STEP-SIZE CONTROL IN BLIND SOURCE SEPARATION

Thomas P. von Hoff [1], *Allen G. Lindgren* [2], *and August N. Kaelin* [1]

[1] Swiss Federal Institute of Technology,
Laboratory for Signal and Information Processing,
Zurich, Switzerland
{vonhoff, kaelin}@isi.ee.ethz.ch

[2] University of Rhode Island,
Departement of Electrical Engineering,
Kingston, RI 02881, USA
lindgren@ele.uri.edu

ABSTRACT

The behavior of the classic algorithm for blind source separation (BSS) is detailed for a fixed step size. To improve the algorithm in speed and exactness, essential in tracking a time-varying mixing environment, a variable step size must be employed. The ideal step size should decrease or increase as the overall system error decreases or increases. It is shown analytically that the coefficients of the estimating function provide a "measure of error" that is available to automatically control the algorithm step size. This paper proposes a self-adjusting, time-varying step size that is built from the square of the running average of the coefficients of the estimating function. Error free convergence is achieved for a time-invariant environment. The ability of the algorithm to improve the convergence in a time-invariant mixing environment and to track a changing mixing environment is demonstrated by extensive simulation results.

1. INTRODUCTION

Blind Source Separation (BSS) is the problem of extracting independent components from an observed mixture. The model considered in this paper is described by

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

where $\mathbf{s}$ is the vector of source signals, $\mathbf{A}$ the mixing matrix, and $\mathbf{x}$ the vector containing the observed signals. The

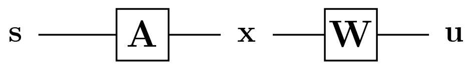

Figure 1: Model for BSS.

matrix $\mathbf{W}$ is to be adjusted such that $\mathbf{u}$ recovers the source signals up to ambiguities of permutation and scaling. This problem has been well studied for the recent years resulting in the approaches of Entropy Maximization [1], ML Estimation [2], and Independent Component Analysis [3]. Using the relative gradient [4] or the natural gradient [5], all approaches lead to the same update equation for $\mathbf{W}_{k+1}$, namely

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \frac{\mu_k}{B} \sum_t \left[\mathbf{I} - \mathbf{\Phi}(\mathbf{u}_t)\mathbf{u}_t^T \right] \mathbf{W}_k \ , \tag{2}$$

where μ_k the step size and B the block size. The so-called score function $\mathbf{\Phi}$ is defined by

$$\Phi_i(u_i) = -\frac{p_{S_i}(u_i)'}{p_{S_i}(u_i)} \ . \tag{3}$$

Unless s_i is Gaussian, the score function $\Phi_i(u_i)$ is a nonlinearity. A separating solution yields

$$\mathbf{V} = \mathbf{W}\mathbf{A} = \mathbf{P}\mathbf{D} \ , \tag{4}$$

where $\mathbf{P}$ is a permutation that results from not ordering the sources correctly and $\mathbf{D}$ is a diagonal scaling matrix. Since the solution can be arbitrarily arranged and rescaled, for convenience it is assumed in this paper $\mathbf{PD} = \mathbf{I}$. If the score functions are replaced by other nonlinearities $\phi_i(u_i)$, the separation of the ith and the jth source is still possible as long as following conditions [6] are fulfilled:

$$\mathsf{E}\left\{u_i^2 \phi_i'(u_i)\right\} + 1 \quad > \quad 0 \tag{5}$$

$$\mathsf{E}\left\{\phi_i'(u_i)\right\} \quad > \quad 0 \tag{6}$$

$$\mathsf{E}\left\{u_i^2\right\} \mathsf{E}\left\{u_j^2\right\} \mathsf{E}\left\{\phi_i'(u_i)\right\} \mathsf{E}\left\{\phi_j'(u_j)\right\} \quad > \quad 1 \tag{7}$$

In case of instability, where the condition in Eq. (7) are not met, procedures exist to stabilize the algorithm [6, 7]. The step size μ defines the dynamic behavior and final error levels achieved by the algorithm.

2. BEHAVIOR OF THE BSS ALGORITHM

We define the estimating function

$$\mathbf{H}_k = \frac{1}{B} \sum_t \left[\mathbf{I} - \phi(\mathbf{u}_t)\mathbf{u}_t^T \right] \tag{8}$$

with t in the kth block. Using the estimating function $\mathbf{H}_k$ the update equation for the overall matrix $\mathbf{V}_k = \mathbf{W}_k\mathbf{A}$ can be formulated as

$$\mathbf{V}_{k+1} = \mathbf{V}_k + \mu_k \mathbf{H}_k \mathbf{V}_k \ . \tag{9}$$

The error performance is expressed by the Interchannel Interference (ICI_k)

$$\mathrm{ICI}_k = \frac{1}{N}\sum_i \sum_{j \neq i} \frac{\mathsf{E}\left\{\widetilde{V}_{ij_k}^2\right\}}{\mathsf{E}\left\{V_{ii_k}^2\right\}} \ , \tag{10}$$

where N the number of sources, V_{ii_k} the ith diagonal element of the overall matrix $\mathbf{V}_k = \mathbf{W}_k\mathbf{A}$, and $\widetilde{V}_{ij_k}$ the ijth element of the error matrix $\widetilde{\mathbf{V}} = \mathbf{V} - \mathbf{I}$. For a constant step size μ and same probability distributions, the final ICI is [7]

$$\mathrm{ICI} = (N-1)\frac{\mu}{2B}\frac{\gamma_+/\kappa_+ + \gamma_-/\kappa_-}{2} \tag{11}$$

with

$$\kappa_+ = \mathsf{E}\left\{\phi_i'(u_i)\right\}\mathsf{E}\left\{u_i^2\right\} + 1 \ , \tag{12}$$
$$\kappa_- = \mathsf{E}\left\{\phi_i'(u_i)\right\}\mathsf{E}\left\{u_i^2\right\} - 1 \ , \tag{13}$$
$$\gamma_+ = \mathsf{E}\left\{\phi_i^2(u_i)\right\}\mathsf{E}\left\{u_i^2\right\} + 1 \ , \tag{14}$$
$$\gamma_- = \mathsf{E}\left\{\phi_i^2(u_i)\right\}\mathsf{E}\left\{u_i^2\right\} - 1 \ . \tag{15}$$

Both the convergence speed and the steady-state error performance are influenced by the ratio μ/B. A small μ/B results in a slow convergence and a small error performance. Thus, for a fixed step size, there is a trade-off between convergence time and error performance.

Considering the behavior of the ICI during the adaptation with a constant step size one notes that the ICI improves very slowly in the beginning and faster as the error matrix gets smaller. Fig. 2 demonstrates this for a separation problem with 10 Laplacian sources. Obviously, the convergence behavior far from equilibrium is different than that in the vicinity of the separating solution. The convergence near an equilibrium for a fixed step size is well studied [6, 7]. There, it is characteristic, that the ICI decreases exponentially down to the steady-state level.

3. THE CONCEPT OF A SELF-ADJUSTING STEP SIZE

With a time-varying step size one can take into account the dependence of the convergence behavior on whether $\mathbf{V}_k$ is close to an equilibrium or not. Such an algorithm could also respond to a change in the mixing environment. The principle of adaptive algorithms asks that the step size should be chosen big if the distance between the estimated parameter

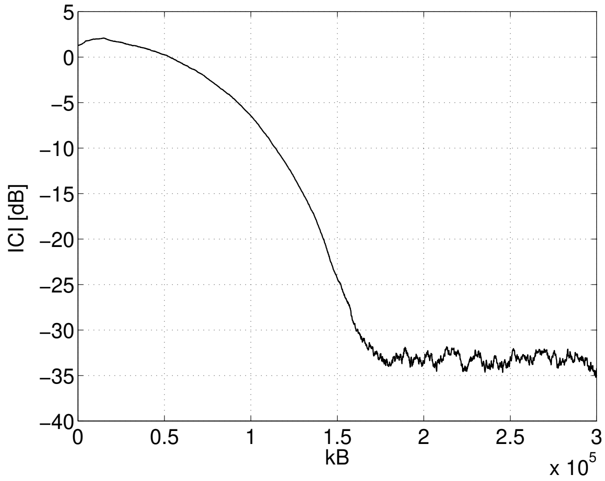

Figure 2: Separation of 10 Laplacian sources with $\phi_i(u_i) = \sqrt{2}\mathrm{sign}(u_i)$, $\mu = 0.01$, and $B = 100$.

and its optimal value is big and smaller if the distance decreases. In the BSS, while the actual distance is not directly available to control the step size, a representative measure is required. While Cichocki et al [8] use the low-pass filtered gradient and Murata et al [9] the averaged flow, $\mathbf{H}_k\mathbf{W}_k$, we propose to utilize a squared norm of the estimating function, $\mathbf{H}_k$, as an appropriate measure. To demonstrate that $\mathbf{H}_k$ is a measure of the distance between $\mathbf{V}_k = \mathbf{W}_k\mathbf{A}$ and $\mathbf{V}_{\mathrm{eq}} = \mathbf{W}_{\mathrm{eq}}\mathbf{A}$, the estimating function is examined when $\mathbf{W}$ is in the neighborhood of $\mathbf{W}_{\mathrm{eq}}$. The estimating function in Eq. (8) is then described to first order by [7]

$$\begin{aligned}\mathbf{H}_k &= \mathbf{I} - \boldsymbol{\phi}(\mathbf{u})\mathbf{u}^T \\ &\approx \mathbf{I} - \boldsymbol{\phi}(\mathbf{s})\mathbf{s}^T - \mathbf{D}_{\phi'}\widetilde{\mathbf{V}}_k\mathbf{s}\mathbf{s}^T - \boldsymbol{\phi}(\mathbf{s})\mathbf{s}^T\widetilde{\mathbf{V}}_k^T \ ,\end{aligned} \tag{16}$$

where

$$\mathbf{D}_{\phi'} = \mathrm{diag}([\phi_1'(s_1), \ldots, \phi_N'(s_N)]) \ .$$

and, for simplicity, the block size B is 1. Now, if $\widetilde{\mathbf{V}}_k$ is held fixed, the expected value of $\mathbf{H}_k$ is given by

$$\mathsf{E}\left\{\mathbf{H}_k\right\} = -\mathsf{E}\left\{\mathbf{D}_{\phi'}\widetilde{\mathbf{V}}_k\mathbf{D}_{s^2}\right\} - \mathbf{D}_{\phi s}\mathsf{E}\left\{\widetilde{\mathbf{V}}_k^T\right\} \ , \tag{17}$$

where

$$\begin{aligned}\mathbf{D}_{s^2} &= \mathrm{diag}([s_1^2, \ldots, s_N^2]) \\ \mathbf{D}_{\phi s} &= \mathsf{E}\left\{\boldsymbol{\phi}(\mathbf{s})\mathbf{s}^T\right\}\end{aligned}$$

or, in terms of the elements of $\mathsf{E}\{\mathbf{H}_k\}$, Eq. (17) can be written as

$$\begin{aligned}\mathsf{E}\left\{H_{ij_k}\right\} = &- \mathsf{E}\left\{\phi_i'(s_i)\right\}\mathsf{E}\left\{s_j^2\right\}\tilde{V}_{ij_k} \\ &- \mathsf{E}\left\{\phi_i(s_i)s_i\right\}\tilde{V}_{ji_k}, \quad i \neq j\end{aligned} \tag{18}$$

$$\mathsf{E}\left\{H_{ii_k}\right\} = -\,\mathsf{E}\left\{\phi_i'(s_i)s_i^2 + \phi_i(s_i)s_i\right\}\tilde{V}_{ii_k} \tag{19}$$

and $\mathsf{E}\{\mathbf{H}_k\}$ is seen to be a measure of the error matrix, $\widetilde{\mathbf{V}}_k = \mathbf{V}_k - \mathbf{V}_{\text{eq}}$. An algorithm that utilizes a step size dependent on the error, $\widetilde{\mathbf{V}}_k$, now becomes available. That concept uses the square of a norm of matrix $\mathsf{E}\{\mathbf{H}_k\}$ to control the step size of the BSS algorithm. To implement this concept of a self-adjusting step size, a smoothed version (running average) of the estimating function, denoted by $\bar{\mathbf{H}}_k$, is used to approximate $\mathsf{E}\{\mathbf{H}_k\}$. It works with a step size equal to the step size in Eq. (9) and given by the recursion

$$\bar{\mathbf{H}}_k = (1 - \mu_{k-1})\bar{\mathbf{H}}_{k-1} + \mu_{k-1}\mathbf{H}_k \ . \tag{20}$$

A squared norm of $\bar{\mathbf{H}}_k$ then serves as the measure of error to control the step size, μ_k, in Eqs. (9) and (20). Many possible squared norms can serve as an appropriate measure: the normalized squared Frobenius norm, $\text{tr}\{\bar{\mathbf{H}}_k\bar{\mathbf{H}}_k^T\}/N^2$, the maximum eigenvalue of $\bar{\mathbf{H}}_k\bar{\mathbf{H}}_k^T$, etc. After many simulations the measure chosen for the control of the step size was the maximum of the squared values of the elements of $\bar{\mathbf{H}}_k$, emphasizing the error of the worst adjusted coefficient in $\mathbf{V}$. That is,

$$f_k = \max_{i,j} \bar{H}^2_{ij_k} \ . \tag{21}$$

The step size, μ_k, is now made proportional to f_k

$$\mu_k = \rho f_k \ , \tag{22}$$

where ρ is a constant. As the algorithm adjusts $\mathbf{W}_k$ toward $\mathbf{W}_{\text{eq}}$, the error, $\widetilde{\mathbf{V}}_k$, in $\mathbf{H}_k$ decreases. From Eq. (20), $\bar{\mathbf{H}}_k$ reflects this change and f_k decreases and, in turn, decreases μ_k. Since the error level decreases with decreasing μ_k, see for example Eq. (11) and [7] for details, the error in $\mathbf{H}_k$ decreases further. This process continues and the error, for a time-invariant mixing matrix $\mathbf{A}$, decreases toward zero. For a changing environment, as $\mathbf{A}$ changes, the overall matrix $\mathbf{V}_k = \mathbf{W}_k\mathbf{A}$, in general, moves further from $\mathbf{V}_{\text{eq}}$ and the increased error is reflected in $\mathbf{H}_k$. The change in $\mathbf{H}_k$ impacts $\bar{\mathbf{H}}_k$ increasing μ_k. Increased μ_k allows the algorithm to adjust $\mathbf{W}_k$ more quickly to move $\mathbf{V}_k$ toward $\mathbf{V}_{\text{eq}}$. This behavior is illustrated quite dramatically in the simulations of Section 5. Hence, a self-adjusting step size results that allows $\mathbf{W}_k$ to track a changing environment.

4. THE SELF-ADJUSTING STEP-SIZE ALGORITHM

In the final algorithm an upper bound on the step size μ_k is necessary. A bad match of the demixing matrix $\mathbf{W}_k$ with the mixing matrix $\mathbf{A}$ adversely impacts the error $\widetilde{\mathbf{V}}_k$ and, in turn, generates an estimating matrix $\mathbf{H}_k$ via Eqs. (20) to (22) that results in a large step size. If μ_k is sufficiently large, algorithm stability becomes an issue. To prevent this a time-variable upper bound

$$\mu_{\text{up}_k} = \frac{1}{2}\frac{1}{\max_i \sum_l |H_{il_k}|} \tag{23}$$

is used to limit μ_k. An intuitive argument for this bound is given in Appendix A. Extensive simulation studies demonstrate that use of this bound provides a stable algorithm. Limiting the step size μ_k can be implemented by a clipping function

$$b_k(x) = \begin{cases} x \ , & x \leq \mu_{\text{up}_k} \\ \mu_{\text{up}_k} \ , & x > \mu_{\text{up}_k} \ . \end{cases} \tag{24}$$

Second, additional smoothing of μ_k in Eq. (22) was found by simulations to be helpful. This is achieved by the recursion

$$\mu_k = \alpha\mu_{k-1} + (1-\alpha)\rho f_k \ , \tag{25}$$

where simulations show that good results are achieved for an $\alpha = 0.998$.

Third, a block-wise processing with block size $B > 1$ is computationally less burdensome. Choosing $B > 1$ implies a multiplication of the measure in Eq. (21) with B and replacing α by α^B in Eq. (25). That is

$$f_k = B\max_{i,j} \bar{H}^2_{ij_k} \tag{26}$$

$$\mu_k = \alpha^B\mu_{k-1} + (1-\alpha^B)\rho f_k \ . \tag{27}$$

The self-adjusting step-size algorithm is depicted in Fig. 3 and summarized in Eqs. (28) to (32).

$$\mathbf{H}_k = \frac{1}{B}\sum_t \left(\mathbf{I} - \phi(\mathbf{u}_t)\mathbf{u}_t^T\right) \tag{28}$$

$$\bar{\mathbf{H}}_k = (1-\mu_{k-1})\bar{\mathbf{H}}_{k-1} + \mu_{k-1}\mathbf{H}_k \tag{29}$$

$$f_k = B\max_{i,j} \bar{H}^2_{ij_k} \tag{30}$$

$$\mu_k = b_k(\alpha^B\mu_{k-1} + (1-\alpha^B)\rho f_k) \tag{31}$$

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu_k\mathbf{H}_k\mathbf{W}_k \ . \tag{32}$$

5. SIMULATIONS

In case of a time-invariant mixing environment, simulations show the step size is controlled such that the ICI_k approaches the lower bound given by Pham and Garat [10] for large k. A simulation of a separation problem with 10 Laplacian sources is depicted in Fig. 4. Fig. 5 shows the long-term behavior of the same situation as in Fig. 4. The log-log plot demonstrates that the ICI exhibits approximately a $1/k$ behavior for large k.

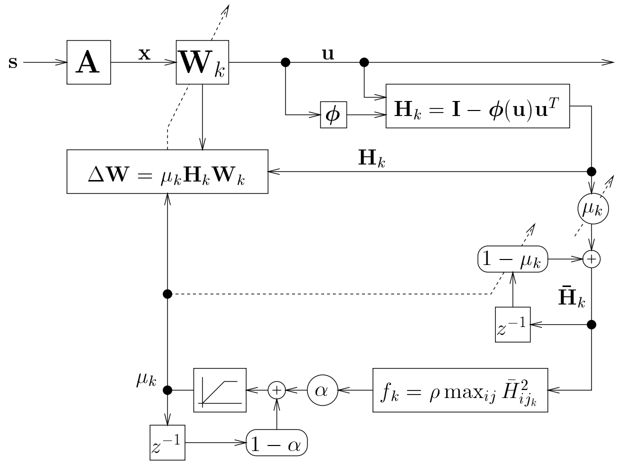

Figure 3: Concept of the step-size control, where for clearity the block size B is 1.

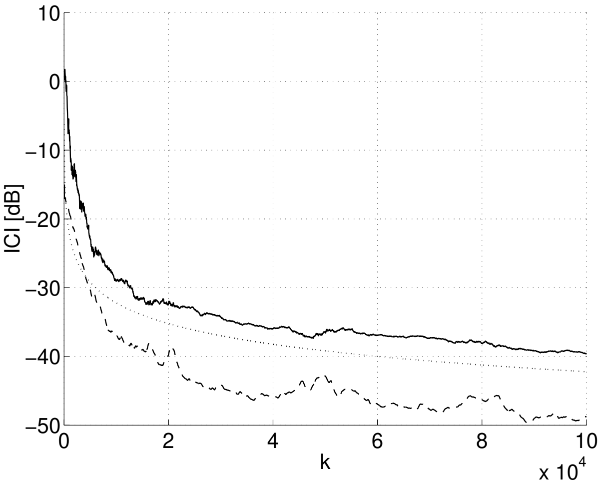

Figure 4: Separation of 10 Laplacian sources with a self-adjusting step size and nonlinearity $\phi(.) = \sqrt{2}\text{sign}\,(.)$. The mixing matrix is time-invariant. Processing was sample-wise ($B = 1$), $\alpha = 0.998$, and $\rho = 0.25$. ICI_k (solid), μ_k dashed, and the lower bound by Pham and Garat [10] (dotted).

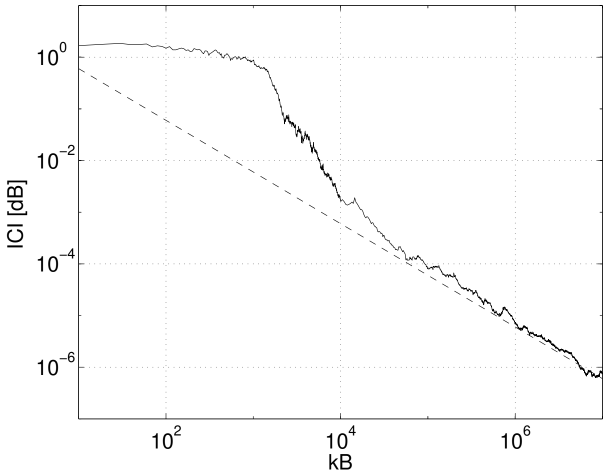

Figure 5: Separation of 10 Laplacian sources with a self-adjusting step size and nonlinearity $\phi(.) = \sqrt{2}\text{sign}\,(.)$. The mixing matrix is time-invariant and the step size was smoothed with $\alpha = 0.998$. $\rho = 0.25$. Processing is block-wise ($B = 10$). Both the time vs. ICI plot (solid) and the time vs. lower bound by Pham and Garat [10] (dashed) are depicted logarithmically.

Next, simulations were done for a separation problem of 10 Laplacian sources where the mixing environment changes abruptly. The new mixing matrix $\mathbf{A}$ is generated by

$$\mathbf{A} = \dot{\mathbf{A}} x^2 \,, \tag{33}$$

where the coefficients $\dot{A}_{ij}$ are Gaussian distributed, i.e.,

$$\dot{A}_{ij} \sim \mathcal{N}(0, 1) \tag{34}$$

and the scalar factor x is uniformly distributed between 0 and 1. This can produce large differences in the level of the coefficients. The step size is seen to respond immediately, rapidly increasing to allow the algorithm to adjust the demixing matrix $\mathbf{W}$ to reduce error levels. The step size then continues to decrease until the next abrupt change as depicted in Fig. 6. Simulation results for the separation of 2 Laplacian sources and a mixing environment that behaves in the same way as described above are depicted in Fig. 7.

If the environment suddenly experiences small continuing changes in the mixing matrix, the step size also responds appropriately. This is simulated by defining the mixing matrix recursively by

$$\mathbf{A}_{k+1} = \begin{cases} \mathbf{A}_k + \mathbf{A}_k^{\Delta} \,, & 20000 < k < 60000 \\ \mathbf{A}_k \,, & \text{elsewhere,} \end{cases} \tag{35}$$

where

$$A^{\Delta}_{ij_k} \sim \mathcal{N}(0, 0.001) \,.$$

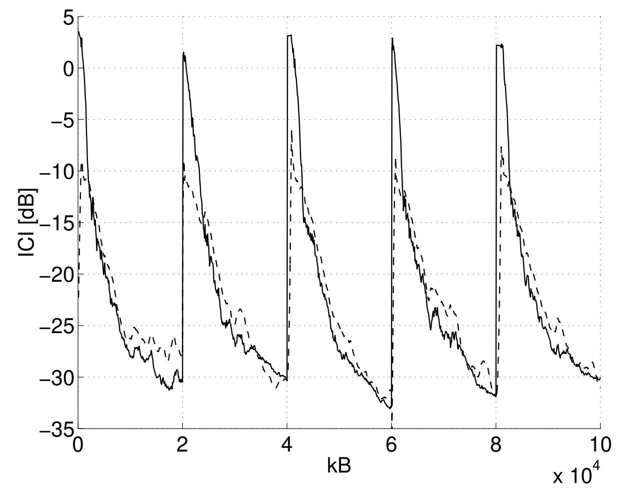

Figure 6: Separation of 10 Laplacian sources with a self-adjusting step size and nonlinearity $\phi(.) = \sqrt{2}\text{sign}\,(.)$. The mixing matrix changes abruptly every 20000 samples. The step size was smoothed $\alpha = 0.998$. Processing was block-wise with $B = 10$ and $\rho = 0.25$. ICI$_k$ (solid), μ_k dashed.

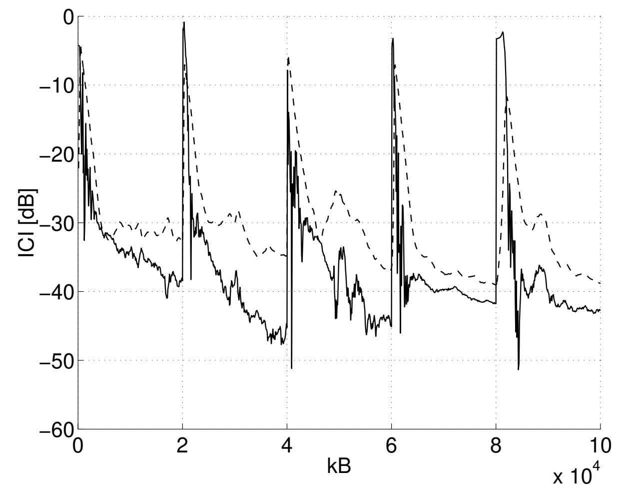

Figure 7: Separation of 2 Laplacian sources with a self-adjusting step size and nonlinearity $\phi(.) = \sqrt{2}\text{sign}\,(.)$. The mixing matrix changes abruptly every 20000 samples. The step size was smoothed $\alpha = 0.998$. Processing was block-wise ($B = 10$). ICI$_k$ (solid), μ_k dashed.

During the period, $20000 < k < 60000$, the mixing environment is changing, otherwise, it is time-invariant. The overall behavior is depicted in Fig. 8. When the mixing matrix is time-invariant the step size decreases as before, then as $\mathbf{A}$ starts to drift the error increases and, in response, the step size increases allowing the algorithm to adjust the demixing matrix, $\mathbf{W}_k$. For this environment, a steady error level is reached and the step size maintains an increased level. After 60000 samples the drift stops and the error level and step size again decrease.

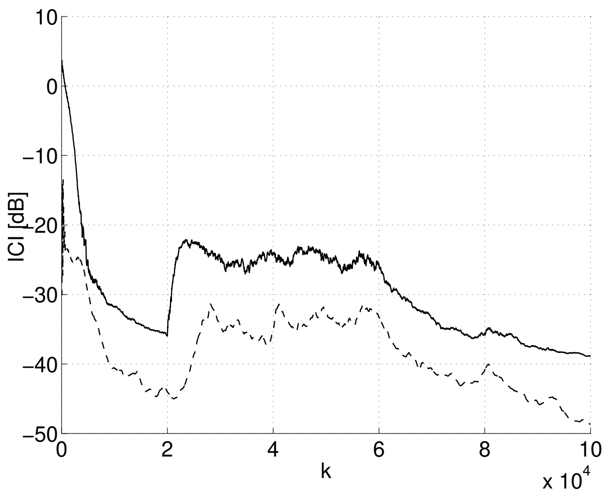

Figure 8: Separation of 10 uniform sources with a self-adjusting step size and nonlinearity $\phi(u_i) = u_i^3$. Sample-wise processing ($B = 1$) and $\alpha = 0.998$. ICI$_k$ (solid), μ_k dashed. The mixing matrix drifts for $20000 < k < 60000$.

6. CONCLUSIONS

This paper has presented a self-adjusting step size that allows error-free convergence of the classic BSS algorithm in a time-invariant mixing environment and that has the ability to track a changing mixing environment. The step size is controlled with the help of a squared norm of the estimating function. Not all design parameters were optimized in the current effort. In addition to the simulation results presented, extensive trials were conducted on many time-varying scenarios. The algorithm performed well in all cases. While the studies of this paper assume constant source power, future research is required for the case of power-varying signals.

7. REFERENCES

[1] Anthony J. Bell and Terrence J. Sejnowski, "An information-maximisation approach to blind separation

and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[2] Barak Pearlmutter and Lukas Parra, "Maximum likelihood blind source separation: A context-sensitive generalization of ICA," *Advances in Neural Information Processing Systems*, vol. 9, pp. 613–619, 1997, MIT Press, Cambridge, MA, USA.

[3] Pierre Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1994.

[4] Jean-François Cardoso, "Performance and implementation of invariant source separation algorithms," in *ISCAS '96*, 1996, vol. 4, pp. 85–88.

[5] Shun-ichi Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, pp. 251–276, 1998.

[6] Shun-ichi Amari, Tian-Ping Chen, and Andrzej Cichocki, "Stability analysis of adaptive blind source separation," *Neural Networks*, vol. 10, no. 8, pp. 1345–1351, 1997.

[7] Thomas P. von Hoff, Allen G. Lindgren, and August N. Kaelin, "Transpose properties in the stability and performance of the classic adaptive algorithms for blind source separation and deconvolution," submitted to Elsevier, Oct. 1999.

[8] Andrzej Cichocki, Shun-ichi Amari, Masaharu Adachi, and Wlodzimierz Kasprzak, "Self-adaptive neural networks for blind separation of sources," in *International Symposium Circuits and Systems*, May 1996, vol. 2, pp. 157–160.

[9] Noboru Murata, Klaus-Robert Müller, Andreas Ziehe, and Shun-ichi Amari, "Adaptive on-line learning in changing environments," in *Advances in Neural Information Processing Systems*, D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, Eds., Cambridge, MA, 1997, vol. 9, The MIT Press.

[10] Dinh Tuan Pham and Philippe Garat, "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Transaction on Signal Processing*, vol. 45, no. 7, pp. 1712–1725, 1997.

APPENDIX A. UPPER BOUND FOR THE STEP SIZE

The step size must be upper bounded to prevent $\mathbf{W}_k$, $\mathbf{H}_k$, and μ_k in the algorithm given by Eqs. (28) to (32) from "pumping up". An approach that has been shown by simulations to be effective requires that the update term for the ijth element in $\mathbf{V}_k$, $\Delta V_{ij_k} = V_{ij_{k+1}} - V_{ij_k}$ must be smaller than the biggest magnitude of the coefficients in $\mathbf{V}_k$. Mathematically, this is

$$|\Delta V_{ij_k}| \ll \max_{p,q} |V_{pq_k}| \tag{A-1}$$

for each $1 \leq i, j \leq N$. Dividing (A-1) by $\max_{p,q} |V_{pq_k}|$ yields

$$\frac{|\Delta V_{ij_k}|}{\max_{p,q} |V_{pq_k}|} \ll 1 \tag{A-2}$$

From Eq. (9),

$$|\Delta V_{ij_k}| \leq \mu_k \sum_l |H_{il_k}||V_{lj_k}| \leq \mu_k \sum_l |H_{il_k}| \max_{p,q} |V_{pq_k}| \tag{A-3}$$

and the left-hand side of (A-2) can be replaced by a bigger μ_k-dependent term

$$\frac{\mu_k \sum_l |H_{il_k}| \max_{p,q} |V_{pq_k}|}{\max_{p,q} |V_{pq_k}|} \ll 1\ . \tag{A-4}$$

Setting the term on the left-hand side of (A-4) equal to a value $r < 1$ yields an upper bound for μ_k

$$\mu_{\text{up}_k} = \frac{r}{\max_i \sum_l |H_{il_k}|}\ . \tag{A-5}$$

With μ_k bounded by μ_{up_k}, the left-hand side in (A-2) is at most r. Applied to the nonlinear function given by (24) the parameter r was chosen to be 0.5 yielding

$$\mu_{\text{up}_k} = \frac{1}{2} \frac{1}{\max_i \sum_l |H_{il_k}|}\ . \tag{A-6}$$