On Coefficient Delay in Natural Gradient Blind Deconvolution and Source Separation Algorithms

Scott C. Douglas[1], Hiroshi Sawada[2], and Shoji Makino[2]

[1] Department of Electrical Engineering
Southern Methodist University
Dallas, Texas 75275 USA
[2] NTT Communication Science Laboratories
NTT Corporation
Kyoto 619-0237 Japan

Abstract. In this paper, we study the performance effects caused by coefficient delays in natural gradient blind deconvolution and source separation algorithms. We present a statistical analysis of the effect of coefficient delays within such algorithms, quantifying the relative loss in performance caused by such coefficient delays with respect to delayless algorithm updates. We then propose a simple change to one such algorithm to improve its convergence performance.

1 Introduction

The related problems of blind source separation and multichannel deconvolution of convolutive signal mixtures have received much attention recently in the signal processing literature [1]–[4]. Interest in such tasks has been largely driven by the development of useful and powerful algorithms for separating and deconvolving such signal mixtures without specific knowledge of the source signals or the mixing conditions. Several of these techniques employ a natural gradient modification in which the estimated inverse model is applied to the parameter updates to improve convergence performance [5]. Since the system being estimated is a multichannel filter, such algorithms invariably employ filtered-gradient updates that have strong ties to classic procedures in model-reference adaptive control [6].

In developing solutions to blind deconvolution and source separation tasks, system designers have many design choices to make. It is often convenient to choose an algorithm structure that re-uses existing computed quantities to minimize the overall complexity of the coefficient updates. Another important design tool is the approximation of two-sided infinite-impulse response (IIR) systems by truncated finite-impulse response (FIR) models. Moreover, when a derived procedure requires signals that are non-causally related to the system's operation at any given time, signal and coefficient delays are often introduced within the algorithm updates to maintain causal operation. A combination of these design

C.G. Puntonet and A. Prieto (Eds.): ICA 2004, LNCS 3195, pp. 634–642, 2004.

choices is often required to achieve a practical algorithm from its theoretical derivation. The effects of these design choices on overall system performance, however, is unclear. A careful study of any of their effects in a specific context would help system designers understand the tradeoffs involved in building blind deconvolution and source separation algorithms that are efficient, useful, and practical.

In this paper, we study the performance effects caused by coefficient delays in one well-known natural gradient multichannel blind deconvolution and source separation procedure [1]. Using a simplified adaptation model, we demonstrate through both analysis and simulation that, for a given convergence rate, algorithms that have coefficient delays within their updates exhibit worse performance than those without such delays. Simulations and analysis also show that recomputing the delayed equalizer output with the most-recent equalizer coefficients within the algorithm nonlinearity can improve this procedure's performance.

2 Coefficient Delay in Natural Gradient Adaptation

As described in [5], natural gradient adaptation is a modified gradient search in which the gradient search direction is modified by the Riemannian metric tensor for the associated parameter space. In [1], a simple but powerful algorithm for multichannel blind deconvolution and source separation is derived using the Kullback-Leibler divergence measure as the optimization criterion. This algorithm has been derived assuming a particular set of coefficient delays for the various signal quantities within the updates to minimize the number of arithmetic operations needed for its implementation. In order to account for the coefficient delays in this algorithm, we present a generalized version of this procedure for which the use of coefficient delays is carefully delineated. For notational simplicity, we shall focus on the single-channel blind deconvolution task in this paper, although our discussions could be easily extended to the multichannel case with minor effort.

Let $s(k)$ denote a sequence of i.i.d. random variables. We observe a filtered version of this sequence given by

$$x(k) = \sum_{i=0}^{\infty} a_i s(k-i), \tag{1}$$

where a_i is the impulse response of an unknown mixing filter. We desire a linear filter $W(z)$ that extracts a scaled, time-delayed version of $s(k)$ from $x(k)$. A single-channel generalized version of the natural gradient algorithm derived in [1] computes an estimated source sequence as

$$y_n(k) = \sum_{l=0}^{L} w_l(n) x(k-l), \tag{2}$$

where n denotes the time index of the equalizer filter coefficients $\{w_l(n)\}$, $0 \leq l \leq L$, and k the time-shift of the input signal. We compute a set of filtered output signals, given by

$$u_{n_1,n_2,\ldots n_{2L+2}}(k) = \sum_{q=0}^{L} w^*_{L-q}(n_{q+L+2}) y_{n_{q+1}}(k-q), \tag{3}$$

where $\{n_1, n_2, \ldots, n_{2L+2}\}$ denote time indices for the filter coefficients used in this calculation. Then, the $(L+1)$ coefficients $\{w_l(k)\}$ are updated as

$$w_l(k+1) = (1+\mu)w_l(k) - \mu f(y_{n_0^{(l)}}(k-L)) u^*_{n_1^{(l)},n_2^{(l)},\ldots n_{2L+2}^{(l)}}(k-l), \tag{4}$$

where $\{n_i^{(l)}\}$ denote the time indices of the coefficients used to update the lth equalizer tap. The above description employs $(2L+3)(L+1)$ different time indices for the coefficient updates, and a practical algorithm requires careful choice of the values of $\{n_i^{(l)}\}$ to allow both a computationally-efficient and statistically-effective algorithm. The only constraint imposed on the values of $\{n_i^{(l)}\}$ is that $n_i^{(l)} \leq k$ to maintain causality of the overall system.

The algorithm in [1] employs the following choices for the coefficient delays in (2)–(4): $n_0^{(l)} = k - L$ for all $0 \leq l \leq L$, $n_{q+1}^{(l)} = k - q - l$ and $n_{q+L+2}^{(l)} = k - l$ for $0 \leq q \leq L$. With these choices, both $y_.(k)$ and $u_.(k)$ become one-dimensional signals, such that delayed versions of $y_.(k)$ and $u_.(k)$ are all that are needed to implement the algorithm. The resulting procedure requires about four multiply/adds per filter tap to implement, not counting the nonlinearity computation $f(y_{k-L}(k-L))$. These choices, however, are not the best from the standpoint of system performance. As is well-known in adaptive control [6], algorithms that have the least adaptation delay within the coefficient updates usually perform the best, implying that $n_i^{(l)} = k$ should be chosen for the above procedure. The computational penalty paid for such an update is severe–the algorithm would require $(3L+4)$ multiply/adds per filter coefficient to implement. Clearly, a trade-off between algorithm complexity and algorithm performance must be made for practical reasons, especially when L is large. But how do coefficient delays affect overall performance of the system in a continuously-adapting scenario?

3 A Simplified Adaptation Model and Its Analysis

To better understand the performance effects caused by coefficient delays in natural gradient algorithms, we propose to study the following four single-coefficient adaptive systems operating on the i.i.d. sequence $x(k)$:

$$w(k+1) = (1+\mu)w(k) - \mu f(w(k)x(k))x(k)|w(k)|^2 \tag{5}$$

$$w(k+1) = (1+\mu)w(k) - \mu f(w(k-D)x(k))x(k)|w(k)|^2 \tag{6}$$

$$w(k+1) = (1+\mu)w(k) - \mu f(w(k)x(k))x(k)|w(k-D)|^2 \tag{7}$$

$$w(k+1) = (1+\mu)w(k) - \mu f(w(k-D)x(k))x(k)|w(k-D)|^2 \tag{8}$$

In these algorithms, D is an integer parameter that sets the coefficient delays within the updates. Eqn. (5) is similar in design to (2)–(4) when $n_i^{(l)} = k$ for all i and l. Eqn. (8) is similar in design to the procedure in [1]. The two algorithms in (6) and (7) are similar to versions of (2)–(4) in which coefficient delays appear in the cost function and in the natural gradient update modification, respectively.

By studying these variants, we can determine through analysis whether algorithms with significant coefficient delay within the updates, represented by $D \gg 1$, cause significant degradation in overall system performance. To make the analysis tractable, we shall make some additional assumptions regarding $x(k)$ and the form of $f(y)$. Specifically,

- $x(k) \sim \mathrm{Unif}(-\sqrt{3}, \sqrt{3})$ is an i.i.d. uniformly-distributed sequence with unit variance, $m_4 = E\{x^4(k)\} = 1.8$, and $m^8 = E\{x^8(k)\} = 9$, and
- $f(y) = y^3$ is a cubic nonlinearity, such that the above procedures are locally-stable for negative-kurtosis $x(k)$.

Specific statistical assumptions and nonlinear update forms are often chosen to perform convergence analyses of linear adaptive filtering algorithms [7]. With these assumptions, we can determine the initial convergence behavior of the mean value of $w(k)$ over time as well as the steady-state value of the variance of $w(k)$ at convergence.

With these choices, the relation in (8) becomes

$$\begin{aligned} w(k+1) &= (1+\mu)w(k) - \mu w^5(k-D)x^4(k) && (9)\\ &= (1+\mu)w(k) - \mu m_4 w^5(k-D) + \mu\left(x^4(k) - m_4\right) w^5(k-D) && (10)\\ &= (1+\mu)w(k) - \mu m_4 w^5(k-D) + \mu\nu(k) && (11) \end{aligned}$$

where we have defined $\nu(k) = \left(x^4(k) - m_4\right) w^5(k-D)$ as a coefficient-dependent noise-like term that drives the deterministic nonlinear system given by

$$\overline{w}(k+1) = (1+\mu)\overline{w}(k) - \mu m_4 \overline{w}^5(k-D). \tag{12}$$

Clearly, the initial convergence behavior of $w(k)$ is dominated on average by the dynamics of the corresponding deterministic system in (12), and the influence of the zero-mean signal $\nu(k)$ is significant only near convergence. Thus, we can simulate the behavior of the deterministic system in (12) with $\overline{w}(0) = w(0)$ to understand how $w(k)$ converges to its optimum value. It is straightforward to show that the stationary point of (12) occurs when $\overline{w}(k) = m_4\overline{w}^5(k)$, or

$$\overline{w}_{ss} = (m_4)^{-1/4} = \pm 0.86334\ldots \tag{13}$$

for uniformly-distributed unit-variance input signals.

To determine the variance of $w(k)$ at convergence, we can used a linearized analysis similar to that employed in [8]. Let $w(k) = \overline{w}_{ss} + \Delta(k)$, where $|\Delta(k)| \ll |\overline{w}_{ss}|$. Then, we can represent (8) as

$$\begin{aligned} \overline{w}_{ss} + \Delta(k+1) = (1+\mu)(\overline{w}_{ss} + \Delta(k)) - \mu m_4(\overline{w}_{ss}^5 + 5\overline{w}_{ss}^4\Delta(k-D))\\ + \mathrm{O}(\mu\Delta^2(k-D)) + \mu\nu(k). \end{aligned} \tag{14}$$

Table 1. Analysis Results for Single-Coefficient Models

Eqn.	$\overline{w}(k+1)$ *Update Relation*	$H(z)$
(5)	$(1+\mu)\overline{w}(k) - \mu m_4 \overline{w}^5(k)$	$\dfrac{\mu z^{-1}}{1-(1-4\mu)z^{-1}}$
(6)	$(1+\mu)\overline{w}(k) - \mu m_4 \overline{w}^2(k)\overline{w}^3(k-D)$	$\dfrac{\mu z^{-1}}{1-(1-\mu)z^{-1}+3\mu z^{-D}}$
(7)	$(1+\mu)\overline{w}(k) - \mu m_4 \overline{w}^3(k)\overline{w}^2(k-D)$	$\dfrac{\mu z^{-1}}{1-(1-2\mu)z^{-1}+2\mu z^{-D}}$
(8)	$(1+\mu)\overline{w}(k) - \mu m_4 \overline{w}^5(k-D)$	$\dfrac{\mu z^{-1}}{1-(1+\mu)z^{-1}+5\mu z^{-D}}$

Using the relationship for $\overline{w}_{ss}$ in (13), (14) simplifies to

$$\Delta(k+1) - (1+\mu)\Delta(k) + \mu 5 m_4 \overline{w}_s^4 \Delta(k-D) = \mu\nu(k). \tag{15}$$

Taking z-transforms of both sides, we can relate $D(z)$, the z-transform of $\Delta(k)$, to $N(z)$, the z-transform of $\nu(k)$, as

$$D(z) = H(z)N(z), \tag{16}$$

where the transfer function $H(z)$ is given by

$$H(z) = \frac{\mu z^{-1}}{1-(1+\mu)z^{-1}+5\mu z^{-D}}. \tag{17}$$

Assuming that each $\nu(k)$ is i.i.d., the power of $\Delta(k)$ in steady state is given by

$$E\{\Delta^2(k)\}_{ss} = E\{\nu^2(k)\}_{ss} \sum_{l=0}^{\infty} h^2(l), \tag{18}$$

where $h(l)$ is the inverse z-transform of $H(z)$ and

$$E\{\nu^2(k)\}_{ss} = E\{(x^4(k) - m_4)^2\} w_{ss}^{10} \tag{19}$$

$$= (m_8 - m_4^2) w_{ss}^{10} = 1.325077\ldots \tag{20}$$

for uniformly-distributed input signals.

Similar analyses can be carried out for the algorithms in (5), (6), and (7), respectively. Table 1 shows the forms of the update relations for $\overline{w}(k)$ and the corresponding $H(z)$ in each case. The specific derivations are omitted for brevity.

These results allow us to fairly and accurately compare the performances of the different single-coefficient procedures in (5)–(8) by carefully maintaining certain convergence relationships between them. For example, we can choose different step size values μ for each procedure to maintain an identical convergence rate from a specific initial $w(0)$ and determine analytically the values of $E\{\Delta^2(k)\}_{ss}$ in steady-state. These results can also be compared with simulations of both the single-coefficient algorithms and their related blind deconvolution procedures in (2)–(4) for specific forms of coefficient delay.

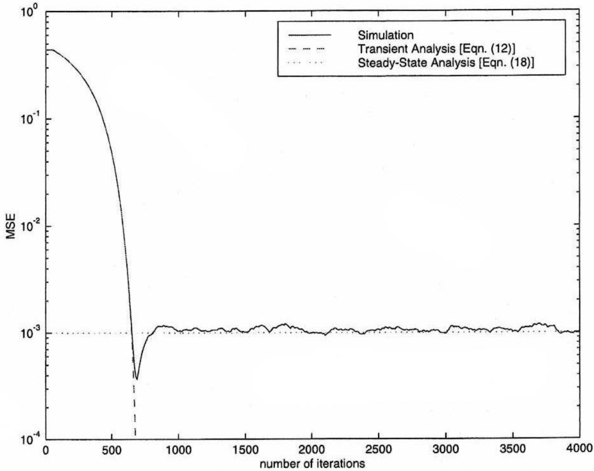

Fig. 1. Comparison of analysis and simulation results for the single-coefficient update in Eqn. (11).

4 Simulations

We now verify the analytical results presented in the previous section and compare these with simulation results from single-channel blind deconvolution tasks. From these results, we can gauge what performance degradations are caused by coefficient delays within the algorithm updates.

Our first set of simulations is designed to verify the analytical results for the single-coefficient systems in (5)–(8). For these simulations, $x(k)$ is a unit-variance Unif$[-\sqrt{3}, \sqrt{3}]$ random sequence, and we have arbitrarily chosen $D = 50$, $w(0) = 0.2$ and $\mu = 2.7062 \times 10^{-3}$. With these choices, our analyses predict that the procedure in (8) will converge to a steady-state variance of $E\{\Delta^2(k)\}_{ss} = 0.001$ in 586 iterations. Shown in Fig. 1 are (a) the evolution of $(\overline{w}(k) - |m_4^{-1/4}|)^2$, (b) the value of $E\{\Delta^2(k)\}_{ss}$ predicted from the analysis, and (c) the evolution of the coefficient MSE $E\{(w(k) - |m_4^{-1/4}|)^2\}$ as computed from ensemble averages of 1000 simulation runs. As can be seen, both $(\overline{w}(k) - |m_4^{-1/4}|)^2$ and $E\{\Delta^2(k)\}_{ss}$ are accurate predictors of the coefficient MSE during the transient and steady-state phases of adaptation, respectively.

Shown in Fig. 2 are the evolutions of the coefficient MSEs for the four systems in (5)–(8) as well as their predicted steady-state MSEs from (18). Here, we have chosen step sizes for each algorithm such that all of them converge from $w(0) = 0.2$ to an MSE of 0.001 in 563 iterations; thus, we can accurately compare the steady-state MSEs of each approach. As can be seen, the algorithm with

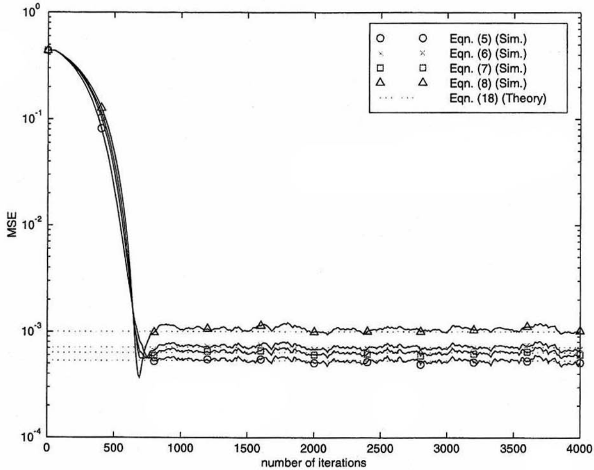

Fig. 2. Comparison of steady-state MSE analysis and simulation results for the single-coefficient update in Eqns. (5)–(8).

no adaptation delay performs the best, and the simulated algorithms in (6), (7), and (8) have steady-state MSEs that are 1.31, 0.78, and 2.99 dB greater, respectively, than that of (5). The steady-state MSE analysis determined from (18) and Table 1 predicts similar performance degradations of 1.26, 0.76, and 2.74 dB, respectively.

An adaptive filtering analysis is only useful if it is an accurate predictor of performance relationships for a given system and scenario. We now explore the simulated performance of the single-channel blind deconvolution procedure in (2)–(4) for various choices of adaptation delay within a particular blind deconvolution task in which $x(k)$ is generated from an i.i.d. uniformly-distributed sequence $s(k)$ as

$$x(k) = 0.7x(k-1) + \sum_{i=0}^{10}(0.4)^{10-i}s(k-i) \tag{21}$$

This non-minimum-phase system cannot be equalized using simple linear prediction. The algorithms studied in this case are:

- **Case 1: All $n_i^{(l)} = k$.** This procedure most closely resembles (5).
- **Case 2: $n_0^{(l)} = k$ for all $0 \leq l \leq L$; $n_{q+1}^{(l)} = k - q - l$ and $n_{q+L+2}^{(l)} = k - l$ for $0 \leq q \leq L$.** This procedure most closely resembles (7) with $D = L$.
- **Case 3: The original procedure in [1].** This procedure most closely resembles (8) with $D = L$.

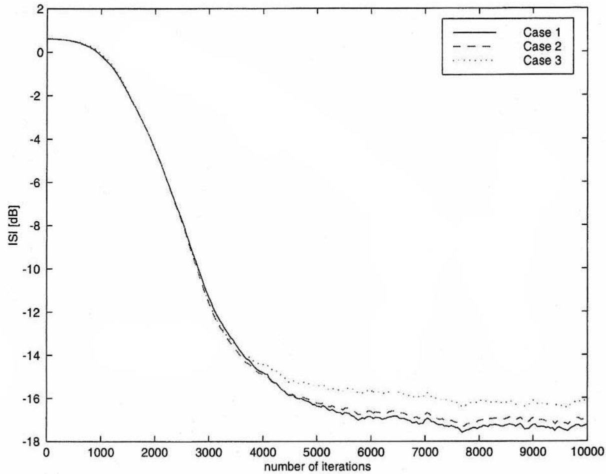

Fig. 3. Algorithm performance for a single-channel blind equalization task.

In each case, we have chosen $L = 50$. Step sizes of $\mu = 0.001, 0.00101$, and 0.00102 were chosen to provide similar convergence rates of the averaged inter-symbol interferences $ISI(k) = (\sum_{i=0}^{1000} c_i^2(k))/(\max_{0\le j\le 1000} c_j^2(k)) - 1$, where $c_i^2(k)$ is the combined channel-plus-equalizer impulse response. Shown in Fig. 3 are the results. The steady-state ISIs of the equalizers for Case 2 and Case 3 are 0.31 and 1.09 dB worse than that for Case 1, respectively. The performance of the Case 2 equalizer is especially noteworthy given its complexity; it only requires five multiply/adds per filter coefficient, a 25% increase over the algorithm in [1].

5 Conclusions

In this paper, we have studied the performance effects caused by coefficient delays in one well-known algorithm for blind deconvolution and source separation tasks. Through a simple analytical model, we show that algorithms with coefficient delays within their updates have worse adaptation performance than those without such delays. We also suggest a simple modification to improve this algorithm's adaptation performance. Simulations have been used to verify the accuracy of the analyses.

References

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

2. R.H. Lambert and A.J. Bell, "Blind separation of multiple speakers in a multipath environment," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing,* Munich, Germany, vol. 1, pp. 423-426, Apr. 1997.
3. L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Processing,* vol. 8, pp. 320-327, May 2000.
4. K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," *Proc. 3rd Int. Workshop Indep. Compon. Anal. Signal Separation,* San Diego, CA, pp. 722–727, Dec. 2001.
5. S.C. Douglas and S. Amari, "Natural gradient adaptation," in *Unsupervised Adaptive Filtering, Vol. I: Blind Signal Separation,* S. Haykin, ed., (New York: Wiley, 2000), pp. 13-61.
6. B. Widrow and E. Walach, *Adaptive Inverse Control* (Upper Saddle River, NJ: Prentice-Hall, 1996).
7. S.C. Douglas and T.H.-Y. Meng, "Stochastic gradient adaptation under general error criteria," *IEEE Trans. Signal Processing,* vol. 42,pp. 1335-1351, June 1994.
8. S.C. Douglas, "Self-stabilized gradient algorithms for blind source separation with orthogonality constraints," *IEEE Trans. Neural Networks,* vol. 11, pp. 1490-1497, Nov. 2000.