),(32)] to $\mathbf{x}(k)$, where $m = 2$, $n = 4$, $L = 6$, $\mu(k) = 0.01$, and each $w_{ijp}(0)$ is initialized to a small random value. Shown in Fig. 1 are the evolutions of

$$\rho(k) = ||\mathbf{x}(k-L) - \mathbf{u}(k)||^2 \quad (42)$$

and
$$\eta(k) = \sum_{l=-L}^{L} \left|\left| \mathbf{I}\delta_l - \sum_{p=0}^{L} \mathbf{W}_p(k)\mathbf{W}_{p+l}^T(k) \right|\right|_F^2 \quad (43)$$

as averaged over twenty different simulation runs. As can be seen, the proposed algorithm effectively adapts to the signal subspace while ensuring the system's paraunitariness without any special coefficient initializations.

5. CONCLUSIONS

In this paper, we have proposed gradient adaptive methods for paraunitary filter banks. When applied to spatio-temporal subspace analysis, the proposed algorithms are computationally-simple and effective. The techniques are expected to be useful for signal-adaptive filter bank design, coding, and equalization tasks.

REFERENCES

[1] P.P. Vaidyanathan *Multirate Systems and Filter Banks* (Englewood Cliffs, NJ: Prentice-Hall, 1993).

[2] M.K. Tsatsanis and G.B. Giannakis, "Principal component filter banks for optimum multiresolution analysis," *IEEE Trans. Signal Processing,* vol. 43, pp. 1766-1777, Aug. 1995.

[3] A. Kiraç and P.P. Vaidyanathan, "Theory and design of optimum FIR compaction filters," *IEEE Trans. Signal Processing,* vol. 46, pp. 903-919, Apr. 1998.

[4] P. Moulin and M.K. Mthçak, "Theory and design of signal-adapted FIR paraunitary filter banks," *IEEE Trans. Signal Processing,* vol. 46, pp. 920-929, Apr. 1998.

[5] B. Xuan and R.H. Bamberger, "FIR principal component filter banks," *IEEE Trans. Signal Processing,* vol. 46, pp. 930-940, Apr. 1998.

[6] 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, pp. 385-399, June 1980.

[7] B. Farhang-Boroujeny and S. Nooshfar, "Adaptive phase equalization using all-pass filters," *Proc. IEEE Int. Conf. Commun.,* Denver, CO, vol. 3, pp. 1403-1407, June 1991.

[8] T.J. Lim and M.D. Macleod, "Adaptive allpass filtering for nonminimum-phase system identification," *IEE Proc.— Vision, Image, Signal Processing,* vol. 141, pp. 373-379, Dec. 1994.

[9] S.C. Douglas and S.-Y. Kung, "KuicNet algorithms for blind deconvolution," *Proc. Workshop Neural Networks Signal Processing,* Cambridge, UK, pp. 3-12, Aug. 1998.

[10] B. Widrow and E. Walach, *Adaptive Inverse Control* (Upper Saddle River, NJ: Prentice-Hall, 1996).

[11] 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, pp. 69-84, 1985.

[12] K.I. Diamantaras and S.-Y. Kung, *Principal Component Neural Networks* (New York: Wiley, 1996).

[13] S.C. Douglas, S. Amari, and S.-Y. Kung, "Gradient adaptation under unit-norm constraints," in *Proc. IEEE Workshop Statistical Signal Array Processing,* Portland, OR, Sept. 1998.

[14] S.C. Douglas, S.-Y. Kung, and S. Amari, "A self-stabilized minor subspace rule," *IEEE Signal Processing Lett.,* to appear.

[15] H.G. Grassman, *Die Ausdehnungslehre,* (Berlin: Enslin, 1862).

[16] E. Stiefel, "Richtungsfelder und fernparallelismus in n-dimensionalem mannig faltigkeiten," *Commentarii Math. Helvetici,* vol. 8, pp. 305-353, 1935-1936.

[17] L. Xu, "Least-mean-square error recognition principle for self-organizing neural nets," *Neural Networks,* vol. 6, pp. 627-648, 1993.

[18] S.T. Smith, *Geometric Optimization Methods for Adaptive Filtering,* Ph.D. thesis, Harvard Univ., Cambridge, MA, 1993.

[19] U. Helmke and J.B. Moore, *Optimization and Dynamical Systems* (New York: Springer-Verlag, 1994).

[20] A. Edelman, T. Arias, and S.T. Smith, "The geometry of algorithms with orthogonality constraints," preprint.

[21] D.R. Fuhrmann, "A geometric approach to subspace tracking," *Proc. 31st Asilomar Conf. Signals, Syst., Comput.,* Pacific Grove, CA, vol. 1, pp. 783-787, Nov. 1997.

[22] T. Chen, S. Amari, and Q. Lin, "A unified algorithm for principal and minor components extraction," *Neural Networks,* vol. 11, pp 385-390, 1998.

[23] 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.

[24] S.C. Douglas and S. Haykin, "On the relationship between blind deconvolution and blind source separation," *Proc. 31st Ann. Asilomar Conf. Signals, Syst., Comput.,* Pacific Grove, CA, vol. 2, pp. 1591-1595, Nov. 1997.

[25] I. Sabala, A. Cichocki, and S. Amari, "Relationships between instantaneous blind source separation and multichannel blind deconvolution," *Proc. IEEE Int. Joint Conf. Neural Networks,* Anchorage, AK, vol. 1, pp. 39-44, May 1998.

[26] S. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang, "Multichannel blind deconvolution using the natural gradient," *Proc. 1st IEEE Workshop Signal Proc. Adv. Wireless Commun.,* Paris, France, pp. 101-104, Apr. 1997.

[27] S.C. Douglas, A. Cichocki, and S. Amari, "Quasi-Newton filtered-error and filtered-regressor algorithms for adaptive equalization and deconvolution," *Proc. 1st IEEE Workshop Signal Proc. Adv. Wireless Commun.,* Paris, France, pp. 109-112, Apr. 1997.

[28] 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," *Proc. 11th IFAC Symp. Syst. Ident.,* Kitakyushu, Japan, vol. 3, pp. 1057-1062, July 1997.

[29] S.C. Douglas, A. Cichocki, and S. Amari, "Self-whitening algorithms for adaptive equalization and deconvolution," *IEEE Trans. Signal Processing,* to appear.