MULTICHANNEL BLIND DECONVOLUTION AND EQUALIZATION USING THE NATURAL GRADIENT

S. Amari†, *S.C. Douglas*‡, *A. Cichocki*†, *and H.H. Yang*†

†Brain Information Processing Group, Frontier Research Program, RIKEN, Saitama 351-01 JAPAN

‡Department of Electrical Engineering, University of Utah, Salt Lake City, UT 84112 USA

ABSTRACT

Multichannel deconvolution and equalization is an important task for numerous applications in communications, signal processing, and control. In this paper, we extend the efficient *natural gradient* search method in [1] to derive a set of on-line algorithms for combined multichannel blind source separation and time-domain deconvolution/equalization of additive, convolved signal mixtures. We prove that the doubly-infinite multichannel equalizer based on the maximum entropy cost function with natural gradient possesses the so-called "equivariance property" such that its asymptotic performance depends on the normalized stochastic distribution of the source signals and not on the characteristics of the unknown channel. Simulations indicate the ability of the algorithm to perform efficient simultaneous multichannel signal deconvolution and source separation.

1. INTRODUCTION

In multichannel blind deconvolution and equalization, an n-dimensional vector of received discrete-time signals $\mathbf{x}(k) = [x_1(k) \cdots x_n(k)]^T$ is assumed to be produced from an m-dimensional vector of source signals $\mathbf{s}(k) = [s_1(k) \cdots s_m(k)]^T$, $m \leq n$ using the mixture model

$$\mathbf{x}(k) = \sum_{p=-\infty}^{\infty} \mathbf{H}_p \mathbf{s}(k-p), \tag{1}$$

where $\mathbf{H}_p$ is an $(n \times m)$-dimensional matrix of mixing coefficients at lag p. The goal is to calculate possibly scaled and/or delayed estimates of the source signals from the received signals using approximate knowledge of the source signal distributions and statistics. Typically, each source signal $s_i(k)$ is an i.i.d. (independent and identically-distributed) sequence that is stochastically-independent of all other source sequences.

For the most part, past methods for multichannel blind deconvolution have focused on estimating the channel impulse response $\{\mathbf{H}_p\}$ from the received signals $\mathbf{x}(k)$ and then determining the source signals from these estimates [2, 3, 4, 5, 6]. Many of these methods use off-line or batch calculations and are not amenable to real-time processing. In addition, several of the algorithms require extensive matrix computations [2, 4, 5] or root finding [3]. Moreover, the estimation of the channel impulse response must occur before the equalized signals can be calculated. In this paper, we consider alternative methods that estimate the source signals directly using a truncated version of a doubly-infinite multichannel equalizer of the form

$$\mathbf{y}(k) = \sum_{p=-\infty}^{\infty} \mathbf{W}_p(k) \mathbf{x}(k-p), \tag{2}$$

where $\mathbf{y}(k) = [y_1(k) \cdots y_m(k)]^T$ is a m-dimensional vector of outputs and $\{\mathbf{W}_p(k), -\infty \leq p \leq \infty\}$ is a sequence of $(m \times n)$-dimensional coefficient matrices. In operator form, the input and output of the equalizer are

$$\begin{aligned} \mathbf{x}(k) &= \mathbf{H}(z)[\mathbf{s}(k)] && (3) \\ \mathbf{y}(k) &= \mathbf{W}(z,k)[\mathbf{x}(k)] = \mathbf{C}(z,k)[\mathbf{s}(k)], && (4) \end{aligned}$$

where

$$\mathbf{W}(z,k) = \sum_{p=-\infty}^{\infty} \mathbf{W}_p(k) z^{-p}, \quad \mathbf{H}(z) = \sum_{p=-\infty}^{\infty} \mathbf{H}_p z^{-p},$$
$$\text{and} \quad \mathbf{C}(z,k) = \mathbf{W}(z,k)\mathbf{H}(z), \tag{5}$$

are the z-transforms of the channel, equalizer, and combined channel-plus-equalizer impulse responses, respectively, z^{-1} is the delay operator, and $z^{-p}[s_i(k)] = s_i(k-p)$. Then, the goal of the deconvolution or equalization task is to adjust $\mathbf{W}(z,k)$ such that

$$\lim_{k \to \infty} \mathbf{C}(z,k) = \mathbf{P}\mathbf{D}(z), \tag{6}$$

where $\mathbf{P}$ is an $(m \times m)$-dimensional permutation matrix with a single unity entry in any of its rows or columns, $\mathbf{D}(z)$ is a diagonal matrix whose (i,i)th entry is $c_i z^{-\Delta_i}$, c_i is a non-zero complex scalar weighting, and Δ_i is an integer delay value. This channel equalization methodology is the multichannel equivalent of traditional Bussgang blind equalization schemes. A significant shortcoming of Bussgang techniques in the multichannel case is their relatively-slow convergence speeds [7], and thus few reports of algorithms that successfully adapt a multichannel equalizer of the form in (2) have appeared in the literature.

When $\mathbf{H}_p = \mathbf{H}_0 \delta_p$, where δ_p is the Kronecker delta function, the task of deconvolution is reduced to the simpler task of multichannel source separation of instantaneous signal mixtures. Several useful, iterative algorithms have been developed for this task [1, 8, 9]. Moreover, the asymptotic convergence behavior of some of these algorithms is independent of the form of H_0, a property known as *equivariance* [8]. Although it might appear that these algorithms

[0] Portions of this research have been supported by the Federal Bureau of Investigation under Contract No. JFBI95017.

0-7803-3944-4/97/$10.00 © 1997 IEEE

could be extended to solve the deconvolution problem, to our knowledge, no description of a time-domain version of a multichannel deconvolution algorithm having equivariant properties has been derived.

In this paper, we extend previously proposed techniques for blind signal separation of additive instantaneous mixtures [1, 8] to the task of joint signal separation and deconvolution. In particular, we determine an extension of the natural gradient search method described in [1] and apply it to a novel blind deconvolution algorithm based on a maximum entropy problem formulation. We prove that our algorithm possesses the equivariance property, such that its statistical performance is independent of any ill-conditioning in the received signal statistics caused by the channel. Simulations indicate the capability of the new algorithm to quickly equalize an unknown non-minimum phase channel.

2. MULTICHANNEL BLIND DECONVOLUTION ALGORITHMS

2.1. The Natural Gradient

Stochastic gradient optimization methods for parameterized systems suffer from slow convergence due to the statistical correlations of the processed signals. While quasi-Newton methods can be used to overcome the performance limitations of these schemes [10], they are often costly in terms of computation and can suffer from numerical problems if not implemented properly.

The natural [1, 11] (or relative [8]) gradient search method is a particularly-useful technique for solving iterative estimation problems. In its simplest form, the natural gradient method alters the true gradient search direction according to the local Riemannian structure of the parameter space. The Riemannian metric tensor measures the degree to which infinitesimal parameter variations change according to the underlying statistical characteristics of the parameters. The natural gradient search direction is given by the inverse of the Riemannian metric tensor as applied to the standard gradient direction. Natural gradient learning gives asymptotically-efficient performance in the Fisher information sense; details of this result and of the forms of the natural gradient algorithms for various tasks can be found in [11]. For linear inverse problems such as source separation or deconvolution, the inverse of the Riemannian metric tensor is a quadratic function of the parameter values, and thus the calculation of the natural gradient algorithm is particularly simple, as we shall indicate.

2.2. Derivation of the Deconvolution Algorithm

We now derive a natural gradient algorithm for adapting $\mathbf{W}(z,k)$ in (4) for the multichannel deconvolution and equalization task. For this derivation, both $\mathbf{H}(z)$ and $\mathbf{W}(z,k)$ are assumed to be stable with no zeros on the unit circle $|z|=1$. In addition, derivatives of quantities with respect to $\mathbf{W}(z,k)$ can be understood as a series of matrices indexed by the lag value p of $\mathbf{W}_p(k)$ in (5). We also assume that the number of real-valued sources m equals the number of sensors n, although the final form of the algorithm is described for the most general case.

For our derivation, we consider the cost or loss function $\phi(\mathbf{W}(z,k))$ of the form

$$\phi(\mathbf{W}(z,k)) = -\sum_{i=1}^{m}\log p_i(y_i(k)) - \frac{1}{2\pi j}\oint \log|\det \mathbf{W}(z,k)|z^{-1}dz, \quad (7)$$

where $p_i(y_i)$ is any valid p.d.f. and $j=\sqrt{-1}$. The expectation of the first term in (7) is an affine measure of the negative of the joint entropy of the time series $\mathbf{z}(k)=[z_1(k)\cdots z_m(k)]^T$, where $z_i(k)=P_i(y_i(k))$ and $p_i(y_i)=P_i'(y_i)$. The second term on the RHS of (7) is a constraint term that guarantees that $\mathbf{W}(z,k)=\mathbf{0}$ is not a minimizing point of (7).

We now determine an algorithm for minimizing $E\{\phi(\mathbf{W}(z,k))\}$ with respect to $\mathbf{W}(z,k)$ iteratively over time. To do so, we determine the total differential $d\phi$ when $\mathbf{W}(z,k)$ undergoes an infinitesimal change $d\mathbf{W}(z,k)$ as

$$d\phi(\mathbf{W}(z,k)) = \phi(\mathbf{W}(z,k)+d\mathbf{W}(z,k)) - \phi(\mathbf{W}(z,k)), \quad (8)$$

as this corresponds to the gradient of the cost function with respect to $\mathbf{W}(z,k)$. Let us define

$$f_i(y_i) = -\frac{d}{dy_i}\log p_i(y_i). \quad (9)$$

Then,

$$d\left(-\sum_{i=1}^{m}\log p_i(y_i(k))\right) = \sum_{i=1}^{m} f_i(y_i(k))dy_i(k) \quad (10)$$
$$= \mathbf{f}^T(\mathbf{y}(k))d\mathbf{y}(k), \quad (11)$$

in which $\mathbf{f}(\mathbf{y}(k))=[f_1(y_1(k))\cdots f_m(y_m(k))]^T$ and $d\mathbf{y}(k)$ is given in terms of $d\mathbf{W}(z,k)$ as

$$d\mathbf{y}(k) = d\mathbf{W}(z,k)[\mathbf{x}(k)] = d\mathbf{W}(z,k)\mathbf{W}^{-1}(z,k)[\mathbf{y}(k)], \quad (12)$$

by nature of the relationship in (4).

Define a modified coefficient differential $d\mathbf{X}(z,k)$ as

$$d\mathbf{X}(z,k) = \sum_{p=-\infty}^{\infty} d\mathbf{X}_p(k)z^{-p} = d\mathbf{W}(z,k)\mathbf{W}^{-1}(z,k). \quad (13)$$

With this definition, we have

$$d\left(-\sum_{i=1}^{m}\log p_i(y_i(k))\right) = \mathbf{f}^T(\mathbf{y}(k))d\mathbf{X}(z,k)[\mathbf{y}(k)]. \quad (14)$$

Similarly, it can be seen that

$$d\left(\frac{1}{2\pi j}\oint \log|\det \mathbf{W}(z,k)|z^{-1}dz\right)$$
$$= \frac{1}{2\pi j}\oint \mathrm{tr}\left(d\mathbf{W}(z,k)\mathbf{W}^{-1}(z,k)\right)z^{-1}dz \quad (15)$$
$$= \mathrm{tr}\left(d\mathbf{X}_0(k)\right), \quad (16)$$

where $\mathrm{tr}(\cdot)$ denotes the trace operation.

Thus, combining (14) and (16) gives

$$d\phi(\mathbf{W}(z,k)) = \mathbf{f}^T(\mathbf{y}(k))d\mathbf{X}(z,k)[\mathbf{y}(k)] - \mathrm{tr}\left(d\mathbf{X}_0(k)\right). \quad (17)$$

The differential in (17) is in terms of the modified coefficient differential matrix $d\mathbf{X}(z,k)$. Note that $d\mathbf{X}(z,k)$ is a linear combination of the coefficient differentials $dW_{ij}(z,k)$ in the matrix polynomial $d\mathbf{W}(z,k)$. Thus, so long as $\mathbf{W}(z,k)$ is non-singular, $d\mathbf{X}(z,k)$ represents a valid search direction to minimize (7), because $d\mathbf{X}(z,k)$ spans the same tangent space of matrices as spanned by $d\mathbf{W}(z,k)$. For these reasons, one could use an alternative stochastic gradient search method of the form

$$\mathbf{W}_p(k+1) = \mathbf{W}_p(k) - \mu(k)\left[\frac{d\phi\left(\mathbf{W}(z,k)\right)}{d\mathbf{X}_p(k)}\right]\mathbf{W}(z,k), \quad (18)$$

where the right-sided operator $\mathbf{W}(z,k)$ acts on the gradient term in brackets only in the time dimension p. The search direction employed by (18) is nothing more than the natural gradient search direction using the Riemannian metric tensor of the space of all matrix filters of the form of $\mathbf{W}(z,k)$ in (5) [11].

Now, we note from (17) and (18) that

$$\left[\frac{d\phi\left(\mathbf{W}(z,k)\right)}{d\mathbf{X}_p(k)}\right]\mathbf{W}(z,k) = \mathbf{f}(\mathbf{y}(k))[\mathbf{y}^T(k-p)]\mathbf{W}(z,k) - [\mathbf{I}\delta_p]\mathbf{W}(z,k). \quad (19)$$

The second term on the RHS of (19) simplifies to $\mathbf{W}_p(k)$. To express the first term, define

$$\mathbf{u}_p(k) = \sum_{q=-\infty}^{\infty} \mathbf{W}_q^T(k)\mathbf{y}(k-p+q). \quad (20)$$

Then, substituting (19) and (20) into (18) gives the coefficient updates as

$$\mathbf{W}_p(k+1) = \mathbf{W}_p(k) + \mu(k)\left[\mathbf{W}_p(k) - \mathbf{f}(\mathbf{y}(k))\mathbf{u}_p^T(k)\right]. \quad (21)$$

2.3. Practical Implementation Issues

In practice, the doubly-infinite non-causal equalizer cannot be implemented, and thus we approximate it by the finite-impulse-response (FIR) causal equalizer given by

$$\mathbf{y}(k) = \sum_{p=0}^{L} \mathbf{W}_p(k)\mathbf{x}(k-p). \quad (22)$$

However, even with this restriction, the pth coefficient matrix update in (21) depends on future equalizer outputs $\mathbf{y}(k+q)$, $q \leq L-p$ through the definition of $\mathbf{u}_p(k)$ in (20) for a truncated equalizer. Instead of using approximations and data storage to estimate $\mathbf{u}_p(k)$, we delay the last term on the RHS of (21) by L samples. This delayed update maintains the same statistical relationships between the signals in the updates and provides similar performance to (21) for small step sizes. In addition, we can assume that $\mathbf{W}_p(k) \approx \mathbf{W}_p(k-1) \approx \cdots \approx \mathbf{W}_p(k-2L)$, such that

$$\mathbf{u}_p(k) \approx \mathbf{u}_0(k-p). \quad (23)$$

With these changes, the proposed algorithm in the general case of complex-valued signals and $m \leq n$ is

$$\mathbf{W}_p(k+1) = \mathbf{W}_p(k) + \mu(k)\left[\mathbf{W}_p(k) - \mathbf{f}(\mathbf{y}(k-L))\mathbf{u}^{*T}(k-p)\right], \quad (24)$$

where $*$ denotes complex-conjugate and

$$\mathbf{u}(k) = \sum_{q=0}^{L} \mathbf{W}_{L-q}^{*T}(k)\mathbf{y}(k-q). \quad (25)$$

Note that this algorithm is computationally-simple, requiring approximately $4mn(L+1)+m$ multiplications per time instant, and it requires approximately $(mn+m+2n)(L+1)$ memory locations to implement.

As noted in [1, 8], the optimum choice for each $f_i(y_i)$ depends on the statistics of each $y_i(k)$ at convergence. The optimal choices $f_i(y_i) = -d\log(p_i(y_i))/dy_i$ yield the fastest convergence behavior; however, suboptimal choices for these nonlinearities still allow the algorithm to perform separation and deconvolution of the sources. Since typical baseband digital communication signals are complex-valued sub-Gaussian with a negative kurtosis, the choices $f_i(y_i) = |y_i|^2 y_i$ yield adequate separation and deconvolution. Similarly, choosing $f_i(y_i) = y_i/|y_i|$ or $f_i(y_i) = \tanh(\gamma y_i)$ for $\gamma > 2$ enables the algorithm to deconvolve mixtures of super-Gaussian sources with a positive kurtosis. In situations where $\mathbf{x}(k)$ contains mixtures of both sub- and super-Gaussian sources, additional techniques are required to enable the system to adapt properly [12].

3. EQUIVARIANT PERFORMANCE

We now study the algorithm in (21) in the combined channel-plus-equalizer parameter space of $\mathbf{C}(z,k)$ defined in (5). In source separation of instantaneous additive mixtures, algorithms whose behaviors do not depend explicitly on the form of the mixing matrix $\mathbf{H}_0$ are equivariant [8]. We extend this definition by stating that a deconvolution algorithm is equivariant if its behavior only depends on the combined filter $\mathbf{C}(z,k)$. Such a result can only hold for the doubly-infinite equalizer in (2), however, as the ability of an FIR equalizer to properly compensate for an IIR channel depends on the length of the equalizer and the initial values of the equalizer coefficients.

Let us consider the algorithm in the form

$$\Delta\mathbf{W}(z,k) = -\mu(k)\frac{d\phi\left(\mathbf{W}(z,k)\right)}{d\mathbf{X}(z,k)}\mathbf{W}(z,k), \quad (26)$$

Post-multiplying both sides of this equation by $\mathbf{H}(z)$ and noting the definition in (5), we have

$$\Delta\mathbf{C}(z,k) = -\mu(k)\frac{d\phi\left(\mathbf{W}(z,k)\right)}{d\mathbf{X}(z,k)}\mathbf{C}(z,k), \quad (27)$$

We note from (17) and (4) that

$$d\phi\left(\mathbf{W}(z,k)\right) = \mathbf{f}^T(\mathbf{W}(z,k)[\mathbf{x}(k)])d\mathbf{X}(z,k)\mathbf{W}(z,k)[\mathbf{x}(k)] - \mathrm{tr}\left(d\mathbf{X}_0(k)\right) \quad (28)$$

$$= \mathbf{f}^T(\mathbf{C}(z,k)[\mathbf{s}(k)])d\mathbf{X}(z,k)\mathbf{C}(z,k)[\mathbf{s}(k)] - \mathrm{tr}\left(d\mathbf{X}_0(k)\right) \quad (29)$$

Thus, both $d\phi\left(\mathbf{W}(z,k)\right)/d\mathbf{X}(z,k)$ and the update in (27) can be expressed in a form that is independent of the channel $\mathbf{H}(z)$. This result proves the equivariance of the proposed multichannel deconvolution algorithm.

Although equivariance indicates the useful convergence behavior of the algorithm, it does not guarantee that $\mathbf{W}(z,k)$ is adequate for equalizing the channel, nor does it guarantee good performance when the equalizer filter has a poor initialization. In fact, initializations that cause $\mathbf{C}(z,0)$ to be nearly zero for some $|z|=1$ can cause convergence problems, as the equalizer effectively has a weak signal excitation in the space of $\mathbf{W}(z,k)$ associated with these values of $|z|=1$. For this reason, we employ a center-tap initialization scheme where

$$\mathbf{W}(z,0) = \mathbf{I}z^{-j} \quad (30)$$

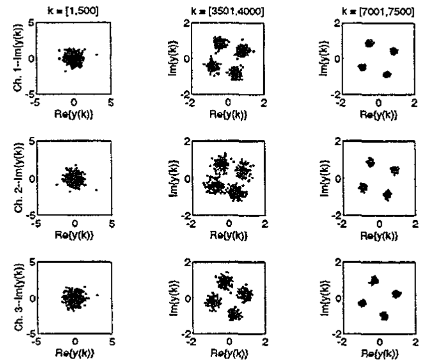

Fig. 1: Output constellations for the blind equalizer.

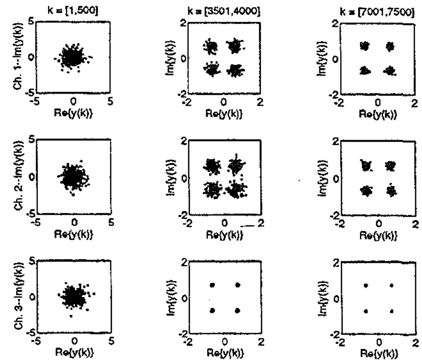

Fig. 2: Output constellations for an LMS equalizer.

for some $0 \leq j \leq L$. It also may be necessary to use tap-centering schemes not unlike those used in standard Bussgang equalizers to obtain good performance from this system.

4. SIMULATIONS

We now illustrate the performance of the multichannel equalization algorithm in (24) via simulations. Consider a three-input, three-output equalization task in which each $s_i(k)$ is an i.i.d. quadrature-amplitude-modulated (QAM) signal with $|s_i(k)| = 1$. The received signals are given by

$$\mathbf{x}(k) = \sum_{i=1}^{2} \mathbf{A}_i \mathbf{x}(k-i) + \sum_{j=0}^{1} \mathbf{B}_j s(k-j), \quad (31)$$

$$\mathbf{A}_1 = \begin{bmatrix} -0.56+0.35j & 0.34+0.08j & -0.14-0.40j \\ -0.28+0.08j & 0.18+0.43j & -0.66-0.15j \\ -0.08-0.27j & -0.40+0.15j & 0.16-0.36j \end{bmatrix} \quad (32)$$

$$\mathbf{A}_2 = \begin{bmatrix} 0.04+0.03j & 0.02+0.16j & 0.03-0.09j \\ 0.08-0.05j & 0.04+0.08j & -0.01j \\ 0.03 & 0.06+0.03j & 0.06+0.01j \end{bmatrix} \quad (33)$$

$$\mathbf{B}_0 = \begin{bmatrix} 0.02+0.09j & 0.07+0.01j & 0.05+0.07j \\ 0.08j & 0.09+0.07j & 0.08+0.09j \\ 0.07+0.05j & 0.04+0.04j & 0.08j \end{bmatrix} \quad (34)$$

$$\mathbf{B}_1 = \begin{bmatrix} 0.1+0.3j & 0.3j & 0.4+j \\ 0.5 & 0.4+0.6j & 0.7+0.4j \\ 0.7+0.7j & 0.1+0.8j & 0.6+0.2j \end{bmatrix}. \quad (35)$$

The channel in (31) is non-minimum phase and cannot be equalized using decision-directed adaptation. We process $\mathbf{x}(k)$ using the proposed multichannel equalizer with non-zero matrices $\mathbf{W}_p(k)$, $0 \leq p \leq 6$, where $\mathbf{W}_p(0) = \delta_{p-3}\mathbf{I}$ and $\mu(k) = 0.001$. Although single simulation runs are shown, the results are indicative of the general behavior of the system in this situation.

Figure 1 shows the constellations of the three output signals $y_i(k)$ of the proposed equalizer for the time intervals $1 \leq k \leq 500$, $3501 \leq k \leq 4000$, and $7001 \leq k \leq 7500$, respectively. Each of the equalizer outputs converges to the characteristic QAM constellation, up to an amplitude and phase rotation factor. Moreover, a comparison of the source and output signals shows that $y_1(k) \approx e^{j\pi 0.41} s_3(k-4)$, $y_2(k) \approx e^{j\pi 0.41} s_2(k-4)$, and $y_3(k) \approx e^{j\pi 0.33} s_1(k-4)$, such that the channel has indeed been equalized. For comparison, Figure 2 shows the output signal constellations for an LMS-based multichannel equalizer trained using $d_i(k) = s_i(k-4)$, $1 \leq i \leq 3$. The blind equalizer provides similar initial convergence to an open-eye condition as does an equalizer with training signals, suggesting that the proposed equalizer can be successfully used for initial channel acquisition.

5. SUMMARY AND CONCLUSIONS

In this paper, we have presented a derivation of an algorithm based on a maximum entropy formulation for multichannel blind deconvolution of mixed and convolved sources. In addition, we have extended the natural gradient method to the multichannel deconvolution task, which provides fast and accurate adaptation that is independent of the characteristics of the unknown channel for sufficiently-long equalizers. We have discussed the choice of nonlinearities $f_i(y_i)$ for blind deconvolution for different source distributions. Simulations indicate that the multichannel blind deconvolution and equalization technique provides excellent convergence behavior for these tasks.

REFERENCES

[1] S. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems 1995* (Boston, MA: MIT Press, 1996), pp. 752-763.

[2] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory,* vol. 40, no. 2, pp. 340-349, Mar. 1994.

[3] M.I. Gurelli and C.L. Nikias, "EVAM: An eigenvector-based algorithm for multichannel bind deconvolution of input colored signals," *IEEE Trans. Signal Processing,* vol. 43, no. 1, pp. 134-149, Jan. 1995.

[4] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing,* vol. 43, no. 12, pp. 2982-2993, Dec. 1995.

[5] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Processing,* vol. 44, no. 3, pp. 661-672, Mar. 1996.

[6] A. Gorokhov, P. Loubaton, and E. Moulines, "Second order blind equalization in multiple input multiple output FIR systems: A weighted least squares approach," *Proc. ICASSP,* Atlanta, GA, vol. 5, pp. 2415-2418, May 1996.

[7] J.K. Tugnait, "Blind equalization and channel estimation for multiple-input multiple- output communications systems," *Proc. ICASSP,* Atlanta, GA, vol. 5, pp. 2443-2446, May 1996.

[8] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing,* vol. 44, no. 12, pp. 3017-3030, Dec. 1996.

[9] A.J. Bell and T.J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Computation,* vol. 7, no. 6, pp. 1129-1159, November 1995.

[10] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification* (Cambridge, MA: MIT Press, 1983).

[11] S. Amari, "Natural gradient works efficiently in learning," submitted to *Neural Computation.*

[12] S. Amari, T.-P. Chen, and A. Cichocki, "Stability analysis of adaptive blind source separation," submitted.

[13] S.C. Douglas, A. Cichocki, and S. Amari, "Quasi-Newton filtered-error and filtered-regressor algorithms for adaptive equalization and deconvolution," presented at *IEEE Workshop Signal Proc. Adv. Wireless Comm.,* Paris, France, Apr. 1997.