NATURAL GRADIENT BLIND DECONVOLUTION AND EQUALIZATION USING CAUSAL FIR FILTERS

Scott C. Douglas

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

Hiroshi Sawada and Shoji Makino

NTT Communication Science Laboratories
NTT Corporation
Kyoto 619-0327 Japan

ABSTRACT

Natural gradient adaptation is an especially convenient method for adapting the coefficients of a linear system in inverse filtering tasks such as blind deconvolution and equalization. Practical implementations of such methods require truncation of the filter impulse responses within the gradient updates. In this paper, we show how truncation of these filter impulse responses can create convergence problems and introduces a bias into the steady-state solution of one such algorithm. We then show how this algorithm can be modified to effectively mitigate these effects for estimating causal FIR approximations to doubly-infinite IIR equalizers. Simulations indicate that the modified algorithm provides the convergence benefits of the natural gradient while still attaining good steady-state performance.

1. INTRODUCTION

The goal of blind deconvolution and equalization is to recover as accurately as possible a desired discrete-time signal sequence $s(k)$ from a linearly-filtered and noisy version of this sequence given by

$$x(k) \;=\; \sum_{l=-\infty}^{\infty} a_l s(k-l) + \nu(k), \qquad (1)$$

where $\{a_l\}$ is the impulse response of the unknown channel and $\nu(k)$ is measurement noise. Typically, this recovery is performed using an adaptive finite-impulse-response (FIR) filter of the form

$$y(k) \;=\; \sum_{l=0}^{L} w_l(k) x(k-l), \qquad (2)$$

where $y(k)$ is the recovered sequence, L is the filter length, and the coefficients $\{w_l(k)\}$ are adapted such that $y(k)$ approaches a delayed and scaled version of $s(k)$. Blind deconvolution and equalization are important for tasks in areas ranging from geophysical exploration to wireless communications.

Recently, a novel blind deconvolution procedure has been developed that is based on a minimum mutual information criterion [1, 2]. This procedure assumes that the source sequence $s(k)$ is both non-Gaussian and independent and identically-distributed (i.i.d.), and it employs a modified natural gradient search procedure [3, 4] to both simplify the coefficient updates and improve convergence performance. The algorithm is given by

$$w_l(k+1) \;=\; w_l(k) + \mu\left[w_l(k) - f(y(k-L))u(k-l)\right] \qquad (3)$$

$$u(k) \;=\; \sum_{q=0}^{L} w_{L-q}(k) y(k-q), \qquad (4)$$

where $f(y)$ is a nonlinearity that depends on the probability density function (p.d.f.) of $s(k)$ and μ is a positive step size. This algorithm requires about four multiply/adds per adaptive filter coefficient, and it has been shown successfully deconvolve source signals without exact knowledge of the p.d.f. of $s(k)$ [5]. The natural gradient procedure on which this update is based can also be extended to a wider class of algorithms, such as Bussgang approaches [6, 7].

The procedure in (2)–(4) was derived in [1] as an approximation to a two-sided infinite-impulse-response (IIR) blind deconvolution procedure, where signal windowing and truncation were used to make the input-output relations and updates causal and of finite complexity. It is not clear, however, how signal windowing and filter truncation affect the convergence performance of the scheme, especially for shorter equalizer filter lengths L that preclude an accurate inverse of the linear measurement model. Recently, similar issues were raised in the design of bin-normalized frequency-domain adaptive filters [8, 9], where it was shown that causality plays an important role in achieving an unbiased and fast-converging procedure. A study of these issues could lead to better procedures for a wide class of filtered-gradient algorithms, including the general class of natural gradient methods in [6, 7].

In this paper, we study the performance effects of signal windowing and filter truncation in natural gradient methods for blind deconvolution and equalization tasks. We show that the windowing approximations used in the derivation of (2)–(4) 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 this natural gradient method for blind deconvolution and equalization that does not suffer from these performance limitations. The proposed algorithm requires about 63% more multiply/adds than the original implementation on a per-sample basis for equivalent filter lengths. Simulations show that the proposed algorithm performs better than the procedure in (2)–(4) for practical situations involving short equalizer lengths.

2. THE PROBLEM

In this section, we identify the issues associated with inadequate signal windowing and filter truncation that are present in the procedure in (2)–(4). This algorithm is designed to iteratively minimize the cost function

$$\mathcal{J}(W_k(z)) \;=\; -E\{\log \widehat{p}(y(k))\} - \frac{1}{2\pi j}\oint \log|W_k(z)| z^{-1} dz, \qquad (5)$$

$$\overline{\mathbf{x}}(k)=[x(k+L)\ x(k+L-1)\ \cdots\ x(k)\ x(k-1)\ \cdots\ x(k-L)\ \cdots\ x(k-2L)]^T \tag{20}$$

$$\overline{\mathbf{R}}(k)=\begin{bmatrix} r_k(-L) & r_k(-L+1) & \cdots & r_k(0) & r_k(1) & \cdots & r_k(L) & 0 & \cdots & 0 \\ 0 & r_k(-L) & \cdots & r_k(-1) & r_k(0) & \cdots & r_k(L-1) & r_k(L) & \cdots & 0 \\ \vdots & & \ddots & & & & & & \ddots & \vdots \\ 0 & 0 & \cdots & r_k(-L) & r_k(-L+1) & \cdots & r_k(0) & r_k(1) & \cdots & r_k(L) \end{bmatrix} \tag{21}$$

---

where

$$W_k(z) \quad = \quad \sum_{l=0}^{L} w_l(k) z^{-l} \tag{6}$$

is the z-transform of the adaptive equalizer's impulse response, $E\{\cdot\}$ denotes statistical expectation, and $\widehat{p}(y)$ is a model of the p.d.f. of the source to be deconvolved. It can be shown that (5) is, up to a constant independent of the equalizer, proportional to the mutual information of the output signal sequence $\{y(k)\}$ when $\widehat{p}(y)$ is the p.d.f. of the source sequence [10]. Minimizing this measure results in a sample sequence that is most independent from sample to sample. When $x(k)$ is a noiseless linearly-filtered version of an i.i.d source sequence $\{s(k)\}$, minimizing (5) results in deconvolution of the filtered source sequence.

The natural gradient procedure used in (2)–(4) to approximately minimize (5) 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 (5) 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

$$\begin{aligned} \frac{\partial \mathcal{J}(W_k(z))}{\partial w_l(k)} \quad &= \quad E\{f(y(k))x(k-l)\} \\ &\quad - \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{W_k(e^{-j\omega})} e^{j\omega l} d\omega, \end{aligned} \tag{7}$$

where $f(y) = -\partial \log \widehat{p}(y)/\partial y$ and we have used the substitution $z = e^{j\omega}$ to transform the contour integral on the right-hand side of (5) into a Fourier integral before taking derivatives of this term with respect to $w_l(k)$. Standard steepest descent minimization of (5) would adjust the sequence $\{w_l(k)\}$ as

$$w_l(k+1) \quad = \quad w_l(k) - \mu \frac{\partial \mathcal{J}(W_k(z))}{\partial w_l(k)}, \tag{8}$$

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

$$\mathbf{w}(k) \quad = \quad [w_0(k)\ w_1(k)\ \cdots\ w_L(k)] \tag{9}$$

$$\mathbf{x}(k) \quad = \quad [x(k)\ x(k-1)\ \cdots\ x(k-L)]^T \tag{10}$$

$$\widetilde{\mathbf{w}}_{inv}(k) \quad = \quad [w_{L,inv}(k)\ w_{L-1,inv}(k)\ \cdots\ w_{0,inv}(k)] \tag{11}$$

$$w_{inv,l}(k) \quad = \quad \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{W_k(e^{-j\omega})} e^{j\omega(L-l)} d\omega, \tag{12}$$

we obtain the standard stochastic gradient minimization procedure

$$y(k) \quad = \quad \mathbf{w}(k)\mathbf{x}(k) \tag{13}$$

$$\mathbf{w}(k+1) \quad = \quad \mathbf{w}(k) + \mu\left[\widetilde{\mathbf{w}}_{inv}(k) - f(y(k))\mathbf{x}^T(k)\right] \tag{14}$$

To better see the connection between the coefficient updates in (3) and the standard gradient procedure in (14), we shall write (3) in its delayless and non-causal form [1], such that

$$w_l(k+1) \quad = \quad w_l(k) + \mu\left[w_l(k) - f(y(k))\overline{u}_k(k-l)\right], \tag{15}$$

where

$$\overline{u}_i(k) \quad = \quad \sum_{q=0}^{L}\sum_{p=0}^{L} w_q(i) w_p(i) x(k+q-p) \tag{16}$$

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

$$r_l(i) \quad = \quad \sum_{p=0}^{L-|l|} w_p(i) w_{p+|l|}(i), \quad -L \leq l \leq L. \tag{17}$$

Then, it is straightforward to show that

$$\overline{u}_i(k) \quad = \quad \sum_{p=-L}^{L} r_p(i) x(k-p). \tag{18}$$

Thus, the update in (15) can be written as

$$\begin{aligned} w_l(k+1) \quad &= \quad w_l(k) + \mu\Big[w_l(k) \\ &\quad - f(y(k)) \sum_{p=-L}^{L} r_p(k) x(k-p-l)\Big]. \end{aligned} \tag{19}$$

This update can be written in vector form by defining $\overline{\mathbf{x}}(k)$ and $\overline{\mathbf{R}}(k)$ as shown at the top of this page. Then, (19) becomes

$$\mathbf{w}(k+1) \;=\; \mathbf{w}(k) + \mu\left[\mathbf{w}(k) - f(y(k))\overline{\mathbf{x}}^T(k)\overline{\mathbf{R}}^T(k)\right]. \tag{22}$$

Comparing (22) with (5) and (14), we make a striking discovery: the update in (22) 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 (5). Later, we shall illustrate the potential problems of these terms through simple numerical examples.

3. A 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 (22) to remove these additive terms.

Such a modification yields a new algorithm with potentially better convergence properties. Define the coefficient autocorrelation matrix $\mathbf{R}(k)$ as

$$\mathbf{R}(k) = \begin{bmatrix} r_k(0) & r_k(1) & \cdots & r_k(L) \\ r_k(-1) & r_k(0) & \cdots & r_k(L-1) \\ \vdots & & \ddots & \vdots \\ r_k(-L) & r_k(-L+1) & \cdots & r_k(0) \end{bmatrix} \quad (23)$$

Unlike $\overline{\mathbf{R}}(k)$ in (21), 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

$$\mathbf{z}(k) = \mathbf{R}(k)\mathbf{x}(k). \quad (24)$$

Then, the proposed algorithm update in vector form is

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\left[\mathbf{w}(k) - f(y(k))\mathbf{z}^T(k)\right] \quad (25)$$

We can make several comments regarding the proposed algorithm in (25):

1. The proposed procedure is similar, but not identical to, the modified stochastic gradient procedure given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\left[\widetilde{\mathbf{w}}_{inv}(k) - f(y(k))\mathbf{x}^T(k)\right]\mathbf{R}(k). \quad (26)$$

The difference is due to the fact that $\widetilde{\mathbf{w}}_{inv}(k)\mathbf{R}(k) \neq \mathbf{w}(k)$ when $\widetilde{\mathbf{w}}_{inv}(k)$ is as defined in (11)–(12). It can be shown that the vector $[w_L(k)\ w_{L-1}(k)\ \cdots\ w_0(k)]\mathbf{R}^{-1}(k)$ is an L-coefficient least-squares estimate of the impulse response of the inverse of the FIR system $W_k(z)$ [11]. Hence, for reasonable filter lengths L, the difference between $\widetilde{\mathbf{w}}_{inv}(k)\mathbf{R}(k)$ and $\mathbf{w}(k)$ will be small in practice.

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 [1]. The difference is in the way truncation is used within the derivation. In [1], 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 (25), the input sequence is truncated to a finite L-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 (5).

3. 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 [1] for identical step sizes, filter lengths, and numbers of iterations. Simulations appear to indicate this fact as well.

4. AN 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 (15)–(16) to one that is proportional to the equalizer length.

In most deconvolution and equalization tasks, the non-quadratic 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 time instants as opposed to every time instant. Thus, when n is an integer, we set

$$\mathbf{R}(nL) = \mathbf{R}(nL-1) = \cdots = \mathbf{R}(nL-L). \quad (27)$$

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

$$\mathbf{t}(k) = [t_0(k)\ t_1(k)\ \cdots\ t_{2L-1}(k)\ t_{2L}(k)]^T \quad (28)$$

where

$$t_p(k) = \sum_{j=0}^{\min\{p,L\}} r_{j+L-p}\,x(k-j). \quad (29)$$

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

$$t_p(k) = \begin{cases} x(k)r_L & \text{if } p=0 \\ t_{p-1}(k-1) + x(k)r_{L-p} & \text{if } 1 \leq p \leq L \\ t_{p-1}(k-1) + x(k)r_{L-p} & \\ \quad -x(k-L-1)r_{2L+1-p} & \text{if } L+1 \leq p \leq 2L. \end{cases} \quad (30)$$

The update in (30) 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 (30) 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 elements of $\mathbf{t}(k)$ will no longer be accurate. We could use a restart procedure to zero-out the errors every L 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 $\widehat{z}_p(k)$ in place of $z_p(k)$ in (25), such that the algorithm becomes

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu[\mathbf{w}(k) - f(y(k))\widehat{\mathbf{z}}^T(k)] \quad (31)$$

$$\widehat{z}_p(k) = \widehat{t}_{L+p}(k) \quad (32)$$

$$\widehat{t}_p(k) = \begin{cases} x(k)r_L(k) & \text{if } p = 0 \\ \widehat{t}_{p-1}(k-1) + x(k)r_{L-p}(k) & \text{if } 1 \leq p \leq L \\ \widehat{t}_{p-1}(k-1) + x(k)r_{L-p}(k) & \\ \quad -x(k-L-1)r_{2L+1-p}(k-L) & \text{if } L < p \leq 2L \end{cases} \quad (33)$$

and $\mathbf{R}(k)$ satisfies the block constraint in (27). Notice that the last term on the last L elements of $\mathbf{t}(k)$ depends on elements within $\mathbf{R}(k-L)$. It can be shown that this procedure produces $\widehat{\mathbf{z}}(k) = \mathbf{R}(k-L)\mathbf{x}(k)$ *exactly* whenever $k = nL$. Thus, the numerical error associated with the sliding-window computation is "zeroed-out" every L samples. For $k \neq nL$, the last $(L+1)$ elements of $\mathbf{t}(k)$ do not match $\mathbf{R}(k-L)\mathbf{x}(k)$ 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 (2) and (31)–(33) define the final form of the simplified blind deconvolution algorithm, where the autocorrelation sequence $r_p(k)$ is updated every L 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 (2)–(4) uses $4L+1$ multiply/adds, the new approach uses approximately 63% more multiply/adds than the original approach.

5. NUMERICAL SIMULATIONS

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

- The proposed algorithm is purported to have less bias in its converged solution than that produced by the original algorithm in (2)–(4).
- The proposed algorithm is purported to perform better than the original algorithm when equalizer truncation is an issue.
- The simplified update in (31)–(33) is purported to perform similarly to the more-complicated update in (24)–(25) 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 $[-1,1]$. The parameters and values chosen for all algorithms are $f(y) = y^3$, $L = 4$, $w_l(0) = \delta_{l-2}$ (*i.e.* center-spike initialization), $M = 20$, and $\mu = 0.0001$. 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

$$x(k) = -0.7x(k-1) + s(k). \quad (34)$$

This autoregressive channel can be exactly equalized using an FIR filter of length $L \geq 2$ having two consecutive non-zero taps equal to $\{1, 0.7\}$. Shown in Fig. 1 are the evolutions of the ensemble averages of the inter-symbol interference (ISI), defined as

$$ISI(k) = \frac{\sum_{l=0}^{M} c_l^2(k)}{\max_{0\leq j\leq M} c_j^2(k)} - 1 \quad (35)$$

where $c_l(k)$ is the convolution of $w_l(k)$ and the channel impulse response, for the original natural gradient update in (3), the preliminary approach in (25), and the proposed final algorithm update in (31). As can be seen, the original algorithm has the best performance, achieving a steady-state ISI of approximately -24dB. 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 the FIR channel model

$$x(k) = 0.7s(k) + s(k-1). \quad (36)$$

This channel is maximum phase, meaning that an infinitely-non-causal equalizer is required to perfectly equalize it. Any FIR equalizer for this task will exhibit a non-zero residual ISI. Shown in Fig. 2 are the evolutions of $E\{ISI(k)\}$ for the three algorithms. In this case, the behaviors are markedly different. Both new algorithms converge to an ISI level of about -10dB in approximately 3000 iterations. In contrast, the original method initially converges like the new algorithms, but it then diverges to a steady-state ISI of about -2dB. 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.

Figs. 3(a) and (b) show the ensemble-averaged values of $E\{c_l(15000)\}$ for the original algorithm update in (3) as well as the proposed algorithm update in (31), respectively, for the second example. Also shown on this plot are the $\pm 3\sigma$ error bars for each non-zero impulse response value over the one hundred simulations. As can be seen, the deviations of the mean values away from the ideal response $[0 \quad 0 \quad 0 \quad * \quad 0 \quad 0]$ are smaller for the proposed algorithm, and the error bars are also smaller as well, indicating that the convergence of the proposed method is more consistent in this situation.

Our final simulation set considers the noisy measurement model

$$x(k) = 0.7s(k) + s(k-1) + \nu(k), \quad (37)$$

in which $\nu(k)$ is an uncorrelated Gaussian sequence with variance $\sigma_\nu^2 = 1.49 \times 10^{-SNR}$, where $SNR = 15$dB is the signal-to-noise ratio. Shown in Fig. 4 are the evolutions of the averaged normalized MSEs

$$MSE(k) = \frac{\sum_{l=0}^{M} c_l^2(k) + 10^{SNR}\sum_{l=0}^{L} w_l^2(k)}{\max_{0\leq j\leq M} c_j^2(k)} - 1 \quad (38)$$

for the various methods as computed from one hundred simulations with these signals. As can be seen, the original method still suffers from a steady-state bias, although the effect is lessened in the presence of noise. The proposed methods perform in a robust manner when noise is present in this case.

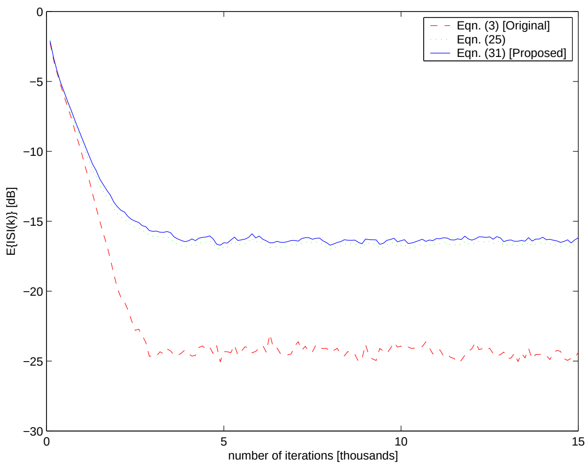

Fig. 1: Evolutions of $E\{ISI(k)\}$ in the first example.

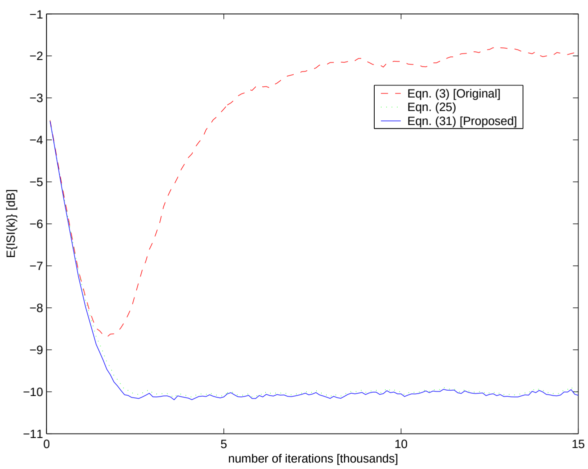

Fig. 2: Evolutions of $E\{ISI(k)\}$ in the second example.

6. CONCLUSIONS

In this paper, we have uncovered a potential problem with a particular natural gradient procedure for blind deconvolution and equalization tasks: The FIR-based filtered-gradient updates can produce a biased solution when perfect equalization is not possible. We have proposed a modification to the algorithm that largely eliminates these effects. The complexity of the algorithm, while about 63% greater than the original approach, is still linear in the number of equalizer filter taps. Simulations show that the proposed method performs better than the original method in situations where an FIR equalizer cannot accurately deconvolve the linear channel.

The results in this paper have a significant impact on the use of the multichannel extension of (2)–(4) for separating convolutive mixtures of acoustic signals [12]. These issues, as well as fast block-based implementations of the methods, are the subject of current work.

7. REFERENCES

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

[2] 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 City, Japan, vol. 3, pp. 1057-1062, July 1997.

[3] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 2, pp. 251-276, Feb. 1998.

[4] S.C. Douglas and S. Amari, "Natural gradient adaptation," in *Unsupervised Adaptive Filtering, Vol. 1: Blind Source Separation*, S. Haykin, ed. (New York: Wiley, 2000), pp. 13-61.

[5] S.C. Douglas and S. Haykin, "Relationships between blind deconvolution and blind source separation," in *Unsupervised Adaptive Filtering, Vol. 2: Blind Deconvolution*, S. Haykin, ed. (New York: Wiley, 2000), pp. 113-145.

[6] S.C. Douglas, A. Cichocki, and S. Amari, "Fast-convergence filtered regressor algorithms for blind equalisation," *Electron. Lett.*, vol. 32, no. 23, pp. 2114-2115, 7 Nov. 1996.

[7] S.C. Douglas, A. Cichocki, and S. Amari, "Self-whitening algorithms for adaptive equalization and deconvolution," *IEEE Trans. Signal Processing*, vol. 47, pp. 1161-1165, Apr. 1999.

[8] S.J. Elliott and B. Rafaely, "Rapid frequency domain adaptation of causal FIR filters," *IEEE Signal Processing Lett.*, vol. 4, pp. 337-339, Apr. 1997.

[9] S.J. Elliott and B. Rafaely, "Frequency domain adaptation of causal digital filters," *IEEE Trans. Signal Processing*, vol. 48, pp. 1354-1364, May 2000.

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

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

[12] S.C. Douglas, H. Sawada, and S. Makino, " Natural gradient multichannel blind deconvolution and speech separation using causal FIR filters," submitted.

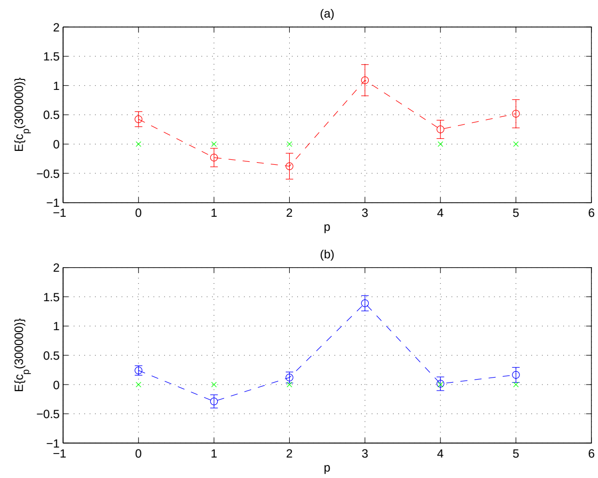

Fig. 3: Plots of $E\{c_l(15000)\}$ for (a) the original and (b) the proposed method in the second example.

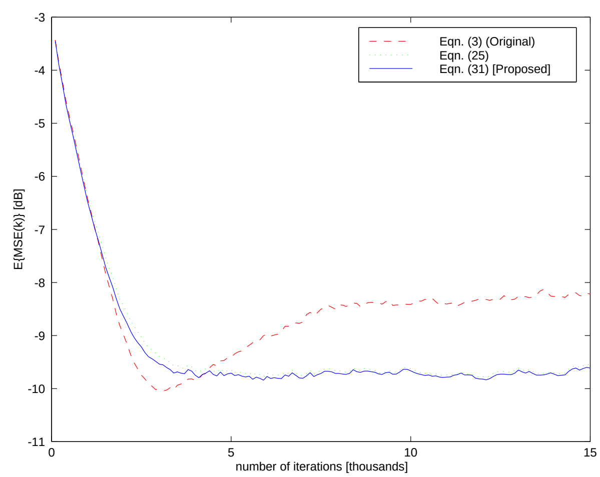

Fig. 4: Evolutions of $E\{MSE(k)\}$ in the third example.