SOURCE DISTRIBUTION INDEPENDENT ADAPTIVE BLIND SIGNAL SEPARATION

Vicente Zarzoso and Asoke K. Nandi

Signal Processing Division, Department of Electronic and Electrical Engineering,
University of Strathclyde, 204 George St., Glasgow G1 1XW, Scotland, UK
Tel: +44 141 548 22 05. Fax: +44 141 552 24 87.
e-mail: vicente@spd.eee.strath.ac.uk

ABSTRACT

In the noiseless two-source two-sensor BSS scenario, the source extraction reduces to the estimation of a Givens rotation matrix from the whitened sensor output. In this contribution an adaptive estimator of the angle which characterizes such a rotation in the real-mixture case is derived. It is shown to converge regardless of the source statistics, provided their kurtosis sum is not null. An asymptotic performance analysis yields a closed-form expression for the asymptotic probability density function of the proposed adaptive angle estimator. It is also demonstrated how the estimator can be incorporated into a complete adaptive source separation system by combining it with an adaptive pre-whitening strategy, and how it can be useful in a general BSS environment composed of more than two sources and two sensors by means of a pairwise approach. Several experiments corroborate the theoretical exposition.

1. INTRODUCTION

A variety of applications require the real-time extraction of the source signals from their linear mixtures observed at the sensor output. In other cases, the parameters of the mixing system or the source signals themselves are not stationary, but are subject to variations with time. As an example, one may consider a multi-user mobile communication system, in which the users' roam alters the channel characteristics, thus modifying the mixing pattern. Even if the users remain still, time variations occur in the channel due to changing atmospheric conditions and environment (fading). All this kind of applications call for adaptive BSS implementations.

This work is supported by the University of Strathclyde and by the Engineering and Physical Sciences Research Council (EPSRC) of the UK under the research grant ref. GR/L/07475.

There is a diverse range of such procedures across the literature. Nevertheless, many existing methods for adaptive BSS manifest two important shortcomings. Firstly, the convergence of some of them is very dependent on the choice and tuning of certain non-linear functions with respect to the source statistics (e.g., [1] and [2]). Secondly, the asymptotic performance of some others is difficult to analyze (e.g., [3] and [4]). This paper puts forward a new method for adaptive blind separation of real mixtures which surmounts such drawbacks. By contrast, the proposed procedure exhibits two very desirable properties: (1) convergence (virtually) independent of the source distribution, and (2) performance analysis easy to carry out. The algorithm is based on the batch-processing method introduced in [5] and [6], from which property 1 is inherited. Effectively, the most appealing feature of this 4th-order method is that it is valid for nearly any source distribution combination, and does not rely on any choice of non-linearities to achieve the source extraction. From the off-line procedure its adaptive version is derived, resulting in an updating rule with the standard stochastic algorithm form. As a consequence, the analysis of its convergence and asymptotic properties is easily accomplished. Most of the derivations are developed in the simplified noiseless two-source two-sensor BSS scenario, but the procedure can also be extended to contexts of more than two signals, in a straightforward pairwise processing fashion.

The exposition of the material is organized as follows. Section 2 summarizes the basic aspects surrounding the BSS problem in an adaptive framework. Next, section 3 reviews the off-line method and condenses its major properties. The adaptive algorithm is developed in section 4 and its asymptotic performance analysis sketched in section 5. Section 6 reports on some experiments aiming to confirm the validity of the theoretical results, as well as to demonstrate the performance of the suggested adaptive procedure in a full separation

In *Proceedings ICA '99*, Aussois, France, 11-15 January 1999, pp. 65-70. (Updated, corrected version.)

system. Concluding remarks are left for section 7.

2. PROBLEM STATEMENT

The goal of BSS consists in recovering a set of q *source signals* $\mathbf{x}_k = [x_1(k), \ldots, x_q(k)]^{\mathrm{t}} \in \mathbb{R}^q$ from a set of p ($p \geqslant q$) instantaneous linear mixtures measured at the *sensor output*, $\mathbf{y}_k = [y_1(k), \ldots, y_p(k)]^{\mathrm{t}} \in \mathbb{R}^p$, in which k is a time index. If the mixture coefficients are contained in the *mixing* or *transfer matrix* $M = (m_{ij}) \in \mathbb{R}^{p\times q}$, then the noiseless BSS model may be stated in matrix notation as:

$$\mathbf{y}_k = M\mathbf{x}_k, \quad k = 1, 2, \ldots \tag{1}$$

The source extraction and transfer matrix identification are carried out relying upon two main assumptions only: (A1) the sources are statistically independent, and (A2) the mixing matrix is full column rank. Indeterminacies of two kinds are inherent to the problem: the ordering of the recovered sources and their magnitude. In most cases both are unimportant. Hence, the convention may be adopted that the source signals are unit-power, so that their covariance matrix becomes $R_x \triangleq \mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{t}}] = I_q$, the $q \times q$ identity matrix.

Adaptively, the source extraction is achieved by estimating a matrix W_k, referred to as *separating matrix*, such that the resultant *global matrix*

$$G_k = W_k M, \tag{2}$$

representing the whole mixing-unmixing system, tends to a matrix with only one non-zero entry per row and per column. That is, in a successful separation the global matrix converges to the product:

$$G_k \to PD, \tag{3}$$

where permutation matrix P models the source ordering indeterminacy, and regular diagonal matrix D is related to the scale-factor indetermination. If the sources are to be recovered with their original power, matrix D must have unit-norm elements and then the associated matrix G becomes a so-called *quasiidentity* matrix [2].

Most approaches to BSS operate in two steps. Algebraically, this is associated with the factorization of the mixing matrix as the product of a regular matrix B and an orthogonal matrix Q (polar factorization):

$$M = BQ. \tag{4}$$

In the first place, a so-called *whitening matrix*, $\widehat{B}_k^*$, is estimated. This matrix should tend to the pseudoinverse of B:

$$\widehat{B}_k^* \to B^*, \tag{5}$$

generating a set of uncorrelated and normalized (unit-power) signals:

$$\mathbf{z}_k = \widehat{B}_k^* \mathbf{y}_k \to Q\mathbf{x}_k, \tag{6}$$

which are termed *whitened observations*. The sources are then recovered by obtaining adaptive estimates $\widehat{Q}_k$ of the rotation matrix Q from the whitened sensor output, and applying them a counter-rotation $\widehat{Q}_k^{\mathrm{t}}$:

$$\hat{\mathbf{x}}_k = \mathbf{s}_k = \widehat{Q}_k^{\mathrm{t}} \mathbf{z}_k = G_k \mathbf{x}_k. \tag{7}$$

The adaptive pre-whitening stage will not be discussed herein (the reader is referred to, e.g., [2], [3], and [4] for various manners of accomplishing this task). Rather, we will focus on the adaptive estimation of orthogonal matrix Q. The obtained results can then be combined with a pre-whitening strategy in order to yield a complete separation system. The described set-up for adaptive BSS is graphically summarized in figure 1.

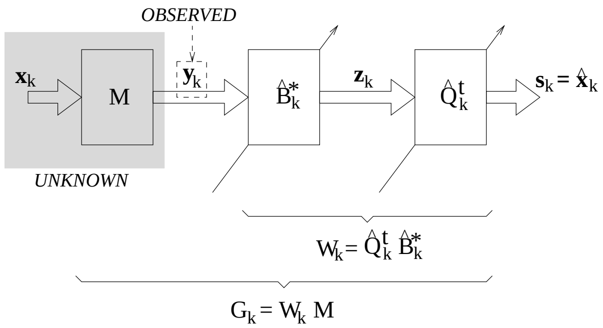

Figure 1: Two-stage adaptive BSS.

3. A BATCH-PROCESSING METHOD

This section reviews a method for the estimation of the orthogonal matrix Q from a batch of whitened-observation samples. Let us begin by considering the batch counterpart of equation (6):

$$\mathbf{z} = Q\mathbf{x} \tag{8}$$

In the two signal case, matrix Q reduces to an elementary Givens rotation matrix:

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \tag{9}$$

and hence the estimation of Q boils down to the estimation of a single relevant parameter, angle θ.

By virtue of equations (8) and (9), the scatter-plot points of the sources and the whitened observations are just rotated versions of each other. In complex form:

$$\left.\begin{array}{l} x_1(k) + jx_2(k) = \rho_k \mathrm{e}^{j\phi'_k} \\ z_1(k) + jz_2(k) = \rho_k \mathrm{e}^{j\phi_k} \end{array}\right\} \quad \phi_k = \phi'_k + \theta. \tag{10}$$

Now, consider the *centroid* computed as the mean value of the whitened-observation scatter points to the fourth power:

$$\xi \triangleq \mathrm{E}[(z_1 + jz_2)^4] = \mathrm{E}[\rho^4 \mathrm{e}^{j4\phi}]. \tag{11}$$

By using (10), expression (11) is easily simplified into

$$\xi = \mathrm{E}[(x_1 + jx_2)^4]\, \mathrm{e}^{j4\theta}. \tag{12}$$

The expectation in (12) turns out to be

$$\mathrm{E}[(x_1 + jx_2)^4] = \kappa_{40}^x + \kappa_{04}^x, \tag{13}$$

in which κ_{mn}^x denotes the $(m+n)$th-order cumulant of the source signals. The source kurtosis sum is unknown, but can be readily determined from the whitened observations as:

$$\gamma \triangleq \mathrm{E}[(z_1^2 + z_2^2)^2] - 8 = \mathrm{E}[\rho^4] - 8 = \kappa_{40}^x + \kappa_{04}^x. \tag{14}$$

From equations (11)–(14), and providing the source kurtosis sum is not null, angle θ can be estimated as:

$$\hat{\theta}_{\mathrm{EML}} = \frac{1}{4}\,\mathrm{angle}(\xi \cdot \mathrm{sign}(\gamma)). \tag{15}$$

A few interesting remarks about the above estimator are summarized next:

- Observe that no assumptions whatsoever have been made about the source statistics in order to arrive at (15): it is valid as long as the source kurtosis sum is different from zero.
- Estimator (15) is referred to as "*EML*", for "*Extended Maximum-Likelihood*". The rationale behind this name is that expression (15) can be regarded as a generalization of the approximate ML criterion developed in [7], and hence the adjective "extended". A more detailed connection between these two approaches is made in [5] and [6].
- The error introduced in estimator (15) due to finite-size sample is only caused by the source signals, via the sample estimates of the expectations in (12) and (14). As a consequence, it is easy to prove [8] that estimator (15) is *orthogonal invariant*, with all the desirable implications that this characteristic conveys (full invariance in the noiseless case yielding uniform performance) [9].
- Both ξ and γ may be expanded as a function of the whitened sensor output cumulants [5]. Hence, by consistency of cumulant sample estimates, estimator (15) is *consistent* and *asymptotically unbiased*. An exhaustive study of the asymptotic properties of this batch angle estimator is carried out in [6], where a very interesting geometrical interpretation is also given.

By design, estimator (15) is only valid to separate two sources from the mixture observed by two sensors. In the general BSS scenario composed of more than two mixtures of more than two sources, expression (15) can be applied in turn to each pair of decorrelated measurements until convergence. This heuristic pairwise procedure was originally introduced in [3], and mathematical grounds were found for it in [10] when polynomial-form contrast functions are employed. Even though there is no proof of the convergence of such iterative scheme, several experiments sustain the validity of this generalization [6]. The number of sweeps over the signal pairs necessary for convergence was empirically found to be around $1+\sqrt{q}$, which coincides with value already proposed in [10].

4. AN ADAPTIVE ALGORITHM

Let ν_k denote the scatter-diagram point of the prewhitened observations to the fourth power at time instant k:

$$\nu_k = (z_1(k) + jz_2(k))^4 = \rho_k^4 \mathrm{e}^{j4\phi_k}. \tag{16}$$

Then, ξ and γ in (11) and (14), respectively, may be obtained in an adaptive fashion from:

$$\xi_k = (1-\alpha_k)\xi_{k-1} + \alpha_k \nu_k \tag{17}$$

$$\gamma_k = (1-\alpha_k)\gamma_{k-1} + \alpha_k(|\nu_k| - 8), \tag{18}$$

where α_k is the adaption coefficient at iteration k, balancing a trade-off between accuracy and speed of convergence. The rotation angle θ is estimated at each iteration by plugging these two parameters into (15):

$$\hat{\theta}_k = \frac{1}{4}\,\mathrm{angle}(\xi_k \cdot \mathrm{sign}(\gamma_k)). \tag{19}$$

From a counter-rotation $\widehat{Q}_k^{\mathrm{t}}$ of this angle estimate, the corresponding source-signal sample at iteration k is obtained as in (7). Estimator (19) is termed "*AEML*", for "*Adaptive EML*". A brief study of the convergence and asymptotic properties of adaptive estimator (19) is outlined in the next section, but can be found in finer detail in [8]. For the sake of simplicity, in the following a constant adaption coefficient will be assumed, $\alpha_k = \alpha$, $\forall k$, which, on the other hand, allows the separator to track non-stationary mixtures.

5. ASYMPTOTIC ANALYSIS

Adaption scheme (17)–(18) accepts the typical adaptive algorithm updating rule [11]:

$$\xi_k = \xi_{k-1} + \alpha_k H(\xi_{k-1}, \nu_k), \tag{20}$$

where, from (17), function $H(\cdot,\cdot)$ is identified as

$$H(\xi, \nu) = \nu - \xi, \tag{21}$$

and accordingly for (18). Therefore, the standard tools for the study of stochastic algorithms of this general form can be borrowed in order to analyze the convergence properties of the algorithm suggested here. Our ultimate objective is the derivation of the asymptotic pdf of the AEML estimator (19).

The term

$$h(\xi) \triangleq \mathrm{E}[H(\xi, \nu_k)] = \xi_* - \xi, \tag{22}$$

is called *mean field*, where

$$\xi_* = \mathrm{E}[\nu_k], \tag{23}$$

is the population centroid location (11) and symbol $\mathrm{E}[\cdot]$ denotes the expectation with respect to the distribution of the state ν_k for a fixed value of ξ when ν_k is asymptotically stationary. The values of ξ that make the mean field $h(\cdot)$ vanish are known as *equilibrium points* or *attractors*. Hence $\xi = \xi_*$ is the unique attractor of (17)/(20). Since

$$h_\xi(\xi_*) \triangleq \left.\frac{\partial h}{\partial \xi}\right|_{\xi=\xi_*} = -1 < 0, \tag{24}$$

the equilibrium point ξ_* is *locally asymptotically stable (LAS)* [11] always, regardless of the source distribution and the mixing structure. Analogously, $\gamma = \gamma_* = \mathrm{E}[|\nu_k|^4 - 8] = \kappa_{40}^x + \kappa_{04}^x$ is the unique LAS equilibrium point of (18).

Under the above conditions, the asymptotic distribution of ξ_k is normal [11, theorem 2, chapter 3]:

$$\tilde{\xi_k} \triangleq \alpha^{-\frac{1}{2}}(\xi_k - \xi(t_k)) \underset{\substack{k\to\infty \\ \alpha\to 0}}{\longrightarrow} N(0, P), \tag{25}$$

where $\xi(t_k)$ is the solution of the associated ODE $\dot{\xi}(t) = h(\xi)$ at $t_k = k\alpha$, and the covariance $P = \mathrm{cov}[\tilde{\xi_k}]$ is given by the solution of the Lyapunov equation:

$$2h_\xi(\xi_*)P + R(\xi_*) = 0, \tag{26}$$

with

$$R(\xi) \triangleq \sum_{k=-\infty}^{\infty} \mathrm{cov}[H(\xi, \nu_k),\, H(\xi, \nu_0)]. \tag{27}$$

In conclusion, the distribution of the centroid ξ_k is asymptotically normal with mean ξ_* and variance

$$\sigma_\xi^2 = \alpha P = \alpha R(\xi_*)/2. \tag{28}$$

From this key result, the asymptotic pdf of $\hat{\theta}_k$ given by (19) can be determined as:

$$\begin{aligned} p_{\hat{\theta}}(\hat{\theta}) = {} & \frac{2}{\pi}\exp\left(\frac{-{\kappa^+}^2}{2\sigma_\xi^2}\right) + \\ & + \sqrt{\frac{2}{\pi}}\frac{\kappa^+}{\sigma_\xi}\cos(4(\hat{\theta}-\theta))\exp\left(\frac{-{\kappa^+}^2\sin^2(4(\hat{\theta}-\theta))}{2\sigma_\xi^2}\right) \times \\ & \times \left\{1 + \mathrm{erf}\left(\frac{1}{\sigma_\xi\sqrt{2}}\,\kappa^+\cos(4(\hat{\theta}-\theta))\right)\right\}, \end{aligned} \tag{29}$$

where the symbol κ^+ denotes the absolute value of the source kurtosis sum, $\kappa^+ \triangleq |\gamma_*| = |\kappa_{40}^x + \kappa_{04}^x|$. Pdf $p_{\hat{\theta}}$ is an even function of $(\hat{\theta} - \theta)$, which means that adaptive estimator (19) is *asymptotically unbiased*:

$$\mathrm{E}[\hat{\theta}] = \theta. \tag{30}$$

Moreover, the smaller the ratio σ_ξ/κ^+, the narrower the pdf (29) becomes. Hence, the asymptotic variance of estimator (19) is directly related to that ratio:

$$\sigma_{\hat{\theta}}^2 \triangleq \mathrm{var}[\hat{\theta}] \sim \frac{\sigma_\xi^2}{{\kappa^+}^2}. \tag{31}$$

In [8] a linear relationship is found between $\sigma_{\hat{\theta}}$ and σ_ξ/κ^+, valid for small values of this latter ratio. According to (28), $\sigma_{\hat{\theta}}^2$ can then be made as small as wished by decreasing α. Of course, this enhanced precision would be attained at the expense of a slower convergence rate.

Note that from equations (10)–(13) and (16), parameter R in (27) can be written as a function of the source signals only. Consequently, the performance of the adaptive estimator (19) does not depend on the mixture, but exclusively on the source signals. The invariance property of batch estimator (15), commented in section 3, is therefore inherited by its adaptive counterpart (19).

6. EXPERIMENTAL RESULTS

The theoretical results presented previously are illustrated with the help of some experiments. In [8] more extensive simulations are carried out and a comparison with other adaptive methods made. Some of such experiments reveal another remarkable attribute of adaptive estimator (19): its fast convergence speed. Due to the lack of space, however, only two issues will be assessed herein. First, the asymptotic results of section 5 are checked out. Later, the performance of a whole separation system made up of (19) is demonstrated in two different BSS environments.

6.1. Asymptotic performance

The accuracy of the asymptotic pdf found in section 5, formula (29), is examined. An orthogonal mixture of a uniformly distributed source and an exponentially distributed source is considered, with a true rotation angle of $\theta = 30°$. The wider solid-line plot in figure 2 displays the empirical pdf of $\hat{\theta}_k$ obtained through the kernel method [12] from $5 \cdot 10^3$ iterations of (19) after convergence, for $\alpha = 10^{-3}$. The dashed line represents analytical pdf (29), where the necessary parameters are calculated from the data by means of their sample estimates. Both curves are very similar. The resemblance increases for smaller adaptation coefficients, as seen in the narrower plots of figure 2, obtained with $\alpha = 10^{-4}$.

6.2. Complete adaptive separation system

Angle estimator (19) can be turned into a complete separation system just by combining it with an adaptive pre-whitening stage, creating a two-step separator as explained in section 2. The pre-whitening serial updating rule

$$\widehat{B}^*_{k+1} = \left[I_q - \lambda_k(\mathbf{z}_k\mathbf{z}_k^{\mathrm{t}} - I_q)\right]\widehat{B}^*_k, \tag{32}$$

was selected for its simplicity. It corresponds to the minimization, via the relative gradient, of the Kullback-Leibler divergence between the whitener output and a Gaussian random vector with identity covariance matrix [2]. Figure 3 shows the evolution of the global matrix elements $G_k = \widehat{Q}_k^{\mathrm{t}}\widehat{B}_k^* M$ for a mixture of two i.i.d. sources with a uniform and an exponential distribution, respectively, with $\lambda = 7.5 \cdot 10^{-3}$ and $\alpha = 5 \cdot 10^{-3}$. A successful separation is attained, since G_k tends to a quasiidentity matrix. The relative lack of accuracy after convergence is due to the impulsive nature of the second source. Smoother results may be achieved by using a decreasing adaption coefficient or, alternatively, some kind of normalization in the updating rule of the algorithm (as done, for instance, in [2]).

Estimator (19) is made applicable in a BSS set-up of more than two signals through an adaptive version of the extension commented at the end of section 3. At each iteration, elementary Givens rotations made up of (19) are applied to each pair of whitened samples in turn over certain number of sweeps. By using this generalization on a three-sensor mixture of three 16-symbol PAM sources, the global matrix trajectories plotted in figure 4 were obtained. The adaption coefficients were chosen as $\lambda = \alpha = 5 \cdot 10^{-3}$. Again, a valid separation solution is achieved, validating the proposed extension.

7. CONCLUSIONS

An adaptive procedure for the estimation of the rotation angle from the whitened sensor output in the noiseless real-mixture two-source two-sensor BSS has been developed. This adaptive estimator is characterized by two interesting features. First, it converges to a right separation solution regardless of the source distribution being treated, provided the source kurtosis sum is different from zero. Second, its asymptotic performance is simple to analyze. Simulations have proven these results, and have demonstrated that this adaptive estimator can also be applied to the BSS scenario of more than two signals through an uncomplicated pairwise extension.

Current research is looking into alternative generalizations to such broader BSS scenario, the extension to complex mixtures, and the effects of noise on the performance of the suggested estimator.

8. REFERENCES

[1] C. Jutten and J. Herault, "Blind Separation of Sources, Part I: An Adaptive Algorithm Based on Neuromimetic Architecture", *Signal Processing*, Vol. 24, No. 1, November 1991, pp. 1–10.

[2] J.-F. Cardoso and B. H. Laheld, "Equivariant Adaptive Source Separation", *IEEE Transactions on Signal Processing*, Vol. 44, No. 12, December 1996, pp. 3017–3030.

[3] P. Comon, "Separation of Stochastic Processes", *Workshop on Higher-Order Spectral Analysis*, Vail, Colorado, June 1989, pp. 174–179.

[4] F. Harroy, J.-L. Lacoume and M. A. Lagunas, "A General Adaptive Algorithm for nonGaussian Source Separation without any Constraint", *Proceedings EUSIPCO-94*, Edinburgh, Scotland, UK, September 1994, pp. 1161–1164.

[5] V. Zarzoso and A. K. Nandi, "Generalization of a Maximum-Likelihood Approach to Blind Source Separation", *Proceedings EUSIPCO-98*, Vol. IV, pp. 2069–2072, Rhodes, Greece, 8–11 September 1998.

[6] V. Zarzoso and A. K. Nandi, "Blind Separation of Independent Sources for virtually any Source Probability Density Function", submitted to *IEEE Transactions on Signal Processing*, 1998.

[7] F. Harroy and J.-L. Lacoume, "Maximum Likelihood Estimators and Cramer-Rao Bounds in Source Separation", *Signal Processing*, Vol. 55, December 1996, pp. 167–177.

[8] V. Zarzoso and A. K. Nandi, "Adaptive Blind Source Separation for virtually any Source Distribution", submitted to *IEEE Transactions on Signal Processing*, 1998.

[9] J.-F. Cardoso, "On the Performance of Orthogonal Source Separation Algorithms", *Proceedings EUSIPCO-94*, Edinburgh, Scotland, UK, September 1994, pp. 776–779.

[10] P. Comon, "Independent Component Analysis, A New Concept?", *Signal Processing*, Vol. 36, No. 3, April 1994, pp. 287–314.

[11] A. Benveniste, M. Métivier and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Berlin: Springer-Verlag, 1990.

[12] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall, 1986.

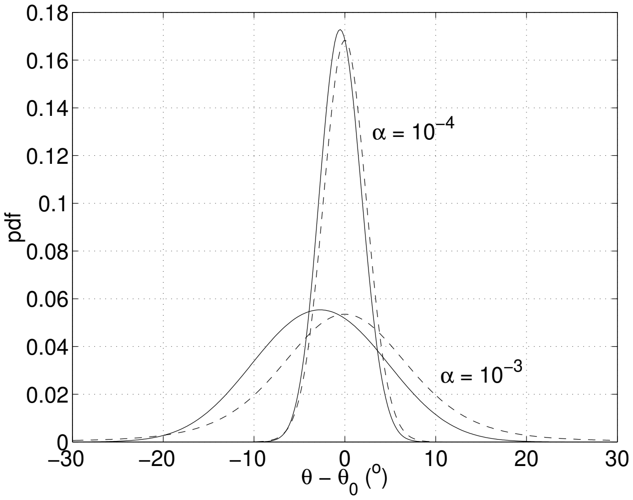

Figure 2: AEML asymptotic pdf for a mixture of uniform–exponential sources, $\theta = 30^o$. a) For $\alpha = 10^{-3}$. Solid line: empirical pdf estimate from $5 \cdot 10^3$ iterations after convergence, kernel method with function width 6^o. Dashed line: analytical pdf (29). b) For $\alpha = 10^{-4}$. Solid line: empirical pdf estimate from last $5 \cdot 10^4$ iterations after convergence, kernel method with function width 2^o. Dashed line: analytical pdf (29).

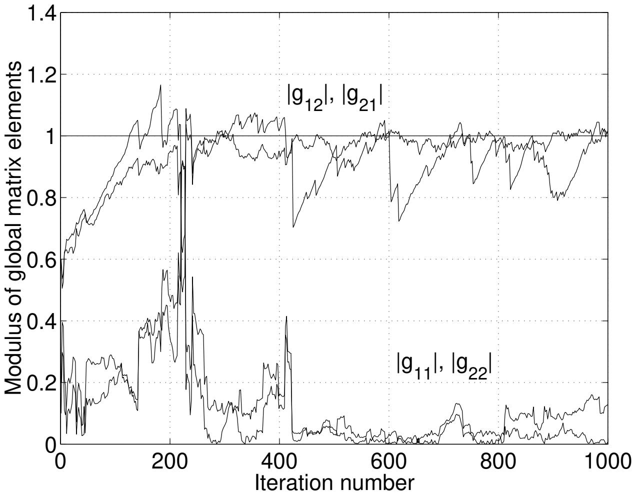

Figure 3: Trajectories of the global matrix coefficients in the separation of uniform–exponential sources from a two-sensor mixture by the AEML method, $\lambda = 7.5 \cdot 10^{-3}$, $\alpha = 5 \cdot 10^{-3}$.

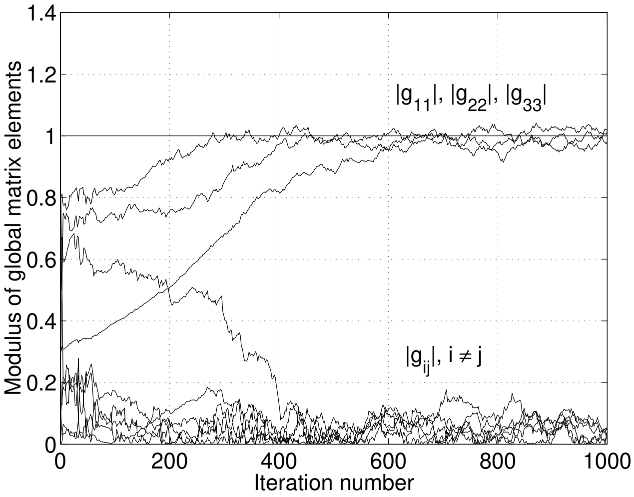

Figure 4: Evolution of the global system in the separation of three PAM16 sources from three observations by the AEML method, $\lambda = \alpha = 5 \cdot 10^{-3}$.