Natural Gradient Multichannel Blind Deconvolution and Speech Separation Using Causal FIR Filters

Scott C. Douglas, *Senior Member, IEEE*, Hiroshi Sawada, *Senior Member, IEEE*, and Shoji Makino, *Fellow, IEEE*

***Abstract*—Natural gradient adaptation is an especially convenient method for adapting the coefficients of a linear system in inverse filtering tasks such as convolutive blind source separation and multichannel blind deconvolution. When developing practical implementations of such methods, however, it is not clear how best to window the signals and truncate the filter impulse responses within the filtered gradient updates. In this paper, we show how inadequate use of truncation of the filter impulse responses and signal windowing within a well-known natural gradient algorithm for multichannel blind deconvolution and source separation can introduce a bias into its steady-state solution. We then provide modifications of this algorithm that effectively mitigate these effects for estimating causal FIR solutions to single- and multichannel equalization and source separation tasks. The new multichannel blind deconvolution algorithm requires approximately 6.5 multiply/adds per adaptive filter coefficient, making its computational complexity about 63% greater than the originally-proposed version. Numerical experiments verify the robust convergence performance of the new method both in multichannel blind deconvolution tasks for i.i.d. sources and in convolutive BSS tasks for real-world acoustic sources, even for extremely-short separation filters.**

***Index Terms*—Blind source separation, multichannel blind deconvolution, natural gradient, speech enhancement.**

I. Introduction

BLIND SOURCE separation (BSS) is a field that has received much attention recently in several research fields. Most formulations to the BSS task assume that a set of sensor signals contain linear mixtures of several source signals of interest. The goal is to process the sensor signals to extract versions of each of the source signals without crosstalk and without precise knowledge of the source signals, the mixing conditions, or any training information.

Initial approaches to BSS assumed that the source signal mixtures were *instantaneous*, in that there were no consistent relationships between the sensor measurements at different times other than those due to the source signals themselves. Recently, however, much research has been devoted to the *convolutive* mixing case, in which a sequence of m sources $\{s_i(k)\}, 1 \leq i \leq m$ is mixed by a causal time-dispersive multichannel system as

$$x_j(k) = \sum_{l=0}^{\infty} \sum_{i=1}^{m} a_{jil} s_i(k-l) \tag{1}$$

where $\{x_j(k)\}, 1 \leq j \leq n$ are the n sensor signals and $\{a_{jil}\}$ are the coefficients of the mixing system. Because the relationship between the $\{x_j(k)\}$ and $\{s_i(k)\}$ signals is time-invariant, it makes sense to construct a separation system in the form of a linear multichannel filter. For practical implementations, this separation system takes on the finite impulse response (FIR) form

$$y_i(k) = \sum_{l=0}^{L} \sum_{j=1}^{m} w_{ijl}(k) x_j(k-l) \tag{2}$$

where $\{w_{ijl}(k)\}$ are the $mn(L+1)$ coefficients of the separation system at time k. The separation system is made time-varying with the assumption that the $\{w_{ijl}(k)\}$ can be iteratively adapted to achieve the separation task.

The goal of blind source separation using the models in (1) and (2) depends on what is known, or can be assumed about the source signals $\{s_i(k)\}$ for $1 \leq i \leq m$ and different k. Almost all separation methods assume that the source signals are spatially-uncorrelated, such that $E\{s_i(k)s_j(l)\} = 0$ for all $i \neq j$, any k, and any l. Certain algorithms require that $s_i(k)$ and $s_j(l)$ are independent, a much stronger condition. All separation methods make use of additional assumptions about the statistics of the sources, such as the non-Gaussianity of their amplitude statistics, the nonstationarity of their second-order statistics, or the uniqueness of their power spectra [1]–[15]. In this paper, we shall be concerned with algorithms that exploit the non-Gaussianity of the source signals. In such cases, two solutions to the separation task are possible

- *Multichannel blind deconvolution:* The goal is to iteratively calculate $\{w_{ijl}(k)\}$ such that

$$y_i(k) = d_{i\Delta_j} s_j(k - \Delta_j) \tag{3}$$

  where d_i is a scaling factor and the one-to-one index assignments $j \rightarrow i$ guarantee that each $s_j(\cdot)$ sequence appears in only one $y_i(\cdot)$ output.
- *Convolutive blind source separation:* The goal is to iteratively calculate $\{w_{ijl}(k)\}$ such that

$$y_i(k) = \sum_{l=0}^{\infty} d_{il} s_j(k-l) \tag{4}$$

Manuscript received February 13, 2004; revised July 28, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Walter Kellermann.

S. C. Douglas is with the Department of Electrical Engineering, School of Engineering, Southern Methodist University, Dallas, TX 75275 USA (e-mail: douglas@engr.smu.edu).

H. Sawada and S. Makino are with the NTT Communications Science Laboratories, NTT Corporation, Kyoto 619-0237, Japan.

Digital Object Identifier 10.1109/TSA.2004.838538

where d_{il} is a sequence of coefficients such that the frequency response $D_i(\omega) \neq 0$ for any $|\omega| \leq \pi$.

Multichannel blind deconvolution is a special case of convolutive BSS in which the time-domain structure of the sources is preserved in the system's outputs.

Numerous algorithms for multichannel blind deconvolution and convolutive BSS have been developed [1]–[15]. These differ in the computational requirements for their implementation. One of the simplest classes of approaches is based on the concept of the *natural gradient*, which is a modified gradient search. In [4], the natural gradient BSS method for instantaneous mixtures was extended to the multichannel blind deconvolution task, resulting in the following algorithm for updating the coefficients $\{w_{ijl}(k)\}$:

$$w_{ijl}(k+1) = w_{ijl}(k) + \mu[w_{ijl}(k) - f_i(y_i(k-L))u_j(k-l)] \quad (5)$$

$$u_j(k) = \sum_{q=0}^{L} \sum_{i=1}^{m} w_{ij(L-q)} y_i(k-q) \quad (6)$$

where $f_i(y)$ is a nonlinear function that is related to the amplitude statistics of the source signal $s_j(k)$ being extracted at the ith output. This algorithm is particularly simple, requiring about four multiply/adds per adaptive filter coefficient. The algorithm also employs linear convolutions, thereby making it amenable to fast convolution techniques.

For convolutive BSS, the assumption of source sequences with time-independent structures is not a good one. Even so, several researchers have indicated that the above algorithm and similar ones to it do an adequate job of separating convolutive signal mixtures [5]–[7]. The main artifact produced by the algorithms appears to be a "whitening" or approximate spectral flattening of the sources at the extracted outputs. Since this operation is linear and time-invariant, it is a straightforward task to impose a spectral weighting via post-filtering of the system's outputs assuming that a desired spectral characteristic has been chosen. Improvements in signal-to-interference (SIR) ratios typically range from 6 dB to 14 dB for real-world two-channel mixtures. This algorithm also forms the basis of several extensions that directly address the time-domain decorrelation of the separated source signals [7], [13], [14].

The multichannel blind deconvolution and source separation algorithm in (2), (5), and (6) has been derived assuming that the separation system is adequate for the chosen task. It is well-known, however, that FIR systems cannot perfectly equalize any causal linear system. In situations where oversampling is not used, only those linear systems that are minimum phase have causal inverses, and only those systems that are purely autoregressive have causal FIR inverses [17]. Moreover, in the multichannel case, FIR systems cannot adequately compensate for channel multipath in general.

Because the approaches in [5]–[14] are based on FIR approximations to doubly infinite impulse response (IIR) filter separation procedures, the way these algorithms have been approximated using signal truncation and windowing could affect their estimating abilities. At best, a procedure with poorly-chosen signal truncation and windowing is likely to require overly-long

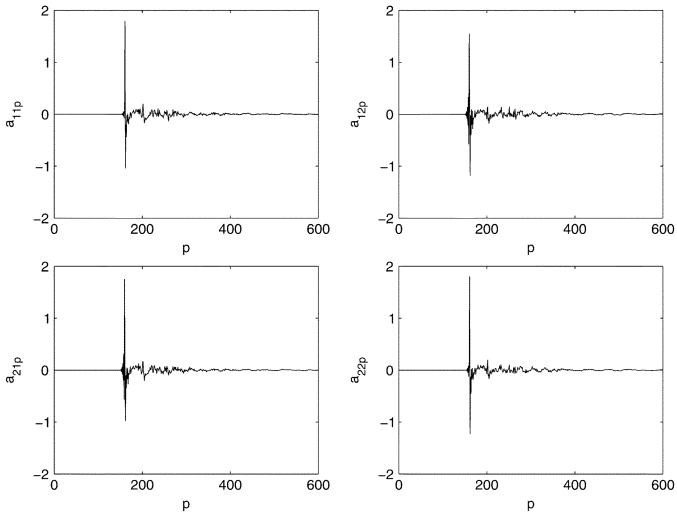

Fig. 1. Impulse responses for the acoustic channel in the multichannel simulation experiments.

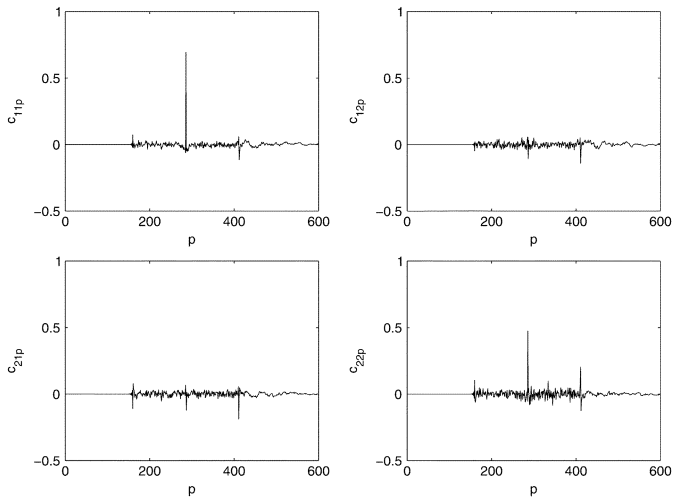

Fig. 2. Combined system impulse responses for the original natural gradient multichannel blind deconvolution procedure for i.i.d. binary source signals, where $L = 250$.

FIR separation filters to achieve adequate performance—even if the acoustic mixing conditions do not require such long filters to achieve adequate separation—leading to systems that are slow to converge and difficult to tune properly for best performance. A system with inadequate signal truncation may require additional mechanisms to achieve adequate separation [16].

We have identified a performance limitation of the algorithm in (2), (5), and (6), as demonstrated in Figs. 1–3. Complete details of this example are provided in Section III-C. Fig. 1 shows the impulse responses $\{a_{ijp}\}$ for a two-loudspeaker, two-microphone laboratory measurement setup at a sampling rate of 8 kHz. Fig. 2 shows the impulse responses $\{c_{ijp}\}$ of the combined system obtained after applying the multichannel blind deconvolution procedure described above to i.i.d. binary-$\{\pm 1\}$-distributed sources that are mixed by the acoustic channel, where each subfilter within the separation system has 251 coefficients. Notice that the edges of the impulse responses of the various subfilters within the combined system exhibits "spikes" at either

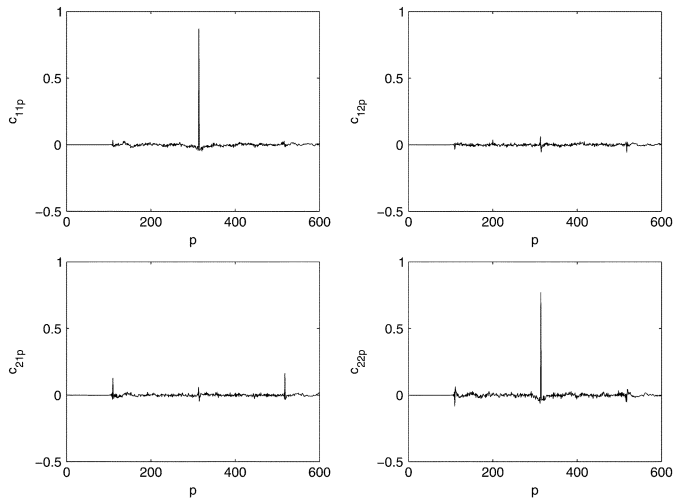

Fig. 3. Combined system impulse responses for the original natural gradient multichannel blind deconvolution procedure for i.i.d. binary source signals, where $L = 407$.

end of their temporal windows. These spikes create a pre- and post-echo that harm the overall separation performance of the scheme. Fig. 3 shows the combined system impulse responses for a separation system with 408-tap subfilters, in which the same spikes are evident at the ends of their temporal windows. Although illustrated with artificial signals, these artifacts appear with most signal sets after an extended number of algorithm iterations, and they cause a significant decrease in overall separation performance. Several researchers have developed extensions of these algorithms in an attempt to mitigate these effects, which generally increase the complexity of the approach [7], [12], [14]. To our knowledge, no researcher has identified the cause of this behavior in the algorithm, and thus, any approach designed to mitigate it is difficult to justify. Addressing this issues could lead to improved steady-state separation and deconvolution performance for this algorithm and others on which it is based.

In this paper, we study the effect that signal windowing and filter truncation play in the design of natural gradient methods for blind deconvolution and source separation tasks. We show that the windowing approximations used in the derivation of (5) and (6) have the potential of introducing a bias into the separating solution, lowering the overall performance of the system in steady-state. We then introduce a new implementation of the natural gradient method for multichannel blind deconvolution that does not suffer from these performance limitations. The proposed multichannel blind deconvolution and source separation algorithm requires about 63% more multiply/adds than the original implementation on a per-sample basis for equivalent filter lengths. Because the steady-state performance of the proposed algorithm is more well-behaved, however, it is possible to choose smaller filter lengths than the original method to tradeoff performance versus complexity, so long as the mixing conditions are not too dispersive. Simulations on real-world acoustical data show that the proposed algorithm performs multichannel blind deconvolution on i.i.d. sources and convolutive BSS on speech signal mixtures.

II. Causal Natural Gradient Algorithms for Single-Channel Systems

A. Problem

In this section, we identify the problems associated with inadequate signal windowing and filter truncation that are present in the procedure in (2), (5), and (6). In this case, the issues surrounding the design of FIR approximations to the doubly-infinite IIR adaptive system on which this algorithm is based do not depend on the number of signal channels. Hence, we shall for simplicity limit our focus in this section to a single-input, single-output blind deconvolution task, in which $m = n = 1$ in the signal model of (1). In later sections, we shall expand our scope to include the more-practical multichannel case.

For a single-channel blind deconvolution task, the natural gradient algorithm in the time domain as specified in [4] is given by the equations

$$y(k) = \sum_{l=0}^{L} w_l(k)x(k-l) \tag{7}$$

$$u(k) = \sum_{q=0}^{L} w_{L-q}(k)y(k-q) \tag{8}$$

$$w_l(k+1) = w_l(k) + \mu[w_l(k) - f(y(k-L))u(k-l)]. \tag{9}$$

This algorithm is designed to iteratively minimize the cost function

$$\mathcal{J}(W_k(z)) = -E\{\log \hat{p}(y(k))\} - \frac{1}{2\pi j}\oint \log|W_k(z)|z^{-1}\,dz \tag{10}$$

where

$$W_k(z) = \sum_{l=0}^{L} w_l(k)z^{-l} \tag{11}$$

is the z-transform of the adaptive equalizer's impulse response, $E\{\cdot\}$ denotes statistical expectation, and $\hat{p}(y)$ is a model of the p.d.f. of the source to be deconvolved. It can be shown that (10) is, up to a constant independent of the equalizer, proportional to the mutual information of the output signal sequence $\{y(k)\}$ when $\hat{p}(y)$ is the p.d.f. of the source sequence [18]. Minimizing this measure results in a sample sequence that is most independent from sample to sample. When $x(k)$ is a linearly-filtered version of an i.i.d source sequence $\{s(k)\}$, minimizing (10) results in deconvolution of the filtered source sequence.

The natural gradient procedure used in (7)–(9) to approximately minimize (10) is a filtered-gradient one, in which an L-sample delay is introduced to make the updating relations causal. It is useful to determine the form of the standard gradient algorithm that minimizes (10) for comparison. The gradient of the cost function $\mathcal{J}(W_k(z))$ is straightforward to calculate assuming that $W_k(z)$ has no zeros on the unit circle; this gradient is

$$\frac{\partial \mathcal{J}(W_k(z))}{\partial w_l(k)} = E\{f(y(k))x(k-l)\} - \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{W_k(e^{-j\omega})}e^{j\omega l}d\omega \tag{12}$$

where $f(y) = -\partial \log \hat{p}(y)/\partial y$ and we have used the substitution $z = e^{j\omega}$ to transform the contour integral on the right-hand side of (10) into a Fourier integral before taking derivatives of this term with respect to $w_l(k)$. Standard steepest descent minimization of (10) would adjust the sequence $\{w_l(k)\}$ as

$$w_l(k+1) = w_l(k) - \mu \frac{\partial \mathcal{J}(W_k(z))}{\partial w_l(k)} \tag{13}$$

where μ is the algorithm step size. Using the stochastic gradient approximation where expectations are replaced by instantaneous values, and defining the quantities

$$\mathbf{w}(k) = [w_0(k) w_1(k) \quad \cdots \quad w_L(k)] \tag{14}$$
$$\mathbf{x}(k) = [x(k) x(k-1) \quad \cdots \quad x(k-L)]^T \tag{15}$$
$$\tilde{\mathbf{w}}_{\text{inv}}(k) = [w_{L,\text{inv}}(k) w_{L-1,\text{inv}}(k) \quad \cdots \quad w_{0,\text{inv}}(k)] \tag{16}$$
$$w_{l,\text{inv}}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{W_k(e^{-j\omega})} e^{j\omega(L-l)} d\omega \tag{17}$$

we obtain the standard stochastic gradient minimization procedure as

$$y(k) = \mathbf{w}(k)\mathbf{x}(k) \tag{18}$$
$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu[\tilde{\mathbf{w}}_{\text{inv}}(k) - f(y(k))\mathbf{x}^T(k)]. \tag{19}$$

Note that $\tilde{\mathbf{w}}_{\text{inv}}(k)$ contains the coefficients of an estimate of the inverse of the separation system in time-reversed order.

To better see the connection between the coefficient updates in (9) and the standard gradient procedure in (19), we shall write (9) in its delayless and noncausal form [4], such that

$$w_l(k+1) = w_l(k) + \mu[w_l(k) - f(y(k))\bar{u}_k(k-l)] \tag{20}$$

where

$$\bar{u}_j(k) = \sum_{q=0}^{L} \sum_{p=0}^{L} w_q(j) w_p(j) x(k+q-p). \tag{21}$$

Define the coefficient autocorrelation function $r_l(j)$ as

$$r_l(j) = \sum_{p=0}^{L-|l|} w_p(j) w_{p+|l|}(j), \quad -L \le l \le L. \tag{22}$$

Then, it is straightforward to show that

$$\bar{u}_j(k) = \sum_{p=-L}^{L} r_p(j) x(k-p). \tag{23}$$

Thus, the update in (20) can be written as

$$w_l(k+1) = w_l(k) + \mu \left[w_l(k) - f(y(k)) \sum_{p=-L}^{L} r_p(k) x(k-p-l) \right]. \tag{24}$$

This update can be written in vector form by defining (25) and (26), as shown at the bottom of the page. Then, (24) becomes

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu[\mathbf{w}(k) - f(y(k))\bar{\mathbf{x}}^T(k)\bar{\mathbf{R}}^T(k)]. \tag{27}$$

Comparing (27) with (10) and (19), we see that the update in (27) depends on signal values that are *not* within the standard gradient-based procedure. Moreover, since the cost function depends only on the signal elements within $\mathbf{x}(k)$, any signal values outside of $\{x(k), x(k-1), \ldots, x(k-L)\}$ used in the coefficient updates are problematic. Introducing such terms could change the gradient search direction for the procedure and ultimately bias the solution obtained by the procedure in steady-state. These arguments are difficult to prove theoretically given the complexity of the cost function in (10). Later, we shall illustrate the potential problems of these terms through a simple numerical example.

B. Proposed Solution

Because the problematic terms in the coefficient updates are additive and easy to identify—they depend on input signal values other than $\{x(k), \ldots, x(k-L)\}$—it is relatively straightforward to modify the algorithm in (27) to remove these additive terms. Such a modification results in a natural gradient procedure that is different from that in [4] with potentially-different convergence properties. Define the coefficient autocorrelation matrix $\mathbf{R}(k)$ as

$$\mathbf{R}(k) = \begin{bmatrix} r_k(0) & r_k(1) & \cdots & r_k(L) \\ r_k(-1) & r_k(0) & \cdots & r_k(L-1) \\ \vdots & & \ddots & \vdots \\ r_k(-L) & r_k(-L+1) & \cdots & r_k(0) \end{bmatrix}. \tag{28}$$

Unlike $\bar{\mathbf{R}}(k)$ in (26), the matrix $\mathbf{R}(k)$ is symmetric, and it is guaranteed to be positive definite because $r_l(k)$ is from a valid FIR autocorrelation function. Define the vector

$$\mathbf{z}(k) = \mathbf{R}(k)\mathbf{x}(k). \tag{29}$$

$$\bar{\mathbf{x}}(k) = [x(k+L) \quad x(k+L-1) \quad \cdots \quad x(k) \quad x(k-1) \quad \cdots \quad x(k-L) \quad \cdots \quad x(k-2L)]^T \tag{25}$$

$$\bar{\mathbf{R}}(k) = \begin{bmatrix} r_k(-L) & r_k(-L+1) & \cdots & r_k(0) & r_k(1) & \cdots & r_k(L) & 0 & \cdots & 0 \\ 0 & r_k(-L) & \cdots & r_k(-1) & r_k(0) & \cdots & r_k(L-1) & r_k(L) & \cdots & 0 \\ \vdots & & \ddots & & & & & & \ddots & \vdots \\ 0 & 0 & \cdots & r_k(-L) & r_k(-L+1) & \cdots & r_k(0) & r_k(1) & \cdots & r_k(L) \end{bmatrix}. \tag{26}$$

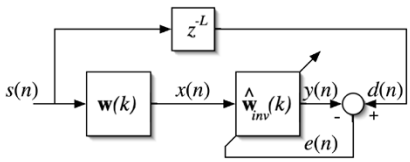

Fig. 4. Inverse system identification task.

Then, the proposed algorithm update in vector form is

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu[\mathbf{w}(k) - f(y(k))\mathbf{z}^T(k)]. \quad (30)$$

We can make several comments regarding the proposed algorithm in (30).

1) If one applies the natural gradient modification originally derived in [4] to the coefficient updates in (19), one obtains the procedure given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu[\tilde{\mathbf{w}}_{\rm inv}(k) - f(y(k))\mathbf{x}^T(k)]\,\mathbf{R}(k). \quad (31)$$

Comparing (30) with (31), we see that these two updates differ in that $\tilde{\mathbf{w}}_{\rm inv}(k)\mathbf{R}(k) \neq \mathbf{w}(k)$ when $\tilde{\mathbf{w}}_{\rm inv}(k)$ is as defined in (16)–(17). Equivalently, since $\mathbf{R}(k)$ is symmetric and Toeplitz, we can state that $\mathbf{w}_{\rm inv}(k) = [w_{\rm inv,0}(k)\ldots w_{\rm inv,L}(k)] \neq \tilde{\mathbf{w}}(k)\mathbf{R}^{-1}(k)$, where $\tilde{\mathbf{w}}(k) = [w_L(k)\ldots w_0(k)]$ is the time-reversed coefficient vector. What then is the meaning of $\hat{\mathbf{w}}_{\rm inv}(k) = \tilde{\mathbf{w}}(k)\mathbf{R}^{-1}(k)$?

As described in [19], the coefficient vector $\hat{\mathbf{w}}_{\rm inv}(k)$ is the mean-square error solution to the inverse system identification task shown in Fig. 4, where $s(n)$ is a stationary sequence of uncorrelated random variables with zero mean and finite variance, $e(n)$ is the estimation error, and n is the time index. Therefore, the algorithm in (30) is a natural gradient procedure in which a time-reversed version of $\hat{\mathbf{w}}_{\rm inv}(k)$ replaces $\tilde{\mathbf{w}}_{\rm inv}(k)$ in the first term within the coefficient updates of (31).

2) The term $\mathbf{z}(k) = \mathbf{R}(k)\mathbf{x}(k)$ that appears with the coefficient updates is consistent with the derivation of the original natural gradient blind deconvolution algorithm in [4]. The difference is in the way truncation is used within the derivation. In [4], the doubly-infinite input sequence $\{\ldots, x(k+1), x(k), x(k-1), \ldots\}$ is filtered by the system $W_k(z)W_k(z^{-1})$, after which it is truncated to finite length to obtain $\{u(k), u(k-1), \ldots, u(k-L)\}$ for the coefficient updates. In (30), the input sequence is truncated to a finite $(L+1)$-sample length, filtered by the system $W_k(z)W_k(z^{-1})$, and finally truncated to finite length again to obtain $\{z_0(k), z_1(k), \ldots, z_L(k)\}$ in $\mathbf{z}(k)$ for the coefficient updates. This extra truncation step guarantees that the coefficient updates depend only on the input signal samples that appear in the cost function of (10).
3) The truncation issue that motivates the use of (30) in place of (27) for inverse filtering tasks is different than the windowing issues facing designers of frequency-domain single- and multichannel blind deconvolution and source separation procedures [5], [9]. For these other methods, signal windowing is employed to reduce or eliminate errors caused by replacing linear convolutions with FFT-based circular convolutions and/or to partially mitigate the permutation ambiguities associated with the frequency-bin-by-frequency-bin processing of the signals when decoupled separation procedures for each input signal frequency are used. See [20] for a discussion of how signal windowing affects the performance of one such frequency-domain convolutive blind source separation method.
4) The proposed method is causal in its operation. Hence, delay need not be introduced into the algorithm updates. It is known that introducing delay into stochastic gradient update terms generally reduces their performance, e.g., by slowing their convergence speeds, limiting the range of stable step sizes, and the like. We can expect that the proposed method will achieve a more-accurate steady-state solution than the method in [4] for identical step sizes, filter lengths, and numbers of iterations. Simulations appear to indicate this fact as well.

C. Efficient Implementation

The main drawback of the proposed method is its computational complexity. It requires forming the matrix $\mathbf{R}(k)$ from $\mathbf{w}(k)$ by calculating the autocorrelation function of the equalizer and then multiplying $\mathbf{x}(k)$ by $\mathbf{R}(k)$. Calculating $\mathbf{R}(k)$ requires approximately $(L+1)(L+2)/2$ multiply/adds, whereas multiplying $\mathbf{x}(k)$ by $\mathbf{R}(k)$ requires $(L+1)^2$ multiply/adds. We would prefer a procedure whose computational complexity in numbers of multiply/adds is proportional to the equalizer length. In what follows, we develop suitable modifications to our proposed approach to obtain this order of complexity. Such approximations are similar to those that were used to reduce the complexity of the original natural gradient procedure in (20) and (21) to one that is proportional to the equalizer length.

In most deconvolution and equalization tasks, the nonquadratic nature of the cost function limits the range of step sizes that can be used to adjust the equalizer coefficients. As such, convergence is not very fast, and the coefficients do not change much from one time instant to the next. Based on this fact, we propose to update $\mathbf{R}(k)$ at every $(L+1)$ time instants as opposed to every time instant. Thus, when n is a positive integer, we set

$$\mathbf{R}(n(L+1)) = \mathbf{R}(n(L+1)-1) = \cdots \cdots = \mathbf{R}((n-1)L+1). \quad (32)$$

In such a scheme, the per-sample computational load of calculating $\mathbf{R}(k)$ is reduced to approximately $(L/2) + 3/2 + 1/L$ multiply/adds at each time instant.

To develop a procedure for updating $\mathbf{z}(k)$, assume for the moment that $\mathbf{R}(k)$ does not change with time, such that $\mathbf{R}(k) = \mathbf{R}$ has elements $r_l, -L \leq l \leq L$. Define a $(2L+1)$-element vector $\mathbf{t}(k)$ as

$$\mathbf{t}(k) = [t_0(k)t_1(k) \ \cdots \ t_{2L-1}(k) \ t_{2L}(k)]^T \quad (33)$$

where

$$t_p(k) = \sum_{j=0}^{\min\{p,L\}} r_{j+L-p}x(k-j). \tag{34}$$

Clearly, $t_{p+L}(k) = z_p(k)$ for $0 \leq p \leq L$ when $r_p(k)$ does not change with time. The vector $\mathbf{t}(k)$ contains the convolution of the $(2L+1)$-element sequence $\{r_k\}$ with the sequence $\{x(k), x(k-1), \ldots, x(k-L)\}$ which has been padded by L zeros on the right. The vector $\mathbf{t}(k)$ is quite similar to $\mathbf{t}(k-1)$ and only differs from it through the addition of terms that depend on $x(k)$ and the subtraction of terms that depend on $x(k-L-1)$. It can be shown that

$$t_p(k) = \begin{cases} x(k)r_L, & \text{if } p = 0 \\ t_{p-1}(k-1) + x(k)r_{L-p}, & \text{if } 1 \leq p \leq L \\ t_{p-1}(k-1) + x(k)r_{L-p} & \\ \quad -x(k-L-1)r_{2L+1-p}, & \text{if } L+1 \leq p \leq 2L. \end{cases} \tag{35}$$

The update in (35) requires $3L+1$ multiply/adds at each time instant to implement, which is much fewer than the $(L+1)^2$ multiply/adds needed to implement the product $\mathbf{R}\mathbf{x}(k)$.

We now show how to combine the above two approximations to obtain a numerically-stable implementation. Since (35) assumes that the autocorrelation sequence r_p is fixed, letting $r_p = r_p(k)$ will introduce errors into these sliding-window calculations, such that the last $L+1$ elements of $\mathbf{t}(k)$ will no longer be accurate. We could use a restart procedure to zero-out the errors every $L+1$ samples, but there is in fact a more ingenious solution. We propose to synchronize the calculation of the r_p sequence with the updating of the $t_p(k)$ values. Specifically, we propose to use $\hat{z}_p(k)$ in place of $z_p(k)$ in (30), such that the algorithm becomes, as shown in (36)–(38) at the bottom of the page. and $\mathbf{R}(k)$ satisfies the block constraint in (32). Notice that the last term on the last L elements of $\mathbf{t}(k)$ depends on elements within $\mathbf{R}(k-L-1)$. It can be shown that this procedure produces $\hat{\mathbf{z}}(k) = \mathbf{R}(k-L-1)\mathbf{x}(k)$ *exactly* whenever $k = n(L+1)$. Thus, the error associated with using both $\mathbf{R}(k)$ and $\mathbf{R}(k-L-1)$ to update $\hat{t}_p(k)$ is "zeroed-out" every $(L+1)$ samples. For $k \neq n(L+1)$, the last $(L+1)$ elements of $\mathbf{t}(k)$ do not match $\mathbf{R}(k-L-1)\mathbf{x}(k)$ or $\mathbf{z}(k)$ exactly, but the differences between these values is of $\mathcal{O}(\mu)$. Thus, they have a negligible effect on the overall performance of the scheme.

Equations (7) and (36)–(38) define the final form of the simplified blind deconvolution algorithm, where the autocorrelation sequence $r_p(k)$ is updated every $L+1$ time instants. The overall complexity of this approach on a per-sample basis is $6.5L + 5/2 + 1/L$ multiply/adds. Since the original procedure in (7)–(9) uses $4L+1$ multiply/adds, the new approach uses approximately 63% more multiply/adds than the original approach.

D. Illustrative Numerical Simulations

The algorithm we have derived in the single-channel case involves some claims as to its performance, namely the following.

- The proposed algorithm is purported to have less bias in its converged solution than that produced by the original algorithm in (7)–(9).
- The proposed algorithm is purported to perform better than the original algorithm when equalizer truncation is an issue.
- The simplified update in (36)–(38) is purported to perform similarly to the more-complicated update in (29) and (30) on which it is based.

It is challenging to justify these claims theoretically, because a full statistical analysis of the algorithm's convergence behavior is difficult to obtain. Instead, we investigate the behaviors of these approaches through numerical simulations. The results observed in these simple single-channel examples will serve to motivate an extension and use of the algorithm in the multichannel case in later sections. In these simulations, $\{s(k)\}$ is generated as a pseudo-random sequence of i.i.d. samples uniformly-distributed in the range $[-1, 1]$. In each case, we compute the ensemble average of the inter-symbol interference (ISI) defined as

$$\text{ISI}(k) = \frac{\sum_{l=0}^{M} c_l^2(k)}{\max_{0 \leq j \leq M} c_j^2(k)} - 1 \tag{39}$$

where $c_l(k)$ is the convolution of $w_l(k)$ and the channel impulse response. The parameters and values chosen for all algorithms are $f(y) = y^3, M = 20$, and $\mu = 0.0001$. We compare the performances of the four following algorithms:

1) preliminary approach in (30) with $L = 4$;
2) proposed final algorithm update in (36) with $L = 4$;
3) original natural gradient update in (9) with $L = 4$;
4) original natural gradient update in (9) with $L = 7$ corresponding to an update whose complexity is similar to that of (36) with $L = 4$.

For each equalizer, "center-spike" initialization was used, such that $w_l(0) = \delta_{l-2}$ or $w_l(0) = \delta_{l-4}$ depending on the equalizer filter length. One hundred simulations have been run and the results averaged in each case. The first example we explore involves a simple autoregressive channel model of the form

$$x(k) = -0.75x(k-1) + s(k). \tag{40}$$

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu[\mathbf{w}(k) - f(y(k))\hat{\mathbf{z}}^T(k)] \tag{36}$$

$$\hat{z}_p(k) = \hat{t}_{L+p}(k) \tag{37}$$

$$\hat{t}_p(k) = \begin{cases} x(k)r_L(k), & \text{if } p = 0 \\ \hat{t}_{p-1}(k-1) + x(k)r_{L-p}(k), & \text{if } 1 \leq p \leq L \\ \hat{t}_{p-1}(k-1) + x(k)r_{L-p}(k) - x(k-L-1)r_{2L+1-p}(k-L-1), & \text{if } L+1 \leq p \leq 2L \end{cases} \tag{38}$$

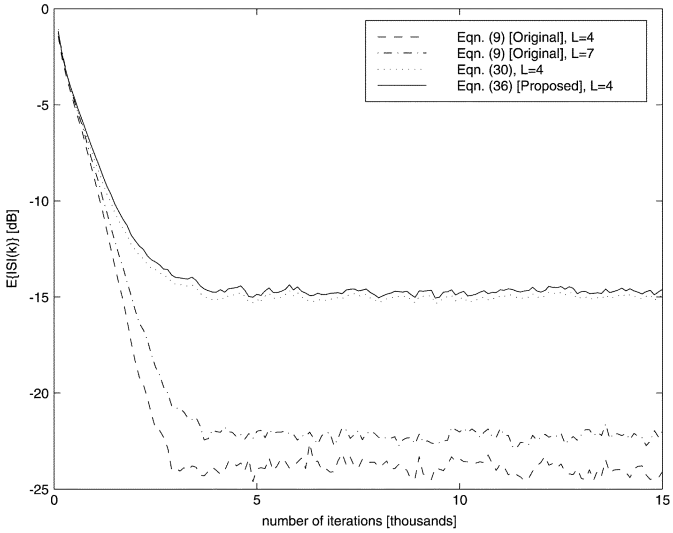

Fig. 5. Evolution of the ensemble-averaged inter-symbol interferences for a simple minimum-phase autoregressive channel model and uniformly-distributed source signals.

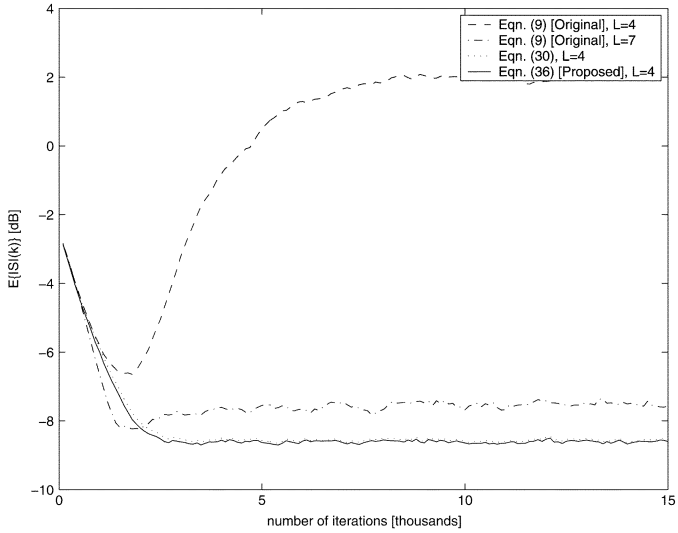

Fig. 6. Evolution of the ensemble-averaged intersymbol interferences for a simple maximum-phase FIR channel model and uniformly-distributed source signals.

This autoregressive channel can be exactly equalized using an FIR filter of length $L \geq 2$ having two consecutive nonzero taps equal to $\{1, 0.75\}$. Shown in Fig. 5 are the evolutions of $E\{\mathrm{ISI}(k)\}$ for the four different algorithm scenarios. As can be seen, the original algorithm has the best performance, achieving a steady-state ISI of approximately -24 dB and -22 dB for $L = 4$ and $L = 7$, respectively. The two new methods do not perform as well as the original approach, which is to be expected given that an FIR equalizer is adequate for this deconvolution task. Signal windowing and filter truncation is unlikely to improve the performance of the original algorithm, which already works quite well in this parsimonious case.

Our second example involves an FIR channel model of the form

$$x(k) = 0.75s(k) + s(k-1). \tag{41}$$

This channel is maximum phase, meaning that an infinitely-noncausal equalizer is required to perfectly equalize it. Any FIR equalizer for this task will exhibit a nonzero residual ISI. Shown in Fig. 6 are the evolutions of $E\{\mathrm{ISI}(k)\}$ for the four scenarios. In this case, the behaviors are markedly different. Both new algorithms converge to an ISI level of about -8.6 dB in approximately 3000 iterations. In contrast, the original method with $L = 4$ initially converges like the new algorithms, but it then diverges to a steady-state ISI of about 2 dB. This divergence is slow and deliberate, suggesting a systematic bias in the coefficient updates. We suspect that the culprit is the addition of the input signal terms outside the interval $\{x(k), \ldots, x(k-L)\}$ within the original algorithm's coefficient updates. The performance of the original method with $L = 7$ is better, since the equalizer is 60% longer; however, the proposed method still outperforms this procedure by about 1 dB.

While these simulations are single-channel ones involving synthetic signals, they support the claims made regarding the behavior of the proposed method. It is expected that the proposed algorithm's robust convergence behavior in situations where a channel inverse is impossible will translate to improved separation performance in multichannel acoustic source separation scenarios.

III. Causal Natural Gradient Algorithms for Multichannel Systems

A. Multichannel Extension

We now extend the approaches described in the previous section to the multichannel case. These extensions are straightforward so long as careful notational similarities are maintained. Define the $m(L+1)$-element input signal vector

$$\underline{\mathbf{x}}(k) = [\mathbf{x}^T(k)\mathbf{x}^T(k-1) \ \cdots \ \mathbf{x}^T(k-L)]^T \tag{42}$$

where $\mathbf{x}(k) = [x_1(k) \ldots x_m(k)]^T$ is the m-element vector containing measurements of the mixtures from all m sensors at time k. Similarly, define the $(m \times m(L+1))$ coefficient matrix

$$\underline{\mathbf{W}}(k) = [\mathbf{W}_0(k) \ \mathbf{W}_1(k) \ \cdots \ \mathbf{W}_L(k)] \tag{43}$$

where $\mathbf{W}_p(k)$ is the $(m \times m)$ tap matrix of the multichannel FIR filter at the pth lag value. Then, we can represent the m output signals at time k using the vector notation

$$\mathbf{y}(k) = \sum_{l=0}^{L} \mathbf{W}_l(k)\mathbf{x}(k-l) \tag{44}$$

$$= \underline{\mathbf{W}}(k)\underline{\mathbf{x}}(k). \tag{45}$$

We first give the multichannel extension of the updates in (29)–(30). Define the m-element vector $\mathbf{z}_l(k)$ and $(m \times m)$ matrix $\mathbf{R}_p(k)$ as

$$\mathbf{z}_l(k) = \sum_{j=0}^{L} \mathbf{R}_{j-l}(k)\mathbf{x}(k-j) \tag{46}$$

$$\mathbf{R}_l(k) = \begin{cases} \displaystyle\sum_{p=0}^{L-l} \mathbf{W}_p^T(k)\mathbf{W}_{p+l}(k), & \text{if } 0 \leq l \leq L \\ \mathbf{R}_{-l}^T(k), & \text{if } -L \leq l \leq -1. \end{cases} \tag{47}$$

These quantities are analogous to $z_l(k)$ and $r_l(k)$ in the single-channel case. Define the $(m(L+1) \times m(L+1))$ matrix $\underline{\mathbf{R}}(k)$ and the $m(L+1)$-element vector $\underline{\mathbf{z}}(k)$ as

$$\underline{\mathbf{R}}(k) = \begin{bmatrix} \mathbf{R}_0(k) & \mathbf{R}_1(k) & \cdots & \mathbf{R}_L(k) \\ \mathbf{R}_{-1}(k) & \mathbf{R}_0(k) & & \mathbf{R}_{L-1}(k) \\ \vdots & & & \vdots \\ \mathbf{R}_{-L}(k) & \cdots & \cdots & \mathbf{R}_0(k) \end{bmatrix} \tag{48}$$

$$\underline{\mathbf{z}}(k) = \left[\mathbf{z}_0^T(k) \cdots \mathbf{z}_L^T(k)\right]^T \tag{49}$$

respectively. Then, it can be shown that

$$\underline{\mathbf{z}}(k) = \underline{\mathbf{R}}(k)\underline{\mathbf{x}}(k). \tag{50}$$

The coefficient update for the multichannel extension of (30) is given by

$$\underline{\mathbf{W}}(k+1) = \underline{\mathbf{W}}(k) + \mu[\underline{\mathbf{W}}(k) - \mathbf{f}(\mathbf{y}(k))\underline{\mathbf{z}}^T(k)] \tag{51}$$

where $\mathbf{f}(\mathbf{y}(k)) = [f_1(y_1(k)) \cdots f_m(y_m(k))]^T$.

The calculation of $\underline{\mathbf{z}}(k)$ in (50) involves $m^2(L+1)^2$ multiply/adds, making it prohibitive for large filter lengths. As such, we look to extend the simplified approach in (36)–(38) to the multichannel case as well. This approach make use of the block Toeplitz structure of $\underline{\mathbf{R}}(k)$ to recursively update the vector

$$\hat{\underline{\mathbf{t}}}(k) = \left[\hat{\mathbf{t}}_0^T(k)\ \ \hat{\mathbf{t}}_1^T(k)\ \ \cdots\ \ \hat{\mathbf{t}}_{2L-1}^T(k)\ \ \hat{\mathbf{t}}_{2L}^T(k)\right]^T \tag{52}$$

whose last $m(L+1)$ entries contained in the vector $\hat{\underline{\mathbf{z}}}(k)$ are approximately equal to $\underline{\mathbf{z}}(k)$. The recursive update for the vector elements of $\hat{\underline{\mathbf{t}}}(k)$ is shown in (53)–(54) at the bottom of the page. The sequence $\{\mathbf{R}_l(k)\}, 0 \le l \le L$ is only updated at the time instants $k = n(L+1)$, where n is an integer. Then, the coefficients are updated as

$$\underline{\mathbf{W}}(k+1) = \underline{\mathbf{W}}(k) + \mu[\underline{\mathbf{W}}(k) - \mathbf{f}(\mathbf{y}(k))\hat{\underline{\mathbf{z}}}^T(k)]. \tag{55}$$

Equations (45) and (53)–(55) define the coefficient updates of the proposed multichannel blind deconvolution and source separation procedure employing causal filters. The complexity of the algorithm is approximately 6.5 multiply/adds per adaptive filter coefficient per time instant, making it approximately 63% more complex than the original multichannel algorithm in (2), (5), and (6).

B. Fast Implementation

In some multichannel blind deconvolution or convolutive BSS tasks, a larger separation filter length value L may be required for adequate performance. In such cases, the complexity of the sample-by-sample iterative update in (45) and (53)–(55) may become prohibitive. In this subsection, we present a fast block-based implementation of this procedure, in which (55) is replaced by the $(L+1)$-sample block update

$$\underline{\mathbf{W}}(k+L+1) = \underline{\mathbf{W}}(k) + \mu\left[(L+1)\underline{\mathbf{W}}(k) - \sum_{l=0}^{L} \mathbf{f}(\mathbf{y}(k+l))\hat{\underline{\mathbf{z}}}^T(k+l)\right]. \tag{56}$$

This implementation assumes the computationally-efficient block size choice of $N = L+1$, although the extension of the algorithm to other block sizes is straightforward. For this implementation, define the time-reversed row signal vectors

$$\tilde{\mathbf{x}}_i(k) = [x_i(k-L)\ldots x_i(k)] \tag{57}$$

$$\tilde{\mathbf{y}}_j(k) = [y_j(k-L)\ldots y_j(k)] \tag{58}$$

$$\tilde{\mathbf{f}}_j(k) = \mu[f_j(y_j(k-L))\ldots f_j(y_j(k))] \tag{59}$$

$$\mathbf{z} = [0 \cdots 0]. \tag{60}$$

Furthermore, let $\mathrm{FFT}(\mathbf{u})$ and $\mathrm{IFFT}(\mathbf{u})$ denote the FFT and inverse FFT, respectively, of a $2(L+1)$-element row vector $\mathbf{u}$, and define $\odot$ as the point-by-point complex multiplication of two vectors $\mathbf{u}$ and $\mathbf{v}$, such that

$$[\mathbf{u} \odot \mathbf{v}]_l = u_l v_l. \tag{61}$$

Then, Table I lists the fast FFT-based implementation of the proposed algorithm. In this algorithm, $\bullet$ corresponds to an $(L+1)$-element vector of "do not care" values to be discarded. The notation used here closely follows that used to describe more-

TABLE I
FAST BLOCK IMPLEMENTATION OF THE PROPOSED ALGORITHM FOR A BLOCK SIZE OF $L+1$ SAMPLES

```
for i = 1 to m do
    X_i(k) = FFT([x̃_i(k)   x̃_i(k+L+1)])
end
for j = 1 to m do
    Y_j(k) = Σ_{i=1}^{m} W_ji(k) ⊙ X_i(k)
    [•   ỹ_j(k)] = IFFT(Y_j(k))
    f̃_j(k) = μ f(ỹ_j(k))
    F_j(k) = FFT([z   f̃_j(k)])
    for i = 1 to m do
        G_ji(k) = F_j(k) ⊙ X_i*(k)
        [g_ji(k)   •] = IFFT(G_ji(k))
        R_ji(k) = Σ_{p=1}^{m} W_jp*(k) ⊙ W_pi(k)
    end
    for i = 1 to m do
        V_ji(k) = Σ_{p=1}^{m} FFT([g_jp(k)   z]) ⊙ R_pi(k)
        [v_ji(k)   •] = IFFT(V_ji(k))
        w_ji(k+L+1) = (1 + μ(L+1)) w_ji(k) − v_ji(k)
        W_ji(k+L+1) = FFT([w_ji(k+L+1)   z])
    end
end
```

$$\hat{\mathbf{z}}_p(k) = \hat{\mathbf{t}}_{L+p}(k) \tag{53}$$

$$\hat{\mathbf{t}}_p(k) = \begin{cases} \mathbf{R}_L^T(k)\mathbf{x}(k) & \text{if } p = 0 \\ \hat{\mathbf{t}}_{p-1}(k-1) + \mathbf{R}_{L-p}^T(k)\mathbf{x}(k) & \text{if } 1 \le p \le L \\ \hat{\mathbf{t}}_{p-1}(k-1) + \mathbf{R}_{p-L}(k)\mathbf{x}(k) - \mathbf{R}_{2L+1-p}^T(k-L-1)\mathbf{x}(k-L-1) & \text{if } L+1 \le p \le 2L \end{cases} \tag{54}$$

traditional block-based adaptive filtering techniques employing fast convolution techniques [22].

The complexity of the proposed algorithm is now considered. For most practical situations in which $L \gg m$, the FFT calculations of a block-based convolutive BSS procedure dominate the system's overall complexity. Our proposed procedure requires $4m^2 + 3m$ FFTs of length $2(L+1)$ for every $(L+1)$ output samples. Thus, the complexity of the algorithm is $\mathcal{O}(m^2 \log_2(L))$ per time sample instant as compared to the $\mathcal{O}(m^2 L)$ complexity of the method in (45) and (53)–(55). For an additional comparison, consider the FDMCBD-I algorithm with 75% overlap described in [23], which is a block-based version of the original time-domain natural gradient convolutive BSS procedure derived in [4]. The FDMCBD-I algorithm requires $2m^2 + 3m$ FFTs of length $4(L+1)$ for every $(L+1)$ output samples. Since an M-element FFT requires $M \log_2 M$ complex-valued operations, the two algorithms have a similar complexity to first order. Considering the number of vector dot-multiplies (*e.g.*, the $\odot$ symbol in Table I), the proposed method has $1.5m^3 + 3.5m^2$ such operations, whereas, the FDMCBD-I algorithm has $4m^2$ such operations. Considering each algorithm as a whole, the proposed method will be similar in complexity to the FDMCBD-I algorithm except for situations in which the number of sensors or sources m is much larger than two and the filter or block length $(L+1)$ is correspondingly small.

For typical step sizes, the algorithm in Table I performs similarly to the procedure in (45) and (53)–(55). Moreover, it appears to be more robust than the FDMCBD-I algorithm in acoustic source separation applications where signal correlations require an overly long filter length for the latter approach to work properly. Simulations illustrating this issue can be found in [21].

C. Illustrative Numerical Simulations

We now explore the behavior of the proposed multichannel blind deconvolution and source separation procedure in Section III-A via numerical simulations. All but the last example in this section use the same two-input, two-output impulse response depicted in Fig. 1. This impulse response was generated from an acoustic laboratory setup consisting of a pair of omnidirectional lapel microphones space 4 cm apart and mounted in a V-configuration approximately 1.5 m from the floor in the center of a 4.45 m x 3.55 m x 2.50 m room. The reverberation time of this room is 130 ms. A pair of loudspeakers located 1.2 m away from the microphones at -40° and $+30^\circ$ from the on-axis direction of the V-configuration microphone array were used as the acoustic sources. Bandlimited white noise played through these loudspeakers was then used to characterize the individual impulse responses of the loudspeaker-to-microphone acoustic paths using standard linear estimation techniques. We then generate the source signal mixtures by filtering recorded signals using these impulse responses digitally. While not exactly identical to real-world acoustic mixtures, these signals allow us to accurately observe and characterize the impulse response distortions due to FIR filter and signal truncation through the combined system impulse responses. These distortions are generally

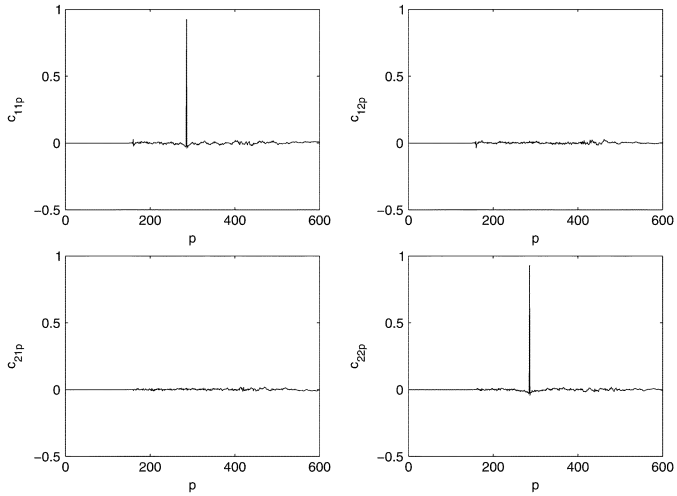

Fig. 7. Combined system impulse responses for the proposed natural gradient multichannel blind deconvolution procedure for i.i.d. binary source signals.

unobservable when looking at the separation system coefficients alone, as shall be indicated. All signals are sampled at 8 kHz.

Our first experiment tests the ability of the algorithms to both separate and deconvolve i.i.d. sources mixed by the acoustic channel. These sources were chosen to be i.i.d. binary-$\{\pm 1\}$ -distributed. The sources were mixed by the acoustic channel to produce the signal sequences $x_1(k)$ and $x_2(k)$. Both the original natural gradient procedure in (5) and the proposed procedure in (55) were applied to this data, where $L = 250, m = 2, \mathbf{W}_l(0) = 0.5\mathbf{I}\delta_{l-125}$, and $\mu = 0.0001$, respectively. Shown in Fig. 2 is the combined impulse responses given by

$$\mathbf{C}_p(k) = \begin{bmatrix} c_{11p}(k) & c_{12p}(k) \\ c_{21p}(k) & c_{22p}(k) \end{bmatrix} = \sum_{l=0}^{L} \mathbf{W}_l(k)\mathbf{A}_{p-l} \qquad (62)$$

at iteration $k = 300\,000$ for the original procedure. As noted in the introduction, the "spikes" present near $p = 0$ and $p = 250$ are troublesome and prevent an adequate equalization result to be achieved. Fig. 3 shows the results of a similar simulation using the original procedure, except that $L = 407$ has been chosen such that the original procedure has a similar complexity to that of the proposed procedure. Again, the "spikes" present near $p = 0$ and $p = 407$ are problematic. Shown in Fig. 7 are the combined impulse responses for the proposed method at iteration $k = 300\,000$ on the same data. As can be seen, the separation and deconvolution performance is nearly perfect, with ideal delta-function responses in $\{c_{11p}\}$ and $\{c_{22p}\}$ and nearly-zero responses in $\{c_{12p}\}$ and $\{c_{21p}\}$, respectively.

The performance difference between these three cases cannot be easily gleaned from the separation system impulse responses. Shown in Fig. 8 are the separation system impulse responses $\{w_{11p}\}$ through $\{w_{22p}\}$ for all three adaptive systems. Although the solutions are different from each other, it is not clear which impulse response set yields the preferable result. We believe that the lack of an observable performance marker in the separation system impulse responses is the reason as to why signal truncation remained an illusive issue for natural gradient multichannel blind deconvolution procedures in the past. Shown in Fig. 9 are

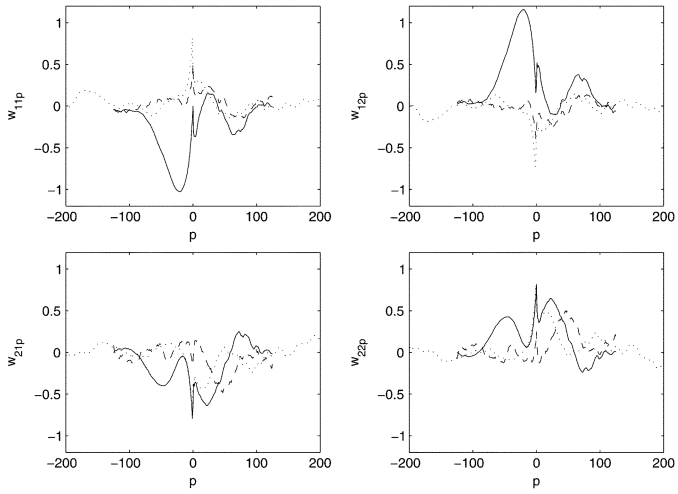

Fig. 8. Estimated separation impulse responses in the multichannel blind deconvolution task: dashed-original method, $L = 250$; dotted-original method, $L = 407$; solid-proposed method.

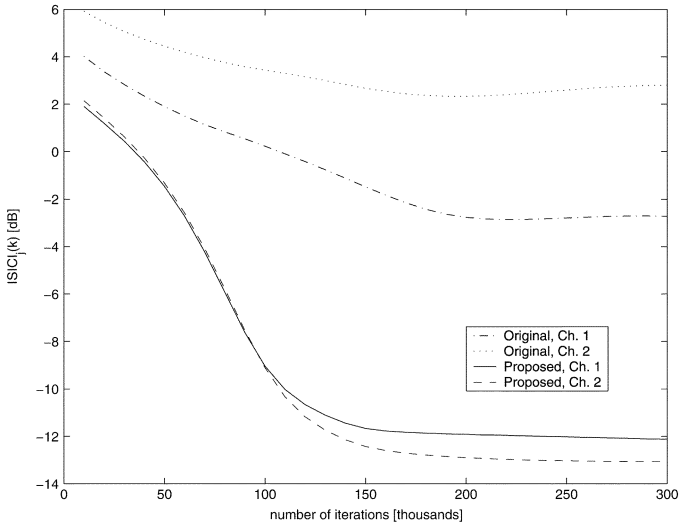

Fig. 9. Evolutions of the joint inter-symbol and inter-channel interferences for the algorithms applied to i.i.d. binary source signals.

the per-channel combined inter-symbol and inter-channel interferences (ISICIs) for the $L = 250$-tap subfilter deconvolution procedures, computed as

$$\text{ISICI}_j(k) = \frac{\sum_{l=0}^{M} c_{j1l}^2(k) + c_{j2l}^2(k)}{\max_{0\le j\le M} c_{jjl}^2(k)} - 1 \qquad (63)$$

for $j \in \{1, 2\}$, respectively, during their respective convergence periods. As can be seen, the original natural gradient approach fail to accurately deconvolve the acoustic channel, whereas, the proposed method is quite effective at reducing the ISICI.

We now turn to examples involving acoustic sources. In this case, we replace the two random uniform sources with two 7-s isolated recordings of a single male speaker from a radio newscast. These signals were repeated six times before being filtered by the acoustic channel to create a pair of signal mixtures with a length of at least 300 000 samples. Both the original and proposed natural gradient algorithms were applied to these signals, where $L = 100, m = 2$, $\mathbf{W}_l(0) = 10\mathbf{I}\delta_{l-50}$, $f_i(y) = \text{sgn}(y)$,

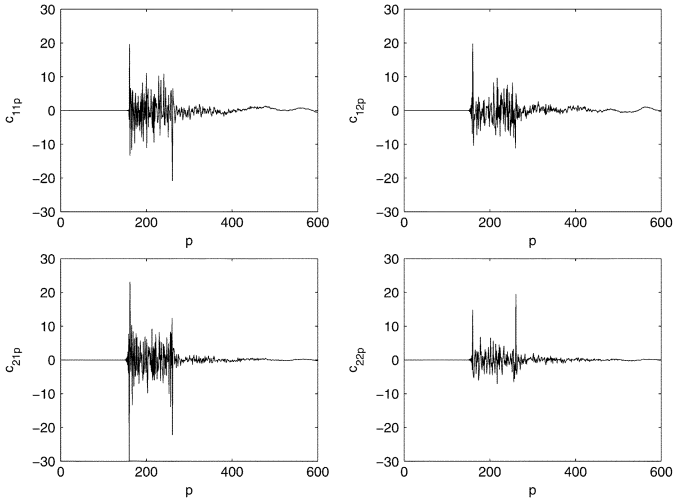

Fig. 10. Combined system impulse responses for the original natural gradient convolutive BSS procedure for speech mixtures, $L = 100$.

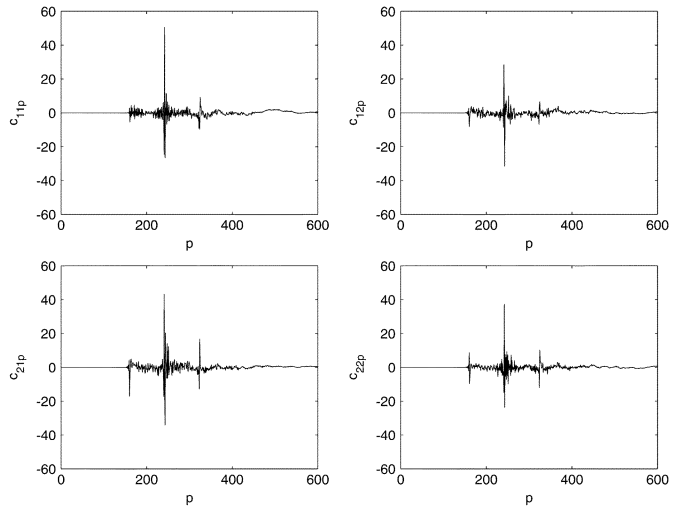

Fig. 11. Combined system impulse responses for the original natural gradient convolutive BSS procedure for speech mixtures, $L = 164$.

and $\mu = 0.0001$, respectively. In this case, we have chosen a shorter separation filter length than in the previous multichannel blind deconvolution task as the system cannot deconvolve the speech signal mixtures in this case. In addition, we applied the original natural gradient algorithm with $L = 164$ and $\mathbf{W}_l(0) = 10\mathbf{I}\delta_{l-82}$, corresponding to a procedure whose complexity is nearly-identical to that of the proposed method with $L = 100$. Shown in Figs. 10–12, are the combined impulse responses for these three methods on this data set. As can be seen, both versions of the original natural gradient algorithm fail to provide any degree of separation with these parameter choices, and large "spikes" can be seen in the $\{c_{ijp}\}$ sequences. In contrast, the proposed method does achieve a reasonable separation solution, as the values of $|c_{12p}|$ and $|c_{21p}|$ are smaller than those of $|c_{11p}|$ and $|c_{22p}|$, respectively. The nonimpulsive shape of c_{11p} and c_{22p} is due to the spectral flattening that all multichannel blind deconvolution procedures impose on the extracted source signals. Using these impulse responses, we compute the overall signal-to-interference ratio (SIR) of Channels 1 and 2 for this coefficient solution and chosen speech signals to be 9.9 and 7.5

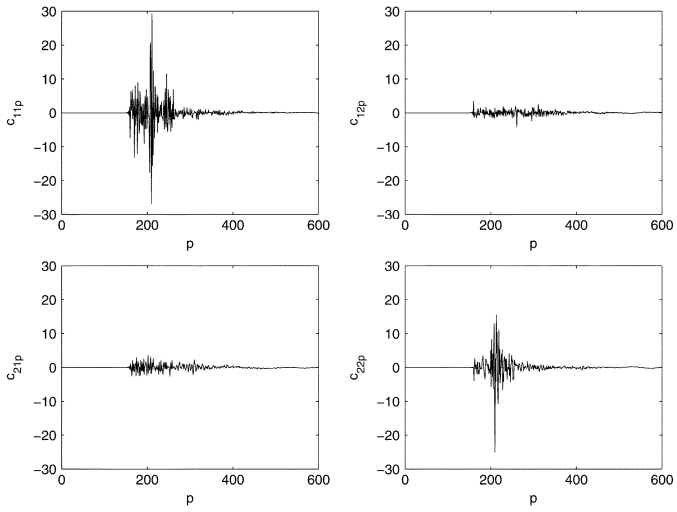

Fig. 12. Combined system impulse responses for the proposed natural gradient convolutive BSS procedure for speech mixtures. $L = 100$.

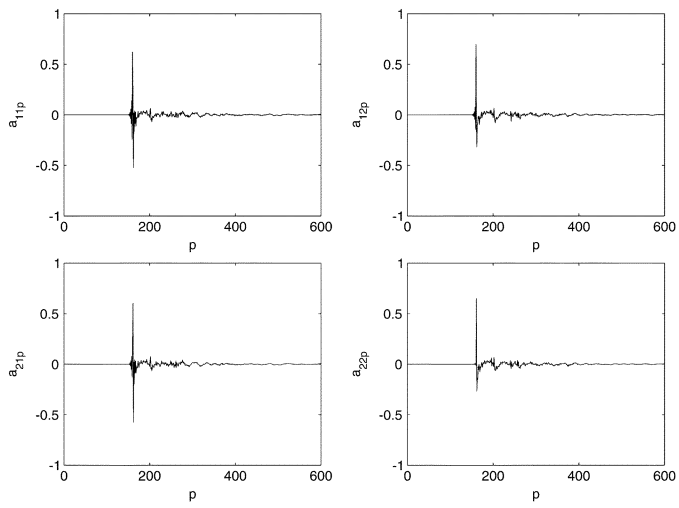

Fig. 13. Impulse responses for the more-channeling acoustic channel in a multichannel simulation experiment.

dB, respectively. The original mixtures had nearly equal SIRs of -0.6 and 0.2 dB, respectively.

Because of its ability to work for shorter filter lengths, the proposed multichannel blind deconvolution procedure has a potential benefit largely unexplored in the convolutive signal separation literature: the ability to select extremely short filters within the separation system. The use of short multichannel FIR filters is quite common for other acoustic signal processing tasks such as adaptive beamforming [24]. To our knowledge, however, they have not been used extensively for convolutive BSS tasks. The capabilities of such short-filter systems are obviously limited to situations where the room reverberation time is short, the microphone array has closely-spaced sensors, and the task is separation as opposed to deconvolution-and-separation. The complexity and memory requirements of the overall scheme, however, are much reduced, making it amenable to implementation on a low-cost, low-memory programmable DSP chip. In addition, the separation performance of the scheme can even be *better* than systems with longer filter lengths in some scenarios. As an example, we applied the proposed separation scheme to the speech mixtures used in the previous example, where filter lengths of $L = 50$ and $L = 24$ were chosen, representing impulse response lengths of 6.25 and 3 ms, respectively, and "center-spike" coefficient initialization was used. Such filter lengths are unheard of in most convolutive BSS approaches, where filters with hundreds or even thousands of taps per channel are often used. After 300 000 iterations, the first system achieved SIRs of 15.5 and 9.5 dB, respectively, and the second system achieved SIRs of 13.8 and 9.8 dB, respectively. These results are extremely promising for practical acoustic source separation in offices and other small-room environments. In contrast, simulations of the original natural gradient procedure on this data for filter lengths of $L = 24, L = 40, L = 50$, and $L = 82$ (not shown) failed to provide any level of separation for this data set.

We now demonstrate the performance of the proposed method for separating a more-challenging source signal mixture in which the source loudspeakers are located at $+20^\circ$ and $+60^\circ$ from the on-axis direction of the V-configuration microphone array in the same acoustic room and speech signals as before. In this case, the initial signal-to-interference ratios for the mixtures are 0.07 dB and -0.85 dB, such that one of the sources is louder than the other source within both mixtures. Shown in Fig. 13 are the impulse responses for this configuration. The proposed algorithm was applied to this data, where $L = 63, \mathbf{W}_l(0) = 10\mathbf{I}\delta_{l-32}$, and $\mu = 0.000\,05$, respectively. Shown in Fig. 14 are the combined impulse responses for the proposed method after convergence. The final signal-to-interference ratios for the separated sources are 11.3 and 6.7 dB, respectively.

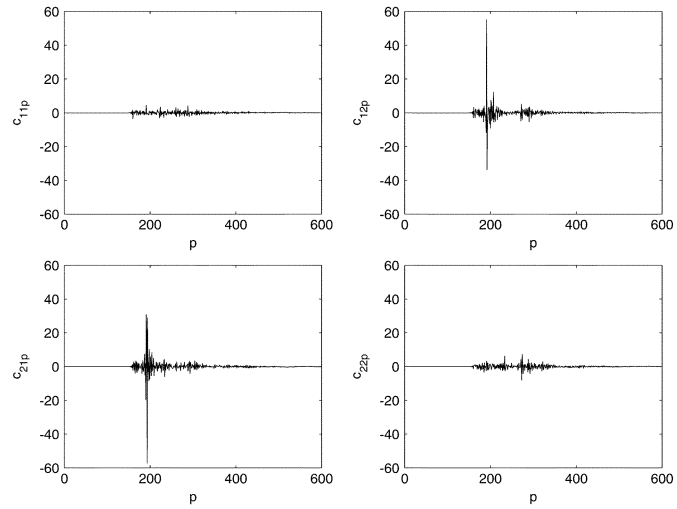

Fig. 14. Combined system impulse responses for the proposed natural gradient convolutive BSS procedure for a more-challenging speech mixture, $L = 63$.

While the reasons for the better performance of the proposed scheme for shorter filter lengths on this data set is not entirely clear at the present time, we conjecture that the careful use of time-domain signal truncation with the algorithm makes the proposed method less sensitive to residual output signal correlation at convergence.

IV. Conclusion

In this paper, we have identified a critical issue associated with a well-known algorithm for multichannel blind deconvolution and convolutive blind source separation tasks. We have demonstrated through argument and simulation that the natural gradient procedure derived in [4] and other variants of this approach [5]–[7] can achieve a biased result when the channel to be inverted is not minimum phase and/or the separation filter length is too short. We then propose a new natural gradient procedure that does not exhibit the detrimental effects of the former approach in these situations. The complexity of the new algorithm, while somewhat greater than the original approach, is still proportional to the number of parameters in the separation system, and it uses only multiplies and adds in its operation. We have demonstrated its accurate capabilities both in multichannel blind deconvolution tasks involving synthetic signals as well as convolutive BSS tasks involving speech signals. Moreover, the algorithm functions in a reasonable manner even when the filter lengths chosen are much shorter than would be required for an accurate channel inverse.

The causality issues that we have identified in the natural gradient algorithm are likely to be similar in other filtered-gradient schemes for equalization, source separation, phase modeling, and paraunitary filter bank design. Causal implementations of these other approaches are the subject of current work.

References

[1] E. Weinstein, M. Feder, and A. V. Oppenheim, "Multi-channel signal separation by decorrelation," *IEEE Trans. Signal Process.*, vol. 4, no. 7, pp. 405–413, Jul. 1993.

[2] S. Van Gerven and D. Van Compernolle, "Signal separation by symmetric adaptive decorrelation: Stability, convergence, and uniqueness," *IEEE Trans. Signal Processing*, vol. 43, no. 7, pp. 1602–1612, Jul. 1995.

[3] R. H. Lambert, "Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures," Ph.D. dissertation, Dept. Elec. Eng., Univ. Southern California, Los Angeles, CA, May 1996.

[4] S. Amari, S. Douglas, A. Cichocki, and H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. 1st IEEE Workshop Signal Processing Advanced Wireless Commun.*, Paris, France, Apr. 1997, pp. 101–104.

[5] R. H. Lambert and A. J. Bell, "Blind separation of multiple speakers in a multipath environment," in *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Processing*, vol. 1, Munich, Germany, Apr. 1997, pp. 423–426.

[6] 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 IFAC Symp. System Identification*, vol. 3, Kitakyushu City, Japan, Jul. 1997, pp. 1057–1062.

[7] T.-W. Lee, A. Ziehe, R. Orglmeister, and T. J. Sejnowski, "Combining time-delayed decorrelation and ICA: Toward solving the cocktail party problem," in *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Processing*, Seattle, WA, May 1998, pp. 1089–1092.

[8] L. Parra, C. Spence, and B. De Vries, "Convolutive blind source separation based on multiple decorrelation," in *Proc. IEEE Workshop Neural Networks Signal Processing*, Cambridge, MA, Sep. 1998, pp. 23–32.

[9] L. Parra and C. Spence, "Convolutive blind separation of nonstationary sources," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 3, pp. 320–327, May 2000.

[10] J. Anemuller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," in *Proc. 2nd Int. Workshop Indep. Compon. Anal. Signal Separation*, Helsinki, Finland, Jun. 2000, pp. 215–220.

[11] X. Sun and S. C. Douglas, "Multichannel blind deconvolution of arbitrary signals: Adaptive algorithms and stability analyzes," in *Proc. 34th Asilomar Conf. Signals, System, Computers*, vol. 2, Pacific Grove, CA, Nov. 2000, pp. 1412–1416.

[12] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *Proc. 3rd Int. Workshop Independent. Compon. Anal. Signal Separation*, San Diego, CA, Dec. 2001, pp. 722–727.

[13] S. C. Douglas, "Blind separation of acoustic signals," in *Microphone Arrays: Techniques and Applications*, M. Brandstein and D. Ward, Eds. New York: Springer-Verlag, 2001, pp. 355–380.

[14] S. C. Douglas and X. Sun, "Convolutive blind separation of speech mixtures using the natural gradient," *Speech Commun.*, vol. 39, pp. 65–78, Dec. 2002.

[15] T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "SIMO-model-based independent component analysis for high-fidelity blind separation of acoustic signals," in *Proc. 4th Int. Symp. Indep. Compon. Anal. Signal Separation*, Nara, Honshu, Japan, Apr. 2003, pp. 993–998.

[16] L. Parra and C. Alvino, "Geometric source separation: Merging convolutve source separation with geometric beamforming," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 352–362, Sep. 2002.

[17] J. F. Claerbout, *Fundamentals of Geophysical Data Processing*. Palo Alto, CA: Blackwell Scientific, 1976.

[18] D.-T. Pham, "Mutual information approach to blind separation of stationary sources," in *Proc. 1st Workshop Indep. Compon. Anal. Signal Separation*, France, Jan. 1999, pp. 215–220.

[19] Z. Gulboy and N. C. Geckinli, "On FIR filters having approximate FIR inverses with a specified LMS error," *Signal Process.*, vol. 19, no. 1, pp. 9–15, Jan. 1990.

[20] M. Z. Ikram and D. R. Morgan, "Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment," in *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Processing*, vol. 2, Istanbul, Turkey, June 2000, pp. 1041–1044.

[21] S. C. Douglas, H. Sawada, and S. Makino, "A causal frequency-domain implementation of a natural gradient multichannel blind deconvolution and source separation algorithm," in *Proc. 18th Int. Congr. Acoustics*, vol. I, Kyoto, Japan, Apr. 2004, pp. 85–88.

[22] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

[23] M. Joho and P. Schniter, "Frequency domain realization of a multichannel blind deconvolution algorithm based on the natural gradient," in *Proc. 4th Int. Symp. Independent. Compon. Anal. Signal Separation*, Nara, Honshu, Japan, Apr. 2003, pp. 543–548.

[24] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, pp. 4–24, Apr. 1988.

[25] S. C. Douglas, H. Sawada, and S. Makino, "Natural gradient multichannel blind deconvolution and source separation using causal FIR filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 5, Montreal, QC, Canada, May 2004, pp. 477–480.

Scott C. Douglas (S'88–M'92–SM'98) received the B.S. (with distinction), the M.S., and the 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 (SMU), Dallas, TX, 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. He is the author or coauthor 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 is associate editor of the *Journal of VLSI Signal Processing Systems*. 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.

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 was associate editor for the IEEE Transactions on Signal Processing and the IEEE Signal Processing Letters. 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. He is a member of both Phi Beta Kappa and Tau Beta Pi.

Hiroshi Sawada (M'02–SM'04) received the B.E., M.E, and Ph.D. degrees in information science from Kyoto University, Kyoto, Japan, in 1991, 1993, and 2001, respectively.

In 1993, he joined NTT Communication Science Laboratories, Kyoto. From 1993 to 2000, he was engaged in research on the computer aided design of digital systems, logic synthesis, and computer architecture. Since 2000, he has been engaged in research on signal processing and blind source separation for convolutive mixtures using independent component analysis.

Dr. Sawada received the 9th TELECOM System Technology Award for Students from the Telecommunications Advancement Foundation in 1994, and the Best Paper Award of the IEEE Circuit and System Society in 2001. He is a member of the IEICE and the ASJ.

Shoji Makino (A'89–M'90–SM'99–F'04) was born in Nikko, Japan, on June 4, 1956. He received the B.E., M.E., and Ph.D. degrees from Tohoku University, Sendai, Japan, in 1979, 1981, and 1993, respectively.

He joined NTT, Kyoto, Japan, in 1981. He is now an Executive Manager at the NTT Communication Science Laboratories. His research interests include blind source separation of convolutive mixtures of speech, acoustic signal processing, and adaptive filtering and its applications such as acoustic echo cancellation. He is the author or coauthor of more than 170 articles in journals and conference proceedings and has been responsible for more than 140 patents.

Dr. Makino received the TELECOM System Technology Award from the Telecommunications Advancement Foundation in 2004, the Best Paper Award of the IWAENC in 2003, the Paper Award of the IEICE in 2002, the Paper Award of the ASJ in 2002, the Achievement Award of the IEICE in 1997, and the Outstanding Technological Development Award of the ASJ in 1995. He is a member of the Conference Board of the IEEE Signal Processing Society and an Associate Editor of the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING. He is a member of the ASJ and the IEICE. He is a member of the Technical Committee on Audio and Electroacoustics, as well as Speech of the IEEE Signal Processing Society. He is a member of the International ICA Steering Committee and the Organizing Chair of the 2003 International Conference on Independent Component Analysis and Blind Signal Separation. He is the General Chair of the 2003 International Workshop on Acoustic Echo and Noise Control. He was a Vice Chair of the Technical Committee on Engineering Acoustics of the IEICE.