SCALED NATURAL GRADIENT ALGORITHMS FOR INSTANTANEOUS AND CONVOLUTIVE BLIND SOURCE SEPARATION

Scott C. Douglas and Malay Gupta

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

ABSTRACT

This paper describes a novel modification to the well-known natural gradient or INFOMAX algorithm for blind source separation that largely mitigates its divergence problems. The modified algorithm imposes an *a posteriori* scalar gradient constraint that adds little computational complexity to the algorithm and exhibits fast convergence and excellent performance for fixed step size values that are largely independent of input signal magnitudes and initial separation matrix estimates. Evaluation of the approach for the separation of instantaneous and convolutive source mixtures using both time- and frequency-domain implementations shows its excellent separation behavior.

Index Terms— blind source separation, independent component analysis, natural gradient algorithm.

1. INTRODUCTION

The goal of blind source separation is to extract individual source signals from linear mixtures without specific knowledge of the mixing process and with limited knowledge about the individual source signals. Algorithms for spatial-only (instantaneous) and spatio-temporal (convolutive) mixing conditions have been developed and usually leverage some prior knowledge about the source signals, such as their statistical independence, amplitude variations, or correlation statistics.

One of the oft-cited methods for blind source separation is the natural gradient or information-maximization (INFOMAX) algorithm [1, 2]. For linear mixtures of the form

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k), \tag{1}$$

where $\mathbf{A}$ is an $(m \times m)$ mixing matrix and $\mathbf{s}(k) = [s_1(k) \cdots s_m(k)]^T$ contains independent source signals, this algorithm adapts a separation matrix $\mathbf{W}(k)$ as

$$\mathbf{G}(k) = \frac{1}{N}\sum_{l=N(k-1)+1}^{Nk} \mathbf{f}(\mathbf{y}_k(l))\mathbf{y}_k^H(l) \tag{2}$$

$$\mathbf{W}(k+1) = (1+\mu)\mathbf{W}(k) - \mu\mathbf{G}(k)\mathbf{W}(k), \tag{3}$$

where $\mathbf{y}_k(l) = \mathbf{W}(k)\mathbf{x}(l)$ contains the estimated source signals at time l and iteration k, $\mathbf{f}(\mathbf{y})$ is a vector-valued nonlinearity function, and μ is the algorithm step size. For appropriate choices of the nonlinearites in $\mathbf{f}(\mathbf{y})$, the algorithm is a natural gradient procedure that approximately minimizes the mutual information of the output vector sequence $\mathbf{y}_k(l)$ over all non-singular matrices $\mathbf{W}(k)$ [2]. Simulations show that the algorithm can separate mixtures of sources so long as the source distributions do not differ too much from those used to design the nonlinearities in $\mathbf{f}(\mathbf{y})$.

Although useful, the natural gradient BSS algorithm in (2)–(3) suffers from a lack of robustness to the magnitudes of the mixtures in $\mathbf{x}(k)$. Convergence is governed both by the initial coefficient matrix $\mathbf{W}(0)$ and the step size μ. It can be quite challenging to choose $\mathbf{W}(0)$ and μ to obtain fast convergence with this algorithm without explosive divergence, particularly if $\mathbf{f}(\mathbf{y})$ is highly nonlinear. The end result is usually slow initial convergence using $\mathbf{W}(0) = \delta\mathbf{I}$ with both μ and δ chosen to be small. Moreover, our experience with this algorithm indicates that for larger-order mixtures (*e.g.* $m \geq 10$), it is nearly impossible to choose $\mathbf{W}(0)$ to obtain convergence for *any* constant step size μ. Although some work has been done in time-varying step size sequences for this algorithm class [3], initial convergence is still slow, and $\mu(k) \leq \mu_{max}$ must be maintained for some unknown μ_{max}.

This paper presents presents a novel, simple modification to (2)–(3) that alleviates the aforementioned difficulties. Since separation performance does not depend explicitly on the scaling of $\mathbf{W}(k)$, a scaling constraint is used to maintain a constant (natural) gradient magnitude for the algorithm. The modification adds little computational complexity and achieves significant practical advantages. In particular, the algorithm exhibits fast convergence and excellent performance for a fixed step size μ, independent of the magnitude of $\mathbf{x}(k)$ and $\mathbf{W}(0)$. Extensive simulations indicate that the algorithm is robust and does not appear to diverge suddenly for any step size, unlike the original unscaled natural gradient algorithm. It typically converges in only 40 to 150 iterations under mixing conditions in which the original algorithm does not converge reliably for any fixed step size. Performance comparisons of the convolutive extensions of the algorithm in both time- and frequency-domain forms with other competing approaches show the superior performance of the proposed methods.

2. DERIVATION

The proposed algorithm attempts to solve the following optimization problem via natural gradient descent:

$$\text{minimize} \quad -\log|\det(\mathbf{W}(k))| - \sum_{l=N(k-1)+1}^{Nk} \sum_{i=1}^{m} \log p(y_{ik}(l)) \quad (4)$$

$$\text{such that} \quad \frac{1}{mN}\left\|\sum_{l=N(k-1)+1}^{Nk} \mathbf{f}(\mathbf{y}_{k+1}(l))\mathbf{y}_{k+1}^H(l)\right\|_1 = 1, \quad (5)$$

where $p(y)$ is a chosen source p.d.f. model. The cost function in (4) is affinely-related to the mutual information of the output vector sequence $\{\mathbf{y}_k(l)\}$ when $p(y)$ accurately models the source p.d.f.'s [2, 4]. Eq. (5) is a scalar L_1-norm constraint on the *a posteriori* or $(k+1)$st gradient data matrix that is satisfied at algorithm convergence for separated sources. A scalar constraint is chosen to stabilize the algorithm as it does not alter the goal of making the output signals statistically-independent and is expected to have a minimal effect on the algorithm's local convergence behavior.

The adaptation procedure used to achieve (4)–(5) jointly adjusts $\mathbf{W}(k)$ in the natural gradient direction in an attempt to minimize (4) while imposing a multiplicative scaling of all of the elements of the updated matrix to satisfy (5) exactly. This algorithm has the form

$$\overline{\mathbf{W}}(k) = c(k)\mathbf{W}(k) \quad (6)$$

$$\overline{\mathbf{y}}_k(l) = \overline{\mathbf{W}}(k)\mathbf{x}(l) \quad (7)$$

$$\overline{\mathbf{G}}(k) = \frac{1}{N}\sum_{l=N(k-1)+1}^{Nk} \mathbf{f}(\overline{\mathbf{y}}_k(l))\overline{\mathbf{y}}_k^H(l) \quad (8)$$

$$\mathbf{W}(k+1) = (1+\mu)\overline{\mathbf{W}}(k) - \mu\overline{\mathbf{G}}(k)\overline{\mathbf{W}}(k), \quad (9)$$

where $c(k)$ is a scaling factor sequence. In order to impose (5), we compute $c(k)$ as

$$c(k) = \frac{1}{h\left(d(k)\right)} \quad (10)$$

$$d(k) = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{m}|g_{ij}(k)|, \quad (11)$$

where $h(d)$ is the inverse function of the magnitude of $y^* f(y)$; i.e. $h(y^* f(y)) = |y|$. The complete algorithm in its simplest form is

$$\mathbf{y}_k(l) = \mathbf{W}(k)\mathbf{x}(l) \quad (12)$$

$$\mathbf{W}(k+1) = (1+\mu)c(k)\mathbf{W}(k) - \mu\frac{c(k)}{d(k)}\mathbf{G}(k)\mathbf{W}(k), \quad (13)$$

where $\mathbf{G}(k)$ is as computed in (2). Note that this algorithm absorbs the scaling factor $c(k)$ into the update for $\mathbf{W}(k)$, such that little computational complexity is added.

As for choices of algorithm nonlinearity $f_i(y)$, typical choices and their associated $h(d)$ functions are

$$f_i(y) = |y|^2 y \quad \Rightarrow \quad h(d) = d^{\frac{1}{4}} = \sqrt{\sqrt{d}} \quad (14)$$

$$f_i(y) = \frac{y}{|y|} \quad \Rightarrow \quad h(d) = d \quad (15)$$

$$f_i(y) = \tanh(\alpha|y|)\frac{y}{|y|} \quad \Rightarrow \quad h(d) = d, \quad (16)$$

where $\alpha > 1$. The last choice uses an approximation for $h(d)$, as the inverse of $|y|\tanh(\alpha|y|)$ approaches $|y|$ as α gets large.

The proposed scaling strategy can be used in frequency-domain blind source separation algorithms involving convolutive mixtures [5], where measured signals from each input signal frequency bin is treated as an instantaneous complex mixture. Permutation issues must be addressed to achieve good separation, however [6]. We can also extend this method to the time-domain algorithm for multichannel blind deconvolution and convolutive BSS in [7]. The complete algorithm for $(L+1)$-samples-long FIR multichannel separation filters in this case is

$$\mathbf{y}_k(l) = \sum_{p=0}^{L}\mathbf{W}_p(k)\mathbf{x}(l-p) \quad (17)$$

$$\mathbf{F}_p(k) = \frac{1}{N}\sum_{l=N(k-1)+1}^{Nk}\mathbf{f}(\mathbf{y}_k(l))\mathbf{x}^H(l-p), 0 \le p \le L \quad (18)$$

$$\mathbf{G}_q(k) = \sum_{p=0}^{L}\mathbf{F}_{q-p}(k)\mathbf{W}_{L-p}^H(k), 0 \le q \le 2L \quad (19)$$

$$\mathbf{H}_p(k) = \sum_{q=0}^{L}\mathbf{G}_{L+p-q}(k)\mathbf{W}_q(k), 0 \le p \le L \quad (20)$$

$$d(k) = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{q=0}^{2L}|g_{ijq}(k)| \quad (21)$$

$$\mathbf{W}_p(k+1) = (1+\mu)c(k)\mathbf{W}_p(k) - \mu\frac{c(k)}{d(k)}\mathbf{H}_p(k). \quad (22)$$

In the above relations, both $\mathbf{W}_p(k)$ and $\mathbf{F}_p(k)$ are assumed to be zero outside the range $p \in [0, L]$.

Simulations of the scaled natural gradient algorithms above in instantaneous and convolutive BSS tasks indicate that the algorithm achieves fast convergence and a low inter-channel interference (ICI) for a range of step sizes that are independent of the initial scaling of either the separation system or the input signal mixtures so long as the amount of data samples N in the gradient calculation are large enough. Data reuse is recommended, such that $\mathbf{x}(k+iN) = \mathbf{x}(k)$ for $i \in \{1, 2, \ldots\}$ and the entire data set $\mathbf{x}(k)$, $k \in [1, N]$ is used at each algorithm iteration. Typically, step sizes in the range $0.1 \le \mu \le 0.5$ have been chosen and appear to work well.

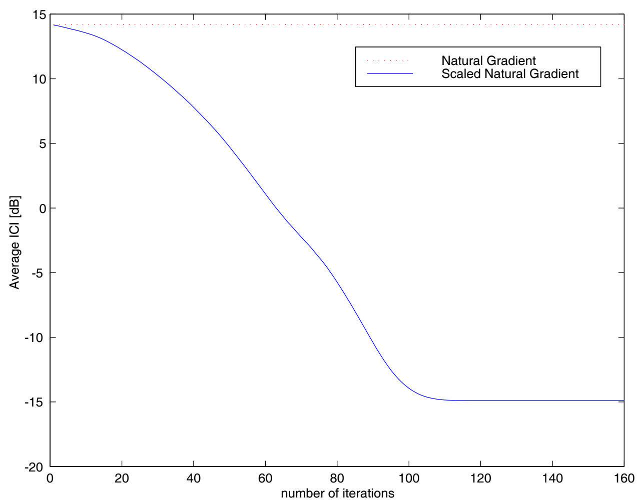

Fig. 1. Performance of the original natural gradient and proposed scaled natural gradient algorithms in a ten-source instantaneous BSS task.

3. NUMERICAL EVALUATIONS

We first explore the behavior of the scaled natural gradient algorithm in an instantaneous BSS task. Complex-valued mixtures of $N = 1000$ samples of 10 complex-valued sources – five 4-QAM and five uniform-$[-1-j, -1+j, 1-j, 1+j]$ – are created using random mixing matrices $\mathbf{A}$ with i.i.d. complex Gaussian entries with $E\{|a_{ij}|^2\} = 1$ and $E\{a_{ij}^2\} = 0$. Both the original (unscaled) and scaled natural gradient algorithms are applied with $\mathbf{W}(0) = 0.001\mathbf{I}$. For each algorithm, step sizes were chosen to achieve fastest convergence for each initial condition. One hundred simulations were run and the convergence curves were averaged in each case. Fig. 1 shows the behaviors of the algorithms in terms of ICI for each block iteration, computed as

$$ICI_k = \left(\sum_{i=1}^{m} \sum_{l=1}^{m} \frac{|c_{il}(k)|^2}{\max_p |c_{ip}(k)|^2} \right) - m. \qquad (23)$$

As can be seen, the original (unscaled) algorithm does not converge, and its convergence behavior does not improve even after 1000 iterations of the algorithm. In addition, increasing the step size for the original unscaled algorithm did not improve its convergence speed; in fact, it caused divergence. The scaled version of the algorithm achieves convergence in about 100 iterations on average.

We now consider a convolutive BSS task involving speech signal mixtures. We used an acoustically-isolated laboratory environment to collect three-channel data using loudspeakers to play recordings of talkers (one female and two male) as the sources. The sources were located 127 cm away from three omnidirectional microphones in a nearly-uniform linear array with 4 cm spacing and had angles of incidence of

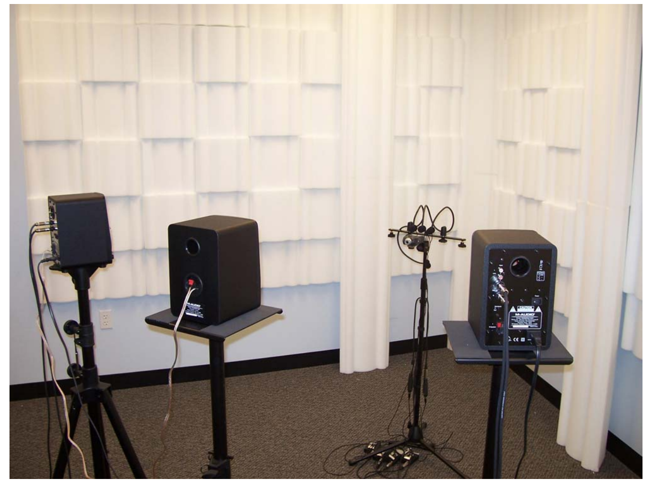

Fig. 2. Laboratory measurement environment used for numerical evaluations.

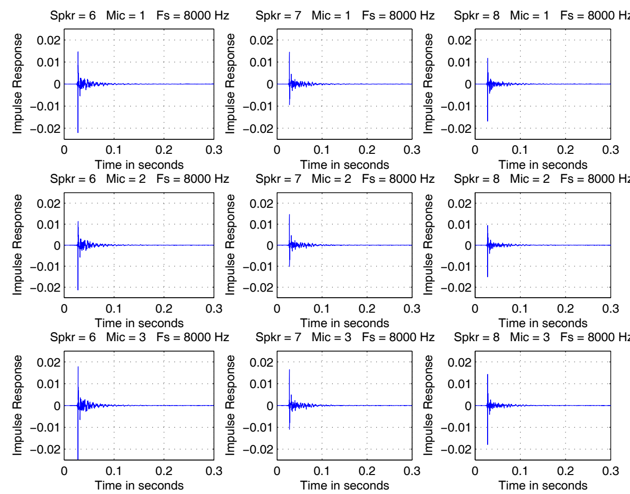

Fig. 3. Impulse responses measured from the uniform linear array, $RT = 300$ ms condition.

-30 degrees, 0 degrees, and 27.5 degrees. The reverberation time of the room was adjusted by placement or removal of foam tiles on the walls of the room. Fig. 2 shows a photograph of the laboratory setup. Figs. 3 and 4 show the impulse responses of the loudspeaker/microphone paths as calculated using pseudo-random noise sequences for the two reverberation conditions corresponding to 300 ms and 425 ms, respectively. For purposes of correlating these plots with the photograph in Fig. 2, the microphones are labeled as 1, 2, and 3 from right to left in the photograph, and the loudspeakers are labeled as 6, 7, and 8 from right to left in the photograph with directions of arrival of $27.5°$, $0°$, and $-30°$, respectively. All speech measurements were made using 7 seconds of data per channel and a 48kHz sampling rate and were downsampled to an 8kHz sampling rate for processing.

We compare the separation performance of five algorithms:

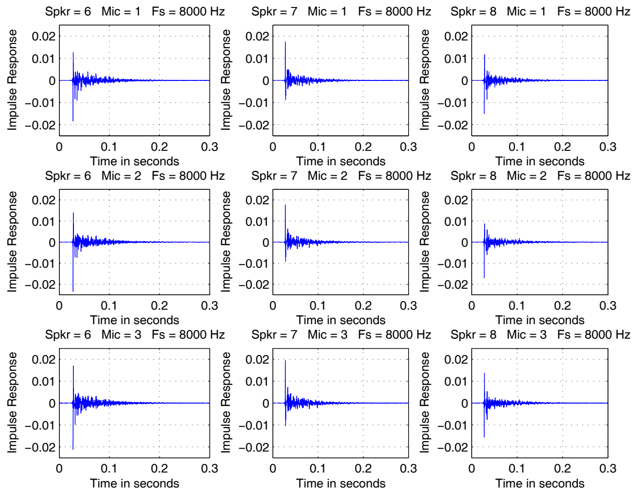

Fig. 4. Impulse responses measured from the uniform linear array, $RT = 425$ ms condition.

(1) Scaled natural gradient time domain (NGTD) method [this paper] using two stages of least-squares prewhitening according to the method in [8], $L = 1024$, $\mu = 0.2$; (2) Scaled natural gradient frequency-domain (NGFD) bin-wise method [this paper] with beamforming initialization, $L = 1024$, $\mu = 0.2$; (3) Spatio-temporal FastICA with two stages of least-squares prewhitening, $L = 300$, $M = 400$ [8]; (4) Parra's decorrelation-based method with the following choices: $\{$number of diagonalized matrices$\} = 5$, $\{$number of data blocks averaged$\} = 10$, $\{$FFT size$\} = 1024$, $L' = 400$-tap time-domain filters per input-output channel, and a 1000-iteration limit [9]; and (5) Parra's decorrelation-based method with beamforming initialization using the above parameter settings [10]. After separation is performed, least-squares methods are used to estimate the contributions of the source recordings to each of the recorded mixtures as well as the separated results from each algorithm. By calculating power ratios, we compute the average improvement in signal-to-interference-plus-noise ratio (SINR) for each algorithm in each data case.

Fig. 5 shows the performance of these algorithms on this data. The time-domain scaled natural gradient algorithm with prewhitening performed the best on both data sets, achieving 13.8dB and 9.5dB SIR improvement in the two data cases, respectively. The frequency-domain scaled natural gradient algorithm with beamforming outperformed Parra's methods by a significant margin and performed nearly as well as the spatio-temporal FastICA algorithm on this data without requiring a prewhitening step. Both of the scaled natural gradient algorithms never diverged in any of our numerical experiments.

4. CONCLUSIONS

This paper describes a simple modification of the well-known natural gradient or INFOMAX algorithm for blind source separation that (a) stabilizes its behavior regardless of initial conditions and input signal scaling, (b) achieves fast convergence for a fixed range of step size values, and (c) produces better separation quality for convolutive BSS of real-world acoustic mixtures as compared to competing methods. An analytical study of the convergence behavior of the scaled natural gradient algorithm is currently underway.

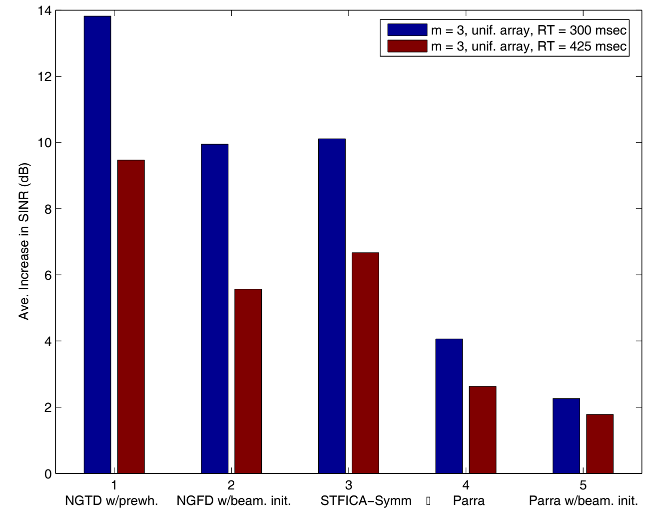

Fig. 5. Performance of various separation methods under two reverberation conditions (RT = 300 msec and 425 msec) for a uniform linear microphone array, $m = 3$ sources.

5. REFERENCES

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

[2] S. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind signal separation," *Adv. Neural Inform. Proc. Sys. 8* (Cambridge, MA: MIT Press, 1996), pp. 757-763.

[3] S.C. Douglas and A. Cichocki, "Adaptive step size techniques for decorrelation and blind source separation," *Proc. 32nd Ann. Asilomar Conf. Signals, Syst., Comput.,* Pacific Grove, CA, vol. 2, pp. 1191-1195, Nov. 1998.

[4] P. Comon, "Independent component analysis: A new concept?" *Signal Processing,* vol. 36, no. 3, pp. 287-314, Apr. 1994.

[5] N. Mitianoudis and M.E. Davies, "Audio source separation of convolutive mixtures," *IEEE Trans. Speech Audio Processing,* vol. 11, pp. 489-497, Sept. 2003.

[6] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans. Speech Audio Processing,* vol. 12, pp. 530-538, Sept. 2004.

[7] S.C. Douglas, H. Sawada, and S. Makino, "Natural gradient multichannel blind deconvolution and speech separation using causal FIR filters," *IEEE Trans. Speech Audio Processing,* vol. 13, pp. 92-104, Jan. 2005.

[8] S.C. Douglas, H. Sawada, and S. Makino, "A spatio-temporal FastICA algorithm for separating convolutive mixtures," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing,* Philadelphia, PA, vol. 5, pp. 165-168, Mar. 2005.

[9] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Processing,* vol. 8, pp. 320-327, May 2000.

[10] L. Parra and C. Alvino, "Geometric source separation: Merging convolutve source separation with geometric beamforming," *IEEE Trans. Speech Audio Processing,* vol. 10, no. 6, pp. 352-362, Sept. 2002.

[11] H. Buchner, R. Aichner, and W. Kellermann, "Blind source separation for convolutive mixtures exploiting nongaussianity, nonwhiteness, and nonstationarity," *Conf. Rec. IEEE Intl. Workshop on Acoustic Echo and Noise Control (IWAENC),* Kyoto, Japan, pp. 275-278, Sept. 2003.